APPLICATIONS OF CAUSAL BAYESIAN NETWORKS ON URBAN PLANNING

A Dissertation

by

ALEXANDER ABUABARA

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Walter Gillis Peacock
Co-Chair of Committee,	Zenon Medina-Cetina
Committee Members,	Douglas Frank Wunneburger
	David Henry Bierling
Head of Department,	Galen Daniel Newman

May 2022

Major Subject: Urban and Regional Sciences

ABSTRACT

This research seeks to improve the current state of knowledge about risks related to natural hazards, particularly those hazards affecting coastal communities. The inquiry focuses on a broad question: how can I help communities *better* understand and assess hurricane-related risks? To answer this question, this research explores how the intuitive format of Bayesian networks can be useful for disaster planning applications. Previous research showed that Bayesian networks are probabilistic models with a relatively easy graphical interpretation (yet still having a solid statistical basis) that are widely used in different fields, although they have a limited application in planning to date. This research comprises three studies. The first study uses Bayesian networks as an exploratory tool to estimate economic and social costs when assessing hurricane risks in a typical single-family home in a coastal community. That study shows that Bayesian networks can be flexible when combining hazards and vulnerabilities to estimate risks and are useful even when only limited information and resources are available, or the data format is heterogeneous. The second study examines in elementary but practical examples and experiments the advantages and limitations of the use of Bayesian networks to model household hurricane evacuation for descriptive and predictive analysis. The third study examines hurricane household evacuation choices using Bayesian networks to isomorphically model the complexities of an established conceptual model for studying protective action decisions such as hurricane household evacuation. The results of these studies indicate that Bayesian networks can flexibly

integrate multiple fields in a complex structure of influence, which is the very nature of planning activities. Therefore, these studies show the potential of Bayesian networks to be used more frequently in future disaster preparedness and planning by facilitating the cooperation of specialists from several disciplines and providing a large potential for the engagement of citizens, policymakers, decision-makers, researchers, and other stakeholders to better understand local risks, which will ultimately foster participatory planning processes.

DEDICATION

This dissertation is dedicated to

my beautiful wife, Gabriela, and our brave son, Nicholas Bento.

ACKNOWLEDGMENTS

I would like to thank my co-chairs, Walt Peacock and Zenon Medina, and my committee members, Doug Wunneburger and David Bierling, for being supportive and encouraging, for listening to my ideas, and for being there for me. They offered me their steady guidance and thought-provoking suggestions over the course of this research. I am truly fortunate to have had the opportunity to work with such inspiring scholars.

I am grateful to my friends and colleagues in the Hazard Reduction and Recovery Center and the faculty and staff of the Department of Landscape Architecture and Urban Planning for their friendship and immense support. Special thanks to Fayola Jacobs, Bara Safarova, Han John Park, You Jung Kim, Nathanael Rosenheim, and Wayne Day for their infinite kindness, generosity, encouragement, and lengthy conversations. Special thanks to Michael Lindell, Carla Prater, Sherry Bame, Shannon van Zandt, Cecilia Giusti, Galen Newman, and Daniel Goldberg for taking extra steps to support and include me whenever possible. My deep gratitude to Joe Feagin, Mark Fossett, Nehemia Geva, and George Rogers, all of whom I have had the fortune of taking courses from, courses that prepared me for this task. Finally, special thanks to Thena Morris and Susie Billings for helping me above and beyond their responsibilities.

I also extend my gratitude to the Texas Coastal Bend area community and hope that one day this research can contribute to their development, as well as the development of all coastal communities around the world. Equally, I wish to thank the entire Brazos County community that is home now. I'm especially appreciative to Blue Baker for making such great breads, pizzas, and pecan pies along this journey.

Thanks also go to my friend Luis Marcelo Zeri for always being willing to discuss ideas with me. I also extend my gratitude to the entire R project for statistical computing community, which facilitates many analytical investigations and made me a coding and reproducible research enthusiast, intensely employing open-source software and data whenever possible. In special, thanks to Edzer Pebesma, Robin Lovelace, and Marco Scutari for being open and for their willingness to share knowledge and hope with the world.

Thanks to all my friends and family, with all my heart. I especially thank Lorena and Marcel for always being there and taking care of us with love; the *pinguiponeros*, for the good times; and my mom Maria, my sisters Leila and Daniela, my brothers Allan, Albert, and Rafael, and my aunts and uncles—Margarida and Carlos Paranhos and Heloisa and Oscar Pion— for their support and cheer along the way. I owe my dad and granddad, Teofilo and Oscar, who have passed, a deep debt of gratitude; though only in memory, they still remain present. I am also grateful to my grandma, Mercedes, who was the first person of so many who patiently taught me with affection. Last, to my wife Gabriela and to our son, Nicholas Bento, thank you for always believing in me and sharing your infinite joy, kindness, hope, patience, and love.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Dr. Walter Gillis Peacock and Dr. Douglas Frank Wunneburger of the Department of Landscape Architecture and Urban Planning, Dr. Zenon Medina-Cetina of the Department of Civil & Environmental Engineering, and Dr. David Henry Bierling of the Texas A&M Transportation Institute. All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by graduate assistantships and funds from (a) the United States Army Corps of Engineers Coastal Bend Hurricane Evacuation Study: Evacuation Zone Development, Vulnerability and Transportation Analysis, (b) the National Science Foundation RAPID: Risk Area Residents' Response to Hurricane Harvey Award (CMMI #1760766), (c) the Sandy and Bryan Mitchell Master Builder Endowed Chair at Texas A&M University, and (d) resources from the Hazards Reduction and Recovery Center and Department of Landscape Architecture and Urban Planning at Texas A&M University. Any opinions, findings, and conclusions or recommendations expressed in this dissertation are those of the authors and do not necessarily reflect the views of the funding sources.

NOMENCLATURE

AME	Average Marginal Effect
AUC	Area Under the Curve
BIC	Bayesian Information Criterion
BNs	Bayesian Networks
CBA	Texas Coastal Bend Area
CPT	Conditional Probability Table
DAG	Directed Acyclic Graph
EV	Expected Value
GIS	Geographical Information Systems
HHEBS	Hurricane Harvey Evacuation Behavior Survey
LR	Logistic Regression
NS	Network Scores
OR	Odds Ratio
PADM	Protective Action Decision Model
PDF	Probability Density Function
ROC	Receiver Operating Characteristic
SD	Standard Deviation
SE	Standard Error

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

1. INTRODUCTION

A broad increase in the frequency of natural hazards and damages has occurred in the United States. The United States accounted for over one-third (38%) of global economic losses caused by weather, climate, and water hazards; in particular, storms are associated with the greatest loss of life (71%) and economic losses (78%) in the region (National Oceanic and Atmospheric Administration, 2021; World Meteorological Organization, 2021). Moreover, the record of Atlantic tropical storms or hurricanes (1878 to present) is on an upward trend, and modeling studies show that the proportion of storms that reach very intense levels should increase in frequency and destructive potential per storm over the 21st century based on warmer ocean water temperatures and higher sea levels (Knutson, 2021).

When striking coastal communities, hurricanes (or even less intense storms) typically leave paths of destruction (Keim et al., 2007), thereby implying that the storm strength was the main cause of damage, economic loss, and loss of lives (Pielke & Landsea, 1998; Rappaport, 2014). However, in general, beyond the overwhelming forces of nature, disasters are also a function of human failures, poor decisions about where and how communities are developed, social and structural vulnerabilities, and/or inadequacy in preparation and mitigation (Cutter et al., 2003; Masterson et al., 2014; Strobl, 2011; Yoon, 2012).

My goal when conceptualizing this dissertation was twofold: first, to understand how communities can better understand and assess hurricane risks by using Bayesian networks (BNs); and second, to evaluate whether BNs, a probabilistic graphical modeling method, is appropriate for such a task. To achieve these goals, I was interested in identifying and assessing hurricane-related risks and risk countermeasures in a way that enables the understanding of this process to be available to anyone, even when knowledge in statistical methods is limited, or resources and data are limited or not perfectly comprehensible, or when it is not possible to fully recognize the influence on impacts of all risk factors working individually.

This dissertation comprises three studies. The first study explores the use of BNs for disaster planning, demonstrating its potential and some of the advantages for the risk assessment of natural hazards. The study explores the intuitive format of BNs, which can facilitate the cooperation of specialists from several disciplines and provide great potential for enabling the engagement of citizens, policymakers, decision-makers, and other stakeholders to better understand and quantify their local risks, and therefore foster participatory planning on communities.

The second study examines the use of BNs to model and predict household hurricane evacuation. Although most household hurricane evacuation studies use logistic regression (LR), research has shown that BNs can be a valuable tool for modeling complex decision problems. This research uses data collected in a survey after Hurricane Harvey. The comparison between BNs with traditional approaches uses only two of the main and most recurrent reasons that directly affect household evacuation, according to research to date: (a) receiving an official warning, and (b) expecting personal and household impacts. The results show that BNs can represent conceptual models more explicitly than traditional methods, offering a graphical representation of a model that has a solid statistical basis, and can, again, facilitate uptake by researchers, communities, and disaster planning practitioners.

The third study examines hurricane household evacuation using BNs. To develop the analysis, I explored a conceptual model recognized from the literature and a large set of variables from the same survey on households' choices after Hurricane Harvey. The results list factors that influence evacuation and indicate that BN can isomorphically model complexities of conceptual models without major statistical complications, thus demonstrating a potential to be used more frequently in future disaster preparedness and planning.

The three papers indicate that BNs are a suitable tool to study disaster planning problems in special hurricane risks and evacuations. This research reveals that although the use of graphs for modeling is not necessarily new to research and applications, a BN offers a friendly approach for modeling the complex relationships, and the calculations involved are relatively easy to estimate and interpret. This research as a whole opens the possibility for future inclusion and tests of a more diverse set of variables and domains in the study of hurricane risks and risk countermeasures, such as evacuation.

2. RISK ASSESSMENT THROUGH BAYESIAN NETWORKS: HELPING COMMUNITIES TO BECOME AWARE OF HURRICANES' MULTIPLE HAZARDS

2.1. Abstract

Although Bayesian networks (BNs) are integrated modeling tools that handle complex analysis with various interdependencies caused by common influencing variables—which are typical for risk assessment involving natural hazards—their application in the context of urban planning is still limited. This study explores the use of BNs for disaster planning, demonstrating its potential and some of the advantages for the risk assessment of natural hazards. For this purpose, a framework for risk assessment is presented and a brief introduction to BNs is provided. The methodology is applied in two illustrative examples as an exploratory tool to estimate economic and social costs when assessing hurricane risks in a typical single-family home in a coastal community. These examples explore the intuitive format of BNs and show how they can be used to model and quantify risks associated to natural hazards. The results of this research indicate that BNs can facilitate the cooperation of specialists from several disciplines and provide great potential for enabling the engagement of citizens, policymakers, decision-makers, and other stakeholders to better understand their local risks and therefore foster participatory planning on communities. Future research can expand this approach in spatial and temporal scales and verify the application in pilot projects developed together with the communities.

2.2. Introduction

Damage and loss mechanisms related to natural hazards are complex and typically involve various influencing variables, which make them difficult to estimate for specific locations and purposes (Meyer et al., 2013; Wright et al., 2012). Because of those issues and the continued exposure to natural hazards, notably hurricanes and hazards triggered by them, a large part of coastal communities is in a critical and recurrent disaster condition. Several researchers have dedicated their work to understanding how these communities can better plan and prepare accordingly (e.g., Berke, 1998; Burby et al., 1999; Dickson et al., 2012; Quay, 2010).

However, in reality, natural hazards and risk cannot be eliminated, and (financial) resources for the protection of communities and individuals are limited (Burby et al., 1999; Morrow, 1999). Suitable tools are needed for an integrated and participatory community level risk assessment, which can become an important first step for local and regional policymakers, citizens, and stakeholders to better understand their community risk profiles and stimulate the development of strategic and disaster planning, including hazard mitigation and climate adaptation policies that can reduce impacts and enhance resilience (Comfort et al., 1999; Godschalk et al., 1989; Tollefson, 2012).

The development of risk assessment may be facilitated by a theoretical framework, especially when it assimilates interdisciplinary challenges and integrated solutions (Gardoni et al., 2016). In line with standards such as the ISO 31000:2018 (International Organization for Standardization, 2021) and most risk analysis textbooks

(Aven, 2012; Smith, 2013), a risk assessment framework and tool should have the following properties:

- It should include entire systems and networks with dependent elements.
- It should allow for combining different models and data.
- It should be applicable to different types of hazards.
- It should be easy to understand and communicate.

Although few studies have sought to develop a probabilistic approach to risk assessment that is consistent under these premises, Bayesian networks (BNs), also called belief networks or probabilistic causal networks, have been used as a suitable tool to handle complex analysis by integrating interdependencies caused by common influencing variables and allowing the update of joint states of information in several circumstances and fields (Bayraktarli et al., 2005; Medina-Cetina & Nadim, 2008; Pourret et al., 2008; Sperotto et al., 2017; Straub, 2005; Uusitalo, 2007). However, despite this advantageous usage, the application of BNs in the context of urban planning is still a limited field of exploration. For example, the *Journal of the American Planning Association* currently has no study that explores the application of BNs in urban planning and related fields.

This study tries to address this gap in knowledge and aims to demonstrate how BNs can be applied to help communities involved in the process of modeling natural hazard risks, even when data are limited and/or heterogeneous. A Bayesian formulation for risk assessments in planning can allow for a systematic and logical update of joint states of information on multiple spatial and temporal scales while integrating both quantitative and qualitative data and knowledge, thereby relating multiple hazards and vulnerabilities, which is the very nature of natural hazards, particularly hurricanes. Most importantly, the visual nature of BNs and the design features associated with their development have the potential to link elements of probability theory with relatively easy interpretation and modification by using graphical models. This aspect of BNs can facilitate the cooperation of experts from several disciplines and provides great potential to enable the engagement of citizens, policymakers, decision-makers, and stakeholders to codevelop models that can provide (a) a better understanding of local risks from natural hazards and (b) costs and benefits of alternative mitigation and adaptation approaches, both of which are expected to foster participatory planning in communities.

This study starts with the presentation of a general framework for risk assessment in an urban environment. It then introduces BNs and outlines some of the advantages of utilizing that approach for modeling natural hazard risks. The use of BNs is illustrated by two examples. Both examples explore the intuitive format of BNs to identify, model, and assess risks. To demonstrate the usefulness of BN, only belief probabilities and estimates are employed in the examples. This process can be useful for several communities, especially for communities most vulnerable to natural disasters that have limited surveyed data on the exposures and vulnerabilities. The proposed approach can help communities generate risk models and estimate the state of risk, and outcomes can be used to guide future data collection, which will serve as inputs for a more accurate estimate of risk.

The first example applies BNs to exploratively estimate the expected economic loss of a typical single-family home in a coastal community, given the probability of wind intensity related to hurricanes and the vulnerability of the building (established by its fragility based on a typology). The first example is then extended to demonstrate the usefulness and flexibility of BNs by including multiple hazards and impacts in the same risk assessment arrangement. The probability of flood due to hurricanes and the estimated number of displaced residents are integrated into the model, even while it still considers hurricane-related winds and the expected loss of structural damage.

2.3. A General Framework for Risk Assessment of Natural Hazards

Figure 1 presents a general theoretical framework for the risk assessment of natural hazards (adapted from Faizian et al. (2005) and Straub (2005)) to help structure the risk assessment of natural hazards on communities. It provides an overview on the involved critical processes and facilitates a rational and consistent approach that can be implemented using BNs. The components of the framework comprise multidisciplinary participation from natural scientists, engineers, social scientists, economists, other experts, and input from local knowledge and tries to address an integrated solution by means of planning and policy for the risk assessment (and management) problem.

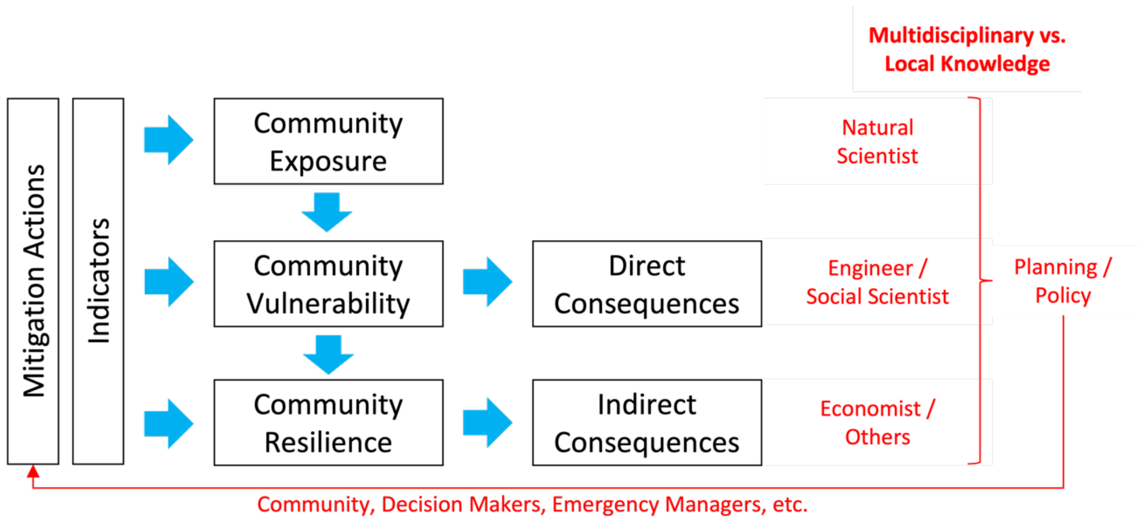


Figure 1. A framework for the risk assessment of natural hazards.

In the framework, three core components are distinguished, namely *community exposure*, *community vulnerability*, and *community resilience*, which can lead notably to *direct consequences* or *indirect consequences*. These components can be described by means of scientific knowledge (i.e., models, either physical or empirical) or by indicators that represent the available information (local knowledge or historical distributions) for specific cases. Although not directly part of the risk assessment (but part of the risk management), *mitigation actions*—which are achieved by the community, decision-makers, emergency managers, and other stakeholders—can also be considered in terms of potential measures influencing risk. Each of these framework components is next examined in greater detail.

Community exposure can be considered an indicator of the hazard potential for a given element or system of the community and estimated as the probability (P) that a particular threat (T) with a given intensity is exceeded within a given time and space. For hurricanes, the exposure is an inherently uncertain phenomenon with probabilistic characteristics usually provided in terms of threat intensities (hurricane category) and corresponding return periods (Emanuel & Jagger, 2010; Vickery & Twisdale, 1995).

Community vulnerability can be considered an indicator of the immediate consequences (C) (to an element or system of the community) associated with a given exposure event, and it can be assessed through the probability $P(C|T)$. In the event of a hurricane approaching a community, the vulnerability is associated with significant uncertainty and is appropriately described by a probability distribution of different damage states conditional on the exposure event—for example, the hurricane intensity, duration, source-to-site distance, and so forth.

A direct consequence can be described as the possible estimated loss associate to each different and mutually exclusive damage state ($u(C)$). Due to the stochastic nature of exposure and vulnerability, risk as an expected value of loss can be described in a probabilistic and quantifiable form based on certain components of the framework. As a generalization of the weighted mean, the equation

$$R = EV[C_D] = P(T) \times P(C|T) \times u(C) \tag{1}$$

can define the assessment of the state of risk (R) by estimating the expected value (EV) of a direct consequence (C_D) in terms of impacts as local social losses (including loss of lives as well as economical losses) of specific threats at a given location and time. By considering changes in these components, communities can understand how to better manage risk relating to specific impacts (Dickson et al., 2012; Medina-Cetina & Nadim, 2008).

Community resilience is an indicator of an indirect consequence (C_{ID}) due to damage or loss in a system or element of the community. It can be assessed through the complement of the probability $P(C_{ID}|T,C)$ and thus associated with the conditional probability of losses of various degrees—conditional on the exposure and a given damage state and also dependent on the location and time the event takes place.

Mitigation actions are specific actions, projects, activities, or processes taken to reduce or eliminate long-term risk to people and community elements from specific community exposure and predicted impacts. There are examples of mitigation actions for hurricanes at community and household levels (e.g., the creation of hurricane evacuation zones, the determination whether to evacuate, and the strengthening of roof and house structures; Godschalk et al., 2000; Sadowski & Sutter, 2008).

As will be outlined in detail in the following section, BNs can be applied to establish a general model of the causal relations between the hazard event itself and the possible consequences. This risk assessment model can also involve several observable characteristics that comprise so-called indicators. Only retrofitting (Jasour et al., 2018; Stewart et al., 2003) is considered in the following examples as a possible risk-reducing measure, and the risk associated with that measure will be compared to the risk associated with doing nothing, and the comparison can then constitute the basis for the decision-making.

2.4. Bayesian Networks for Modeling Natural Hazard Risks

Studies show that BNs can be used as tools to describe and assess natural hazards and to quantify related risks (Bayraktarli et al., 2005; Faizian et al., 2005; Straub, 2005; Uusitalo, 2007). In recent years, the potential of BNs for risk assessments and as a decision support tool for urban systems has increased interest among some in their practical application. For example, Medina-Cetina and Nadim (2008) and Blaser et al. (2011) applied BNs to the stochastic design of early warning systems for natural threats such as landslides and earthquake-triggered tsunamis. Balbi et al. (2016) used BNs to assess flood risk to people by integrating people's vulnerability and ability to cushion hazards through coping and adapting.

This section presents a brief introduction to BNs and the rationale of using BNs for risk assessment. A concise overview on BNs is provided in Pearl (2011). More extensive publications on BNs include Pearl (1995) and Jensen and Nielsen (2007). In addition, many software packages, both commercial and freeware, are available for the computation of BNs, as discussed in Scutari and Denis (2014).

BNs are probabilistic models based on directed acyclic graphs (DAGs) that help the representation of priori assumptions about the relationships between and among variables in causal structures. Figure 2 illustrates a simple BN that consists of three variables (X_1, X_2, X_3). The variable X_1 is a parent of X_2 and X_3, which are children of the former. Note that the common influencing variable X_1 introduces a probabilistic dependency between X_2 and X_3. This feature can be a typical situation in natural hazards modeling. For example, X_1 can represent the hurricane wind intensity, and X_2 and X_3 are damage conditional on the wind intensity on buildings at two different locations.

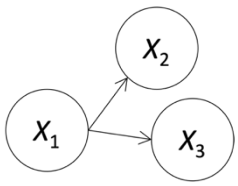


Figure 2. A simple Bayesian network of three nodes.

The BN model describes the joint probability distribution P(X) of a set of variables $X = X_1 \ldots X_n$. The size of P(X) increases exponentially with the number of variables (n), but BNs can allow efficient modeling by factoring the joint probability distributions into conditional (local) distributions for each variable given its parents. The joint probability distribution for any BN can be described by $P(X) = P(X_1 \ldots X_n) = \Pi_{i = 1 \ldots n} P(X_i|p_i)$, where p_i is a set of values for the parents of X_i. The joint probability distribution of the BN illustrated in Figure 2 is defined as $P(X_1,X_2,X_3) = P(X_1)\ P(X_2|X_1)\ P(X_3|X_1)$.

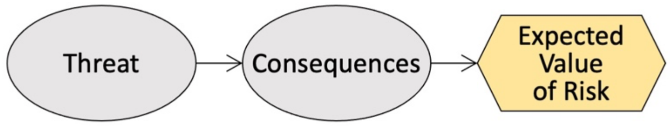
BNs can facilitate the inference of such a probabilistic model using efficient calculation algorithms (see, e.g., Jensen and Nielsen [2007] and Scutari [2010]). For computational reasons, in general, BNs are restricted to variables with discrete states. Therefore, for most applications, all random variables must be discretized in mutually exclusive and exhaustive states. In some applications, continuous variables can be used for Gaussian BNs or hybrid BNs (Scutari & Denis, 2014).

BNs allow entering evidence; that is, probabilities in the network are updated when new information is available for any of the variables. For example, when the state of X_2 (in the graph in Figure 2) is observed to be e, this information will propagate through the network, and the joint prior probabilities of X_1 and X_3 will change to the joint posterior probabilities, or $P(X_1, X_3|e) = \frac{P(X_1,e,X_3)}{P(e)}$. Consequently, the marginal posterior probabilities of X_1 and X_3 are also updated.

An interesting aspect of BNs is that they can be extended to decision graphs by including decision nodes and utility nodes in the network (Jensen & Nielsen, 2007). This enables the assessment (and the optimization of possible actions) in the framework of decision theory. Such decision graphs can describe a concise representation of decision trees commonly applied for the optimization in the framework of Bayesian decision theory. The optimal action decision on an action is the one yielding the maximal expected utility. If no actions are considered, the expected utility represents a measure of the total risk (the expected value of risk; Dyckman, 1961; Straub, 2005).

2.4.1. Rationale of Bayesian Networks for Risk Assessment

Figure 3 shows a simple BN model that represents the main components of the risk assessment framework and directly expresses risk (Equation 1). The model describes the expected value of risk (EV[R]) representing (a) hazard as the probability that a threat T with a given intensity is exceeded within a given time and space, and (b) consequences as the multiplication of the vulnerability (probability of a consequence given a threat intensity) times a set of values of consequence certain to happen for each vulnerability state (Medina-Cetina & Nadim, 2008).



Risk = Hazard x Vulnerability x Consequence

$EV[R] = P(T) \times P(C|T) \times u(C)$

Figure 3. BN representing the key components of the risk assessment framework.

Two illustrative examples are presented next. The first example applies BNs to estimate the expected economic cost of a typical single-family home in a coastal community given the probability of intense hurricane-related winds and the vulnerability of the building established by the fragility likelihood based on a typology. The second example extends the BN by integrating multiple hazards and impacts into a single model topology and includes (into the model) the probability of flood related to a hurricane and the estimated number of displaced residents.

2.4.2. First Illustrative Example of Application

To demonstrate the use of BNs in risk assessment of natural hazards, this example considers a simplified situation of a single-family home located in a typical coastal community on the U.S. Gulf of Mexico or Atlantic coast. Initially, the only information available is that this structure is exposed to an eventual hurricane of still unknown category. It has been hypothesized that the probability of structural damage of the building is conditional to the hurricane wind intensity and that the value of building damage is a result of the structural damage. Without loss of generalization, such an example can be adjusted to capture risk for multiple structures by changing the utility value and the building damage probabilities accordingly.

Figure 4 presents a BN for the risk assessment using an illustrative network. Different sources of knowledge can be used for the development of such network topology, such as modeled data, documented literature, experts' opinion, and local experience. Herein, the associated causal relationship between the probability of different hurricane wind intensities (hazard) and the building vulnerability associated with a given wind intensity (vulnerability) seeks to capture risk in terms of an expected cost of damage (consequence). Once the topology of the network is outlined, the next step is to define the conditional probability distributions associated to each node.

Figure 4. BN for the risk assessment of the hurricane wind on home structures.

The input data for each variable or combination of any two variables are defined in a form of a probability density function (PDF) or a conditional probability distribution, respectively. Applicable to a given spatial and time domain, these probabilities can be obtained from local or expert knowledge, historical distributions (or return periods), or performance-based fragility functions of modeled data or empirical observations. Data and probabilities can be updated with evidence or deliberate changes in conditional probabilities and probabilities. Figure 5 exemplifies a probability table taken from a wind speed frequency distribution. In this example, for illustrative purpose and easy interpretation, the probability states for the wind speed node are defined from a generic wind speed distribution considering the availability of evidence in a few states: low, moderate, and high.

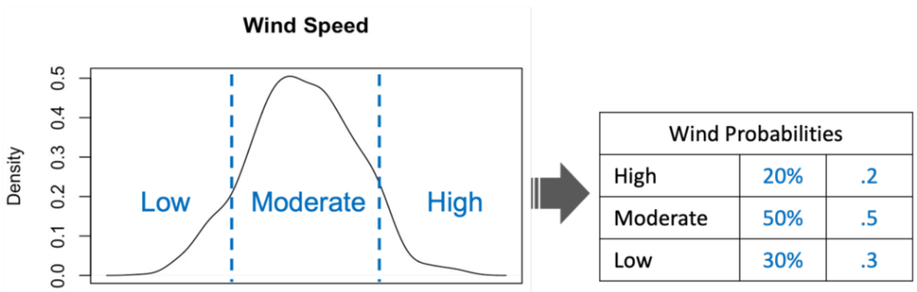


Figure 5. Probability density and probability table for an illustrative hurricane-related wind intensity.

With values conditional to the wind status, the building vulnerability node defines the probability states of the building damage in a few states as well: none, low, moderate, and high. Probabilistic values for the combination of wind state and building damage state can be observed in the vulnerability table in Figure 6. For example, if the wind speed is high, there is 20% chance of no damage, 25% chance of low damage, 30% chance of moderate damage, and 25% chance of high damage. If considering a lower wind intensity, the chances of high or moderate damage are reduced. Those values can be assessed specifically for the buildings of a community or generally obtained from the literature (e.g., Kopp et al., 2012; Masoomi et al., 2018; Pinelli et al., 2004; van de Lindt & Dao, 2009).

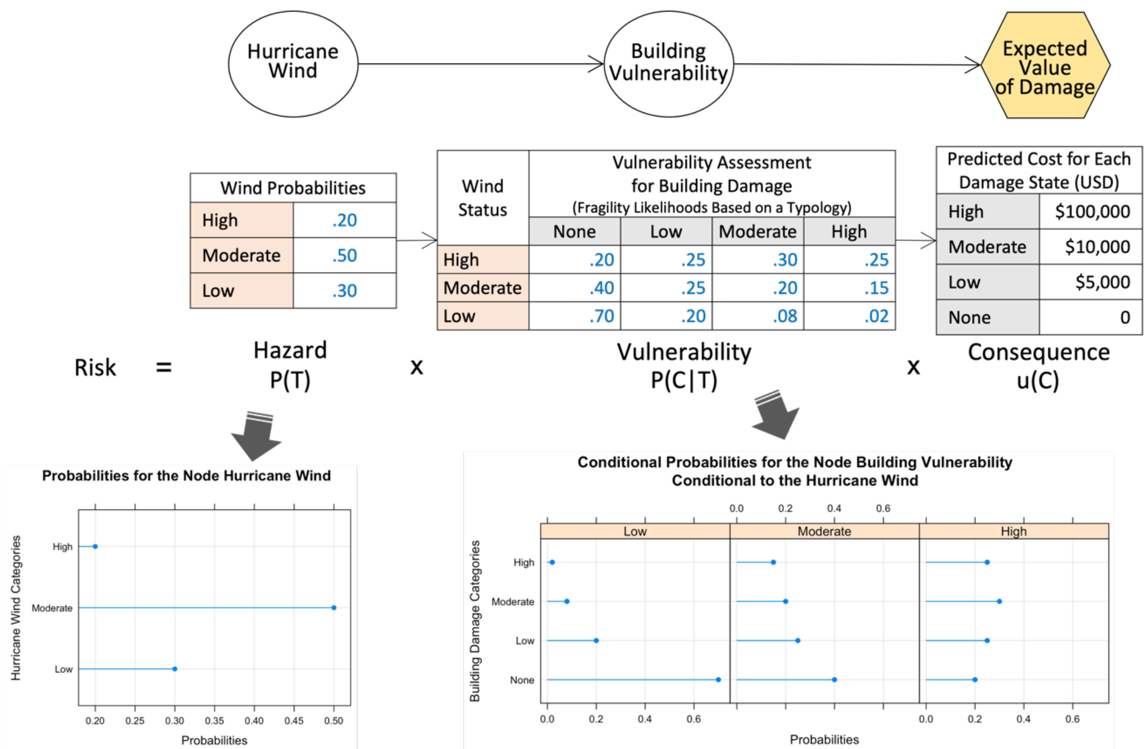


Figure 6. Probabilities and values for the nodes of the BN.

Figure 6 also includes charts to illustrate the distribution of values of the PDF and corresponding conditional probability tables (CPTs) as described above, and a utility table for the consequences node that shows the predicted cost assigned for each building damage state. For example, no cost is projected if there is no damage to the building structure, but the predicted cost is $100,000 for a level of high damage.

Figure 7 exemplifies the computation of this simple BN model. The matrix multiplication calculates the probability of each structural damage conditional to the wind state. The expected value of risk is given by multiplying the probability of each state of damage (red circle) by the respective loss value it represents. In this example, the expected value of damage for the building structure is $16,115. Figure 7 also includes the visualization of the marginal probabilities in the form of the network probabilities.

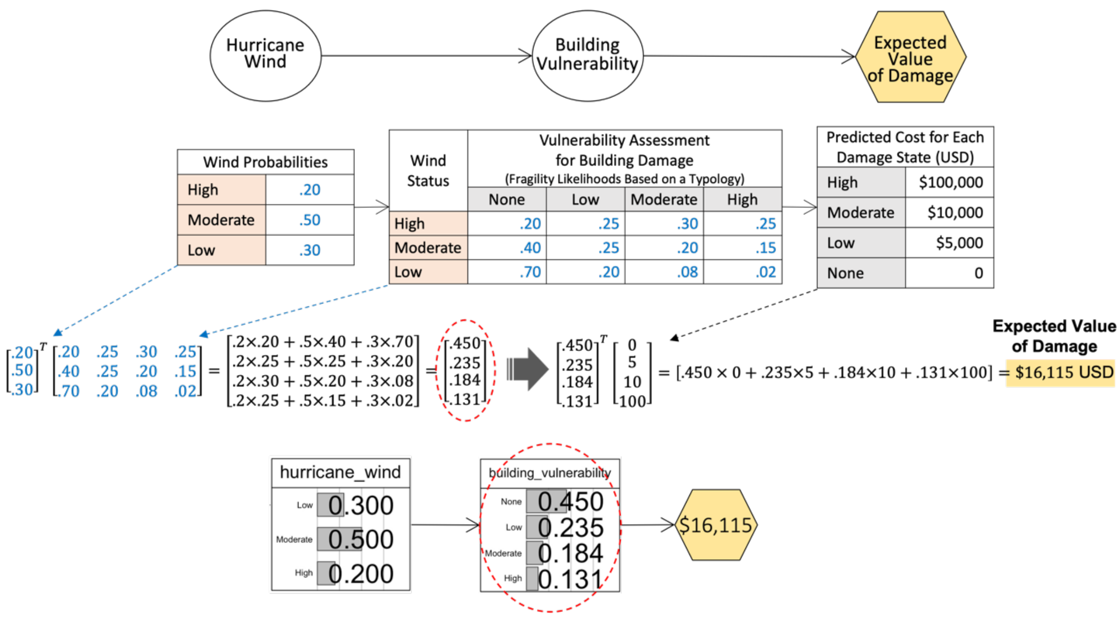


Figure 7. Illustrative computation of the simple BN.

2.4.3. Second Illustrative Example of Application

Typical housing damage models (e.g., Highfield et al., 2014; Zhang & Peacock, 2009) consider variables in systematic domains. These domains are usually (a) location and hazard exposure, (b) demographic/socioeconomic characteristics, (c) mitigation status, and (d) impacts on housing. Integrating these domains in a BN model can assist in the visualization and identification of how additional information may be efficiently used to reduce the risks. Figure 8 presents a graphical representation of a risk assessment model departing from these key domains and including direct economic cost and social cost as a measure of risk. This specific topology is based on hypothesized relationships between and among these domains.

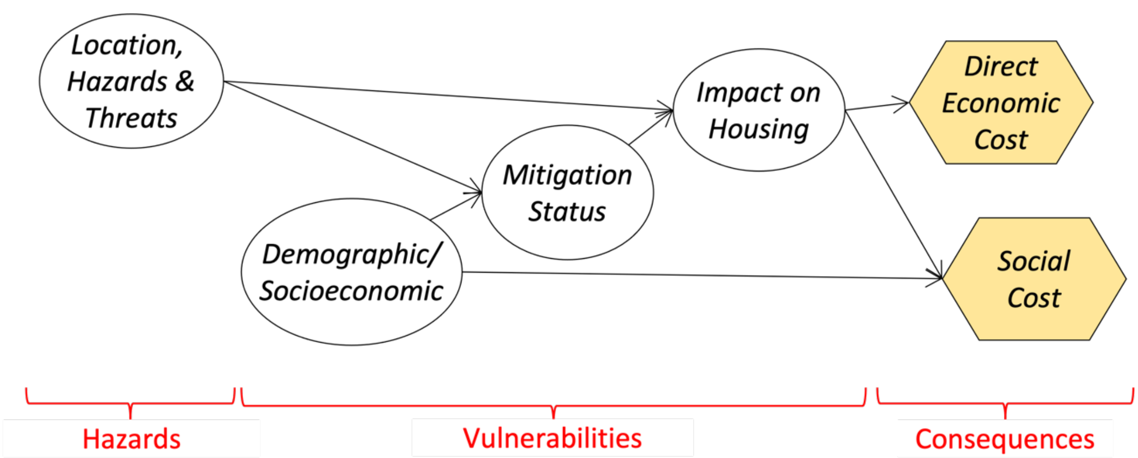


Figure 8. Risk assessment model using systematic domains of a typical housing damage model.

The model assumes that mitigation status is conditional to the location and hazard status and the demographic and socioeconomic characteristics, and impacts on housing are conditional to the location and hazard status and mitigation status. The model also assumes that social costs can be associated with physical impacts (impacts on housing) and demographic and socioeconomic characteristics.

After defining a BN model topology that represents a conceptual model of housing damage, variables can be elected (by researchers, experts, members of the community, and other stakeholders) to capture these domains and create new networks that synthesize such an arrangement. Figure 9 presents a simple risk assessment model with three common variables, in essence adding hurricane-related flood to the previous model (Figure 4). The model assumes that the building vulnerability is conditional to the wind intensity and flooding related to hurricanes, and the expected damage cost is a consequence that associates the hazard exposure with the vulnerability status.

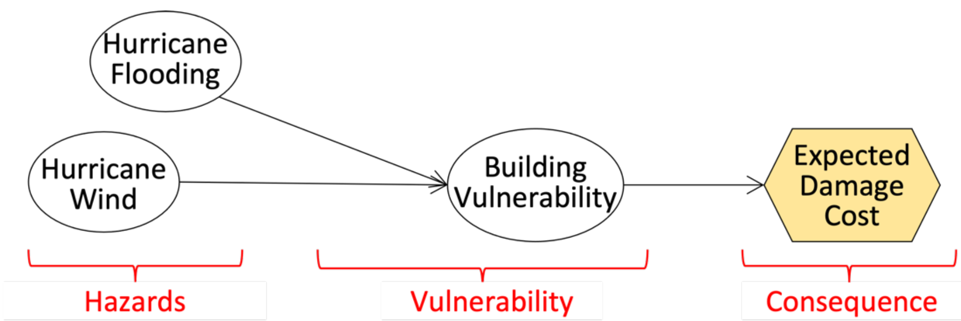


Figure 9. A basic risk assessment model.

Still following the conceptual model, domains can be further extended. Typical variables such as housing tenure status or describing owner or rental occupancy can characterize a sociodemographic condition and have probabilistic effects on the fragility of the structure, thereby influencing its maintenance and protection status (Peacock et al., 2014). The housing tenure status can also have a direct influence on the estimate of displaced residents (Lee & Van Zandt, 2019).

Figure 10 presents an extended arrangement of the former model (Figure 9), including tenure status as a possible probabilistic indicator of building vulnerability and tenure status and building vulnerability as a possible probabilistic indicator of the expected number of displaced residents (as a social measure of risk). Although still a simple model with few variables, this BN exemplifies multiple hazards and impacts in the same network arrangement.

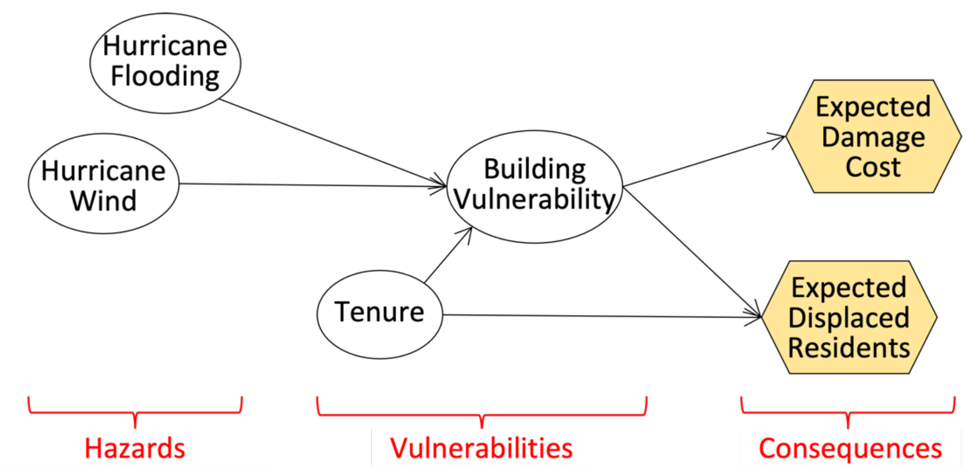


Figure 10. An extended risk assessment model.

A difficulty of having multiple variables with multiple levels conditioning a node is finding out probabilities for all the possible combinations. When obtaining data from a survey using a large number of observations, the various combinations of levels will facilitate this task. However, when setting up an exploratory network, belief probabilities can be difficult to infer, and this task can be simplified by reducing the possible levels for each variable (because it reduces the number of possible combinations).

Figure 11 presents a contingency table with a set of belief probabilities for each combination of building damage, tenure status, hurricane flooding, and hurricane wind levels, as defined by the researcher. As already mentioned, such values can be obtained in a variety of ways, including by expert opinion, scientific models, and local knowledge. When creating a BN with communities, local knowledge comes from the input and iteration of the parties involved with the risk assessment.

Figure 11 also provides a graphical visualization of such a combination of probabilities for a presumed easier interpretation. These data illustrate rental structures with greater probability of structural damage for the same combinations of threat intensities due to lesser mitigation adoption (e.g., Ge et al., 2011). Changing the probabilities in this table through various risk mitigation and preventive actions (e.g., retrofitting of different classes of structures in accordance with specific exposure) will directly impact the expected value of risk, which can be assessed and reexamined with the associated baseline risk of no intervention.

Tenure	Hurricane Flooding	Hurricane Wind	Vulnerability Assessment for Building Damage (Likelihoods Based on a Typology)			
			None	Major	Minor	None
Owner	None	Low	.99	.01	.00	.00
		Moderate	.90	.08	.02	.00
		High	.75	.10	.10	.05
	Low	Low	.80	.10	.08	.02
		Moderate	.70	.15	.10	.05
		High	.55	.20	.15	.10
	High	Low	.50	.10	.30	.10
		Moderate	.40	.12	.33	.15
		High	.30	.15	.35	.20
Renter	None	Low	.94	.05	.01	.00
		Moderate	.80	.12	.05	.03
		High	.50	.20	.20	.10
	Low	Low	.80	.10	.05	.05
		Moderate	.60	.15	.15	.10
		High	.50	.20	.20	.10
	High	Low	.20	.20	.30	.30
		Moderate	.15	.15	.32	.38
		High	.05	.20	.35	.40

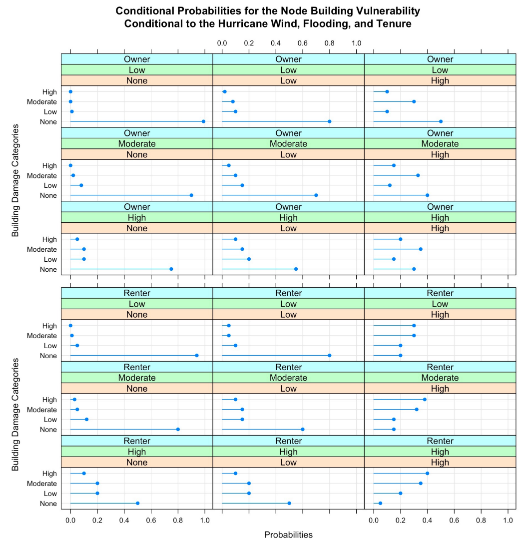


Figure 11. Conditional probabilities for the building vulnerability.

Finally, Figure 12 presents the network probabilities and the values assumed for the utility nodes. Such values can also have multiple sources, such as surveys, census averages, or experts and community members. These values are confined to the spatial scope that the network represents, which can be a specific structure or any combination of multiple structures in a domain. By multiplying the probability of each updated state of structural damage by the respective assessed loss value it represents, the expected damage cost for a single-family home is determined to be $7,055, and the expected number of displaced residents per structure is 0.62.

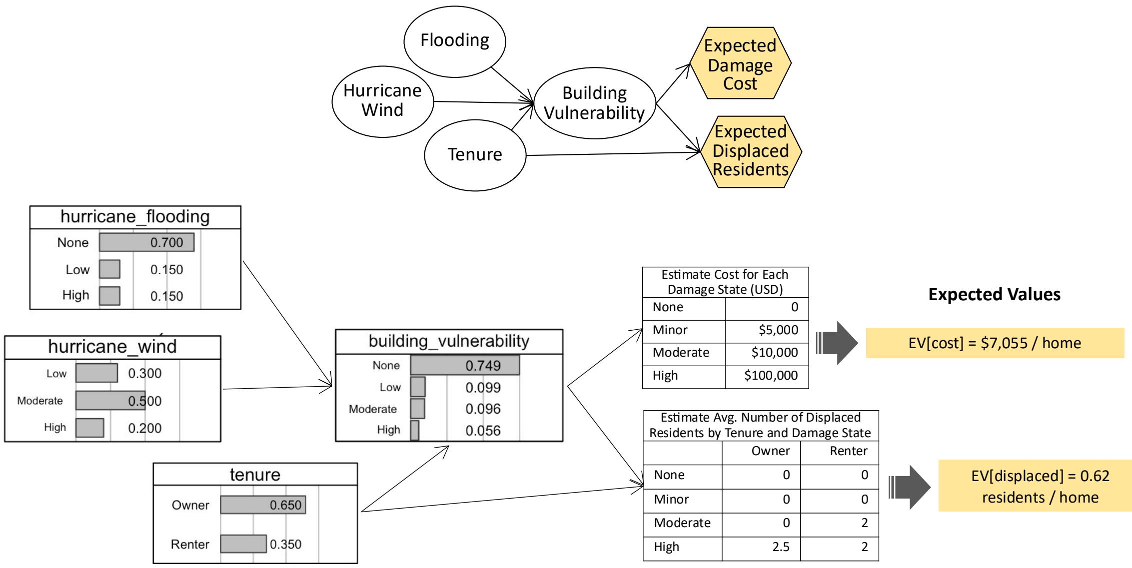


Figure 12. Network probabilities and values of the utility nodes.

Compared to the first illustrative example, the overall expected damage cost is lower ($16,115 versus $7,055) due to a detailed vulnerability assessment of the structures and the proportion of structures occupied by owners and renters. These values reflect the overall expected damage costs in opposition to an exclusive assessment for owners and renters, which could still be implemented in the same network arrangement.

2.5. Discussion and Conclusions

This study addresses the problem of risk assessment in coastal communities in the face of natural hazards, specifically hurricanes. To offer a consistent approach to the risk assessment problem using BNs, this study adapts a general framework from the literature that facilitates the structure of critical processes. Although this approach can serve as a guideline to specific management of risks due to natural hazards, this study focuses on the representation of a part of the system. The proposed framework is hierarchically structured based on how exposure to natural hazards, risk, and different sources of knowledge interact and have vulnerability as the link to the magnitude of direct consequences. BNs allow a detailed evaluation of the joint influence of the different indicators on the risk, providing results that, in contrast to traditional methodologies, are consistent with the mathematical (probabilistic) concept of risk and can be directly used for optimization purposes.

In the illustrative examples, BNs can represent some of the complexities of the urban environment, such as the combination of physical and social vulnerabilities, while predicting economic and social losses. The BN modeling ensures that the models can be further extended when additional (or complementary) information is included or examination by different stakeholders is performed, and a potential unavailability of indicators can be assessed by prior beliefs. For example, a city planner may want to include the age of the building in the fragility assessment, while emergency managers may want to estimate the number of displaced children. Even if these variables cannot be primarily assessed, prior probability distributions can be applied and included in the network. BNs are flexible enough to be extended to model an entire community system based on evidence of local and expert knowledge, and different levels of detailing can be integrated into a common model (e.g., a forecast model that considers different scales of time).

It is a challenge for planners and other professionals to provide methods and tools for communities to improve understanding and decision-making for the effective assessment and management of local natural hazard risks. In contrast to traditional approaches, BNs can be used for communication because they are graphically based and allow explicit documentation of assumptions and uncertainties, thereby facilitating the interaction of the various parties of the community and the identification of decisions of mitigation and preparedness for optimal cost-efficient improvements. When the large effects that can be associated with climate change are considered, it is obvious that more research is needed to support decision-making on how to cope with the increasing frequency and strength of hurricanes and other natural hazards, and the associated consequences for the coastal communities. In this scenario, BNs can provide an appropriate approach for probabilistically modeling these problems for communities so that the communities are better aware of and better prepared for natural hazards.

2.5.1. Limitations and Future Research

For BNs to go beyond an exploratory risk assessment, calibration and data support is necessary to better capture risk estimates—as is true of any mathematical modeling approach. Moreover, the design of the network can be supported by data and statistical tests, for example, using network scores (a measure of how well the model fits the data) and link strength (a measure of the probabilistic dependence corresponding to each link of the network). Accordingly, various conditional independence tests can be used to test for the existence of each individual link strength by removing that link from the network and quantifying the change with some probabilistic analysis (Scutari & Denis, 2014). To perform these tests, surveyed data are needed. The models proposed by the presented methodology can serve as a starting point to identify variables of interest.

The proposed approach is illustrated with examples that consider the assessment of hurricane risks on a house structure scale. Future studies can process data on aggregated scales and apply the same process through the support of geographical information systems (GISs). GIS tools can facilitate the management of relevant information for the assessment of risks in a specific area and a large range of assets.

3. A STUDY OF BAYESIAN NETWORKS TO MODEL HOUSEHOLD HURRICANE EVACUATION

3.1. Abstract

In many coastal communities, household hurricane evacuation is an important protective action taken by local authorities and residents to reduce risk. For such a critical problem, it is important to continuously review and analyze modeling tools. Although most household hurricane evacuation studies use logistic regression, research has shown that probabilistic graphical techniques such as Bayesian networks (BNs) can be a valuable tool for modeling these complex decision problems. The aim of this study is to introduce and examine the use of BNs to model and predict household hurricane evacuation. This study uses data collected in a survey after Hurricane Harvey passed through the Texas Coastal Bend area in August 2017. To facilitate the examination of BN and the comparison with traditional approaches, this study uses only two of the main and most recurrent reasons that directly affect household evacuation, according to research to date: (a) receiving an official warning, and (b) expecting personal and household impacts. The results show that BNs can represent conceptual models more explicitly than traditional methods by offering a graphical representation of a model that can facilitate the uptake by researchers, communities, and disaster planning practitioners while still having a solid statistical basis. This study indicates that BNs are a suitable tool to study disaster planning problems; moreover, future studies may include a greater number of predictors in the modeling of such problems.

3.2. Introduction

Although household evacuation is an important risk countermeasure in many coastal communities by which local authorities and residents can reduce exposure and prevent loss of life due to threats triggered by hurricanes or even less intense tropical storms, research to date indicates that receiving an official warning and expecting personal and household impacts are both the main and most recurrent *reasons* (i.e., predictors) for this protective action (Baker, 1991; Dash & Gladwin, 2007; Huang et al., 2016; Lindell & Perry, 2012).

Although most studies use logistic regression (LR) to examine a wide range of factors that affect evacuation (Yang et al., 2016), probabilistic graphical techniques such as Bayesian networks (BNs) may offer an alternative method to model the so-called "complexity involved in the household evacuation decision-making process" (Hasan et al., 2011, p. 341). Generally, the formulation of these problems involves a chain of selection of actions from a set of alternatives, each of which is evaluated against multiple and often conflicting criteria.

Indeed, BNs' modeling can take into account hierarchical and probabilistic interrelation of variables (Li et al., 2014), which also allows the learning of causal effects from observational data in which collecting experimental data is often not possible (Ramanan & Natarajan, 2020; Zheng & Pavlou, 2010). While these complex structures of variables can be used for descriptive analysis, they also have predictive capability, thereby allowing the analysis of scenarios in various planning activities (e.g., Cinar & Kayakutlu, 2010). However, the use of BNs to study responses to environmental hazards and disasters is still limited. Additionally, no research has been conducted using BNs to model household behaviors within the need to evacuate from hurricanes.

To better understand BNs for modeling household evacuation during hurricanes, this study presents a sequence of analyses and four experiments that aim to demonstrate the utility of this novel approach in the disaster planning field. The development of the analyses and experiments compares BN with traditional approaches and reveals the method's adequacy, effectiveness, and limitations. The data for the analyses were collected from a survey of households in the Texas Coastal Bend area, a large coastal region of Texas impacted by Hurricane Harvey in August 2017.

To facilitate the examination of BN and the comparisons with traditional approaches, this researcher uses only two of the main and most recurrent variables to predict household evacuation: (a) receiving an official evacuation warning, and (b) expecting personal and household impacts. Research has shown that both an official evacuation order and expecting personal and household impacts have a positive and significative correlation with the evacuation of households (Baker, 1991; Hasan et al., 2011; Huang et al., 2016; Tanim et al., 2022).

The next section presents a short introduction to BNs and the rationality to model hurricane household evacuation. Many other researchers provide an enhanced overview on BNs—for example, Pearl (1995, 2011) and Jensen and Nielsen (2007). In addition, many software packages, both commercial and freeware, are available for the computation of BNs, as discussed in Scutari and Denis (2014). The code of this research is implemented in R (R Core Team, 2021) and uses the following R packages: *bnlearn* (Scutari, Silander, & Ness, 2021) and *tidyverse* (Wickham, 2019). The code and data to reproduce all analyses are presented in the appendix.

3.2.1. Bayesian Networks to Model Decision Problems

A BN is a representation of a joint probability distribution of a set of random variables with a possible mutual causal relationship (Pearl, 2011), but also the association that can represent containment, ownership, part, requirements, or any other connection that has meaning within the context of the domain being modeled (Achumba et al., 2013). Generally, the method of modeling using BN can be described in the following three steps:

1. Setting up a model (i.e., the graphical representation of a model). In BNs, the network is more precisely a *directed acyclic graph* (DAG). In research, DAGs have been used to help choose which covariates to include in traditional statistical approaches, thus helping to minimize bias in estimates (Shrier & Platt, 2008). A DAG consists of nodes that represent random variables and of edges between pairs of nodes, which represent the relationship of nodes. A network can be built manually with knowledge of the underlying domain, consistent with knowledge about the underlying scientific problem and the data collection process (Ellis & Wong, 2008).
2. Estimate of probabilities and conditional probabilities. A variety of numerical methods are available to estimate marginal distributions (Bernardo, 1979). Typically, the marginal distributions are calculated by dividing the range for

the quantity of interest into a few discrete bins of equal width, conditional on any preceding variable proposed by the network structure, if applicable.

3. Evaluating the fit of the model and the implications of the joint probability distribution. The assumptions of the model (i.e., each one of the linkages) can be tested through conditional independence tests (such as mutual information) on each arc conditional to the network (Scutari & Denis, 2014). Two random variables are independent if the occurrence of one variable does not affect the probability of occurrence, and therefore, the probability distribution, of the other variable (Jensen & Nielsen, 2007). Besides the examination of each pair of nodes, network scores can be used to select alternative network structures that best fit the data (Scutari & Denis, 2014).

Importantly, BNs can be useful in a wide range of applications (Pourret et al., 2008). Research has shown that BNs are a valuable tool for modeling complex decision-making problems, including when considering individual's concern in choosing to define the importance of criteria according to the disposition of the information (Sedki et al., 2010).

Specifically, the potential of BNs to model decisions under uncertainty and as integrative decision support tools has increased interest among some for its practical applications. For example, Jager et al. (2018) developed a BN approach to integrate the separate models that support decision-making in the risk management of coastal areas in the United Kingdom. The researchers effectively applied BN to integrate the output from storm simulations with land use data, vulnerability relationships, and different levels of disaster risk reduction measures.

However, despite this favorable basis, the use of BNs to study household hurricane evacuation still represents a limited explored field. The author found no study that explores the application of BNs to investigate household responses to environmental hazards and disasters, and more specifically, no research has been conducted using BNs to model household behaviors within the need to evacuate from hurricanes.

3.3. Data and Methods

This study examines the use of BNs to model household hurricane evacuation based on data collected in a survey after Hurricane Harvey passed through the Texas Coastal Bend area in August 2017. To validate the application of BNs and create a baseline for comparison, this section briefly presents how the data were obtained and the descriptive statistics of the selected variables. Also, this section presents empirical cumulative distributions, Pearson’s correlation, and chi-squared tests among the selected variables, develops an LR, and implements a simple BN model that can directly be compared to the regression. For BNs, conditional independent tests and network scores are also explained. The following section develops four experiments that test the BN model in specific disaster planning circumstances, and then the results are discussed.

3.3.1. Hurricane Harvey and the Texas Coastal Bend Area

According to the National Hurricane Center, Harvey was the first Category 4 hurricane to hit the coast of Texas since Hurricane Carla in 1961 and the first major hurricane to hit the middle of the coast of Texas since Hurricane Celia in 1970. Hurricane Harvey first made landfall in the middle of Aransas County and then moved on to its second landfall near the State of Louisiana. Figure 13 shows a part of the

Hurricane Harvey Final Best Track (National Hurricane Center, 2022) over the Texas Coastal Bend area, which is a significant geographic region along the coast of Texas that lies exposed to hurricanes. The region consists of eight counties (Aransas, Calhoun, Jackson, Matagorda, Nueces, Refugio, San Patricio, and Victoria).

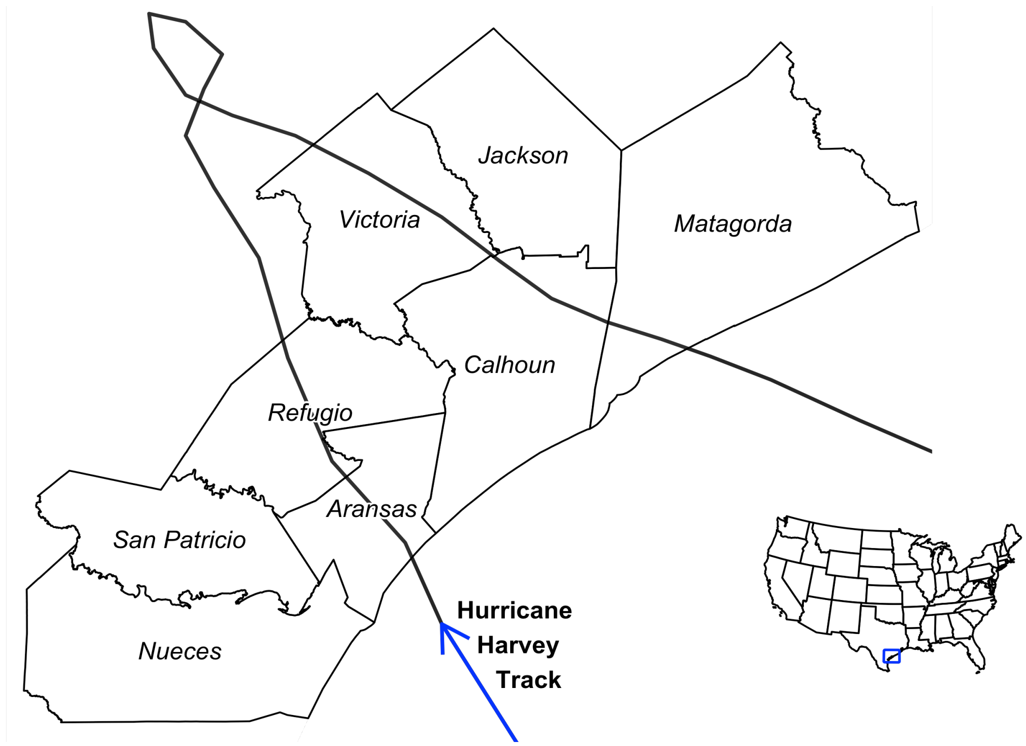


Figure 13. Texas Coastal Bend area and Hurricane Harvey track.

In 2019, to better understand household-level evacuation experiences, researchers from the Hazard Reduction & Recovery Center at Texas A&M University, the Texas A&M Transportation Institute, and the Institute for Hazard Mitigation Planning and Research at University of Washington conducted the Hurricane Harvey

Evacuation Behavior Survey (HHEBS). Households were randomly selected, and the survey distribution was administered in three mailout waves that were nonproportional random samples based on address sample frames. Wave 1 of the survey was online only and Waves 2 and 3 included paper copies of the survey instrument with postage-paid return (Bierling et al., 2020).

The HHBES behavioral survey dataset contained 958 observations. After excluding duplicated entries because of the multiple waves of the survey distribution and responses of nonresidents in the area at the time of Hurricane Harvey, 907 observations remained to be analyzed. Out of the 41 questions that comprised the survey instrument, six questions are used in this study, two of which are used directly, and four of which are used to compose a mean score that captures one of the model's predictors.

The next sections provide a concise description of each of the variables and the information provided by Table 1 and Table 2. The first table presents a summary of the variables, the descriptions, the proportion of the sample with data, and the categories of how the variables are initially coded. After a description of the selection of residents at the time of the hurricane (*evacuation* = 2) is given and a listwise deletion to handle the missing data is made, which is illustrated by Figure 14, the second table presents the variables and respective categories used in the analysis. After that, the number of observations in the dataset (*n*) drops to 826 from the 907 previously obtained.

3.3.1.1. Evacuation

The survey asked if the household evacuated from Hurricane Harvey. This question presents four possible outputs: (a) No, (b) Yes, (c) Not Resident, or (d) No

Answer. For analysis, only the observations that answered Yes and No are selected; No is coded as 0, and Yes is coded as 1. In the initial dataset, 2.5% of respondents were not residents at the time, and 0.5% did not answer the question. These samples were excluded from the dataset. For the analysis, 63% of the households evacuated from Harvey at some point in time, and 37% did not evacuate. This variable is labeled as *evacuation*.

3.3.1.2. Expected Impacts

Expected impacts can be considered a form of risk perception and assessed in several ways (e.g., wind damage, surge damage, flood damage, casualties, job disruption, and service disruption; Huang et al., 2016). In this study, expected impacts capture the expectations of personal and/or household impacts and are assessed by the mean score of four questions of the survey: how likely did the participant think, as the storm was approaching, (1) that they or household members would be injured or killed if they stayed? (2) that their home would be inundated by storm surge? (3) that their home would be exposed to inland flooding? or (4) that their home would be severely damaged or destroyed by storm wind? These questions were answered on a scale of 1 to 5, where 1 represented *not at all likely* and 5 represented *almost certain*. In the initial dataset, 4.7% of the participants did not respond to any of these questions. If only one or more of the questions were answered, it was computed as the average of the answers. This variable is coded as *expected_hh_impacts*.

3.3.1.3. Hurricane Evacuation Orders

Typically, in a hurricane threat, the official evacuation request starts as a call for voluntary evacuations and, at some point, is elevated to a mandatory evacuation order (McCausland & Chuck, 2017). A hurricane evacuation order can allow residents to identify, and possibly better understand, that the risk of a hurricane is imminent and thus lead the public to a life-saving decision-making process. One question asked the informants to what extent they considered the official recommendation of local authorities to evacuate when deciding whether to evacuate. Answers to this question were scaled from 1 to 5, in which 1 represented *not at all considered* and 5 represented *very great extent considered*. In the initial dataset, 7.4% of the observations did not respond to this question. This variable is coded as *signif_evac_orders*.

Table 1. Variables, Descriptions, Proportion of Sample with Data, and Coding.

Variables ($n_{\text{total}} = 907$)	**Description**	**Proportion of Sample with Data** (n_{missing})	**Coding** (levels/count/percent)
evacuation (1 question)	If household evacuated from Hurricane Harvey.	0.994 (5)	**No** = 0: 336 (37%) **Yes** = 1: 543 (60%) **Not resident at time** = 2: 23 (2.5%) **No answer**: 5
expected_hh_impacts (Mean score of 4 questions)	How likely the informant thought, as the storm was approaching, that they or household members would be injured or killed if they stayed, or that home would be inundated by storm surge or inland flooding, or severely damaged or destroyed by storm wind.	0.953 (43)	**Not at all likely** = 1: 171 (19%) 2: 405 (45%) 3: 146 (16%) 4: 108 (12%) **Almost certain** = 5: 34 (3.7%) **No answer**: 43 (4.7%)
signif_evac_orders (1 question)	To what extent was the recommendation of local authorities to evacuate considered when deciding whether to evacuate.	0.926 (67)	**Not at all considered** = 1: 141 (16%) 2: 90 (9.9%) 3: 129 (14%) 4: 159 (18%) **Very great extent** = 5: 321 (35%) **No answer**: 67 (7.4%)

Figure 14 presents the missing data in the survey. The missing data per variable are clustered, so in this way, one can approximately identify the frequency with which missing questions happen in the same observation. It can be noted that those household informants who did not answer the question about evacuating also did not answer the other questions, and only a small part of individuals who did not answer the question about expected personal and household impacts answered the other two questions. Figure 14 also includes the number of the missing data for each variable and the percentage they represent.

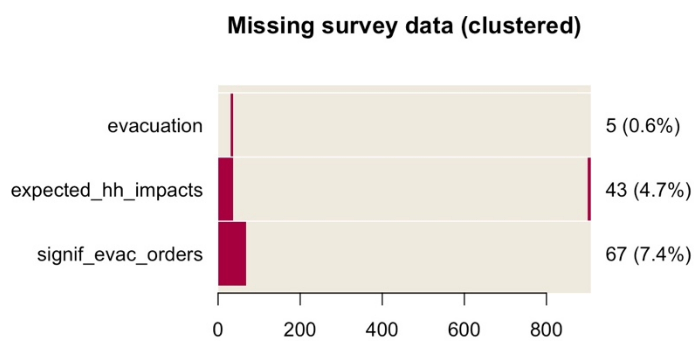


Figure 14. Missing survey data.

Table 2 provides the mean, standard deviation (SD), and counts for each of the levels and percentage for each of the variables. It can be observed that 63% of the participants evacuated from the hurricane. An average of 2.34 out of 5 people sampled expected personal and household impacts. The average rate of participants who considered local authorities' official recommendations to evacuate when deciding whether to evacuate was 3.53 out of 5.

Table 2. Variables' Mean, Standard Deviation (SD), and Coding for the Analysis.

Variables (n = 826)	**Mean (SD)**	**Coding** (count percent)
evacuation	0.63 (0.48)	0: 306 (37%) 1: 520 (63%)
expected_hh_impacts	2.34 (1.05)	1: 163 (20%) 2: 386 (47%) 3: 140 (17%) 4: 106 (13%) 5: 31 (3.8%)
signif_evac_orders	3.53 (1.49)	1: 134 (16%) 2: 89 (11%) 3: 127 (15%) 4: 158 (19%) 5: 318 (38%)

A general goal of this analysis is to study the relationship between the independent variables and evacuation, the dependent variable. For an initial visualization of how this relationship can happen, an empirical cumulative distribution of the two independent variables conditional on the evacuation response was made (see Figure 15). The cumulative percentage of responses is shown on the y-axis and each level of the variables on the x-axis.

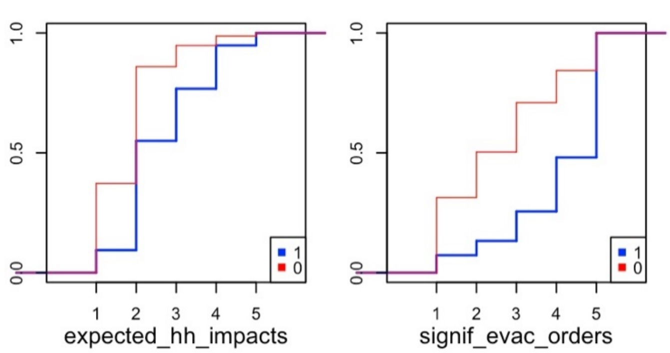


Figure 15. Empirical cumulative distribution function conditional to evacuation.

A possible indication of dependence can be observed for both variables since the distributions for evacuees and non-evacuees (1 and 0, respectively) seem to be different and do not overlap. The accumulated percentage for non-evacuees (0) increases faster at the lower levels of expecting impacts and considering evacuation orders.

3.3.2. Pearson Correlation Coefficient

Table 3 presents the Pearson correlation coefficients (*r*) to verify the linear relationship between evacuation and the two independent variables. The correlation between evacuation and considering local authorities' official recommendations to evacuate is 0.46, and between evacuation and expecting personal and household impacts is 0.37. Both values are significant, as can be seen in the test statistic (*z*) used to compute the *p*-value (*p*), the *p*-value itself, and the lower and upper bounds on the 95% confidence interval for the correlation values.

Table 3. Pearson Correlation Coefficients.

Variable 1	Variable 2	*r*	Statistic	*p*-value	Conf. lower	Conf. upper
evacuation	signif_evac_orders	0.46	15.0	≤ 0.000	0.408	0.516
evacuation	expected_hh_impacts	0.37	11.5	≤ 0.000	0.311	0.429

3.3.3. Chi-Square Test of Independence

Although the way the variables are coded allows the estimation of the correlation among them, because the variables are all categorical, the Chi-square test of independence is more suitable to show whether a relationship exists between the variables. Table 4 and Table 5 are contingency tables that present the cross-tabulation of the data. The levels for evacuation are shown in the rows, and the levels for the independent variable are shown in the columns.

Table 4. Cross-Tabulation of Evacuation and Considering Evacuation Orders.

$N = 826$		**signif_evac_orders**				
		1	**2**	**3**	**4**	**5**
evacuation	**1**	38	31	64	117	270
	0	96	58	63	41	48

Table 5. Cross-Tabulation of Evacuation and Expected Impacts.

$N = 826$		**expected_hh_impacts**				
		1	**2**	**3**	**4**	**5**
evacuation	**1**	49	237	113	94	27
	0	114	149	27	12	4

The calculated value of Chi-square for Table 4 is 182, with degrees of freedom equal to 4 and a *p*-value much smaller than 0.05 ($p \leq 0.000$). The calculated value of Chi-square for Table 5 is 133, with the same degrees of freedom equal to 4, and a *p*-value that is also much smaller than 0.05 ($p \leq 0.000$). A *p*-value smaller than 0.05 is the usual test for dependence. In both cases, the *p*-values are much smaller than 0.05, so there are reasons to believe that the variables are not independent (i.e., they are linked together). In other words, considering evacuation orders and expecting impacts likely makes a difference to evacuation.

Also, descriptive statistics can verify the multicollinearity between the variables. Multicollinearity is when two or more predictors are linearly dependent. The *eigenvalue* (λ) stands for the variance of the linear combination of the variables and estimates a vector of values (eigenvalues) such that the sum must equal the number of independent variables. In such a vector, a very small eigenvalue (close to 0.05) is an indicator of multicollinearity (Shrestha, 2020). For the data, the vector of eigenvalues is [1.8461, 0.6315, 0.5224], which does not indicate multicollinearity.

Therefore, considering the statistics above and the dichotomy of the *evacuation*, an LR can be applied for modeling the probability of the variable. It should be noted that dependent variables in LRs are not necessarily measured on an interval or ratio scale. Moreover, LRs do not make assumptions about distributions of variables and do not require a linear relationship between the dependent and independent variables. Additionally, in the logistic models, the error terms (residuals) do not need to be normally distributed, and homoscedasticity is also not a requirement (Hosmer et al., 2013), which implies flexibility for the application of the method and its wide popularity in various fields.

3.3.4. Logistic Regression

In regression analysis, LR is used to create a statistical model that uses a sigmoid function (also called logistic function) for estimating the odds ratio (calculated from the exponentiated coefficients) of a binary dependent variable for one or more explanatory variables, although several other complex extensions exist. Many textbooks and publications provide enhanced learning on LR—for example, Kleinbaum et al. (2002) and Hosmer et al. (2013). The application and outputs of the method are provided below.

As a very simple example of LR, to predict if *evacuation* was affirmative (1) or negative (0), if expectation of personal and/or household impacts (*expected_hh_impacts*) was not at all likely (0) to almost certain (5), and if local authorities' official recommendations to evacuate (*signif_evac_orders*) were not at all considered (0) or considered to a very great extent (5), the logistic function is of the form:

$$p(\text{evacuation}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \text{expected\_hh\_impacts} + \beta_2 \text{signif\_evac\_orders})}}$$

or,

$$\ln\left(\frac{p(\text{evacuation})}{1 - p(\text{evacuation})}\right) = \beta_0 + \beta_1 \text{expected\_hh\_impacts} + \beta_2 \text{signif\_evac\_orders} \qquad (1)$$

Figure 16 illustrates the logistic model just described, where *evacuation* is given as a probability that assumes a value between 0 and 1.

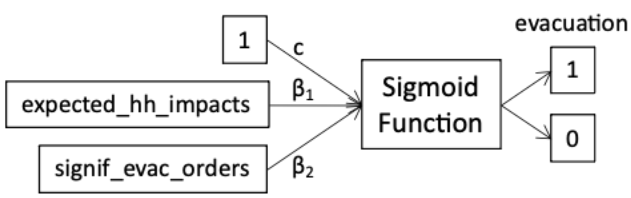


Figure 16. An LR model to predict evacuation.

Table 6 shows the results for the LR model (equation 1). Both predictors are significant (large *z*-values, *very small* p-values), and the parameter estimates are positive, as expected. The *z*-values are the regression coefficients divided by standard error (SE). A large *z*-value (in magnitude) indicates that the corresponding predictor matters. A common rule of thumb is to use a cut-off value of 2, which approximately corresponds to a two-sided hypothesis test with a significance level of 0.05. The estimated parameters are the expected change in the *log* odds of *evacuation* for a unit increase in the corresponding predictor variable holding the other predictor variable constant at a certain value. The odds ratio (OR) that are greater than 1 indicate that the event (*evacuation*) is more likely to occur as the predictor increases. In these results, the odds ratio indicates that for every 1-point increase in considering evacuation orders (*signif_evac_orders*), the likelihood that the household evacuates from the hurricane increases by 1.76 times ($e^{0.5642} = 1.76$). For more detailed information on odds ratio interpretations, see McHugh (2009).

Table 6. Results of the LR to Predict Evacuation.

Variable	Estimate (β)	OR	SE	Statistic	p-value
Intercept	-2.751	0.064	0.264	-10.42	≤ 0.000
expected_hh_impacts	0.624	1.870	0.102	6.14	≤ 0.000
signif_evac_orders	0.564	1.760	0.061	9.33	≤ 0.000

Average marginal effect (AME) is an alternative metric in LR models that can be used to describe the impact of a predictor on the outcome variable (Norton et al., 2019). Table 7 presents the results for the AME analysis. The AME value of *expected_hh_impacts* is 0.1082, which can be interpreted as meaning that a unit increase in the variable value increases the probability of *evacuation* by 10.82%. Again, both predictors are significant (large *z*-values, *very small p*-values), and the effects are positive, as expected. The lower and upper bounds are presented on the 95% confidence interval for the AME estimates.

Table 7. Average Marginal Effects of the LR.

Predictor	AME	SE	Statistic	p-value	Conf. lower	Conf. upper
expected_hh_impacts	0.1082	0.016	6.585	≤ 0.000	0.076	0.141
signif_evac_orders	0.0979	0.008	11.948	≤ 0.000	0.082	0.114

3.3.5. Bayesian Networks

The development of a BN starts with the design of the graphical model, or more precisely, a DAG. Variables are represented by nodes, and their conditional relationships by directed edges (arrows). Note that the terms node and variable can be used interchangeably.

Occasionally, there are misperceptions between a BN and a DAG. A DAG expresses only the conditional independence structure of a BN via the graph structure, which encodes conditional dependencies between random variables. The DAG representation is useful for discussing the construction and the interpretation of a BN.

Figure 17 shows, on the left, the DAG that represents that the probability of evacuation is conditional to the probabilities of expecting impacts and considering official evacuation orders. This representation of the model is equivalent to that made in LR (Figure 16). On the right, Figure 17 presents the probabilities for each level of *expected_hh_impacts* and *signif_evac_orders* and the conditional probability for each level of *evacuation*.

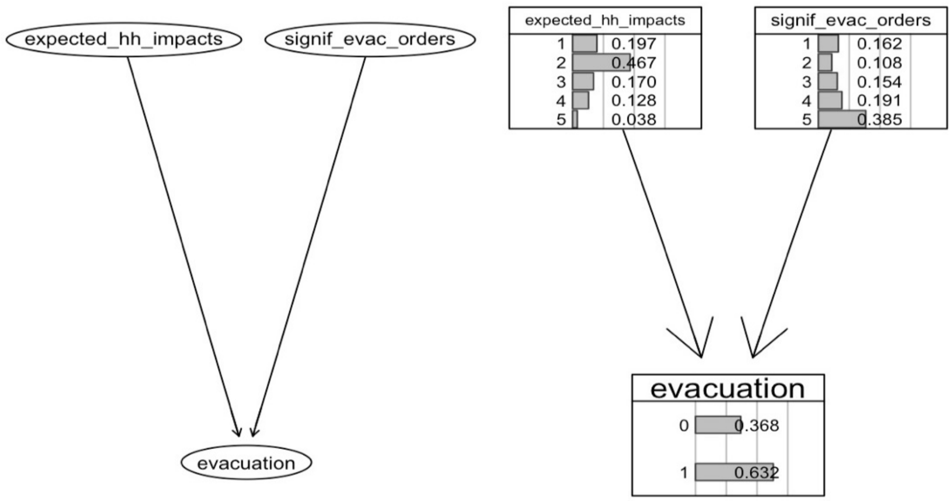


Figure 17. The DAG expressing the conditional structure of evacuation (on the left) and the joint probabilities for each node (on the right).

Every joint distribution on *n* variables factorizes over a DAG with *n* nodes such that there is a directed edge between every pair of vertices (i.e., the vertices are numbered from 1 to *n*). A joint probability distribution factorizes with respect to the DAG, generally, in the form $P(x_1, x_2, \dots, x_n) = P(x_1) \cdot \prod_{i=2}^{n} P(x_i | x_{i-1}, \dots, x_1)$. This yields the classic *frequentist* and *maximum likelihood* estimates (Scutari & Denis, 2014).

For example, Table 8 shows the count of each combination of *signif_evac_orders* and expected*_hh_impacts* and *evacuation*. A prior probability of *evacuation* is estimated by dividing the total number of observations that evacuate (*evacuation* = 1) by the general total number of observations $\left(\text{i.e., } \frac{520}{826} = 0.6295\right)$.

Estimated likelihoods from Table 8 can also assist in estimating the posterior probability for evacuation given *signif_evac_orders* and *expected_hh_impacts*. For example, *evacuation* = 1 conditional to *signif_evac_orders* and *expected_hh_impacts* (model of Figure 17) are computed using the Bayes' theorem (P(A|B) = P(B|A) P(A) / P(B) (Stone, 2013). Therefore,

$P(\text{evacuation} = 1|\text{expected\_hh\_impacts,signif\_evac\_orders}) =$

$\sum_{i,j} P(\text{expected\_hh\_impacts}_i)P(\text{signif\_evac\_orders}_j)P(\text{expected\_hh\_impacts}_i\text{,signif\_evac\_orders}_j|\text{evacuation} = 1) = 0.632,$

where *I* represents each of the levels of *expected_hh_impacts*, and *j* represents each of the levels of *signif_evac_orders*.

Table 8. Combination of All Levels of *evacuation*, *signif_evac_orders* and *expected_hh_impacts*.

		evacuation		
signif_evac_orders	expected_hh_impacts	0	1	Total
1	1	56	8	64
	2	34	20	54
	3	4	6	10
	4	2	4	6
	5	0	0	0
2	1	19	8	27
	2	34	21	55
	3	4	1	5
	4	0	1	1
	5	1	0	1
3	1	23	13	36
	2	32	34	66
	3	5	11	16
	4	3	3	6
	5	0	3	3
4	1	7	10	17
	2	22	57	79
	3	8	28	36
	4	4	18	22
	5	0	4	4
5	1	9	10	19
	2	27	105	132
	3	6	67	73
	4	3	68	71
	5	3	20	23
	Total	306	520	826

Every BN model demands a particular factorization of a joint probability distribution. This factorization implies certain independent assumptions about the underlying model. These assumptions can be tested through conditional independence tests on each arc conditional to the network. Two random variables are independent if the occurrence of one variable does not affect the probability of occurrence and therefore the probability distribution of the other variable (Gelman & Speed, 1993; Holmes, 2008; Jensen, 1996). Generally, $x_1 \perp x_2 | x_3 \Leftrightarrow P(x_1, x_2 | x_3) = P(x_1 | x_3) P(x_2 | x_3)$. In the example, *expected_hh_impacts* and *signif_evac_orders* are independent random variables; thus, *P*(expected_hh_impacts, signif_evac_orders | evacuation) = *P*(*expected_hh_impacts* | evacuation) *P*(signif_evac_orders | evacuation).

Instead of using marginal independence, the independence between two random variables can be assessed through conditional independence tests by adapting either the log-likelihood ratio G^2 or Pearson's X^2 to test for conditional independence (Agresti, 2003; Burkart & Király, 2018). Pearson's X^2 test is preferable for continuous data, and the log-likelihood ratio G^2 test, equivalent to the mutual information (MI) test from information theory, is preferable for discrete data (Scutari & Denis, 2014). The MI between two random variables (x_1, x_2) can be estimated through the following equation:

$$MI(x_1, x_2) = E\left(ln \frac{p(x_1, x_2)}{p(x_1)p(x_2)}\right) = \sum_{x_1, x_2} p(x_1, x_2)[\ln p(x_1, x_2) - \ln p(x_1)\, p(x_2)] \quad (2)$$

Table 9 shows the MI tests for the example. Both predictors have a very small value of MI (<< 0.05), which suggests strong evidence of the connections given the structure of the network. The strength of the connection can be represented by the width of the lines in the DAG (see Figure 18, on the right).

Table 9. Mutual Information Tests.

From	To	MI *p*-value
expected_hh_impacts	evacuation	≤ 0.000
signif_evac_orders	evacuation	≤ 0.000

The task of testing all the links in a network can be automatized. The sequential mutual information Monte Carlo permutation test is one of the conditional independence tests for discrete BNs that uses categorical variables and is implemented in R. Based on the MI of two random variables, this test is proportional to the log-likelihood ratio and produces a measure of the mutual dependence between the two variables (Tsamardinos & Borboudakis, 2010).

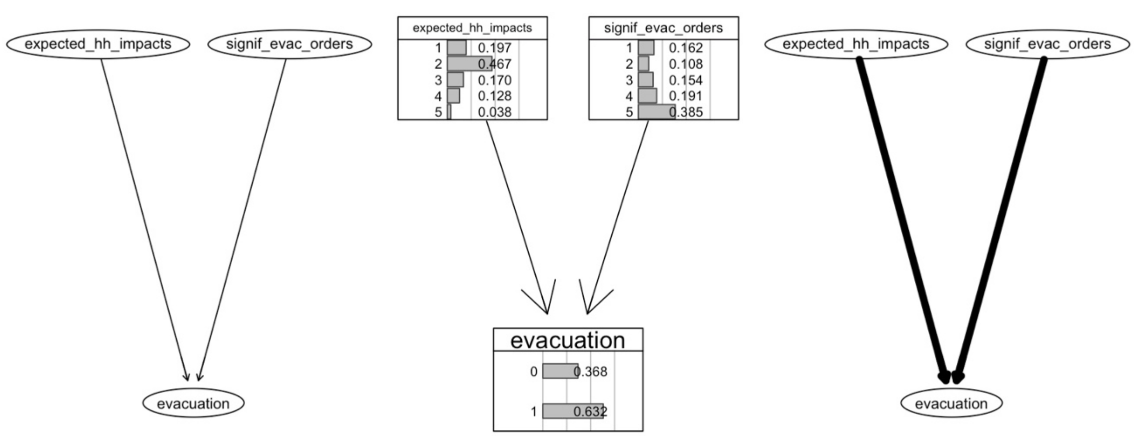


Figure 18. The DAG expressing the conditional structure of evacuation (on the left), the joint probabilities for each node (in the center), and the indication of connections strength (on the right).

Unlike conditional independence tests, network scores (NS) can test the DAG as a whole. NS are goodness-of-fit statistics that measures how well the structure of the DAG mirrors the dependence structure of the data. Bayesian information criterion (BIC) is one of the most popular tests. In the example, it takes the form of:

$$BIC = logLik - \mathrm{d} \cdot \frac{\log(n)}{2} = -2790 - 3.358 \cdot 33 = -2901$$

where *logLik* is log-likelihood of the network considering the data, *n* is the sample size, and *d* is the number of parameters extracted from the network structure and variable levels. The BIC value should only be compared with another network structure that uses the same data and variables, but different connections between the variables. That can be examined in the following experiments below. For a more detailed explanation of BIC and other NS, please see Scutari and Denis (2014).

The fundamentals of BN have been explained in this section. Next, the use of the studied network will be explored in four experiments. The first experiment estimates the evacuation probability given that some information is known about the predictor variables. The second experiment uses a synthetic dataset with a distribution like the data surveyed and compares the predictive capacity of BNs with that of LR. The third experiment studies which of the two predictors has higher influence on the probability to evacuate, which specifically can be useful for practical planning purposes. And finally, the fourth experiment studies an alternative network structure and shows that this network structure can more isomorphically reproduce conceptual models.

3.3.6. Experiment 1

This experiment analyzes the prediction ability of both models in one specific observation. It estimates the probability of evacuation given that information about the state of the predictor variables is known. First, the most adverse condition is assumed on both variables: (a) local authorities issuing official recommendations to evacuate is **not at all considered** (*signif_evac_orders* = 1), and (b) expecting personal or household impacts is **not at all likely** (*expected_hh_impacts* = 1).

Evidence on the state of the variables is commonly called *soft evidence*, in contrast to when there is evidence on the connection of variables, called *hard evidence* (Mrad et al., 2015). By using the results of the LR (see Table 6), a prediction model (i.e., prognostic analysis) takes the form of:

$$y = -2.751 + 0.624 \text{ expected\_hh\_impacts} + 0.564 \text{ signif\_evac\_orders}$$

$$= -2.751 + 0.624 \cdot 1 + 0.564 \cdot 1 = -1.563$$

Accordingly, the probability that such a household does evacuate is given by:

$$P(\text{evacuation} = 1) = \frac{e^y}{1 + e^y} = \frac{e^{-1.563}}{1 + e^{-1.563}} = 0.174$$

Correspondingly, the probability that such a household does not evacuate is given by:

$$P(\text{evacuation} = 0) = 1 - P(\text{evacuation} = 1) = 0.827$$

For predicting the probability of evacuation using a BN, Figure 19 shows the conditional probabilities after fixing both *signif_evac_orders* and *expected_hh_impacts* to 1. The computation of evacuation probabilities is similarly calculated using Bayes' theorem (as shown on page 47). The probability that such a household does evacuate is 0.125, and accordingly, the probability that the household does not evacuate is 0.875.

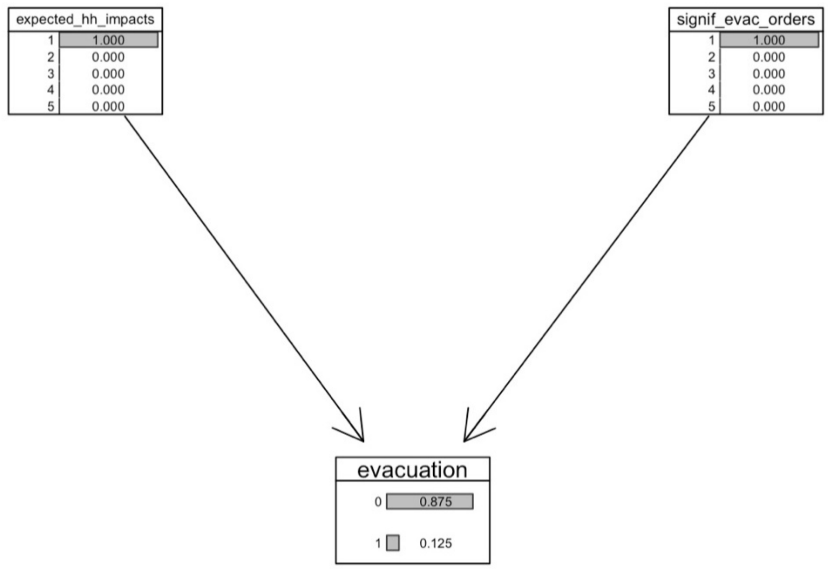


Figure 19. BN predicting evacuation.

The results are slightly different—probabilities of 82.68% versus 87.5% that the household does not evacuate and probabilities of 17.35% versus 12.5% that it does evacuate—for the LR and the BN, respectively. In the surveyed sample, there are 64 observations with such a combination (*signif_evac_orders* = 1 and *expected_hh_impacts* = 1). Of these, 56 (87.5%) did not evacuate, and eight did evacuate (12.5%), which are the same probabilities given by the BN model's prediction.

This experiment may emphasize a limitation on both methods. Logistic regression can only make a prediction if there is concomitantly information available on the state of all variables of the model—that is, complete cases. One way to get around this limitation for LR is to consider a missing value as zero, but that procedure can ultimately cause bias in the prediction.

For BN, the probability can be estimated in incomplete cases. However, the probability of evacuation can only be estimated if there is a combination of predictors in the observed data in order to create a probability estimate for it. For example, in the case of fixing *signif_evac_orders* to 1 and *expected_hh_impacts* to 5, there are no observations in the survey, so no probability is generated for that (see Table 8).

Two possible solutions can be adopted to solve this limitation specifically. First, a *reasonable* belief can be assigned if no other information is available prior (e.g., equal chance for each one of the levels). Second, the number of levels for a particular variable can be decreased, and consequently the combination of observations implies the increase of the number of observations on each level and also increases the chances of having probabilities for all the combinations of levels.

This experiment infers only one conceivable situation to demonstrate how the prediction occurs using both approaches. The next experiment executes the prediction on several circumstances and compares the prediction accuracy of the models.

3.3.7. Experiment 2

An alternative approach to test the inference ability of the models is to use Monte Carlo simulations to randomly generate observations (i.e., synthetic data) and use these observations to estimate and verify the models' prediction capacity. This experiment seeks external validity of the models' inference ability by testing how generalizable the prediction's accuracy is on a large set of conditions. In the literature, the aim of producing synthetic data has been to provide publicly available datasets that can be used for inference in place of the actual data.

The *synthpop* package in R (R Core Team, 2021) is used to create synthetic data that allows inferences from the fitted statistical models and the comparison of the results. The package was first written as part of the United Kingdom Economic and Social Research Council-funded Synthetic Data Estimation for Longitudinal Studies project to allow local researchers to produce synthetic data tailored to the needs of particular projects. For more details on the *synthpop* package, please see Nowok et al. (2016).

The use of *synthpop* requires only a set of observed data that are assumed to be a sample from a population with parameters that can be estimated by the package's *synthesizer*. A set of synthetic data was created having the same number of observations as the original dataset (n = 826). Figure 20 visually compares distributions of synthesized and observed data.

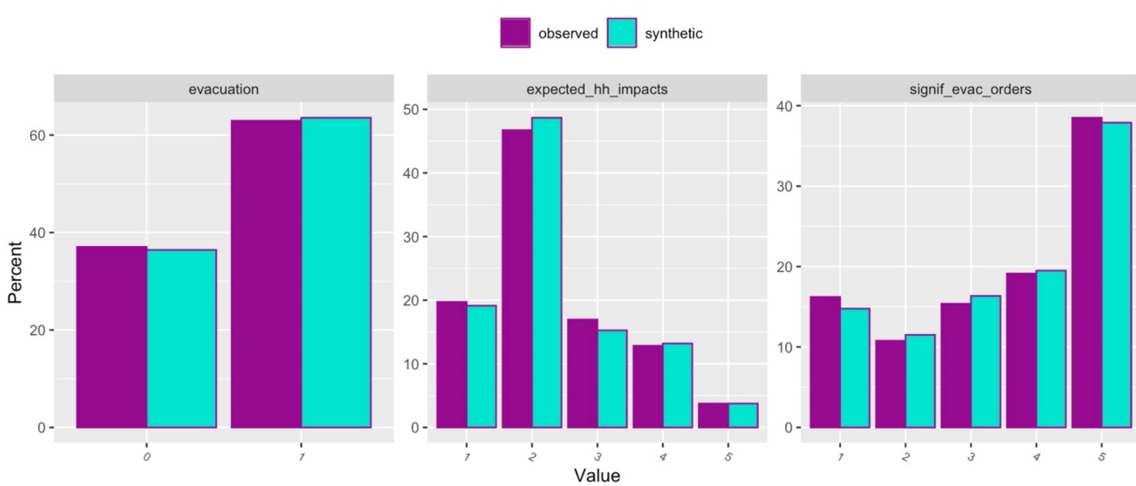


Figure 20. Distribution of a synthetic population.

Based on the synthetic dataset, Table 10 presents a summary of the predictions on both LR and BN models. A traditional classifier threshold of 0.5 is applied. That is, if the model predicts less than this value, it is assumed that the household does not evacuate. Correspondingly, if the probability value of the model output is higher than or equal to the threshold value, it is assumed that the household does evacuate.

The first column in Table 10 (probability) shows the probability inferenced by the models, and each Column 0 or 1 counts the number of times the model predicts correctly or incorrectly. For example, the BN model predicts the probability of 0 percent on two observations that did not evacuate, and on one observation that *actually* evacuated. The LR does not predict 0 percent probability on any of the observations.

Table 10. Inferenced Results Using Synthetic A Dataset.

N = 826	**Counts on Each Method**			
	BN		**LR**	
Probability	**0**	**1**	**0**	**1**
0	2	1	0	0
0.1	27	16	0	0
0.2	7	1	27	16
0.3	6	16	33	40
0.4	79	71	62	50
Subtotal	121	105	122	106
0.5	41	70	31	62
0.6	12	28	19	2
0.7	48	62	41	87
0.8	47	140	44	136
0.9	19	67	39	119
1	13	53	5	13
Subtotal	180	420	179	419
Total	**301**	**525**	**301**	**525**

Both methods are good and very similar in the categorization and prediction accuracy. BN correctly predicts 121 no evacuations, while LR correctly predicts LR 122 no evacuations (highlighted in the table). Therefore, both methods have a true negative rate (i.e., specificity) of approximately 40% (121/301 and 122/301, respectively). BN correctly predicts 420 evacuations, while LR correctly predicts 419 evacuations (also highlighted in the table). Therefore, both methods have a true positive rate (i.e., sensitivity) of approximately 80% (420/525 and 419/525, respectively).

Figure 21 presents the receiver operating characteristic (ROC) curve and the area under the curve (AUC), which illustrates the diagnostic ability of the models as binary classifiers. Many publications provide an enhanced overview on AUC-ROC curves—for example, Marzban (2004) and Hoo et al. (2017). It can be observed in the figures that both the BN and LR models have similar predictive ability, considering the AUC estimates (59.8% versus 59.2%, respectively) and the 95% confidence intervals.

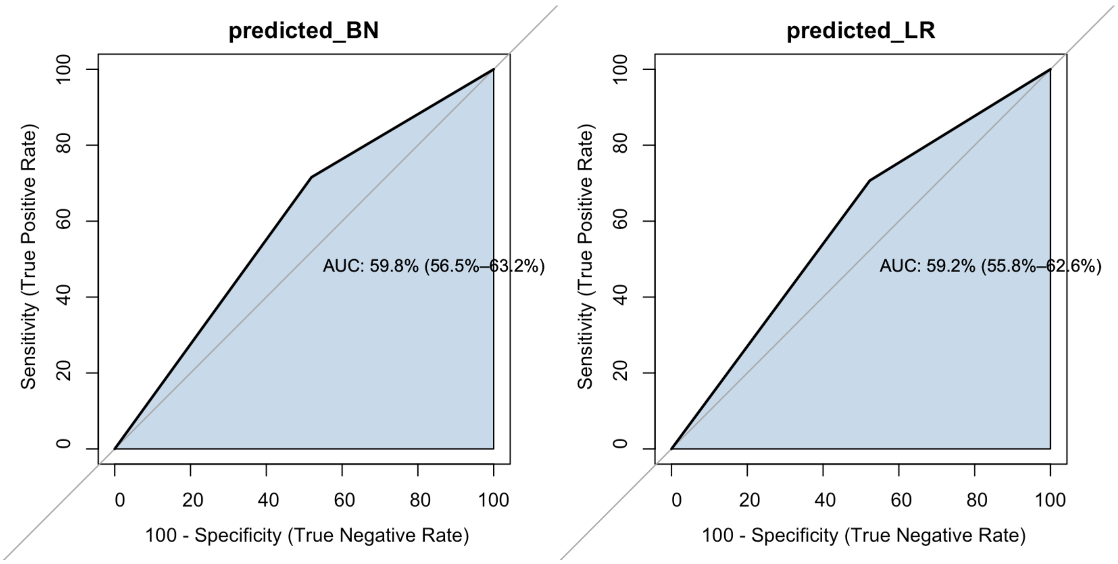


Figure 21. AUC and ROC curve for sensitivity and specificity analysis.

3.3.8. Experiment 3

This experiment explores the possibility of knowing that a household has evacuated, but it is desirable to investigate the influence of each predictor level on such an output. The evidence that the household has evacuated is so-called soft evidence. Although it analyzes the parent nodes of an outcome, it is usually acknowledged in Bayesian literature as a *prognostic analysis*.

Using Bayes' theorem (explained on page 47), probabilities can be estimated for each of the predictors if it is known that the household evacuated (*evacuation* = 1). For example, the probability of expecting impacts is not at all likely (*expected_hh_impacts* = 1), and the household evacuate (*evacuation* = 1) can be estimated using the values from Table 8 in the following equation:

$$P(\text{expected\_hh\_impacts} = 1 \mid \text{evacuation} = 1)$$

$$= \frac{(\text{number of obs. for which evacuation} = 1 \text{ and expected\_hh\_impacts} = 1)}{(\text{number of obs. for which evacuation} = 1\ )}$$

$$= \frac{49}{520} = 0.0942$$

Table 11 presents the results for all levels on both predictors. Based on the evidence that the household did evacuate (*evacuation* = 1), the probability of *expected_hh_impacts* is lower (2 has the greater probability), and *signif_evac_orders* is higher (5 has the greater probability). This information can be very useful for disaster planning. To enhance evacuation, emergency managers need to focus on intensifying the risk perception in relation to *signif_evac_orders*, which most impacts the probability of a household evacuating according to this specific surveyed data and model structure.

Table 11. Prognostic Results on the Predictors.

Levels	**Predictors' Probabilities**	
	expected_hh_impacts	*signif_evac_orders*
1	0.0942	0.0731
2	0.4558	0.0596
3	0.2173	0.1231
4	0.1808	0.2250
5	0.0519	0.5192

3.3.9. Experiment 4

This experiment considers an alternative network structure on the BN model. The initial network structure simply pointed out all the predictive variables for evacuation. Changes in an initial network structure can happen as hard evidence is learned from data or beliefs (i.e., hypotheses; Mrad et al., 2015).

For example, consider the probability of official evacuation orders influencing the probability of expecting personal and household impacts. This hypothesis has already been tested and supported by data (Huang et al., 2017). In the BN model, such a connection will update the prior *expected_hh_impacts*. Figure 22 presents the alternative network structure, in which *signif_evac_orders* still influence *evacuation* but also simultaneously influence *expected_hh_impacts*. Testing the MI of the nodes reveals that the three connections are statistically significant. As can be noted in the *evacuation* node, once again the evacuation probabilities are updated. The BIC of the updated BN is -2850.

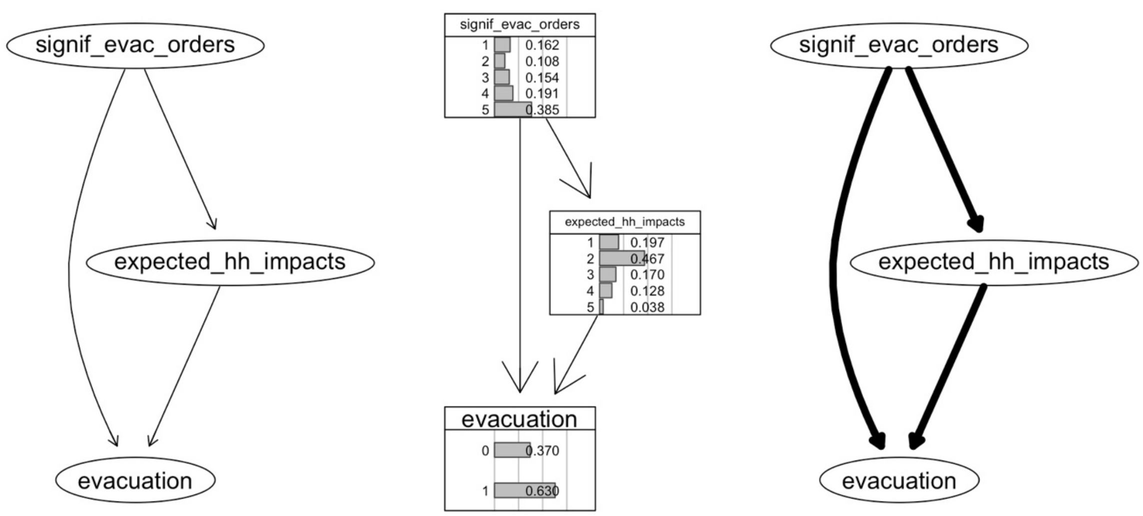


Figure 22. An alternative network structure.

The capability of connecting nodes and forming more complex structures while priors are also updated can help the modeling of conceptual models without other statistical complexities. Figure 23 presents an adapted section of the protective action decision model (PADM; Lindell & Perry, 2012).

The diagram shows a section of the information flow in the PADM that can be directly compared to the alternative proposed BN. Warning messages (equivalent to considering official evacuation orders) are related to the threat perceptions (consistent with expecting impacts), which are related to the decision for and the act of protective measures (in this case, evacuation).

This conceptual model also considers experience in decision-making influence predictors, which is not considered in the proposed BN, although the BN model in its alternative implementation possibly better reflects such a structure and creates the possibility for the straightforward implementation of more complete versions of PADM conceptual models. The next chapter presents a more extensive discussion of this idea.

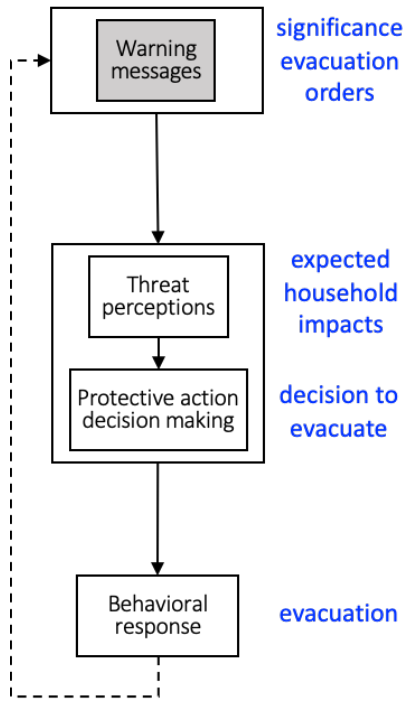


Figure 23. Section of the information flow in the PADM adapted from the model presented by Lindell & Perry (2012).

3.4. Discussion and Conclusions

BNs are used in several fields and have been widely applied to decision-making problems because they combine the benefits of formal probabilistic methods, an engaging visual form, and efficient computational techniques to explore complex arrangements of predictors for an outcome. Hurricane evacuation is a recurrent and important decision-making problem for many coastal communities exposed to this natural hazard, and BNs can be useful in helping to reach that decision even though they currently have had limited exploration.

This study began with the identification of two objectives: first, to understand how BNs could be used to model hurricane household evacuation; and second, to demonstrate the adequacy, effectiveness, and limitations of this novel approach in this specific field of disaster planning. A random sample of 826 completed cases based on address sample frames that were collected in a survey after Hurricane Harvey passed through the Texas Coastal Bend area in August 2017 was used for this study. The analysis considered only two of the main and most recurrent reasons that directly affect household evacuation, based on research to date: (a) receiving an official warning, and (b) expecting personal and household impacts.

Since each relationship between two variables is a hypothesis to be tested, even when not acknowledged, the assumptions that (a) receiving an official warning, and (b) expecting personal and household impacts to influence evacuation were first confirmed (i.e., supported by the surveyed data) using traditional approaches, such as linear correlation, chi-square tests, and LR. Although the correlation among variables is a traditional indicator of similarity or of a relationship between two variables, an empirical cumulative distribution of an independent variables conditional on the dependent variable seems to be more effective in identifying possible conditional relationships.

The same variables and dataset were used for the development of a simple BN, which presented consistent findings with the results and literature. The probabilities were obtained from the observations for all possible level combinations (i.e., categories of variables used), and by using the significant level of 0.05, the MI test estimated the linkage strength (i.e., *p*-values) among the variables. The initial BN only had direct links from the predictors to evacuation; therefore, the conditional MI estimates were equivalent to the Chi-square independence tests of each pair of variables.

The development of the initial BN model explained the fundamentals of the tool, such as the factorization of the probabilities and other core definitions and properties that provided basis for the experiments developed next. The odds ratios of an LR cannot be interpreted as absolute effects, which implies that BN results can have a more direct explanation without requiring additional steps and context. Although odds ratios mean the ratio of occurrence to nonoccurrence, the probability of BNs are the ratio of occurrence to the *whole*. The probabilities of evacuation found by the BN are updated consequences from the network structure, which revises the prior knowledge. The initial dataset showed an estimated 62.9% chance that a random household would evacuate (520/826). Considering such initial network structure, the updated estimate was 63.2%.

To further study the use of BNs, four experiments were developed. Experiment 1 estimated the evacuation probability given that information is known about the predictors' levels. The most adverse condition was assumed on both variables; that is, (a) a household did *not at all consider local authorities issuing official recommendations to evacuate*, and (b) the household *did not at all likely expect personal or household impacts*.

This experiment showed that the probability values are not very different, although they are not the same. This example shows how the prediction calculations are developed by each BN and LR, and no conclusion can be taken from the differences in the probabilities. For that, Experiment 2 sought to compare the predictive capacity of BN in a more generalized and formal procedure.

The way in which the probabilities are obtained exposed a limitation of the BN. The probability of evacuation can only be estimated if there is a combination of predictors' levels in the observed data to create a probability estimate for it. In case there is no such value in the factorization of the observations, a prior must be adopted, either by a belief, or by decreasing the number of levels of a variable, to increase the chances of having probabilities for all the combinations of levels.

The second experiment sought to generalize the predictive capacity of the BN model, and to compare the results when using the LR as a classifier. For the external validity of the models, a synthetic population generator was used to create a large number of simulated observations with the analogous distributions for each variable as the surveyed data.

Both models had similar classifications and were able to classify the evacuation satisfactorily (AUC values were statistically significance greater than 50% on both models, but not statistically significance different between them). Both models had a higher sensitivity rate (i.e., true positive rate) than sensitivity (i.e., true negative rate), approximately 80% versus 40%, respectively.

Experiment 3 explored diagnostic analysis using BN. It showed that knowing the outcome of evacuation, it is possible to investigate the probability levels of the predictors that are more likely to have happened. In the example, given that the household evacuated, it is likely that receiving an official warning was to a very great extend (rate 5) while expecting impacts was more likely to be at the lower of the average rate. This type of analysis can be appropriate and useful for the development of public policies and for a better understanding of factors that influenced evacuation.

Experiment 4 considered an alternative network structure for the BN. This is usually called hard evidence in the Bayesian literature, in contrast to the soft evidence that is related to knowledge in the state of a variable. The new network seems to better mirror the dependence structure of the data because its BIC value is higher than the previous network (-2850 > -2901).

Experiment 4 suggested that BNs can isomorphically represent complexities of conceptual models from the literature without major statistical complications such as the ones found in multistage modeling. However, there is a limitation to using linkages between variables with a single direction that do not form a cycle in the network structure considering that DAG are premises for BN. More advanced analysis, such as a dynamic BN, can accept the modeling of variables at different moments in time and bypass this limitation.

None of the findings of these experiments are unknown in regard to BN. The major methodological contribution of this study is its validation of the application of BN in disaster planning problems, specifically in the analysis of hurricane evacuation. Although household evacuations in natural disasters have often been studied, constant examination of methods to predict evacuation behavior is relevant for more accurate outcomes and for a deeper learning process of the population choices.

Instead of competing with traditional methods, BN can be a complementary method that can help in the examination of gaps in the evacuation literature. For example, usually an expressive quantity of residents at risk do not evacuate from an approaching hurricane when they are advised to by local authorities. This inaction can cause unnecessary suffering, injuries, and ultimately deaths. Non-evacuation is still not fully understood. The use of BN to study household hurricane evacuation can help members of the community, planners, and emergency managers jointly learn and explore factors that affect the evacuation behavior. Thus, BNs can provide tools that can more comprehensively assist in preparedness yet still allow the presence of analysis of alternative risk countermeasures and scenarios by defining specific states for variables.

This study was based on the observations of a particular survey instrument in a specific region during a single event. Future studies should continue testing the use of BNs with different datasets and expand the number of factors that can influence evacuation, thereby further testing the mediating effects between variables and BN's asymptotic assumptions in terms of a larger dataset and a number of variables on complex network structures.

4. APPLICATION OF BAYESIAN NETWORKS TO STUDY HOUSEHOLD HURRICANE EVACUATION

4.1. Abstract

During hurricanes in coastal communities, household evacuation is a critical response by which local authorities and residents can prevent loss of life. Research to date indicates that receiving an official warning and expecting personal and household impacts directly affect evacuation. However, more analyses are needed to understand the mediating and interaction effects of other variables. Although most studies use logistic regression to examine the wide range of factors that affect evacuation, probabilistic graphical techniques such as Bayesian networks (BNs) may provide an alternative that allows learning of causal effects from observational data and creates intuitive graphical representations of probabilistic models that can more directly represent conceptual models and explicitly acknowledge the complexity of evacuation. This study aims to examine hurricane household evacuation using BNs. To develop the analysis, this study considers an established conceptual model from the literature and data collected after Hurricane Harvey passed through the Texas Coastal Bend area in August 2017. The results list factors that influence evacuation and indicate that BN can isomorphically model complexities of conceptual models without major statistical complications, thereby demonstrating a potential to be more frequently used in future disaster preparedness and planning.

4.2. Introduction

Hurricane-related hazards, such as high wind speed and storm surge, can cause some of the most devastating natural disasters in coastal communities (Bengtsson, 2001). Moreover, household evacuations are so critical in saving lives that several studies are still trying to understand the factors that influence this choice (e.g., Baker, 1991; Lazo et al., 2015; Lindell et al., 2005).

The analysis of evacuation behavior is critical for disaster preparedness and planning—particularly for hurricanes, since Baker's (1991) research indicates that receiving an official warning to evacuate and expecting personal and household impacts from the storm directly influence evacuation, and frequent evaluations are necessary to better understand the mediating and interactive effects of many other variables (Tanim et al., 2022).

Modeling the "complexity involved in the household evacuation decision-making process" (Hasan et al., 2011, p. 341) is an *ongoing* challenge effectively captured by conceptual models such as the protective action decision model (PADM) (Heath et al., 2018; Lindell & Perry, 2012).

However, mathematically modeling such problems is not a simple task because of the heterogeneity of the social vulnerabilities of populations impacted (Burton, 2010), the areas affected by those *compounding* threats (Cegan et al., 2022), and the indefinite mediating and interacting effects of diverse variables.

Most studies use logistic regression (LR) to examine the wide range of factors that affect evacuation (Yang et al., 2016), which provides a rigorous analytical framework for modeling the discrete outcome of a choice. However, although many prediction models are developed and available, supposedly the uptake in planning practice and public policy is relatively slow, considering that an expressive number of at-risk residents do not evacuate from an approaching hurricane when advised by local authorities. Some studies (e.g., Laitin, 2003; Waddell, 2011) have tried to identify the challenges that can happen in the process of taking models developed in an academic research setting, where theoretical validity and the advancement of methodology receive high priority, and moving them into public agency settings in which priorities are typically reliability, facility of use, and staff capacity to explain to stakeholders what the models are doing.

More specifically, including amid researchers, probably two important challenges associated with using the LR to model evacuation are as follows: (a) the difficulty to incorporate dependencies among variables, and (b) the presence of numerous (risk) factors with only a small and/or mixed effect across previous decisions and intentions to evacuate. For example, while some studies indicate that gender is significantly related to hurricane evacuation and women are more likely to evacuate (e.g., Bateman & Edwards, 2002; Gladwin et al., 2001), others have found no significant effect of gender on evacuation intention, although men were less likely to evacuate among respondents (Lazo et al., 2015).

This study tries to overcome these challenges by examining hurricane household evacuation through an alternative approach—Bayesian networks (BNs). This probabilistic graphical technique has been used to study decision-making problems in risk management (e.g., Fenton & Neil, 2011; Jager et al., 2018); however, the use of BNs in research to study responses to environmental hazards and disasters, particularly hurricane evacuation, has been limited.

BNs are possibly a more accessible approach to model and examine factors that can directly and indirectly influence evacuation. BNs allow the learning of causal effects from observational data and offer an intuitive graphical representation of probabilistic models with relatively easy interpretation while still having a solid statistical basis. These graphical representations can more explicitly specify conceptual models and acknowledge the complexity of evacuation.

This study develops a BN model that is consistent with an abbreviated version of the PADM and uses data collected in a survey after Hurricane Harvey (that hit Texas in August 2017). The PADM proposes to integrate the influence of environmental and social cues, information sources, preferences, warning messages, and receiver characteristics in the evacuation choice. The general goal of the survey used was to obtain data on household evacuation response and factors influencing household response in the Texas Coastal Bend area, which includes Aransas, Calhoun, Jackson, Matagorda, Nueces, Refugio, San Patricio, and Victoria Counties.

The results find important factors that influence evacuation, but more importantly, the results indicate that BNs can isomorphically model complexities of conceptual models without major statistical complications and are likely to be used more often in future disaster preparedness and planning. Future research can incorporate more variables and domains in the model and, also, include a specific variable that measures the evacuation decision, or choice, previously the evacuation, which tries, in this way, to contrast the difference between choice, decision, or intention to evacuate and its occurrence and success.

4.3. Data and Methods

This section presents a conceptual model that can facilitate the study of factors that influence evacuation and includes the description of the dataset used in the analysis. The specific version of the model presented was chosen from among the many varieties in the literature because it was recently published and used for a storm with similar characteristics in a geographical region near the Coastal Bend. Details are presented below.

4.3.1. Conceptual Model and the Graphical Representation of a Probabilistic Model

Lindell and Perry's (2012) PADM has been extensively used to analyze the behavior and decision-making of people and households subject to threats from a wide range of hazards (e.g., Strahan & Watson, 2019; Terpstra & Lindell, 2013) and has acquired the status of a theory in environmental hazard and disaster planning (Lindell & Perry, 2012).

Generally, slightly different versions of this conceptual model and information flow appear in the literature (e.g., Lazo et al., 2015; Lindell et al., 2005). Figure 24 shows an analogous version of the PADM multistage model of hurricane evacuation presented in Huang et al. (2017) to examine hurricane evacuation decisions during Hurricanes Katrina and Rita.

Hurricane Katrina made landfall off the coast of Louisiana, and Hurricane Rita threatened the Texas coast but gradually curved east toward the coast of Louisiana, making its landfall near the border of Texas and Louisiana. The PADM conceptual model facilitated defining the hypotheses tested in the study in which basically every relationship between the variables was an assumption to be examined.

In the *aforementioned* study, a series of statistical tests and regression analyses showed that two predictive paths affected evacuation decisions. For the first path, the effects of the antecedent variables on evacuation decisions were mediated by expected storm threats and impacts. For the second path, the effects of the antecedent variables on evacuation were mediated by expected evacuation impediments—although expected evacuation impediments affect evacuation decisions indirectly (via expected wind impacts) as well as directly.

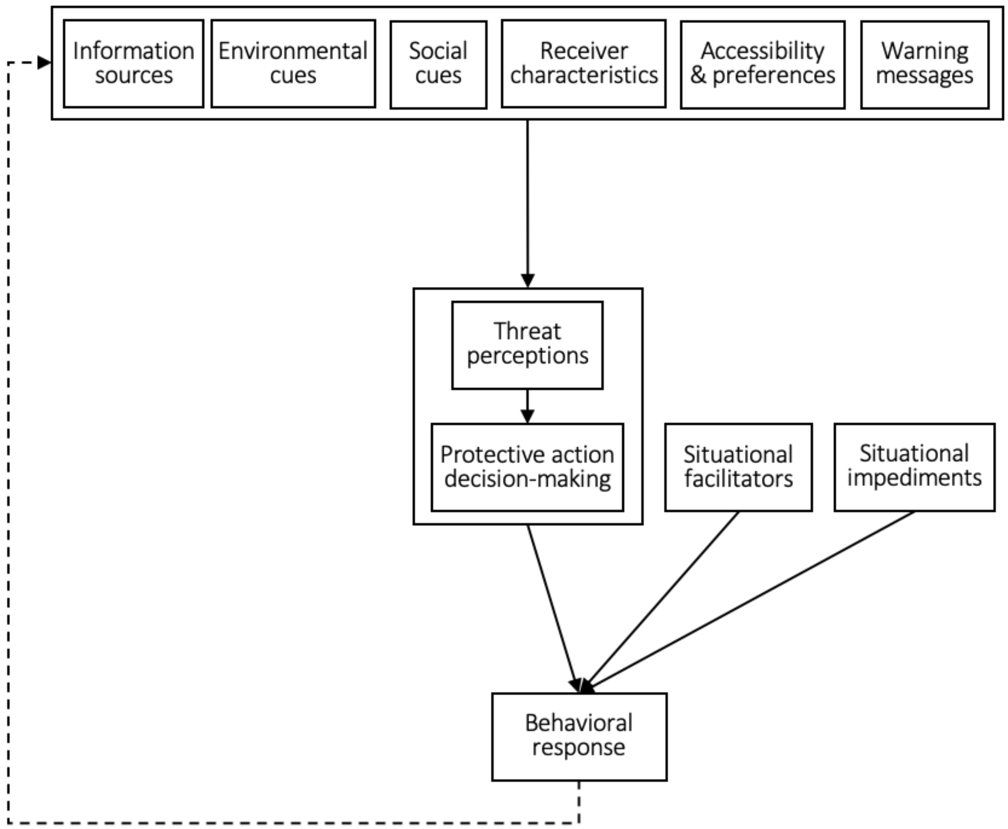


Figure 24. Adapted version of the PADM multistage model for hurricane evacuation.

Apart from the dotted link, this conceptual model is very comparable to a directed acyclic graph (DAG), which may represent a BN model by directly using the same structure.

BNs are a type of probabilistic graphical model that use Bayesian inference for probability computations. BNs aim to model conditional dependence and therefore causation by representing conditional dependence on a DAG. A DAG can be thought of as a kind of flowchart that visualizes a whole causal network linking causes and effects

(Foraita et al., 2014). In a graph that represents a particular model, a set of variables is shown as nodes and their conditional dependencies shown as links (also called arcs or edges).

Every BN demands a particular factorization of a joint probability distribution, and this factorization implies certain independence assumptions about the underlying model that can be found using only the DAG (Jensen, 1996). In other words, the BN model that a specific DAG represents is a joint probability distribution that takes the form of a product of n factors ($p = p_1 p_2 \ ... \ p_n$) wherein these factors need meaningful interpretations as probability densities (for the initial variables of the network that are theoretically not dependent on other influences) or marginal and conditional probabilities (for the other variables that are conditioned to one or more factors).
All the independence relationships implied by the factorization can be found using only the DAG. That is, BNs satisfy the *global* and *local* Markov properties (Kang & Tian, 2012) that state, respectively, that the set of conditional independence relationships encoded in a DAG can be read by *d*-separation criterion (Geiger et al., 1990), and a node is conditionally independent of its non-descendants given its preceding nodes in the network.

To ultimately construct the BN probabilistic model to examine evacuation, variables selected from a survey are presented below.

4.3.2. Hurricane Harvey Household Evacuation Behavior Survey

In 2019, researchers from the Hazard Reduction & Recovery Center at Texas A&M University, Texas A&M Transportation Institute, and the University of

Washington Institute for Hazard Mitigation Planning and Research conducted the Hurricane Harvey Evacuation Behavior Survey (HHEBS), which targeted the eight counties that comprise the Texas Coastal Bend area (Aransas, Calhoun, Jackson, Matagorda, Nueces, Refugio, San Patricio, and Victoria counties). Figure 25 shows a part of the Hurricane Harvey Final Best Track (National Hurricane Center, 2022) over the Texas Coastal Bend area, which is a significant geographic region of the coast of Texas exposed to hurricanes.

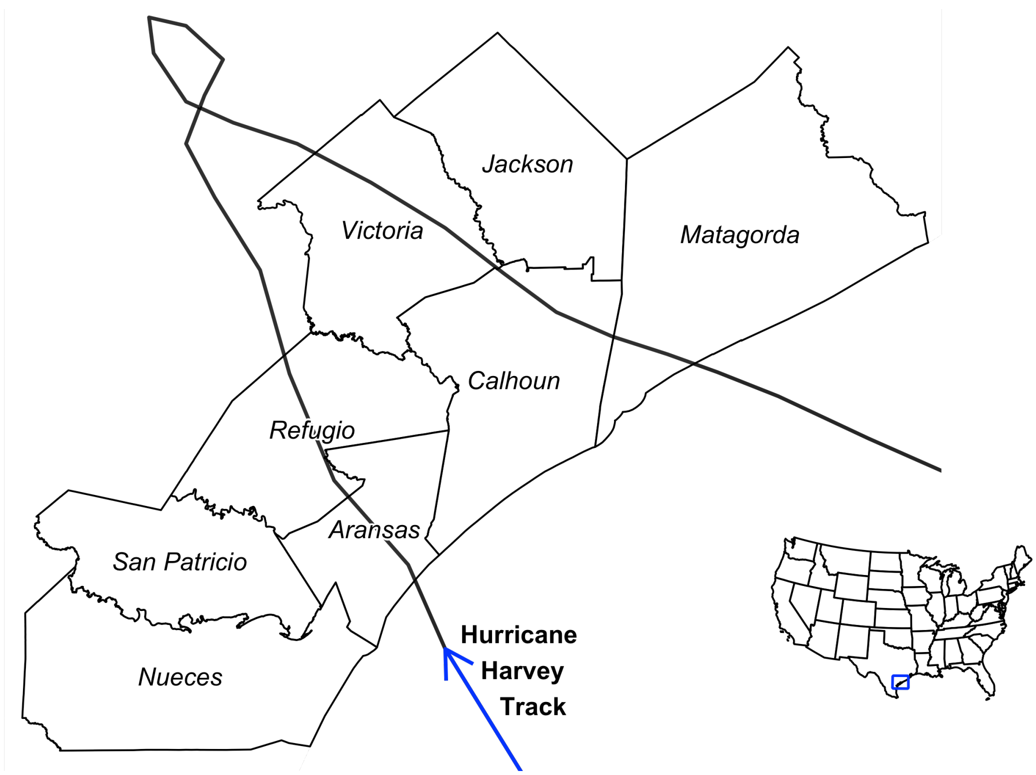


Figure 25. Texas Coastal Bend area and Hurricane Harvey track.

The general goal of the researchers conducting the HHEBS was to obtain data on factors influencing household responses to Hurricane Harvey, which made its first landfall in Aransas County, Texas, on August 25, 2017, as a Category 4 hurricane. Households were randomly selected, and the survey distribution was administered in three waves. Wave 1 of the survey was online only, and Waves 2 and 3 included paper copies of the survey instrument with postage-paid return. To enhance communication and responses from Hispanic residents, 20% of the sample distribution included a bilingual version (English and Spanish). These bilingual surveys were addressed to locations with a higher proportion of Spanish speakers in the study area (Bierling et al., 2020).

The HHBES behavioral survey covered four primary topic areas: household evacuation decisions and associated cues, evacuation preparations and logistics, evacuation route choices, and respondent/household demographics and related characteristics (Bierling et al., 2020). The survey instrument, comprising 41 questions, was based upon earlier questionnaires that the HRRC used in previous hurricane evacuation behavior studies (e.g., Lindell et al., 2001; Lindell et al., 2013), along with items from other evacuation studies summarized by Lindell et al. (2019). The original survey dataset contained 958 observations. After excluding duplicated entries because of the multiple waves of the survey distribution, there remained 907 observations to be initially analyzed.

What follows next is a concise description of each of the selected variables. Table 12 presents a summary of the variables, which includes the proportion of the sample with data and the categories of how the variables are initially coded. The analysis used only the observations of residents at the time of the hurricane.

4.3.2.1. Evacuation

The survey asked if the household evacuated from Hurricane Harvey. This question presents four possible outputs: (a) No, (b) Yes, (c) Not Resident, or (d) No Answer. For analysis, only the observations that answered Yes and No are selected; No is coded as 0, and Yes is coded as 1. In the initial dataset, 2.5% of the observations were not residents at the time, and 0.5% did not answer the question. These samples were excluded from the dataset. For the analysis, 60% of the households evacuated from Harvey at some point in time, and 37% did not evacuate. This variable is labeled *evacuation*.

4.3.2.2. Hurricane Evacuation Orders

Typically, a hurricane evacuation starts as a call for voluntary evacuations and at some point is elevated to a mandatory evacuation order (McCausland & Chuck, 2017). A hurricane evacuation order can allow residents to identify, and possibly better understand, that the risk of a hurricane is imminent and thus lead the public through a life-saving decision-making process.

Using communication from the Aransas Pass Police Department (https://police.aptx.gov/hurricane-harvey/), multiple local news articles published by the Texas Press Association (https://www.texaspress.com), and Alana Rocha from the State Operations Control (https://twitter.com/viaAlana/status/901130510117867521), the researchers could identify whether each of the households was in an area of voluntary or mandatory evacuation orders. This variable is coded as *evac_orders* and can assume the values of *mandatory* (= 1) or *voluntary* (= 0). It is worth noting that this measure is different from whether the respondent considered hurricane evacuation orders and is not necessarily applied to evacuation zones only.

4.3.2.3. Consideration of Hurricane Evacuation Orders

This question asked to what extent the residents weighed the local authorities' official recommendations to evacuate when deciding whether to evacuate. Answers to this question were on a scale of 1 to 5, where 1 represented *not at all considered* and 5 represented *considered to a very great extent*. In the initial dataset, 7.4% of the participants did not respond to this question. This variable is coded as *signif_evac_orders*.

4.3.2.4. Sociodemographic Characteristics (Age, Gender, and Education)

Sociodemographic characteristics not only provide context on the observations, but many studies also show that sociodemographic characteristics such as age, gender, and years of education can influence the perception of risk and, subsequently, the evacuation choice (Huang et al., 2016). These variables were self-reported in the survey. Approximately 58% of the respondents were male (coded as 0) and 42% were females (coded as 1). The variable gender is coded as *gender*. Among respondents, the minimum age was 20 years old, the maximum age was 94 years old, and the average age was 61.7 years old, with a standard deviation (SD) of 13.2. Ages have been reclassified into groups based on decades and the variable is coded as *age*. In regard to education, respondents were asked to reveal their highest level of education based on the following possible answers: some high school, some college/vocational school, college graduate, high school graduate/GED, or graduate school. These answers have been recoded to the highest year of education—up to 10, 12, 16, 18 or 21, respectively. On average, HHEBS respondents had higher levels of formal education than the area population; 47.1% of survey respondents reported having a bachelor's degree or higher, in comparison to 19.9% of the area population (Bierling et al., 2020). This variable is coded as *education*.

4.3.2.5. Risk Area

This variable was identified by the researchers using detailed evacuation maps to discover whether the household of the respondent is in a formal existing hurricane evacuation zone. This variable is coded as *risk_area* and can assume the values of Yes (= 1) or No (= 0). It is worth noting that this measure is different from whether the respondent acknowledged living in an area of high risk for storm surge or an evacuation zone.

4.3.2.6. Consult Sources of Information

In this study, *consult sources of information* assessed how many times per day, on average, the respondent consulted sources for information about the hurricane in the three days before landfall. It is assessed by the mean score of six sources of information: (a) local authorities (e.g., mayor, sheriff or police chief, emergency coordinator); (b) local news media (e.g., newspapers, radio stations, or television stations); (c) national news media (e.g., network news or Weather Channel); (d) the internet (e.g., National Hurricane Center website); (e) social media (e.g., Facebook and Twitter); and (f) phone or face-to-face contact with peers (such as friends, relatives, or neighbors). Answers to these questions were given on a scale of 1 to 5, respectively, based on consulting 0, 1-2, 3-4, 5-6, or 7 times or more on average per day. In the initial dataset, 94.5% of the observations responded to at least one of these questions. The mean of this variable was 3.0, and the SD was 1.20. This variable is coded as *consult_info*.

4.3.2.7. Expected Personal and/or Household Impacts

Although risk perception can be considered a complex analysis (Wachinger & Ren, 2010), expected impacts can be considered a form of risk perception and assessed in several ways (e.g., wind damage, surge damage, flood damage, casualties, job disruption, and service disruption; Huang et al., 2016). In this study, expected impacts try to capture the expectation of personal and/or household impacts and are assessed by the mean score of four questions from the survey: how likely the informant thought, as the storm was approaching, (1) that they or household members would be injured or killed if they stayed? (2) that their home would be inundated by storm surge? (3) that their home would be exposed to inland flooding? or (4) that their home would be severely damaged or destroyed by storm wind? These questions were answered on a scale of 1 to 5, where 1 represented *not at all likely* and 5 represented *almost certain*. In the initial dataset, 4.7% of the respondents did not respond to any of these questions. If only one or more of the questions were answered, it was computed as the average of the answers. This variable is coded as *expected_hh_impacts*.

4.3.2.8. Multiple Concerns

In this study, social cues are assessed by the mean score of five concerns: (a) concern about protecting their home from looters; (b) concern about protecting their home from storm impact; (c) concern about evacuation expenses, such as gas, food, and lodging; (d) concern about traffic accidents during the evacuation; and (e) concern about traffic jams during the evacuation. Answer to these questions were given on a scale of 1 to 5, where 1 represented *not at all considered* and 5 represented *considered to a very great extent*. In the initial dataset, 94.5% of the participants responded to at least one of these questions. The mean of this variable was 3.0, and the SD was 1.20. This variable is coded as *multiple_concerns*.

4.3.2.9. Social Cues

Social cues are forms of risk awareness that use peers as a source of information (Lindell et al., 2005). In this study, social cues are assessed by the mean score of four observations: (a) seeing businesses closing; (b) seeing friends, relatives, neighbors, or coworkers evacuating; (c) hearing announcements of watches and warnings; and (d) hearing local authorities recommending evacuation. Answers to these questions were given on a scale of 1 to 5, where 1 represented *not at all considered* and 5 represented *considered to a very great extent*. In the initial dataset, 93.4% of participants responded to at least one of these questions. The mean of this variable was 2.7, and the SD was 1.41. This variable is coded as *social_cues*.

4.3.2.10. Previous Unnecessary Hurricane Evacuation Experience

This question asked respondents to what extent they considered previous experience with an unnecessary evacuation when deciding whether to evacuate. In the initial dataset, 91.6% of the respondents responded to this question. Answers to this question were given on a scale of 1 to 5, where 1 represented *not at all considered* and 5 represented *considered to a very great extent*. In the initial dataset, 91.6% of the respondents responded to this question. The mean of this variable was 2.8, and the SD was 1.57. This variable is coded as *unnecessary_evac_exp*.

Table 12. Variables' Proportion of Sample with Data, Coding, Mean, and SD.

Variable (n = 907)	**Complete Rate ($n_{missing}$)**	**Coding (levels / categories)**	**Count & percent, or Mean & SD**	
evacuation	0.994 (5)	**1**: yes, **0**: no (drop not living in the place)	**yes**: 543 (60%) **no**: 336 (37%) **not res.**: 23 (2.5%)	
evac_orders	1.000 (0)	**1**: mandatory **0**: voluntary	**mandatory**: 571 (63%) **voluntary**: 336 (37%)	
gender	0.965 (32)	**1**: female **0**: male	**male**: 506 (58%) **female**: 369 (42%)	
risk_area	1.000 (0)	**1**: yes **0**: no	**yes**: 570 (63%) **no**: 337 (63%)	
age	0.963 (34)	min.: **20** max.: **94** (reclassified in decades)	61.7	13.20
consult_info	0.974 (24)	**1**: 0x, **2**: 1-2x, **3**: 3-4x, **4**: 5-6x, **5**: 7+/day	2.9	0.98
expected_hhs_impacts	0.953 (43)	From **1**: not at all likely to **5**: almost certain	2.3	1.04
signif_evac_orders	0.926 (67)	From **1**: not at all to **5**: very great extent	3.5	1.50
education	0.959 (37)	Up to **10**, **12**, **16**, **18** or **21** years of education	16.5	3.10
multiple_concerns	0.945 (50)	From **1**: not at all to **5**: very great extent	3.0	1.20
social_cues	0.934 (60)	From **1**: not at all to **5**: very great extent	2.7	1.41
unnecessary_evac_exp	0.916 (76)	From **1**: not at all to **5**: very great extent	2.8	1.57

A listwise deletion was used to handle the missing data (illustrated by Figure 26—in maroon are the missing data for each one of the variables, clustered, so that the approximate frequency that missing questions happen in the same observation can be visually identified). After that procedure, the number of observations (*n*) in the dataset drops from the original 907 to 771. Figure 26 also includes the number of the missing data for each variable and the percentage they represent.

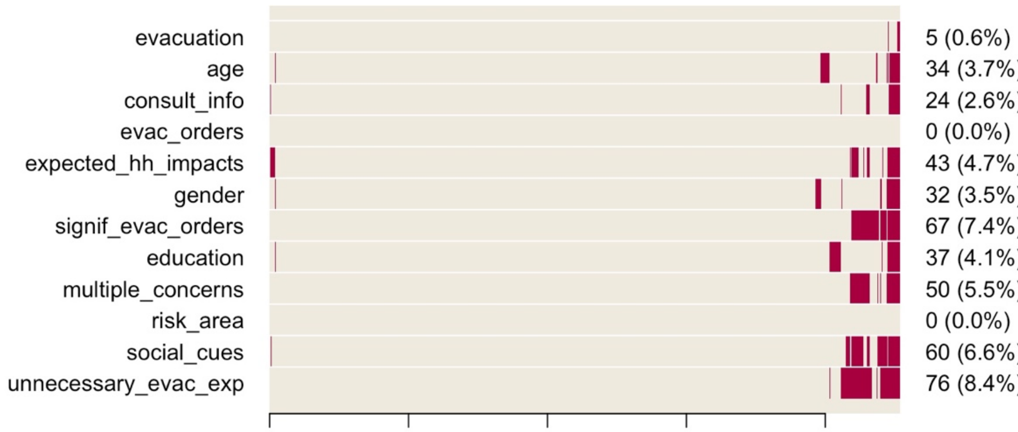


Figure 26. Missing survey data.

Table 13 presents the intercorrelations among the variables (Pearson correlation coefficient). The values in red are nonsignificant at $p \leq 0.05$. The correlation coefficient (r) can take on values between -1 and 1. The further away r is from zero, the stronger the linear relationship between the two variables. The matrix indicates that the highest correlations among these scales are the correlation between *social_cues* and *signif_evac_orders* ($r = 0.61$), *signif_evac_orders* and *expected_hh_impacts* ($r = 0.42$), and *signif_evac_orders* and *evacuation* ($r = 0.46$). Figure 27 presents illustratively the same information and the level of significance for each pair of variables.

Table 13. Intercorrelations Among the Variables. Values in red are nonsignificant at p ≤ 0.05.

	evacuation	expected_hh_impacts	signif_evac_orders	evac_orders	unnecessary_evac_exp	multiple_concerns	risk_area	social_cues	consult_info	age	education	gender
evacuation	1	0.36	0.46	0.28	-0.02	0.02	0.18	0.31	0.04	-0.01	-0.03	0.01
expected_hh_impacts		1	0.42	0.09	0.14	0.32	0.10	0.39	0.20	-0.12	-0.16	0.10
signif_evac_orders			1	0.15	0.15	0.26	0.10	0.61	0.21	-0.10	-0.05	0.14
evac_orders				1	0.03	-0.01	0.13	0.08	0.07	0.10	-0.12	0.02
unnecessary_evac_exp					1	0.41	-0.04	0.19	0.13	-0.03	-0.06	0.07
multiple_concerns						1	-0.02	0.36	0.20	-0.11	-0.19	0.09
risk_area							1	0.05	-0.03	0.13	0.09	-0.03
social_cues								1	0.19	-0.16	-0.11	0.20
consult_info									1	-0.12	-0.08	0.17
age										1	0	-0.12
education											1	-0.04
gender												1

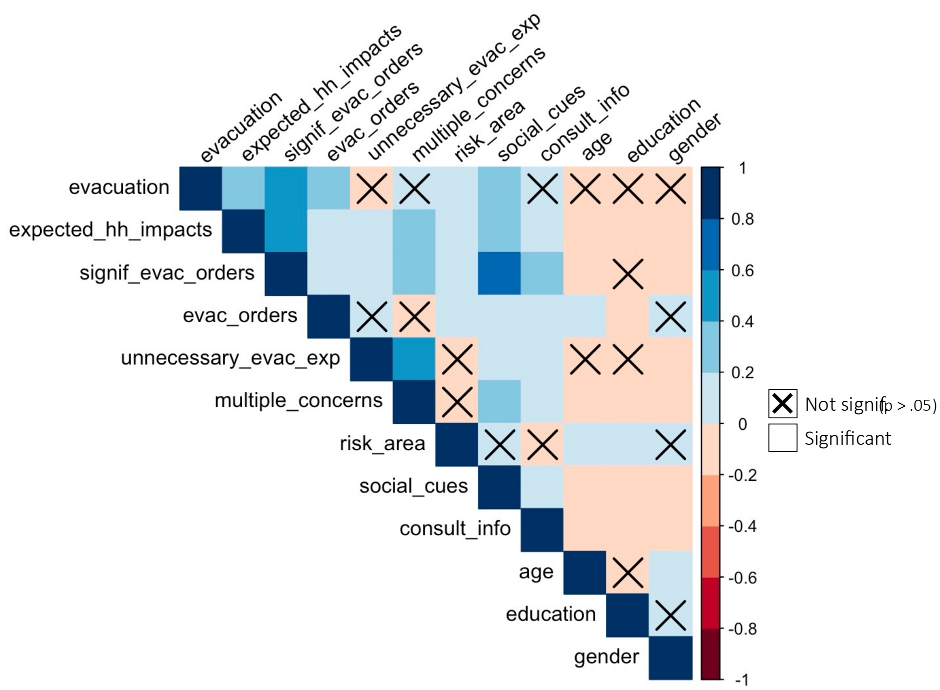


Figure 27. Illustrative of the intercorrelations among the variables.

A goal in this analysis is to study the relationship between the independent variables and evacuation (the dependent variable). For an initial (and exploratory) visualization of how this relationship can happen, an empirical cumulative distribution of each independent variable conditional on the evacuation response was made (see Figure 28, in blue when evacuation = 1 and in red when evacuation = 0). The cumulative percentage of responses is shown on the y-axis, and each level of the variables is shown on the x-axis. The empirical cumulative distribution functions seem to be an effective way of identifying variables whose evacuation may be conditioned. It can be noted that, for example, *signif_evac_orders* seem to be related to *evacuation*, but *age*, *gender*, or *education*, do not seem to be related to *evacuation*.

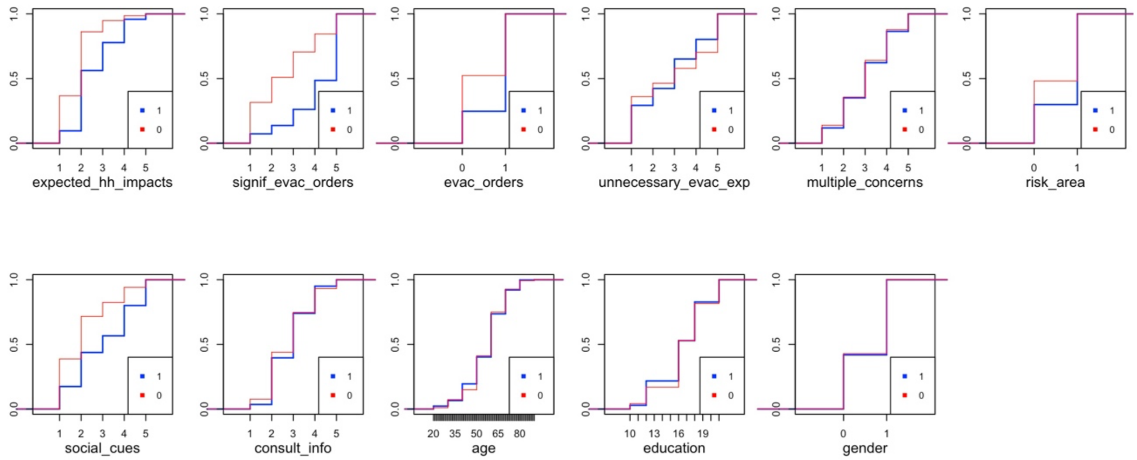


Figure 28. Empirical cumulative distribution function conditional to *evacuation*. In blue, evacuation = 1 (Yes). In red, evacuation = 0 (No).

4.3.3. An Initial Bayesian Network Model

One of the advantages of using probabilistic graphical modeling approaches is the ability to make connections between variables when these linkages represent hypotheses that can be tested by the model while supported by the data. Next, Figure 29 presents a BN model that substitutes the element of the conceptual model (in blue; see Figure 24, the adapted version of the PADM multistage model of hurricane evacuation) with the selected variables from the survey that can at least partially capture and describe each of the phenomena.

In this model, each link represents a conditional dependency and each node a random variable. Nodes that are not connected represent variables that are conditionally independent of each other. Each node is associated with a probability function that takes as input a particular set of values from the preceding nodes, if applicable, and creates as an output the probability distribution of the variable it represents.

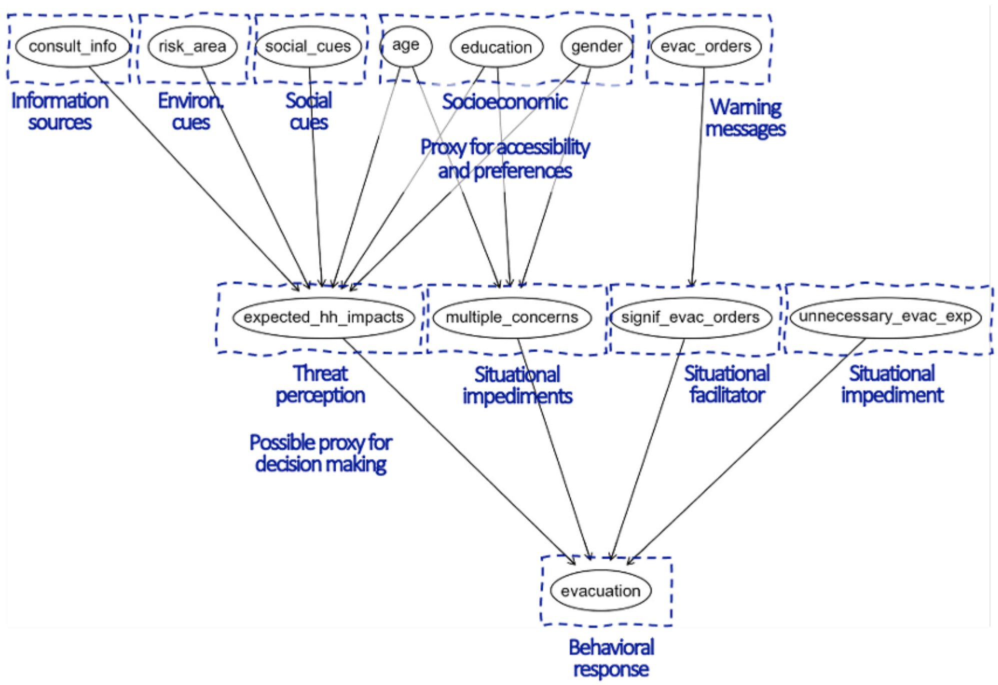


Figure 29. Probabilistic model of hurricane evacuation. Each node represents a variable of the survey that is associated with an element of the conceptual model.

Figure 30 repeats the BN model without the elements of the conceptual model, illustrating the conditional dependency of each of the variables. Evacuation is related to expectation of personal and/or household impacts (*expected_hhs_impacts*), consideration of multiple concerns about the storm and possible consequences (*multiple_concerns*), consideration of official evacuation recommendation (*signif_evac_orders*), and unnecessary evacuation experience (*unnecessary_evac_exp*). Considering official evacuation recommendation (*signif_evac_orders*) is supposedly conditional to having evacuation orders for the household. Considering multiple concerns about the storm and possible consequences (*multiple_concerns*) is conditional to *age*, *education*, and *gender*. And expectation of personal and/or household impacts (*expected_hhs_impacts*) is conditional to the frequency of consulting information about the storm (*consult_info*), the household being in a risk area (evacuation zone – *risk_area*), perceived social cues (*social_cues*), and age, education, and gender.

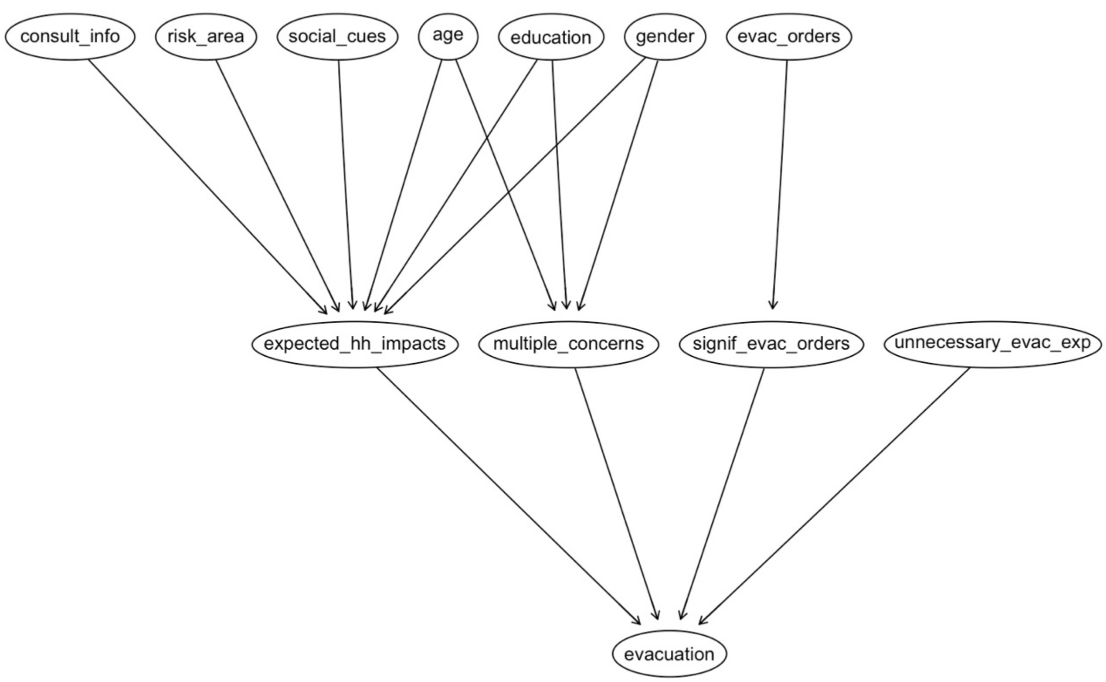


Figure 30. BN model of hurricane evacuation.

Figure 31 presents the probabilities of each variable. The model shows that 59.9% of the households will evacuate, while 40.1% will not evacuate. This estimate updates the prior number of 61.8% for the probability of a household evacuation (i.e., the number of observations that evacuated/total number of observations = 543/(543+336)) and the prior 38.2% of households that will not evacuate.

In addition, the network shows the probabilities for the other variables. The probability of unnecessary previous evacuation experience has 31.8% of the informants rating at the lowest level (1) and 23.5% of the informants rating it at the highest level (5). Considering evacuation orders has a distribution toward great or very great extent. Considering multiple concerns is relatively balanced in the middle range. Expecting personal and/or household impacts is toward the lowest share. It seems that very few people do not consult sources of information. The consideration of social cues is relatively balanced. Approximately 65% of the households were under mandatory evacuation orders, and 63.3% of the households were in an evacuation zone.

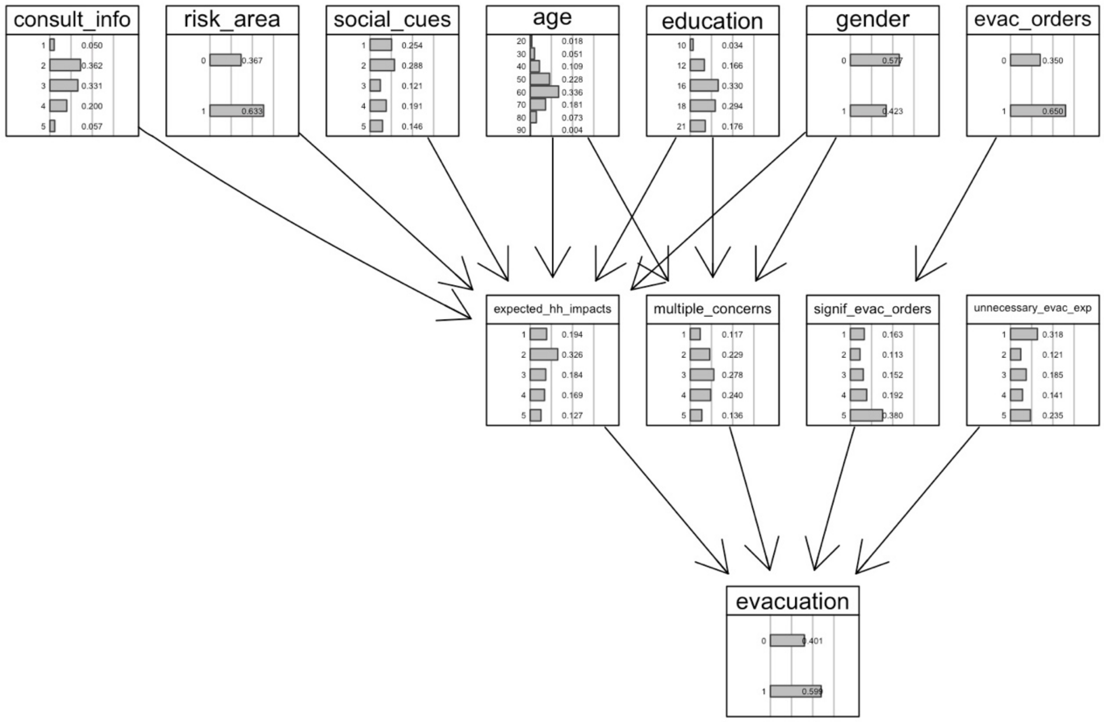


Figure 31. The joint probabilities for each variable.

The estimate network score Bayesian information criterion (BIC) is -66,485. To verify the assumptions of the dependence of the network, Figure 32 presents the strength of each linkage resulting from a mutual information (MI) test on each connection (with a threshold of $p < 0.05$).

MI is a statistic to measure the relatedness between two variables. The concept of mutual information is complex and is the basis of information theory. Compared with traditional measures such as correlation, mutual information can detect a wider range of relationships. For example, a zero-correlation coefficient does not necessarily imply that two variables are independent while zero mutual information is mathematically equivalent to independence. Mutual information between two discrete variables is conventionally calculated by their joint probabilities estimated from the frequency of observed samples in each combination of variable categories (Cover & Thomas, 1990). The MI tests detected that the data did not support the following hypotheses/linkages:

- *unnecessary_evac_exp* was not significant for *evacuation*.
- *risk_area* and *age* were not significant for *expected_hhs_impacts*.
- *age* and *gender* were not significant for *multiple_concerns*.

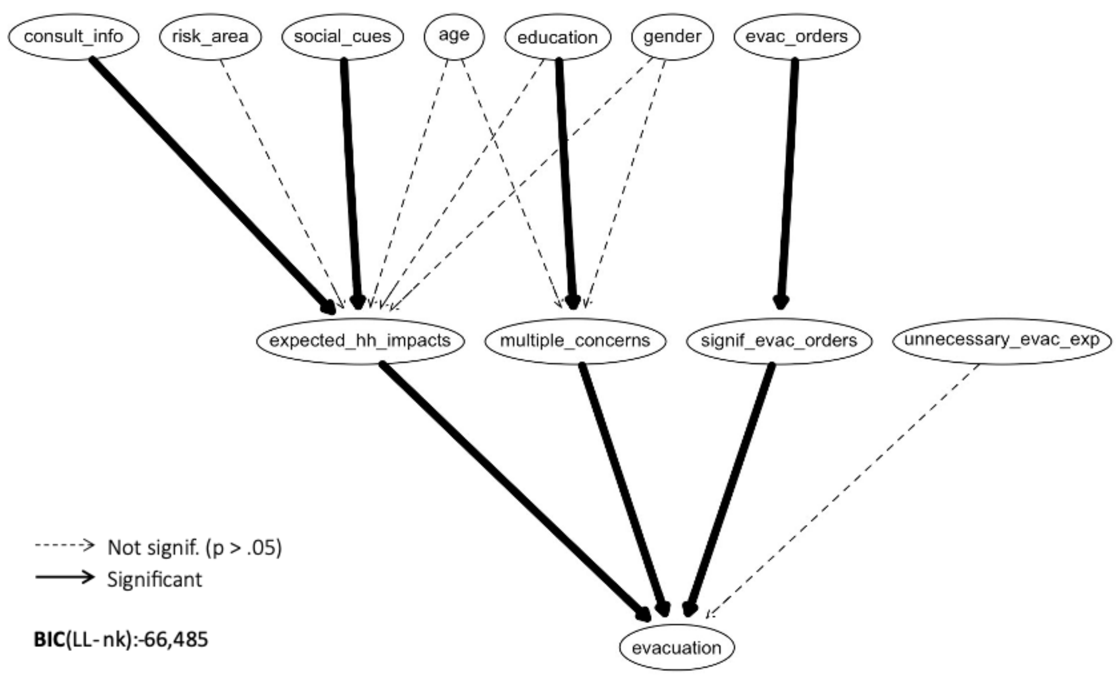


Figure 32. Results of the MI tests to examine linkage strength.

4.3.4. Adjusting the Bayesian Network Model

In Figure 28, it appears that *risk_area* can have dependency on *evacuation* (Figure 33 shows it in detail). This possible direct relationship was not tested in the BN above.

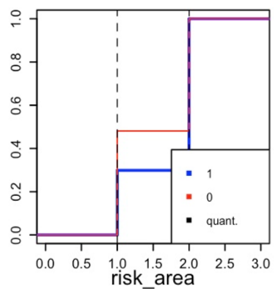


Figure 33. Empirical cumulative distribution function conditional of *risk_area* to *evacuation*.

The linkage between *risk_area* and *expected_hhs_impacts* was not significant. However, *risk_area* is significantly directly connected with *evacuation* (see Figure 34). The network score BIC of the updated version of the network is -68,484. This value is smaller than the previous BIC, which implies that this network structure fits the data worse than the preceding structure.

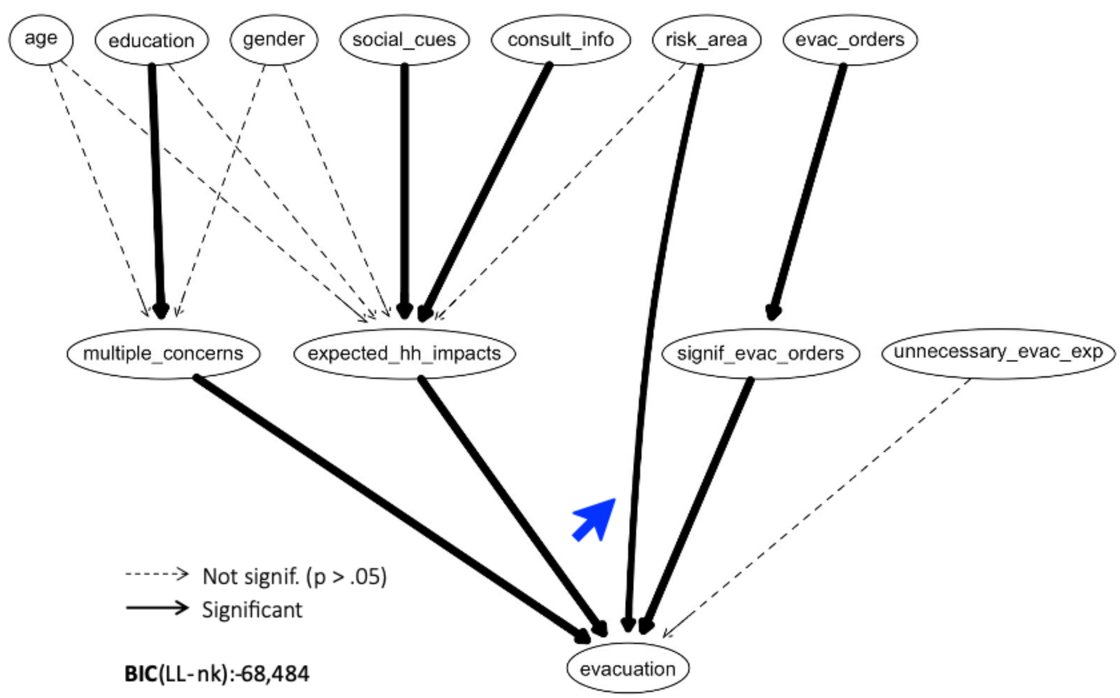


Figure 34. BN model updated: *risk_area* directly influencing *evacuation*.

New connections can be tested. Figure 35 analyzes the possibility that *age* influences *unnecessary_evac_exp* by taking into account that older people may have lived longer in the region and have had a greater chance of experiencing unnecessary evacuation. This network update also tested whether *age* influences *signif_evac_orders*, *multiple_concerns*, and *expected_hhs_impacts*. *Age* was not significant for any of these hypotheses.

This network also tests the hypotheses that *gender* can influence *signif_evac_orders*, *multiple_concerns*, and *expected_hhs_impacts*. *Gender* was significant for *signif_evac_orders* only. *Age* and *gender* were not significant for *education*. *Education* was significant for *multiple_concerns*, but not significant for *signif_evac_orders* and *expected_hhs_impacts*. The updated network score BIC is -70,558, indicating that this network structure fits the data even worse.

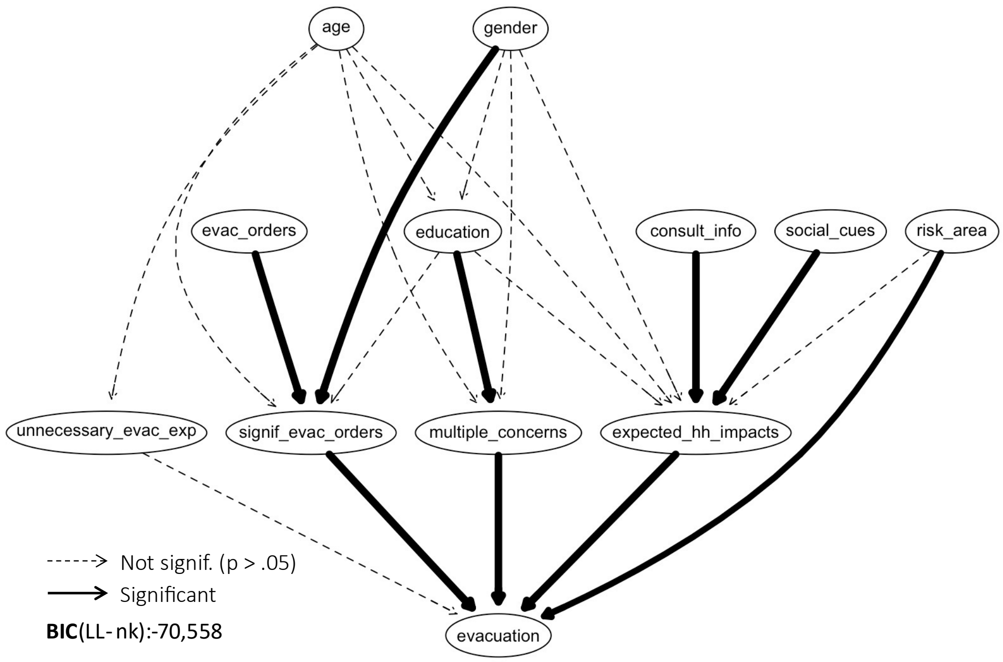


Figure 35. BN model updated: testing multiple hypothesis regarding the sociodemographic variables *age*, *gender*, and *education*.

Based on the results obtained by the updated model above, Figure 36 presents a *fitted* model using only significant linkages, and that model reaches a much higher network score BIC value (-10,041), indicating a network structure that *better* fits the data. *Education* is the only influence in *multiple_concerns*. *Evac_orders* and *gender* influence *signif_evac_orders*. *Consult_info*, *social_cues*, and *risk_area* influence *expected_hhs_impacts*. *Risk_area*, *signif_evac_orders*, and *multiple_concerns* influence *evacuation*.

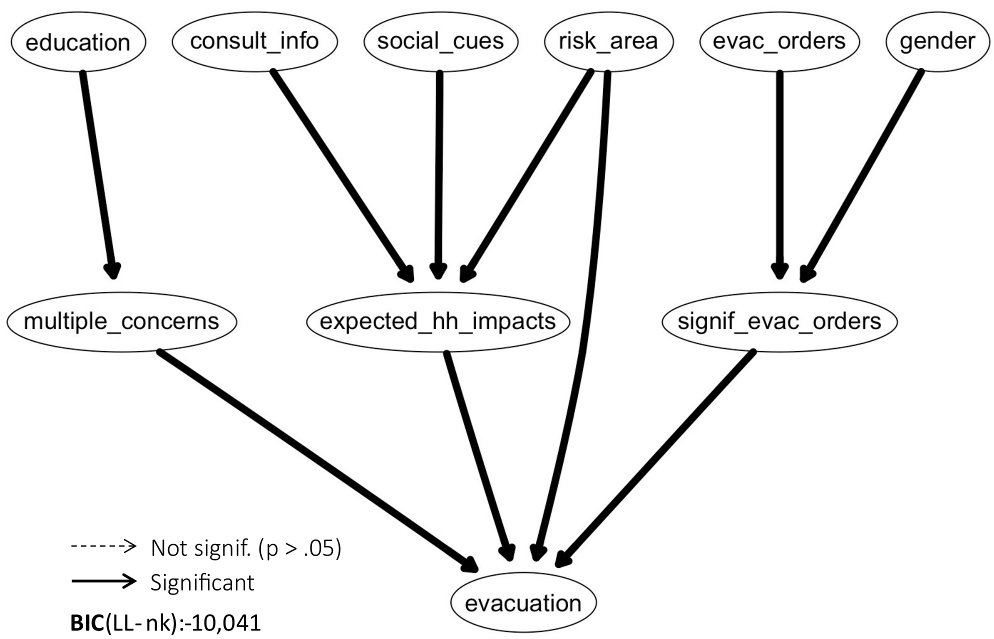


Figure 36. BN model updated: fitted model.

Figure 37 presents the updated joint probabilities of the network. The probability that a household evacuates changes to 63.0% (the first network had this probability at 59.9%).

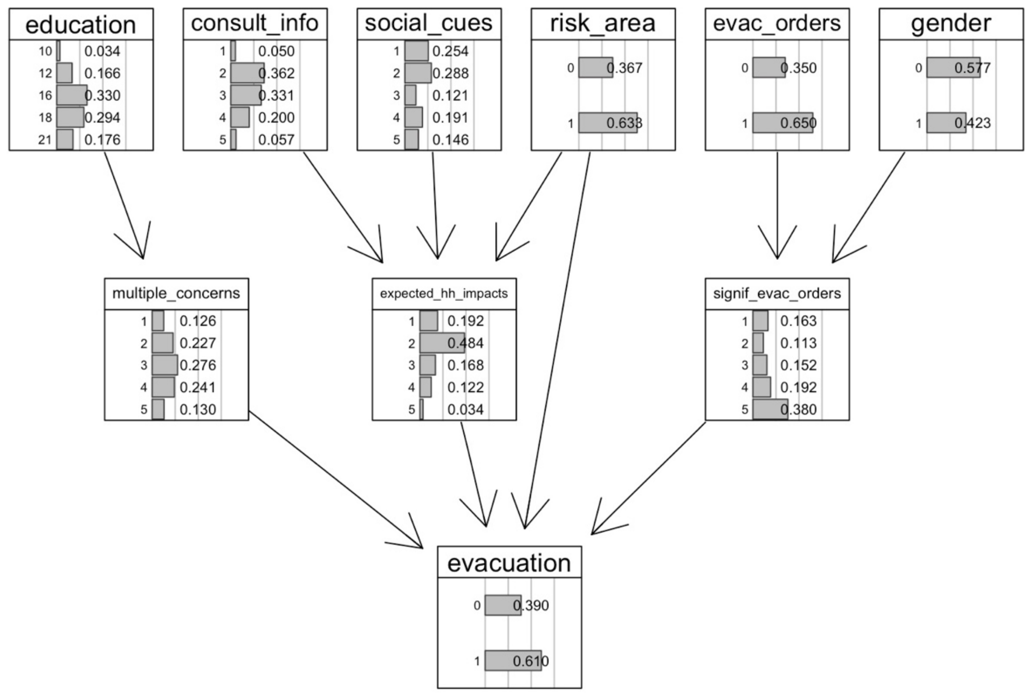


Figure 37. The updated joint probabilities for each variable.

4.3.5. Experiment 1: Probabilities for Fixing multiple_concerns to the Lowest Level

In addition to testing hypotheses in the connection of variables, BNs allow testing changes in joint probability by fixing the state of particular variables. This experiment analyzes the prediction ability of the BN model. To demonstrate a typical application in disaster planning, Figure 38 shows how changes can be examined in the evacuation probability by fixing the state of multiple_concerns to the lowest level (not considered important when deciding to evacuate = 1). Interestingly, the evacuation probability has increased to 73.7% from 61.0%, which may indicate that not having concerns about multiple factors (about protecting home from looters and from storm impact, about evacuation expenses, about traffic accidents and traffic jams during evacuation) can increase the evacuation participation by approximately 21%.

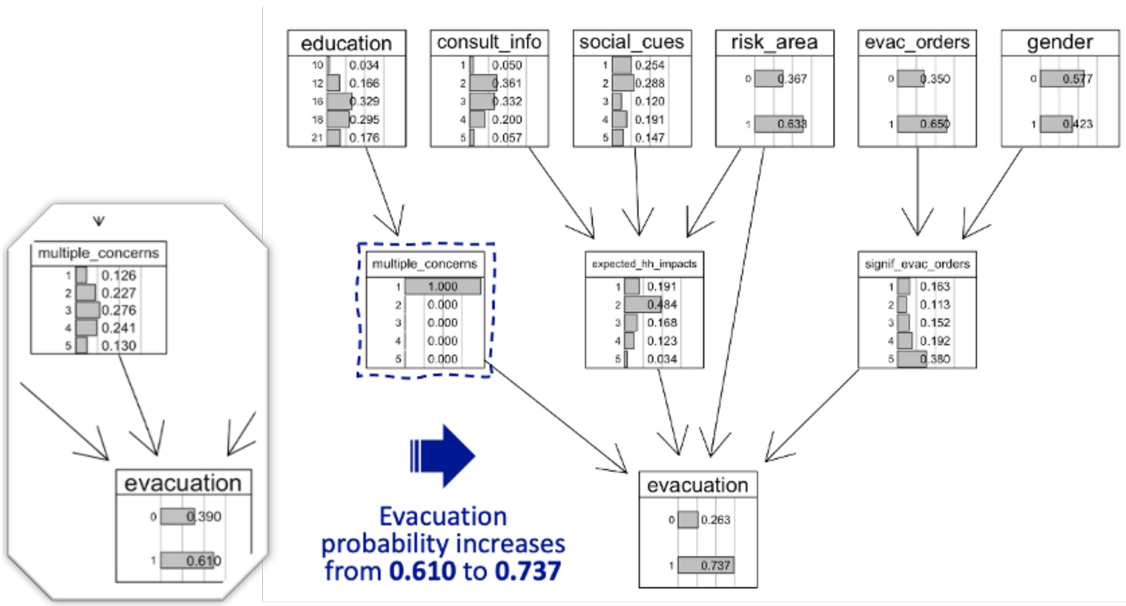


Figure 38. Joint probabilities after fixing *multiple_concerns* to the lowest level.

4.3.6. Experiment 2: Probabilities for Fixing Evacuation

This experiment explores, by using Bayes' theorem, the possibility of knowing that a household has evacuated, but it is desirable to investigate the influence of each predictor level on such an output. The evidence that the household has evacuated is so-called soft evidence. Although it analyzes the preceding nodes of an outcome, it is usually acknowledged in Bayesian literature as a prognostic analysis.

Table 14 presents the results for all levels on the three predictors of evacuation. If the evidence shows that the household did evacuate (*evacuation* = 1), the probability of *expected_hh_impacts* is lower (2 has the greater probability), and *signif_evac_orders* is higher (5 has the greater probability). This information can be very useful for disaster planning. The probability that the household was in the risk area slightly changed to 68.7%. To enhance evacuation, assuming that residents need to evacuate, emergency managers need to focus on intensifying the risk perception in relation to *signif_evac_orders*, which most impacts the probability of a household evacuating, according to the specific surveyed data and model structure.

Table 14. Prognostic Results on the Predictors of Evacuation.

Levels	Predictors' Probabilities		
	expected_hh_impacts	*signif_evac_orders*	*risk_area*
1	0.124	0.107	0.313
2	0.493	0.068	0.687
3	0.194	0.130	–
4	0.154	0.218	–
5	0.036	0.477	–

4.4. Discussion and Conclusions

Analysis of evacuation behavior is critical for disaster preparedness and planning, particularly hurricane household evacuation. Modeling the complexity involved in the household evacuation decision-making process is an ongoing challenge effectively captured by conceptual models such as the PADM. However, the mathematical modeling of such problems is not a simple task due to the heterogeneity of (a) the social vulnerabilities of populations impacted, (b) the areas affected by these compounding threats, and (c) the indefinite mediating and interacting effects of diverse (risk) factors. Most studies use LR to examine the wide range of affective factors, which provides a rigorous analytical framework. However, although many prediction models are available, the uptake in planning practice and public policy is relatively slow. This can happen for different reasons: (a) struggles in the implementation and use of statistical models, and (b) difficulties in interpretation of what the models are doing by researchers and, more frequently, by authorities and the general population.

BNs are used in several fields and have been widely applied to decision-making problems because they combine the benefits of formal probabilistic methods, an engaging visual form, and efficient computational techniques to explore complex arrangements of predictors for an outcome. Hurricane evacuation is a repeated and important decision-making problem for many coastal communities exposed to this natural hazard; importantly, although BNs currently have limited exploration in this field, they may prove to be quite useful.

This study began with the identification of two objectives: first, to model hurricane household evacuation through BNs and acknowledge the factors that affected the evacuation choice in a recent event; and second, to find and demonstrate some advantages and limitations to this novel approach in this specific field of disaster planning. This study develops a BN model that is consistent with an abbreviated (reduced) version of the PADM and uses data collected in a survey after Hurricane Harvey passed through the Texas Coastal Bend area in August 2017, which includes Aransas, Calhoun, Jackson, Matagorda, Nueces, Refugio, San Patricio, and Victoria Counties. These data consist of a random sample of 826 completed cases. The original survey dataset contained 958 observations. After excluding duplicated entries because the multiple waves of the survey distribution, and listwise deletion of missing data, the number of observations in the dataset dropped to 771.

The analysis considered 11 of the main and most recurrent reasons that influence household evacuation, according to research to date: (a) previous unnecessary hurricane evacuation experience, (b) perceived social cues, (c) multiple concerns about the household and/or the evacuation, (d) expected personal and/or household impacts, (e) frequency of consulting sources of information about the storm, (f) the location of the residence in a risk zone (evacuation zone), (g) hurricane evacuation orders in the location of the residence, (h) consideration of hurricane evacuation orders, (i) age, (j) gender, and (k) the highest level of education.

An initial BN model can more directly capture the conceptual model. The results of this model show that the factors significant for the evacuation choice were relatively consistent with the correlation analysis. The intercorrelation test were at a significance level of 0.05, and all tests of statistical significance of the linkages in the network use MI with a significance level of 0.05 as well.

The factors influencing evacuation in this study were not necessarily new, while the survey instrument was based on previous research, and the model development was based on the conceptual model broadly accepted in the literature. However, the proposed model was built graphically and explicitly to acknowledge the conceptual model structure. Later, the adjustments in the model aided the update of the prior evacuation probability and the description of the flow of information by the identification of statistically significant influences on evacuation—both direct and indirect effects.

Unlike previous research (Huang et al., 2017), unnecessary evacuation experience was not correlated with evacuation and did not indicate statistical significance in the BN model. This finding is intriguing, and this indication should be specifically considered in the future.

Interestingly, multiple concerns were not correlated with evacuation; however, it was statistically significant in the BN model. Note that the prior probability distribution of multiple concerns was first adjusted by the demographic factors (age, gender, and education) but later in the fitted network by education only.

The influence of consulting information and the sociodemographic variables on evacuation were not directly tested in the BN model. These variables did not have a significant correlation with evacuation, but consulting information and education had an influence on expected personal and/or household impacts and multiple concerns, respectively. Therefore, the BN model should better analyze the mediating effects on the evacuation.

In addition to the correlation analysis, the visualization of the empirical cumulative distribution function conditional to evacuation displayed a visual indication of possible dependence on evacuation. Thus, the BN model was adjusted to connect the risk area directly to evacuation. Note that the risk area was not significant for expected personal and/or household impacts, as initially hypothesized by the link in the model.

However, the adjusted model showed that the risk area was ultimately statistically significant and influential in the evacuation choice.

A second adjustment to the BN model tested the influence of sociodemographic variables with risk factors, as described below. Gender was statistically significant for considering evacuation orders only. Age and gender were not statistically significant for education. Education was statistically significant for multiple concerns but not significant for considering evacuation orders and expected personal and/or household impacts.

Finally, a fitted model was presented with only statistically significant linkages. As expected, the network score BIC result of this network shows that this model fits the data better than the previous models. With this model, two typical situations in disaster planning were tested. First, an analysis of how much the probability of evacuation was modified by setting multiple concerns to the lowest level (i.e., not considered important in deciding to evacuate) showed that evacuation increases (to 73.7% from 61.0%). It can be assumed that not having concerns about protecting one's home from looters and from the storm, or about evacuation costs, traffic accidents, and traffic jams during evacuation will make residents more disposed to evacuate if instructed to do so. What should be studied in the future are actions that might make this factor fall to such a level.

The second experiment tested the probabilities of factors that most influenced evacuation. The results showed that intensifying the risk perception in relation to considering evacuation orders most influenced the probability of a household evacuating, according to this specific surveyed data and model structure. As can be observed, BN can facilitate the discussion of uncontrolled risk factors (e.g., gender and age) versus controlled risk factors (e.g., establishment of mitigation policies, official warnings, and definition of effective risk areas / evacuation zones). The use of graphs for modeling is not necessarily something new in statistical analysis. However, BNs specifically offer a friendly approach for modeling complex relationships, and the calculations involved are of relatively easy estimation and interpretation, as shown in the previous study (Chapter 3, page 30).

A difficulty in developing a BN model was the discretization of the variables in order to find meaningful levels to the problem of evacuation while creating probabilities for all the combinations of levels given the network structure and the number of observations analyzed without using prior beliefs for missing combinations. Future research should include and test a more diverse set of variables and domains in the BN model while also testing these model structures with different datasets. In addition, future research should include an evacuation decision variable before the evacuation to assess the difference between the decision to evacuate and the evacuation occurrence, which can be limited by many factors, including sociodemographic ones.

5. CONCLUSIONS

To offer a consistent approach to the risk assessment problem using BNs, Chapter 2 adapted a risk framework from the literature that facilitated the structure of critical processes. Although this approach can serve as a guideline to specific management of risks due to natural hazards, this study focused on the representation of a part of the system. The proposed framework was hierarchically structured based on how exposure to natural hazards, risk, and different sources of knowledge interact, and considered vulnerability as the link to the magnitude of direct consequences. BNs allow a detailed evaluation of the joint influence of the different indicators on the risk, providing results that, in contrast to traditional methodologies, are consistent with the mathematical (probabilistic) concept of risk and can be directly used for optimization purposes. In the illustrative examples, BNs can represent some of the complexities of the urban environment, such as the combination of physical and social vulnerabilities, while predicting economic and social losses. BN modeling ensures that models can be further extended when additional (or complementary) information is included or an examination by different stakeholders is performed, and a potential unavailability of indicators can be assessed by prior beliefs.

None of the findings of the experiments in Chapter 3 are unknown regarding BN. The major contribution of this study is the validation of the application of BN in disaster planning problems, more specifically the analysis of hurricane evacuation. Although household evacuations in natural disaster have been often studied, constant examining of methods to predict evacuation behavior is relevant for more accurate outcomes and a deeper understanding of the choices a population makes.

The proposed model in Chapter 4 to study evacuation choices was built graphically and explicitly to acknowledge the conceptual model structure. The adjustments developed in the model allowed for the update of the evacuation's prior probability, and the description of the flow of information by the identification of both direct and indirect effects showed significant influences on evacuation. Different than previous research, unnecessary evacuation experience was not correlated with evacuation and did not indicate statistical significance in the proposed model. This finding is intriguing, and this indication should be specifically considered in the future.

The major methodological contribution of these studies is the validation of the application of BNs in disaster planning problems. The results of this dissertation suggest that BNs can be useful for disaster planning applications in the future by more directly capturing conceptual models and enabling the participation of multiple actors in the analysis of risk and risk countermeasures, thereby not limiting the process of risk assessment to experts only but instead creating a broader and more inclusive understanding of hurricane risks in communities. The use of BNs to study household hurricane evacuation can help members of the community, planners, and emergency managers get involved and jointly learn and explore factors that affect evacuation behavior. Furthermore, rather than competing with traditional methods, BNs can be a complementary and regular method to further assist the examination of gaps in the disaster planning literature.

This investigation also raised some new questions. Recurrently, an expressive number of residents at risk do not evacuate from an approaching hurricane when they are advised to do so by local authorities. This causes unnecessary suffering, injuries, and ultimately deaths. Such situations are likely to intensify based on a predicted increase in frequency and intensity of tropical storms and because of abrupt climate change.

Although there are diverse studies on the evacuation behavior and choices, non-evacuation is still not fully understood; both theoretical and policy framing of the decision to evacuate are mostly centered around logical and socioeconomic approaches that assume that risk is objective, and people will rationally evacuate if they have the material means to do so. Conventional (neoclassical) economic models and most policy makers assume that rational individuals should be able to keep informed of weather conditions, determine the probability of being impacted by a hurricane and wonder how serious the impact would be, and make the decision to evacuate (or not) accordingly.

With more data and prospective surveys, I hope to be able to identify and formulate environmental and social contexts and psychological processes based on people's responses to environmental hazards and disasters. Most studies on this topic still barely explain illogical beliefs (e.g., religiosity, faith, superstitions, and fads), decision paralysis (e.g., procrastination), and certain common and undiagnosed diseases and conditions in the population (e.g., obsessive, and compulsive behaviors). Some irrational behaviors in sports, academics, and economics have been analyzed in regard to prevalence and performance and have been postulated to be positively associated with external locus of control, high role-identity, ambiguous intolerance, and high stress situations. These factors can also be very pertinent to situations when someone (or a household) is threatened by natural hazards. However, to date there has been no research exploring how irrational behaviors affect people's responses to environmental hazards and disasters. In addition, no research has been conducted that analyzes individual irrational behaviors within the need to evacuate from hurricanes.

REFERENCES

Achumba, I., Azzi, D., Ezebili, I., & Bersch, S. (2013). Approaches to Bayesian network model construction. In *IAENG Transactions on Engineering Technologies* (pp. 461-474). Springer.

Agresti, A. (2003). *Categorical data analysis* (Vol. 482). John Wiley & Sons.

Aven, T. (2012). *Foundations of risk analysis: A knowledge and decision-oriented perspective.* John Wiley & Sons.

Baker, E. J. (1991). Hurricane evacuation behavior. *International Journal of Mass Emergencies and Disasters, 9*(2), 287-310.

Balbi, S., Villa, F., Mojtahed, V., Hegetschweiler, K. T., & Giupponi, C. (2016). A spatial Bayesian network model to assess the benefits of early warning for urban flood risk to people. *Natural Hazards and Earth System Sciences, 16*(6), 1323-1337.

Bateman, J. M., & Edwards, B. (2002). Gender and evacuation: A closer look at why women are more likely to evacuate for hurricanes. *Natural Hazards Review, 3*(3), 107-117.

Bayraktarli, Y. Y., Ulfkjaer, J.-P., Yazgan, U., & Faber, M. H. (2005). On the application of Bayesian probabilistic networks for earthquake risk management. In *9th International Conference on Structural Safety and Reliability (ICOSSAR 05)* (pp. 20-23). Millpress.

Bengtsson, L. (2001). Hurricane threats. *Science, 293*(5529), 440-441.

Berke, P. R. (1998). Reducing natural hazard risks through state growth management. *Journal of the American Planning Association, 64*(1), 76-87.

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *Journal of the Royal Statistical Society: Series B (Methodological), 41*(2), 113-128.

Bierling, D. H., Lindell, M. K., Peacock, W. G., Abuabara, A., Moore, R. A., Wunneburger, D. F., Mullins, J. A., Borchardt, D. W. (2020). *Coastal Bend Hurricane Evacuation Study: Hurricane Harvey evacuation behavior survey outcomes and findings.* Texas A&M University Hazard Reduction & Recovery Center.

Blaser, L., Ohrnberger, M., Riggelsen, C., Babeyko, A., & Scherbaum, F. (2011). Bayesian networks for tsunami early warning. *Geophysical Journal International, 185*(3), 1431-1443.

Burby, R., Beatley, T., Berke, P., Deyle, R., French, S., Godschalk, D., . . . Olshansky, R. (1999). Unleashing the power of planning to create disaster-resistant communities. *Journal of the American Planning Association, 65*(3), 247-258.

Burkart, S., & Király, F. (2018). *Predictive independence testing, predictive conditional independence testing, and predictive graphical modelling.* University College London.

Burton, C. G. (2010). Social vulnerability and hurricane impact modeling. *Natural Hazards Review, 11*(2), 58-68.

Cegan, J. C., Golan, M. S., Joyner, M. D., & Linkov, I. (2022). The importance of compounding threats to hurricane evacuation modeling. *Urban Sustainability, 2*(1), 1-4.

Cinar, D., & Kayakutlu, G. (2010). Scenario analysis using Bayesian networks: A case study in energy sector. *Knowledge-Based Systems, 23*(3), 267-276.

Comfort, L., Wisner, B., Cutter, S., Pulwarty, R., Hewitt, K., Oliver-Smith, A., . . . Krimgold, F. (1999). Reframing disaster policy: the global evolution of vulnerable communities. *Global Environmental Change Part B: Environmental Hazards, 1*(1), 39-44.

Cover, T. M., & Thomas, J. A. (1990). *Elements of information theory.* John Wiley & Sons.

Cutter, S., Boruff, B., & Shirley, W. (2003). Social vulnerability to environmental hazards. *Social Science Quarterly, 84*(2), 242-261.

Dash, N., & Gladwin, H. (2007). Evacuation decision making and behavioral responses: Individual and household. *Natural Hazards Review, 8*(3), 69-77.

Dickson, E., Baker, J., Hoornweg, D., & Asmita, T. (2012). *Urban risk assessments: An approach for understanding disaster and climate risk in cities.* The World Bank.

Dyckman, J. W. (1961). Planning and decision theory. *Journal of the American Institute of Planners, 27*(4), 335-345.

Ellis, B., & Wong, W. H. (2008). Learning causal Bayesian network structures from experimental data. *Journal of the American Statistical Association, 103*(482), 778-789.

Emanuel, K., & Jagger, T. (2010). On estimating hurricane return periods. *Journal of Applied Meteorology and Climatology, 49*(5), 837-844.

Faizian, M., Schalcher, H., & Faber, M. (2005). Consequence assessment in earthquake risk management using damage indicators. In *9th International Conference on Structural Safety and Reliability (ICOSSAR 05)* (pp. 19-23). Millpress.

Fenton, N., & Neil, M. (2011). The use of Bayes and causal modelling in decision-making, uncertainty, and risk. *CEPIS Upgrade, 12*(5), 10-21.

Foraita, R., Spallek, J., & Zeeb, H. (2014). Directed acyclic graphs. In W. Ahrens & I. Pigeot (Eds.), *Handbook of epidemiology* (pp. 1481-1517). Springer.

Gardoni, P., Murphy, C., & Rowell, A. (2016). Risk analysis of natural hazards: Interdisciplinary challenges and integrated solutions. In P. Gardoni, C. Murphy, & A. Rowell (Eds.), *Risk analysis of natural hazards* (pp. 1-7). Springer.

Ge, Y., Peacock, W. G., & Lindell, M. K. (2011). Florida households' expected responses to hurricane hazard mitigation incentives. *Risk Analysis, 31*(10), 1676-1691.

Geiger, D., Verma, T., & Pearl, J. (1990). d-separation: From theorems to algorithms. *Machine Intelligence and Pattern Recognition*, *10*, 139-148.

Gelman, A., & Speed, T. (1993). Characterizing a joint probability distribution by conditionals. *Journal of the Royal Statistical Society: Series B (Methodological), 55*(1), 185-188.

Gladwin, C. H., Gladwin, H., & Peacock, W. G. (2001). Modeling hurricane evacuation decisions with ethnographic methods. *International Journal of Mass Emergencies and Disasters, 19*(2), 117-143.

Godschalk, D. R., Brower, D. J., & Beatley, T. (1989). *Catastrophic coastal storms: Hazard mitigation and development management.* Duke University Press.

Godschalk, D. R., Norton, R., Richardson, C., & Salvesen, D. (2000). Avoiding coastal hazard areas: Best state mitigation practices. *Environmental Geosciences, 7*(1), 13-22.

Hasan, S., Ukkusuri, S., Gladwin, H., & Murray-Tuite, P. (2011). Behavioral model to understand household-level hurricane evacuation decision making. *Journal of Transportation Engineering, 137*(5), 341-348.

Heath, R. L., Lee, J., Palenchar, M. J., & Lemon, L. L. (2018). Risk communication emergency response preparedness: Contextual assessment of the protective action decision model. *Risk Analysis, 38*(2), 333-344.

Highfield, W. E., Peacock, W. G., & Van Zandt, S. (2014). Mitigation planning: Why hazard exposure, structural vulnerability, and social vulnerability matter. *Journal of Planning Education and Research, 34*(3), 287-300.

Holmes, D. (2008). *Innovations in Bayesian networks: Theory and applications* (Vol. 156). Springer.

Hoo, Z. H., Candlish, J., & Teare, D. (2017). What is an ROC curve? *Emergency Medicine Journal, 34*(6), 357-359.

Hosmer, D. W., Jr., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression.* John Wiley & Sons.

Huang, S.-K., Lindell, M. K., & Prater, C. S. (2016). Who leaves and who stays? A review and statistical meta-analysis of hurricane evacuation studies. *Environment and Behavior, 48*(8), 991-1029.

Huang, S.-K., Lindell, M. K., & Prater, C. S. (2017). Multistage model of hurricane evacuation decision: Empirical study of Hurricanes Katrina and Rita. *Natural Hazards Review, 18*(3), 05016008.

International Organization for Standardization. (2021). *Popular standards*. https://www.iso.org/iso-31000-risk-management.html

Jager, W., Christie, E., Hanea, A., den Heijer, C., & Spencer, T. (2018). A Bayesian network approach for coastal risk analysis and decision making. *Coastal Engineering, 134*, 48-61.

Jasour, Z. Y., Davidson, R. A., Trainor, J. E., Kruse, J. L., & Nozick, L. K. (2018). Homeowner decisions to retrofit to reduce hurricane-induced wind and flood damage. *Journal of Infrastructure Systems, 24*(4), 04018026.

Jensen, F. (1996). *An introduction to Bayesian networks.* UCL Press.

Jensen, F. V., & Nielsen, T. D. (2007). *Bayesian networks and decision graphs.* Springer.

Kang, C., & Tian, J. (2012). Local Markov property for models satisfying composition axiom. *arXiv e-Print, 1207*(1378), 1-8.

Keim, B., Muller, R., & Stone, G. (2007). Spatiotemporal patterns and return periods of tropical storm and hurricane strikes from Texas to Maine. *Journal of Climate, 20*(14), 3498-3509.

Kleinbaum, D. G., Dietz, K., Gail, M., Klein, M., & Klein, M. (2002). *Logistic regression.* Springer.

Knutson, T. (2021). *Global warming and hurricanes: An overview of current research results.* NOAA's Geophysical Fluid Dynamics Laboratory. https://www.gfdl.noaa.gov/global-warming-and-hurricanes/

Kopp, G. A., Morrison, M. J., & Henderson, D. J. (2012). Full-scale testing of low-rise, residential buildings with realistic wind loads. *Journal of Wind Engineering and Industrial Aerodynamics, 104*, 25-39.

Laitin, D. D. (2003). The Perestroikan challenge to social science. *Politics & Society, 31*(1), 163-184.

Lazo, J. K., Bostrom, A., Morss, R. E., Demuth, J. L., & Lazrus, H. (2015). Factors affecting hurricane evacuation intentions. *Risk Analysis, 35*(10), 1837-1857.

Lee, J. Y., & Van Zandt, S. (2019). Housing tenure and social vulnerability to disasters: A review of the evidence. *Journal of Planning Literature, 34*(2), 156-170.

Li, M., Liu, J., Li, J., & Kim, B. U. (2014). Bayesian modeling of multi-state hierarchical systems with multi-level information aggregation. *Reliability Engineering & System Safety, 124*, 158-164.

Lindell, M. K., & Perry, R. W. (2012). The protective action decision model: Theoretical modifications and additional evidence. *Risk Analysis: An International Journal, 32*(4), 616-632.

Lindell, M. K., Ge, Y., Huang, S.-K., Prater, C. S., Wu, H.-C., & Wei, H.-L. (2013). *Behavioral study: Valley hurricane evacuation study for Willacy, Cameron, and Hidalgo Counties, Texas.* Texas A&M University Hazard Reduction & Recovery Center.

Lindell, M. K., Lu, J. C., & Prater, C. S. (2005). Household decision making and evacuation in response to Hurricane Lili. *Natural Hazards Review, 6*(5), 171-179.

Lindell, M. K., Murray-Tuite, P., Wolshon, B., & Baker, E. J. (2019). *Large-scale evacuation: The analysis, modeling, and management of emergency relocation from hazardous areas.* Routledge.

Lindell, M. K., Prater, C. S., Sanderson, W. G., Lee, H. M., Zhang, Y., Mohite, A., & Hwang, S. N. (2001). *Texas Gulf Coast residents' expectations and intentions regarding hurricane evacuation.* Texas A&M University Hazard Reduction & Recovery Center.

Marzban, C. (2004). The ROC curve and the area under it as performance measures. *Weather and Forecasting, 19*(6), 1106-1114.

Masoomi, H., Ameri, M. R., & van de Lindt, J. W. (2018). Wind performance enhancement strategies for residential wood-frame buildings. *Journal of Performance of Constructed Facilities, 32*(3), 04018024.

Masterson, J., Peacock, W., Van Zandt, S., Grover, H., Schwarz, L., & Cooper, J. (2014). Assessing physical vulnerability. In *Planning for community resilience* (pp. 83-96). Springer.

McCausland, P., & Chuck, E. (2017). *Hurricane Harvey evacuations: Residents warned to leave, stay away*. NBC News. https://www.nbcnews.com/storyline/hurricane-harvey/hurricane-harvey-evacuations-residents-college-students-warned-stay-away-n795891

McHugh, M. L. (2009). The odds ratio: Calculation, usage, and interpretation. *Biochemia Medica, 19*(2), 120-126.

Medina-Cetina, Z., & Nadim, F. (2008). Stochastic design of an early warning system. *Georisk, 2*(4), 223-236.

Meyer, V., Becker, N., Markantonis, V., Schwarze, R., van den Bergh, J., Bouwer, L. M., . . . Green, C. (2013). Assessing the costs of natural hazards: State of the art and knowledge gaps. *Natural Hazards and Earth System Sciences, 13*(5), 1351-1373.

Morrow, B. H. (1999). Identifying and mapping community vulnerability. *Disasters, 23*(1), 1-18.

Mrad, A. B., Delcroix, V., Piechowiak, S., Leicester, P., & Abid, M. (2015). An explication of uncertain evidence in Bayesian networks: Likelihood evidence and probabilistic evidence. *Applied Intelligence, 43*(4), 802-824.

National Hurricane Center. (2022). *Tropical cyclone best track.* https://www.nhc.noaa.gov/gis/archive_besttrack.php?year=2017

National Oceanic and Atmospheric Administration. (2021). *Weather-ready nation*. https://www.weather.gov/wrn/hurricane-hazards

Norton, E. C., Dowd, B. E., & Maciejewski, M. L. (2019). Marginal effects: Quantifying the effect of changes in risk factors in logistic regression models. *JAMA, 321*(13), 1304-1305.

Nowok, B., Raab, G. M., & Dibben, C. (2016). Synthpop: Bespoke creation of synthetic data in R. *Journal of Statistical Software, 74*, 1-26.

Peacock, W. G., Van Zandt, S., Zhang, Y., & Highfield, W. E. (2014). Inequities in long-term housing recovery after disasters. *Journal of the American Planning Association, 80*(4), 356-371.

Pearl, J. (1995). *From Bayesian networks to causal networks.* Springer.

Pearl, J. (2011). *Bayesian networks.* California Digital Library, University of California.

Pielke, R. A., & Landsea, C. W. (1998). Normalized hurricane damages in the United States: 1925-95. *Weather and Forecasting, 13*(3), 621-631.

Pinelli, J.-P., Simiu, E., Gurley, K., Subramanian, C., Zhang, L., Cope, A., . . . Hamid, S. (2004). Hurricane damage prediction model for residential structures. *Journal of Structural Engineering, 130*(11), 1685-1691.

Pourret, O., Na, P., & Marcot, B. (2008). *Bayesian networks: A practical guide to applications.* John Wiley & Sons.

Quay, R. (2010). Anticipatory governance: A tool for climate change adaptation. *Journal of the American Planning Association, 76*(4), 496-511.

R Core Team. (2021). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing.

Ramanan, N., & Natarajan, S. (2020). Causal learning from predictive modeling for observational data. *Frontiers in Big Data*, *3*, 535976.

Rappaport, E. N. (2014). Fatalities in the United States from Atlantic tropical cyclones: New data and interpretation. *Bulletin of the American Meteorological Society, 95*(3), 341-346.

Sadowski, N. C., & Sutter, D. (2008). Mitigation motivated by past experience: Prior hurricanes and damages. *Ocean & Coastal Management, 51*(4), 303-313.

Scutari, M. (2010). Learning Bayesian networks with the bnlearn R Package. *Journal of Statistical Software, 35*(3), 1-22.

Scutari, M., & Denis, J.-B. (2014). *Bayesian networks: With examples in R.* CRC Press.

Scutari, M., Silander, T., & Ness, R. (2021). bnlearn: Bayesian network structure learning, parameter learning and inference. *R Package Version 4.7*. http://cran.uvigo.es/web/packages/bnlearn/

Sedki, K., Delcroix, V., Lepoutre, F.-X., Adam, E., Maquinghen-Godillon, A.-P., & Ville, I. (2010). Bayesian network model for decision problems. *Intelligent Information Systems, 2010*, 285-298.

Shrestha, N. (2020). Detecting multicollinearity in regression analysis. *American Journal of Applied Mathematics and Statistics, 8*(2), 39-42.

Shrier, I., & Platt, R. W. (2008). Reducing bias through directed acyclic graphs. *BMC Medical Research Methodology, 8*(1), 1-15.

Smith, K. (2013). *Environmental hazards: Assessing risk and reducing disaster.* Routledge.

Sperotto, A., Molina, J.-L., Torresan, S., Critto, A., & Marcomini, A. (2017). Reviewing Bayesian network potentials for climate change impacts assessment and management: A multi-risk perspective. *Journal of Environmental Management, 202*, 320-331.

Stewart, M. G., Rosowsky, D. V., & Huang, Z. (2003). Hurricane risks and economic viability of strengthened construction. *Natural Hazards Review, 4*(1), 12-19.

Stone, J. V. (2013). *Bayes' rule with R: A tutorial introduction to Bayesian analysis.* Sebtel Press.

Strahan, K., & Watson, S. J. (2019). The protective action decision model: When householders choose their protective response to wildfire. *Journal of Risk Research, 22*(12), 1602-1623.

Straub, D. (2005). Natural hazards risk assessment using Bayesian networks. In *9th International Conference on Structural Safety and Reliability (ICOSSAR 05).* Millpress.

Strobl, E. (2011). The economic growth impact of hurricanes: Evidence from US coastal counties. *Review of Economics and Statistics, 93*(2), 575-589.

Tanim, S. H., Reader, S., & Hu, Y. (2022). Predictors of hurricane evacuation decisions: A meta-analysis. *Journal of Environmental Psychology, 79*, 101742.

Terpstra, T., & Lindell, M. K. (2013). Citizens' perceptions of flood hazard adjustments: An application of the protective action decision model. *Environment and Behavior, 45*(8), 993-1018.

Tollefson, J. (2012). Hurricane sweeps US into climate-adaptation debate. *Nature News, 491*(7423), 167.

Tsamardinos, I., & Borboudakis, G. (2010). Permutation testing improves Bayesian network learning. In *Machine Learning and Knowledge Discovery in Databases* (pp. 322-337). Springer.

Uusitalo, L. (2007). Advantages and challenges of Bayesian networks in environmental modelling. *Ecological Modelling, 203*(3-4), 312-318.

van de Lindt, J. W., & Dao, T. N. (2009). Performance-based wind engineering for wood-frame buildings. *Journal of Structural Engineering, 135*(2), 169-177.

Vickery, P. J., & Twisdale, L. A. (1995). Prediction of hurricane wind speeds in the United States. *Journal of Structural Engineering, 121*(11), 1691-1699.

Wachinger, G., & Renn, O. (2010). *Risk perception and natural hazards.* DIALOGIK Non-Profit Institute for Communication and Cooperative Research.

Waddell, P. (2011). Integrated land use and transportation planning and modelling: addressing challenges in research and practice. *Transport Reviews, 31*(2), 209-229.

Wickham, H. (2019). Welcome to the tidyverse. Journal of Open-Source Software, 4(43), 1686.

World Meteorological Organization. (2021). *Weather-related disasters increase over past 50 years, causing more damage but fewer deaths*. https://public.wmo.int/en/media/press-release/weather-related-disasters-increase-over-past-50-years-causing-more-damage-fewer

Wright, J. D., Pereira, J. A., & Weber-Burdin, E. (2012). *Victims of the environment: Loss from natural hazards in the United States, 1970-1980.* Springer Science & Business Media.

Yang, H., Morgul, E. F., Ozbay, K., & Xie, K. (2016). Modeling evacuation behavior under hurricane conditions. *Transportation Research Record, 2599*(1), 63-69.

Yoon, D. (2012). Assessment of social vulnerability to natural disasters: A comparative study. *Natural Hazards, 63*(2), 823-843.

Zhang, Y., & Peacock, W. G. (2009). Planning for housing recovery? Lessons learned from Hurricane Andrew. *Journal of the American Planning Association, 76*(1), 5-24.

Zheng, Z., & Pavlou, P. A. (2010). Research note—Toward a causal interpretation from observational data: A new Bayesian networks method for structural models with latent variables. *Information Systems Research, 21*(2), 365-391.

APPENDIX A

CODE FOR CHAPTER 2

```
# author: "Alexander Abuabara"

#### Preamble ####
library(bnlearn)   # Bayesian Network Structure Learning, Parameter Learning
and Inference
library(tidyverse) # Easily Install and Load the 'Tidyverse'

setwd("/Users/alexander/Desktop/Dissertation/P1 script/")
options(digits = 4, scipen = 20)
set.seed(123)

#### Model 1 ####
DiagrammeR::grViz("digraph {
                  # a 'graph' statement
                  graph [layout = dot, # dot, neato, twopi, and circo
                         rankdir = LR,
                         overlap = true,
                         fontsize = 10]

                  # several 'node' statements
                  node [shape = hexagon] # diamond
                  'Damage\nCost'

                  node [shape = oval,
                        fixedsize = false,
                        width = 1.4]
                  'Hurricane\nWind'; 'Building\nVulnerability'

             'Hurricane\nWind' -> 'Building\nVulnerability'
     'Building\nVulnerability' -> 'Damage\nCost'}")

dag <- model2network("[hurricane_wind][building_vulnerability|hurricane_wind]")
par(new = TRUE, bg = "white")
graphviz.plot(dag, layout = "dot")

hurricane_wind.lv <- c("Low", "Moderate", "High")

######################################   Wind
(hurricane_wind.prob <- array(c(0.30,  # Low
                                0.50,  # Moderate
                                0.20), # High
                              dim = 3,
                              dimnames = list(hurricane_wind =
hurricane_wind.lv)))

building_vulnerability.lv <- c("None", "Low", "Moderate", "High")
```

```
############################### Damage: None    Low    Mod   High     Wind
(building_vulnerability.prob <- array(c(0.70,  0.20,  0.08,  0.02,  # Low
                                        0.40,  0.25,  0.20,  0.15,  # Moderate
                                        0.20,  0.25,  0.30,  0.25), # High
                                      dim = c(4,3),
                                      dimnames = list(building_vulnerability =
building_vulnerability.lv,
                                                      hurricane_wind         =
hurricane_wind.lv)))

cpt <- list(hurricane_wind         = hurricane_wind.prob,
            building_vulnerability = building_vulnerability.prob)

bn <- custom.fit(dag, cpt)

bn.fit.dotplot(bn$hurricane_wind,
               ylab = "Hurricane Wind Categories",
               main = "Probabilities for the Node Hurricane Wind")

bn.fit.dotplot(bn$building_vulnerability,
               ylab = "Building Damage Categories",
               main = "Conditional Probabilities for the Node Building
Vulnerability\nConditional to the Hurricane Wind")

par(new = TRUE, bg = "white", mar = c(1, 1, 1, 1))
graphviz.chart(bn,
               type = "barprob",
               grid = TRUE,
               layout = "neato",
               bg = "white",
               bar.col = "black",
               text.col = "black",
               strip.bg = "white",
               draw.levels = TRUE,
               main = "Model 1",
               sub = "")

values_loss <- array(c(0,
                       5000,
                       10000,
                       100000), dim = 4,
                     dimnames = list(building_vulnerability =
building_vulnerability.lv))

values_loss %>% as.data.frame() %>%
   janitor::clean_names() %>% stats::setNames(c("USD"))

junction <- gRain::querygrain(gRbase::compile(as.grain(bn)))
```

```
junction$building_vulnerability

(expected_value <- round(sum(junction$building_vulnerability * values_loss),
2))

(results_table <-
      data.frame("Model1.Limited" =
scales::percent(round(junction$building_vulnerability, 2))) %>%
      rbind.data.frame(., Expected_Loss =
scales::dollar(round((expected_value)))))

#### Model 2* ####
dag <-
model2network("[hurricane_wind][hurricane_flooding][building_vulnerability|hurr
icane_wind:hurricane_flooding]")
par(new = TRUE, bg = "white")
graphviz.plot(dag)

hurricane_flooding.lv <- c("None", "Low", "High")

(hurricane_flooding.prob <- array(c(0.7, 0.15, 0.15),dim = 3,
                                  dimnames = list(hurricane_flooding =
hurricane_flooding.lv)))

################################ Damage: None    Low    Mod   High    Wind
(building_vulnerability.prob <- array(c(0.50,  0.30,  0.15,  0.05,  # Low
                                        0.25,  0.40,  0.20,  0.15,  # Moderate
                                        0.20,  0.35,  0.25,  0.20,  # High

                                        0.25,  0.40,  0.20,  0.15,  # Low
                                        0.20,  0.35,  0.25,  0.20,  # Moderate
                                        0.15,  0.30,  0.30,  0.25,  # High

                                        0.20,  0.35,  0.25,  0.20,  # Low
                                        0.15,  0.30,  0.30,  0.25,  # Moderate
                                        0.01,  0.14,  0.35,  0.50), # High
                                      dim = c(4,3,3),
                                      dimnames = list(building_vulnerability =
building_vulnerability.lv,
                                                      hurricane_wind =
hurricane_wind.lv,
                                                      hurricane_flooding =
hurricane_flooding.lv)))

cpt <- list(hurricane_wind         = hurricane_wind.prob,
            hurricane_flooding     = hurricane_flooding.prob,
            building_vulnerability = building_vulnerability.prob)

bn <- custom.fit(dag, cpt)
```

```
bn.fit.dotplot(bn$hurricane_wind,
               ylab = "Hurricane Wind Categories",
               main = "Probabilities for the Node Hurricane Wind")

bn.fit.dotplot(bn$building_vulnerability,
               ylab = "Building Damage Categories",
               main = "Conditional Probabilities for the Node Building
Vulnerability\nConditional to the Hurricane Wind")

graphviz.chart(bn,
               type = "barprob",
               grid = TRUE,
               layout = "neato",
               bg = "white",
               bar.col = "black",
               text.col = "black",
               strip.bg = "white",
               draw.levels = TRUE,
               main = "Model 1",
               sub = "")

values_loss %>% as.data.frame() %>%
   janitor::clean_names() %>% stats::setNames(c("USD"))

junction <- gRain::querygrain(gRbase::compile(as.grain(bn)))

junction$building_vulnerability

(expected_value <- round(sum(junction$building_vulnerability * values_loss),
2))

(results_table <-
      data.frame("Model1.Limited" =
scales::percent(round(junction$building_vulnerability, 2))) %>%
      rbind.data.frame(., Expected_Loss =
scales::dollar(round((expected_value)))))

#### Model 2 ####
dag <-
model2network("[hurricane_wind][hurricane_flooding][tenure][building_vulnerabil
ity|hurricane_wind:hurricane_flooding:tenure]")
par(new = TRUE, bg = "white")
graphviz.plot(dag)

tenure.lv <- c("Owner", "Renter")

(tenure.prob <- array(c(0.65, 0.35),
                      dim = 2,
                      dimnames = list(tenure = tenure.lv)))
```

```
############################### Damage: None    Low    Mod   High   # Wind
# Flooding
######################################### Tenure: Owner
(building_vulnerability.prob <- array(c(0.99,  0.01,  0.00,  0.00,  # Low
# None
                                        0.90,  0.08,  0.02,  0.00,  # Moderate
                                        0.75,  0.10,  0.10,  0.05,  # High

                                        0.80,  0.10,  0.08,  0.02,
# Low
                                        0.70,  0.15,  0.10,  0.05,
                                        0.55,  0.20,  0.15,  0.10,

                                        0.50,  0.10,  0.30,  0.10,
# High
                                        0.40,  0.12,  0.33,  0.15,
                                        0.30,  0.15,  0.35,  0.20,

                                        # Tenure: Renter
                                        0.94,  0.05,  0.01,  0.00,  # Low
# None
                                        0.80,  0.12,  0.05,  0.03,  # Moderate
                                        0.50,  0.20,  0.20,  0.10,  # High

                                        0.80,  0.10,  0.05,  0.05,
# Low
                                        0.60,  0.15,  0.15,  0.10,
                                        0.50,  0.20,  0.20,  0.10,

                                        0.20,  0.20,  0.30,  0.30,
# High
                                        0.15,  0.15,  0.32,  0.38,
                                        0.05,  0.20,  0.35,  0.40),
                                      dim = c(4,3,3,2),
                                      dimnames = list(building_vulnerability =
building_vulnerability.lv,
                                                      hurricane_wind =
hurricane_wind.lv,
                                                      hurricane_flooding =
hurricane_flooding.lv,
                                                      tenure = tenure.lv)))

cpt <- list(hurricane_wind = hurricane_wind.prob,
            hurricane_flooding = hurricane_flooding.prob,
            tenure = tenure.prob,
            building_vulnerability = building_vulnerability.prob)

bn <- custom.fit(dag, cpt)

bn.fit.dotplot(bn$hurricane_wind,
```

```
               ylab = "Hurricane Wind Categories",
               main = "Probabilities for the Node Hurricane Wind")

bn.fit.dotplot(bn$hurricane_flooding,
               ylab = "Hurricane Flooding Categories",
               main = "Probabilities for the Node Hurricane Flooding")

bn.fit.dotplot(bn$tenure,
               ylab = "Tenure Categories",
               main = "Probabilities for the Node Tenure")

bn.fit.dotplot(bn$building_vulnerability,
               ylab = "Building Damage Categories",
               main = "Conditional Probabilities for the Node Building
Vulnerability\nConditional to the Hurricane Wind, Flooding, and Tenure")

graphviz.chart(bn,
               type = "barprob",
               grid = TRUE,
               layout = "neato",
               bg = "white",
               bar.col = "black",
               text.col = "black",
               strip.bg = "white",
               draw.levels = TRUE,
               main = "Model 1",
               sub = "")

values_loss %>% as.data.frame() %>%
   janitor::clean_names() %>% stats::setNames(c("USD"))

junction <- gRain::querygrain(gRbase::compile(as.grain(bn)))

junction$building_vulnerability

(expected_value <- round(sum(junction$building_vulnerability * values_loss),
2))

(results_table <-
      data.frame("Model1.Limited" =
scales::percent(round(junction$building_vulnerability, 2))) %>%
      rbind.data.frame(., Expected_Loss =
scales::dollar(round((expected_value))))
```

APPENDIX B

CODE FOR CHAPTER 3

Maps

```
# author: "Alexander Abuabara"

###### Preamble ######
library(tidyverse)  # Easily Install and Load the 'Tidyverse'
library(sf)         # Simple Features for R
library(tigris)     # Load Census TIGER/Line Shapefiles
library(shadowtext) # Shadow Text Grob and Layer
library(ggspatial)  # Spatial Data Framework for ggplot2

setwd("/Users/alexander/Desktop/Dissertation/P2 script/")

options(digits = 5,
        scipen = 999,
        tigris_use_cache = TRUE)

###### Old Evac. Zones ######
CBSA_counties <- c("Aransas", "Calhoun", "Matagorda", "Jackson",
                   "Victoria", "Refugio", "San Patricio", "Nueces")

CBSA_shapes <- counties(state = "TX",
                        cb = FALSE,
                        year = 2018) %>%
   filter(NAME %in% CBSA_counties)

CBSA_crop <- CBSA_shapes %>%
   group_by() %>%
   summarise()

TX_cropped <- counties(state = "TX",
                       cb = FALSE,
                       year = 2018) %>%
   st_crop(., st_buffer(CBSA_crop %>% st_transform(3083),
                        dist = units::set_units(10, km)) %>%
              st_transform(st_crs(CBSA_crop)))

CBSA_water <- map_df(CBSA_counties, ~area_water(state = "TX",
                                                county = .x,
                                                year = 2018)) %>%
   st_combine() %>% rmapshaper::ms_simplify()

CBSA_places <- places(state = "TX",
                      cb = TRUE,
                      year = 2018) %>%
```

```
   st_crop(., CBSA_crop) %>%
   st_filter(., CBSA_crop) %>%
   filter(ALAND > 2000000)

hurr_track <- read_sf("./P2 data/Harvey.shp") %>%
   st_transform(st_crs(CBSA_crop)) %>%
   st_crop(., st_buffer(CBSA_crop %>% st_transform(3083),
                        dist = units::set_units(10, km)) %>%
              st_transform(st_crs(CBSA_crop)))

old_evac_zones <- read_sf("./P2 data/old_evac_zones.shp") %>%
   filter(EvacZone != "Out") %>%
   mutate(EvacZoneGen = case_when(EvacZone == "Coastal"  ~ "A",
                                  EvacZone == "A"        ~ "A",
                                  EvacZone == "B"        ~ "A",
                                  EvacZone == "C"        ~ "B",
                                  EvacZone == "D"        ~ "C",
                                  EvacZone == "E"        ~ "C",
                                  EvacZone == "Zone 1-2" ~ "A",
                                  EvacZone == "Zone 3"   ~ "B",
                                  EvacZone == "Zone 4-5" ~ "C",
                                  EvacZone == "Risk 1"   ~ "A",
                                  EvacZone == "Risk 2"   ~ "A",
                                  EvacZone == "Risk 3"   ~ "B",
                                  EvacZone == "Risk 4"   ~ "C",
                                  EvacZone == "Risk 5"   ~ "C",
                                  TRUE ~ "Missing"),
          Cat = case_when(EvacZoneGen == "A" ~ "Cat.1-2-3",
                          EvacZoneGen == "B" ~ "Cat.4",
                          EvacZoneGen == "C" ~ "Cat.5",
                          TRUE ~ "Missing")) %>%
   st_crop(., CBSA_crop)

pts <- do.call(rbind,
               st_geometry(st_centroid(CBSA_shapes)))
CBSA_shapes$X <- pts[,1]
CBSA_shapes$Y <- pts[,2]

ggplot2 <- ggplot() +
   geom_sf(data = TX_cropped,
           fill = "antiquewhite",
           col = "black",
           size = 0,
           inherit.aes = FALSE) +
   geom_sf(data = CBSA_shapes,
           fill = "antiquewhite",
           col = "black",
           size = 0,
           inherit.aes = FALSE) +
   geom_sf(data = old_evac_zones,
```

```
          aes(fill = Cat),
          size = 0,
          alpha = .7,
          inherit.aes = FALSE) +
  geom_sf(data = CBSA_water,
          fill = "lightcyan",
          size = 0,
          alpha = 1,
          inherit.aes = FALSE) +
  geom_sf(data = CBSA_shapes,
          fill = "transparent",
          col = "black",
          size = .5,
          inherit.aes = FALSE) +
  geom_sf(data = CBSA_places,
          aes(color = "Local\nCommunities"),
          fill = "grey20",
          size = 0,
          alpha = .4,
          inherit.aes = FALSE) +
  geom_sf(data = hurr_track,
          col = "blue",
          size = 1,
          alpha = .85,
          inherit.aes = FALSE) +
  geom_shadowtext(data = CBSA_shapes,
                  aes(x=X, y=Y, label=NAME),
                  color="black", bg.color="white", size = 5, fontface =
"italic",
                  inherit.aes = FALSE) +
  annotate(geom = "text",
           label = "Gulf of Mexico",
           x = -96.1, y = 28.1,
           fontface = "italic", color = "cyan3", size = 5) +
  annotate(geom = "text",
           label = "Hurricane\n Harvey\n    Track",
           x = -96.625, y = 27.75,
           fontface = "bold", color = "blue", size = 5) +
  coord_sf(xlim = c(-97.99, -95.46),
           ylim = c(27.51, 29.32),
           expand = FALSE) +
  labs(fill = "Generalized\nHurricane\nEvacuation\nZones",
       color = "", x = "", y = "") +
  annotation_scale(location = "br", width_hint = .2, style = "ticks") +
  annotation_north_arrow(location = "br", which_north = "true", style =
north_arrow_minimal, pad_y = unit(0.2, "in")) +
  theme(
    panel.background = element_rect(fill = "lightcyan"),
    panel.border = element_rect(color = "black", fill = "transparent"),
    panel.grid = element_blank() ,
```

```
      legend.background = element_rect(color = "transparent", fill =
"transparent"),
      legend.key = element_rect(color = "transparent", fill = "transparent"),
      legend.justification = "top",
      legend.title = element_text(size = 13), #text sizes
      legend.text = element_text(size = 11),
      axis.title = element_text(size = 13),
      axis.text = element_text(size = 11),
      element_line(color = "black"))

ggplot2

# ggsave("./P2 images/map_1.png", width = 9, height = 6)

###### Counties ######
ggplot3 <- ggplot() +
   geom_sf(data = TX_cropped,
           fill = "white",
           col = "black",
           size = 0,
           inherit.aes = FALSE) +
   geom_sf(data = CBSA_shapes,
           fill = "white",
           col = "black",
           size = 0,
           inherit.aes = FALSE) +
   geom_sf(data = CBSA_shapes,
           fill = "white",
           col = "black",
           size = .5,
           inherit.aes = FALSE) +
   geom_sf(data = hurr_track,
           col = "black",
           size = 1,
           alpha = .85,
           inherit.aes = FALSE) +
   geom_shadowtext(data = CBSA_shapes,
                   aes(x=X, y=Y, label=NAME),
                   color="black", bg.color="white", size = 5, fontface =
"italic",
                   inherit.aes = FALSE) +
   annotate(geom = "text",
            label = "Hurricane\n Harvey\n    Track",
            x = -96.60, y = 27.75,
            fontface = "bold", color = "black", size = 5) +
   annotate("segment",
            x = -96.581, xend = -96.8,
            y = 27.488, yend = 27.8,
            colour = "blue", size = 1,
            arrow = arrow()) +
```

```
   coord_sf(xlim = c(-97.99, -95.46),
            ylim = c(27.51, 29.32),
            expand = FALSE) +
   theme_void()

ggplot3

# ggsave("./P2 images/map_2.png", width = 9, height = 6)

###### Inset ######
library(rnaturalearth) # World Map Data from Natural Earth
library(cowplot)       # Streamlined Plot Theme and Plot Annotations for
'ggplot2'

data("us_states", package = "spData")

world <- ne_countries(scale = "medium", returnclass = "sf")
states <- map_data("state")

bbox <- st_as_sfc(st_bbox(CBSA_crop))

ggplot1 <- ggplot() +
   geom_sf(data = us_states,
           size = .5,
           fill = "white",
           col = "black") +
   geom_sf(data = st_buffer(bbox %>% st_transform(3083),
                            dist=units::set_units(20, km)),
           size = .7,
           fill = NA,
           col = "blue") +
   coord_sf(crs = "+proj=aea +lat_1=29.5 +lat_2=45.5 +lat_0=37.5 +lon_0=-96
+x_0=0 +y_0=0 +ellps=GRS80 +datum=NAD83 +units=m +no_defs")+
   theme_void()

ggplot1

inset <- ggdraw() +
   draw_plot(ggplot2) +
   draw_plot(ggplot1, x = 0.75, y = 0.175, width = 0.25, height = 0.25) # size
and location

inset

# ggsave("./P2 images/map_3.png", width = 9, height = 6)

inset <- ggdraw() +
   draw_plot(ggplot3) +
   draw_plot(ggplot1, x = 0.75, y = 0.075, width = 0.25, height = 0.25) # size
and location
```

```
inset

# ggsave("./P2 images/map_4.png", width = 9, height = 6)

beepr::beep()
# sessionInfo()

# line = st_sfc(st_linestring(rbind(c(-96.571, 27.474),
#                                   c(-96.581, 27.488))),
#               crs = 3083)
#
# ggplot() +
#    geom_sf(data = line) +
#    annotate("segment", x = -96.571, xend = -96.581, y = 27.474, yend =
27.488,
#             colour = "blue", size = 2, arrow = arrow()) +
#
#    coord_sf(datum = 3083)
#
# dat <- data.frame(x = c(-96.571, -96.581),
#                   y = c(27.474, 27.488))
#
# ggplot() +
#    geom_path(data = dat,
#              aes(x, y),
#              arrow = arrow())
```

The code below requires a dataset file. May be available upon request to the author.

```
# author: "Alexander Abuabara"

###### Preamble ######
library(bnlearn)      # Bayesian Network Structure Learning, Parameter Learning
and Inference
library(corrplot)     # Visualization of a Correlation Matrix
library(DescTools)    # Tools for Descriptive Statistics
library(gRain)        # Graphical Independence Networks
library(gtsummary)    # Presentation-Ready Data Summary and Analytic Result
Tables
library(haven)        # Import and Export "SPSS", "Stata" and "SAS" Files
library(labelled)     # Manipulating Labelled Data
library(modelsummary) # Summary Tables and Plots for Statistical Models and
Data: Beautiful, Customizable, and Publication-Ready
library(plyr)         # Tools for Splitting, Applying and Combining Data
library(rstatix)      # Pipe-Friendly Framework for Basic Statistical Tests
library(tidyverse)    # Easily Install and Load the "Tidyverse"
```

```
setwd("/Users/alexander/Library/Mobile
Documents/com~apple~CloudDocs/TAMU/Research/3-Dissertation/P2 script/")
options(digits = 2, scipen = 99999999, na.strings = "NA")

###### Data ######
survey_ <- read_sav("./P2 data/Coastal Bend Hurricane Evacuation Behavior
Generation 2.sav") %>%
   mutate(county_aux = as.factor(str_remove_all(ActualCounty, " County")),
          # data cleaning for structure type (suggested by Peacock)
          Q31aux    = case_when(Q31 == 1         ~ 1,
                                Q31 %in% c(2,3) ~ 2,
                                Q31 == 4         ~ 3,
                                Q31 == 5         ~ 4),
          Q31aux    = ifelse(InformID %in%
c(576,224,154,613,333,287,672,194,29,58,306,758,885,5,

392,260,458,192,580,834,481,65,805,426,408), 1, Q31aux),
          Q31aux    = ifelse(InformID %in%
c(830,249,272,421,317,446,279,894,78,815,46,
                                             493,764,690,848,782,879,407,502),
2, Q31aux),
          Q31aux    = ifelse(InformID %in%
c(114,881,582,701,127,239,634,877,71,84,261),  3, Q31aux),
          Q31aux    = ifelse(InformID %in% c(666,616,146),
4, Q31aux),
          Q31aux    = ifelse(InformID %in% c(899),
NA, Q31aux),
          structure = labelled(Q31aux, c("Single_family" = 1,
                                         "Multi_family"  = 2,
                                         "Mobile_home"   = 3,
                                         "Other"         = 4)))

survey <- survey_ %>%
   remove_var_label() %>%
   filter(Use == 1) %>%
   remove_attributes("format.spss") %>%
   transmute(evacuation = factor(tolower(as_factor(Q5Mod))),
             age = as.numeric(InfAge),
             consult_info = as.numeric(rowMeans(select(., Q1_1, Q1_2, Q1_3,
Q1_4, Q1_5, Q1_6), na.rm = TRUE)),
             evac_orders =
factor(case_when(grepl(c("Matagorda|Calhoun|Refugio|Aransas|San Patricio"),
county_aux) ~ "mandatory",
                                            grepl("Nueces", county_aux) &
Q31aux == 3 ~ "mandatory",
                                            TRUE ~ "voluntary"), ordered =
TRUE, levels = c("mandatory", "voluntary")),
             expected_hh_impacts = as.numeric(rowMeans(select(., Q3_3, Q3_4,
Q3_5, Q3_6), na.rm = TRUE)),
             gender = factor(case_when(Gender == 1 ~ "male",
```

```
                                       Gender == 0 ~ "female"), ordered = TRUE,
levels = c("male", "female")),
             signif_evac_orders = as.numeric(rowMeans(select(., Q4_4), na.rm =
TRUE)),
             # education = as.numeric(rowMeans(select(., Q36), na.rm = TRUE)),
             education = as.numeric(case_when(Q36 == 1 ~ 10,
                                              Q36 == 2 ~ 12,
                                              Q36 == 3 ~ 16,
                                              Q36 == 4 ~ 18,
                                              Q36 == 5 ~ 21)),
             multiple_concerns = as.numeric(rowMeans(select(., Q4_7, Q4_8,
Q4_9, Q4_11, Q4_12), na.rm = TRUE)),
             risk_area = factor(case_when(EvacZoneOld %in% c("A", "B", "C",
"D", "E",
                                                             "Risk 1", "Risk
2", "Risk 3", "Risk 4", "Risk 5",
                                                             "Coastal", "Zone
1-2", "Zone 3", "Zone 4-5") ~ "yes",
                                          TRUE ~ "no"), ordered = TRUE, levels
= c("yes", "no")),
             social_cues = as.numeric(rowMeans(select(., Q4_1, Q4_2), na.rm =
TRUE)),
             unnecessary_evac_exp = as.numeric(rowMeans(select(., Q4_6), na.rm
= FALSE)),
   ) %>%
   mutate_all(~ case_when(!is.nan(.x) ~ .x),) %>%
   mutate_if(is.numeric, signif, 3)

PlotMiss(survey, main = "Missing survey data (clustered)", clust = TRUE)

dat_bn_dicretized <-
   survey %>%
   as.data.frame() %>%
   transmute(
      evacuation = case_when(evacuation == "yes" ~ 1,
                             evacuation == "no" ~ 0),
      expected_hh_impacts,
      signif_evac_orders,
      evac_orders = case_when(evac_orders == "mandatory" ~ 1,
                              evac_orders == "voluntary" ~ 0),
      unnecessary_evac_exp,
      multiple_concerns,
      risk_area = case_when(risk_area == "yes" ~ 1,
                            risk_area == "no" ~ 0),
      social_cues,
      consult_info,
      age = age,
      education,
      gender = case_when(gender == "male" ~ 1,
                         gender == "female" ~ 0),
```

```
  ) %>%
  filter(if_all(everything(), ~!is.na(.x))) %>% # na.omit()
  mutate_if(is.integer, as.double) %>%
  mutate(age = round_any(age, 10, floor)) %>%
  mutate(across(where(is.numeric), round, 0)) %>%
  mutate_if(is.double, as.ordered)

dat_bn_dicretized %>% glimpse()
dat_bn_dicretized %>% tbl_summary() %>% as_hux_table()

dat_dicretized <-
  dat_bn_dicretized %>%
  mutate_if(is.ordered, as.character) %>%
  mutate_if(is.character, as.double)

dat_dicretized %>%
  tbl_summary(
    type = list(where(is.numeric) ~ "continuous"),
    statistic = list(all_continuous() ~ "mean {mean} (sd {sd})"),
    missing_text = "(Missing)"
  ) %>% as_hux_table()

###### Descriptive ######
# Empirical CDF of discretized data
par(mfrow = c(2, 6), mar = c(1, 3, 1, 1), pty = "s")
for (var in colnames(dat_dicretized %>%
                     # select(-evacuation) %>%
                     mutate_if(is.ordered, as.character) %>%
                     mutate_if(is.character, as.double))){
  x = dat_dicretized[, var]
  plot(ecdf(x),
       col = "black", lwd = 1, lty = 1, xaxt = "n", yaxt = "n",
       verticals = TRUE, do.points = FALSE, col.01line = NULL,
       main = "", xlab = "", ylab = "", add = FALSE)
  axis(1, at = min(x):max(x))
  axis(2, seq(0, 1, by = .5))
  mtext(var, side = 1, line = 2.4)
}
title(main = "Empirical Cumulative Distribution Function of Each Variable",
line = -3, cex.main = 2, outer = TRUE)

for (var in colnames(dat_bn_dicretized %>% select(-evacuation))){
  x = dat_dicretized[dat_dicretized$evacuation == 1, var]
  y = dat_dicretized[dat_dicretized$evacuation == 0, var]
  plot(ecdf(x),
       col = "blue", lwd = 1.5, lty = 1, xaxt = "n", yaxt = "n",
       verticals = TRUE, do.points = FALSE, col.01line = NULL,
       main = "", xlab = "", ylab = "", add = FALSE)
  plot(ecdf(y),
       col = "red", lwd = 0.75, lty = 1, xaxt = "n", yaxt = "n",
```

```
        verticals = TRUE, do.points = FALSE, col.01line = NULL,
        main = "", xlab = "", ylab = "", add = TRUE)
   axis(1, at = min(x):max(x))
   axis(2, seq(0, 1, by = .5))
   mtext(var, side = 1, line = 2.2, cex = .9)
   legend("bottomright",
          legend = c("1", "0"),
          col    = c("blue", "red"),
          pch    = 15, cex = .9)
}
title(main = "Empirical Cumulative Distribution Function of Each Variable
Conditional to Evacuation", line = -3, cex.main = 2, outer = TRUE)

# Correlation
corr <- cor_test(data   = dat_dicretized,
                 vars   = evacuation,
                 method = "pearson",
                 use    = "pairwise.complete.obs") %>% arrange(-cor, p)

# Matrix of p-values
p.mat <- cor.mtest(dat_dicretized, conf.level = 0.95)

# p.mat %>% view()
round(cor(dat_dicretized), 2) %>% view()

par(mfrow = c(1, 1), mar = c(1, 1, 10, 1), bg = "white", pty = "m")
corrplot(cor(dat_dicretized, method = "pearson"),
         method = "color", type = "upper", tl.srt = 40, tl.col = "black",
         p.mat  = p.mat$p, sig.level = 0.05, # insig = "blank",
         order  = "original", col = RColorBrewer::brewer.pal(n = 10, name =
"RdBu"))

# Detect multicollinearity
eigen(cor(dat_dicretized))$values
kappa(cor(dat_dicretized), exact = TRUE)

###### Logistic Regression ######
summary(model_1 <- glm(evacuation ~ social_cues,
                     data = dat_dicretized,
                     family = binomial(link = "logit")))

summary(logit <- glm(evacuation ~ social_cues + unnecessary_evac_exp,
                     data = dat_dicretized,
                     family = binomial(link = "logit")))

summary(logit <- glm(evacuation ~ social_cues + unnecessary_evac_exp +
signif_evac_orders + multiple_concerns,
                     data = dat_dicretized,
                     family = binomial(link = "logit")))
```

```
summary(model_1 <- glm(evacuation ~ expected_hh_impacts + signif_evac_orders,
                       data = dat_dicretized,
                       family = binomial(link = "logit")
                       ))

summary(logit <- glm(evacuation ~ .,
                     data = dat_dicretized,
                     family = binomial(link = "logit")
                     ))

anova(model_1, logit)
# hypothesis test:
# H0 = the two models are equally useful for predicting the outcome
# H1 = the larger model is significantly better than the smaller model
# canno reject the null hypothesis, and prefer to use the first model?

lmtest::lrtest(model_1, logit)
# likelihood-ratio test

###### Standardized and performance ######
summary(lm.beta::lm.beta(logit))
library(tidymodels)
performance::check_model(logit)

broom::tidy(logit, exponentiate = TRUE, conf.level = 0.95)
performance::r2_nagelkerke(logit)
VIF(logit)
epiDisplay::logistic.display(logit, simplified = TRUE)[["table"]] %>%
as.data.frame() %>%
   rownames_to_column("var") %>% mutate(signif = case_when(`Pr(>|Z|)` <= 0.001
~ "Signif. 0.001",
                                                           `Pr(>|Z|)` > 0.001
& `Pr(>|Z|)` <= 0.01  ~ "Signif. 0.01",
                                                           `Pr(>|Z|)` > 0.01
& `Pr(>|Z|)` <= 0.05  ~ "Signif. 0.05",
                                                           TRUE ~ "Not
signif."),
                                        signif = factor(signif, levels = c("Not
signif.", "Signif. 0.05", "Signif. 0.01", "Signif. 0.001"))) %>%
   ggplot(aes(x = OR, y = fct_reorder(var, OR), fill = signif, color = signif))
+
   geom_point(shape = 21, size = 3) +
   geom_errorbar(aes(xmin = lower95ci, xmax = upper95ci), width = .1) +
   scale_colour_manual(values = rev(c("red", "green4", "blue", "black")),
                       breaks = c("Not signif.","Signif. 0.05", "Signif. 0.01",
"Signif. 0.001")) +
   scale_fill_manual(values   = rev(c("red", "green4", "blue", "black")),
                     breaks = c("Not signif.", "Signif. 0.05", "Signif. 0.01",
"Signif. 0.001")) +
   geom_vline(aes(xintercept = 1), size = .25, linetype = "dashed") +
```

```
   coord_trans(x = "log10") +
   scale_x_continuous(breaks = seq(0, 10, 1) ) +
   labs(title = "Logit regression predicting evacuation", x = "Odds ratio and
95% confidence intervals (log scale)",
        y = "", color = "", fill = "") + theme_bw()

effects_logit = margins::margins(logit)
summary(effects_logit)
par(new = TRUE, mfrow = c(1, 1), mar = c(12, 4, 2, 2), bg = "white", pty = "s")
plot(effects_logit) # las = 3

par(new = TRUE, mfrow = c(1, 2), mar = c(7, 3, 3, 2), bg = "white", pty = "s")
for (var in c("expected_hh_impacts", "signif_evac_orders")){
   visreg::visreg(logit, var, scale = "response", partial = FALSE, rug = 2,
xlab = paste(var), ylab = "P(evacuation)")
}

###### BN ######
dag =
model2network("[age][consult_info][evac_orders][gender][education][risk_area][s
ocial_cues][unnecessary_evac_exp][signif_evac_orders|evac_orders][multiple_conc
erns|age:gender:education][expected_hh_impacts|age:consult_info:gender:educatio
n:risk_area:social_cues][evacuation|expected_hh_impacts:signif_evac_orders:mult
iple_concerns:unnecessary_evac_exp]")
par(new = TRUE, mfrow = c(1, 1), bg = "white", pty = "m")
graphviz.plot(dag, shape = "ellipse")
bn = bn.fit(dag, dat_bn_dicretized)
coefficients(bn)
par(mfrow = c(1, 1), bg = "white", pty = "m")
graphviz.chart(bn, type = "barprob", grid = TRUE, draw.levels = TRUE, scale =
c(1, 1.2)) # c(1.2,2))
(pvalues = arc.strength(dag, data = dat_bn_dicretized))
par(new = TRUE, mfrow = c(1, 1), bg = "white", pty = "m")
strength.plot(dag, strength = pvalues, shape = "ellipse")
LL  = logLik(dag, dat_bn_dicretized)
k   = log(nrow(dat_bn_dicretized))/2
N   = nparams(dag, dat_bn_dicretized)
(BIC = LL - N * k)
score(dag, dat_bn_dicretized)

###### Exp.1 ######
dag =
model2network("[age][consult_info][evac_orders][gender][education][risk_area][s
ocial_cues][unnecessary_evac_exp][signif_evac_orders|evac_orders][multiple_conc
erns|age:gender:education][expected_hh_impacts|age:consult_info:gender:educatio
n:risk_area:social_cues][evacuation|expected_hh_impacts:signif_evac_orders:mult
iple_concerns:risk_area:unnecessary_evac_exp]")
par(new = TRUE, mfrow = c(1, 1), bg = "white")
graphviz.plot(dag, shape = "ellipse")
bn = bn.fit(dag, dat_bn_dicretized)
```

```
pvalues = arc.strength(dag, data = dat_bn_dicretized)
par(new = TRUE, mfrow = c(1, 1), bg = "white")
strength.plot(dag, strength = pvalues, shape = "ellipse")
LL  = logLik(dag, dat_bn_dicretized)
k   = log(nrow(dat_bn_dicretized))/2
N   = nparams(dag, dat_bn_dicretized)
(BIC = LL - N * k)
score(dag, dat_bn_dicretized)

###### Exp.2: Risk area ######
dag =
model2network("[age][consult_info][evac_orders][gender][education|age:gender][r
isk_area][social_cues][unnecessary_evac_exp|age][signif_evac_orders|age:gender:
education:evac_orders][multiple_concerns|age:gender:education][expected_hh_impa
cts|age:consult_info:gender:education:risk_area:social_cues][evacuation|expecte
d_hh_impacts:signif_evac_orders:multiple_concerns:risk_area:unnecessary_evac_ex
p]")
par(new = TRUE, mfrow = c(1, 1), bg = "white")
graphviz.plot(dag, shape = "ellipse")
bn = bn.fit(dag, dat_bn_dicretized)
pvalues = arc.strength(dag, data = dat_bn_dicretized)
par(new = TRUE, mfrow = c(1, 1), bg = "white")
strength.plot(dag, strength = pvalues, shape = "ellipse")
LL  = logLik(dag, dat_bn_dicretized)
k   = log(nrow(dat_bn_dicretized))/2
N   = nparams(dag, dat_bn_dicretized)
(BIC = LL - N * k)
score(dag, dat_bn_dicretized)

dag =
model2network("[consult_info][evac_orders][gender][education][risk_area][social
_cues][signif_evac_orders|gender:evac_orders][multiple_concerns|education][expe
cted_hh_impacts|consult_info:risk_area:social_cues][evacuation|expected_hh_impa
cts:signif_evac_orders:multiple_concerns:risk_area]")
par(new = TRUE, mfrow = c(1, 1), bg = "white")
graphviz.plot(dag, shape = "ellipse")
bn = bn.fit(dag, dat_bn_dicretized %>% select(-unnecessary_evac_exp, -age))
par(new = TRUE, mfrow = c(1, 1), bg = "white")
graphviz.chart(bn, type = "barprob", grid = TRUE, draw.levels = TRUE, scale =
c(1, 1.2)) # c(1.2,2))
pvalues = arc.strength(dag, data = dat_bn_dicretized %>% select(-
unnecessary_evac_exp, -age))
par(new = TRUE, mfrow = c(1, 1), bg = "white")
strength.plot(dag, strength = pvalues, shape = "ellipse")
LL  = logLik(dag, dat_bn_dicretized %>% select(-unnecessary_evac_exp, -age))
k   = log(nrow(dat_bn_dicretized))/2
N   = nparams(dag, dat_bn_dicretized %>% select(-unnecessary_evac_exp, -age))
(BIC = LL - N * k)
score(dag, dat_bn_dicretized %>% select(-unnecessary_evac_exp, -age))
```

```
###### Exp.3: Soft-evidence ######
dag =
model2network("[consult_info][evac_orders][gender][education][risk_area][social
_cues][signif_evac_orders|gender:evac_orders][multiple_concerns|education][expe
cted_hh_impacts|consult_info:risk_area:social_cues][evacuation|expected_hh_impa
cts:signif_evac_orders:multiple_concerns:risk_area]")
par(new = TRUE, mfrow = c(1, 1), bg = "white")
graphviz.plot(dag, shape = "ellipse")
bn = bn.fit(dag, dat_bn_dicretized %>% select(-unnecessary_evac_exp, -age))

ev <- list(multiple_concerns = "1")
# evidence vector
updated_dat <- cpdist(bn, nodes = bnlearn::nodes(bn), evidence = ev, method =
"lw", n = 1e6) # draw samples
updated_fit <- bn.fit(dag, data = updated_dat)
# refit: you'll get warnings over missing levels
par(new = TRUE, mfrow = c(1, 1), bg = "white")
# plot
graphviz.chart(updated_fit, type = "barprob", grid = TRUE, draw.levels = TRUE,
scale = c(1, 1.2))

junction <- compile(as.grain(bn))
multiple_concerns_low <- setEvidence(junction,
                                     nodes = "multiple_concerns",
                                     states = "1")
querygrain(multiple_concerns_low)$evacuation
querygrain(multiple_concerns_low)$education

###### Exp.4: Soft-evidence ######
dag =
model2network("[consult_info][evac_orders][gender][education][risk_area][social
_cues][signif_evac_orders|gender:evac_orders][multiple_concerns|education][expe
cted_hh_impacts|consult_info:risk_area:social_cues][evacuation|expected_hh_impa
cts:signif_evac_orders:multiple_concerns:risk_area]")
par(new = TRUE, mfrow = c(1, 1), bg = "white")
graphviz.plot(dag, shape = "ellipse")
bn = bn.fit(dag, dat_bn_dicretized %>% select(-unnecessary_evac_exp, -age))

ev <- list(evacuation = "1")
updated_dat <- cpdist(bn, nodes = bnlearn::nodes(bn), evidence = ev, method =
"lw", n = 1e6)
updated_fit <- bn.fit(dag, data = updated_dat)
par(new = TRUE, mfrow = c(1, 1), bg = "white")
graphviz.chart(updated_fit, type = "barprob", grid = TRUE, draw.levels = TRUE,
scale = c(1, 1.2))

junction <- compile(as.grain(bn))
evac_yes <- setEvidence(junction,
                        nodes = "evacuation",
                        states = "1")
```

```
querygrain(evac_yes)$expected_hh_impacts
querygrain(evac_yes)$signif_evac_orders
querygrain(evac_yes)$risk_area
```

APPENDIX C

CODE FOR CHAPTER 4

The code below requires a dataset file. May be available upon request to the author.

```
# author: "Alexander Abuabara"

###### Preamble ######
library(bnlearn); library(gRain)
library(corrplot); library(DescTools); library(gtsummary);
library(modelsummary)
library(haven); library(labelled); library(plyr); library(rstatix);
library(tidyverse)

setwd("/Users/alexander/Desktop/Dissertation/Paper/")

options(digits = 2, scipen = 99999999, na.strings = "NA")

###### Data ######
survey_ <- read_sav("./Data/Coastal Bend Hurricane Evacuation Behavior
Generation 2.sav") %>%
   mutate(county_aux = as.factor(str_remove_all(ActualCounty, " County")),
          # data cleaning for structure type (suggested by Walt)
          Q31aux    = case_when(Q31 == 1        ~ 1,
                                Q31 %in% c(2,3) ~ 2,
                                Q31 == 4        ~ 3,
                                Q31 == 5        ~ 4),
          Q31aux    = ifelse(InformID %in%
c(576,224,154,613,333,287,672,194,29,58,306,758,885,5,

392,260,458,192,580,834,481,65,805,426,408), 1, Q31aux),
          Q31aux    = ifelse(InformID %in%
c(830,249,272,421,317,446,279,894,78,815,46,
                                             493,764,690,848,782,879,407,502),
2, Q31aux),
          Q31aux    = ifelse(InformID %in%
c(114,881,582,701,127,239,634,877,71,84,261),  3, Q31aux),
          Q31aux    = ifelse(InformID %in% c(666,616,146),
4, Q31aux),
          Q31aux    = ifelse(InformID %in% c(899),
NA, Q31aux),
          structure = labelled(Q31aux, c("Single_family" = 1,
                                         "Multi_family"  = 2,
                                         "Mobile_home"   = 3,
                                         "Other"         = 4)))

survey <- survey_ %>%
   remove_var_label() %>%
```

```
  filter(Use == 1) %>%
  remove_attributes("format.spss") %>%
  transmute(evacuation            = factor(tolower(as_factor(Q5Mod))),
            # location
            evac_orders           =
factor(case_when(grepl(c("Matagorda|Calhoun|Refugio|Aransas|San Patricio"),
county_aux) ~ "mandatory",
                                                     grepl("Nueces",
county_aux) & Q31aux == 3 ~ "mandatory",
                                                     TRUE ~ "voluntary"),
ordered = TRUE, levels = c("mandatory", "voluntary")),
            risk_area             = factor(case_when(EvacZoneOld %in% c("A",
"B", "C", "D", "E",
                                                                       "Risk
1", "Risk 2", "Risk 3", "Risk 4", "Risk 5",

"Coastal", "Zone 1-2", "Zone 3", "Zone 4-5") ~ "yes",
                                                     TRUE ~ "no"), ordered =
TRUE, levels = c("yes", "no")),
            # risk perception
            expected_impacts   = as.numeric(rowMeans(select(., Q3_3, Q3_4,
Q3_5, Q3_6), na.rm = TRUE)),
            effect_evac_order = as.numeric(rowMeans(select(., Q4_4), na.rm =
TRUE)),
            multiple_concerns     = as.numeric(rowMeans(select(., Q4_7, Q4_8,
Q4_9, Q4_11, Q4_12), na.rm = TRUE)),
            social_cues           = as.numeric(rowMeans(select(., Q4_1, Q4_2),
na.rm = TRUE)),
            freq_consult_info          = as.numeric(rowMeans(select(., Q1_1,
Q1_2, Q1_3, Q1_4, Q1_5, Q1_6), na.rm = TRUE)),
            # impediments
            unnecessary_evac_exp  = as.numeric(rowMeans(select(., Q4_6), na.rm
= FALSE)),
            # socio-demographic
            age                   = as.numeric(InfAge),
            education             = as.numeric(case_when(Q36 == 1 ~ 10,
                                                         Q36 == 2 ~ 12,
                                                         Q36 == 3 ~ 16,
                                                         Q36 == 4 ~ 18,
                                                         Q36 == 5 ~ 21)),
            gender                = factor(case_when(Gender == 0 ~ "woman",
                                                     Gender == 1 ~ "man"),
ordered = TRUE, levels = c("woman", "man")),
  ) %>%
  mutate_all(~ case_when(!is.nan(.x) ~ .x),) %>%
  mutate_if(is.numeric, signif, 2)

# PlotMiss(survey, main = "Missing survey data (clustered)", clust = TRUE)

dat_bn_dicretized <-
```

```
   survey %>%
   as.data.frame() %>%
   transmute(
      evacuation = case_when(evacuation == "yes" ~ 1,
                             evacuation == "no" ~ 0),
      evac_orders = case_when(evac_orders == "mandatory" ~ 1,
                              evac_orders == "voluntary" ~ 0),
      risk_area = case_when(risk_area == "yes" ~ 1,
                            risk_area == "no" ~ 0),
      expected_impacts,
      effect_evac_order,
      multiple_concerns,
      social_cues,
      freq_consult_info,
      unnecessary_evac_exp,
      age = age,
      education,
      gender = case_when(gender == "man" ~ 1,
                         gender == "woman" ~ 0),
   ) %>%
   filter(if_all(everything(), ~!is.na(.x))) %>% # na.omit()
   mutate_if(is.integer, as.double) %>%
   mutate(age = round_any(age, 10, floor)) %>%
   mutate(across(where(is.numeric), round, 0)) %>%
   mutate_if(is.double, as.ordered)

# dat_bn_dicretized %>% glimpse()

###### Descriptive ######
dat_bn_dicretized %>% tbl_summary() %>% as_hux_table()

dat_dicretized <-
   dat_bn_dicretized %>%
   mutate_if(is.ordered, as.character) %>%
   mutate_if(is.character, as.double)

dat_dicretized %>%
   tbl_summary(
      type = list(where(is.numeric) ~ "continuous"),
      statistic = list(all_continuous() ~ "mean {mean} (sd {sd})"),
      missing_text = "(Missing)"
   ) %>% as_hux_table()

# Empirical Cumulative Distribution Functions Conditional to Evacuation
dev.off(); par(mfrow = c(2, 6), mar = c(1, 3, 1, 1), pty = "s")
for (var in colnames(dat_bn_dicretized %>% select(-evacuation))){
   x = dat_dicretized[dat_dicretized$evacuation == 1, var]
   y = dat_dicretized[dat_dicretized$evacuation == 0, var]
   plot(ecdf(x),
        col = "blue", lwd = 1.5, lty = 1, xaxt = "n", yaxt = "n",
```

```
        verticals = TRUE, do.points = FALSE, col.01line = NULL,
        main = "", xlab = "", ylab = "", add = FALSE)
   plot(ecdf(y),
        col = "red", lwd = 0.75, lty = 1, xaxt = "n", yaxt = "n",
        verticals = TRUE, do.points = FALSE, col.01line = NULL,
        main = "", xlab = "", ylab = "", add = TRUE)
   axis(1, at = min(x):max(x))
   axis(2, seq(0, 1, by = .5))
   mtext(var, side = 1, line = 2.2, cex = .9)
   legend("bottomright",
          legend = c("1", "0"),
          col    = c("blue", "red"),
          bty    = "n",
          pch    = 15,
          cex    = .8
          )
}
# dev.copy(png, "./Figures/figure2.png", width = 2800, height = 1200, res =
300); dev.off()

# Correlation
corr <- cor_test(data   = dat_dicretized,
                 vars   = evacuation,
                 method = "pearson",
                 use    = "pairwise.complete.obs") %>% arrange(-cor, p)

# Matrix of p-values
p.mat <- cor.mtest(dat_dicretized, conf.level = 0.95)

# p.mat %>% view()
round(cor(dat_dicretized), 2) # %>% view()

par(mfrow = c(1, 1), mar = c(0, 0, 0, 0), bg = "white", pty = "m")
corrplot(cor(dat_dicretized, method = "pearson"),
         method = "color", type = "upper", tl.srt = 40, tl.col = "black",
         p.mat  = p.mat$p, sig.level = 0.05, tl.cex = .85,
         order  = "original",
         col    = RColorBrewer::brewer.pal(n = 10, name = "RdBu"))
dev.copy(png, "./Figures/figure3.png", width = 2000, height = 1200, res = 300);
dev.off()

# Detect multicollinearity
eigen(cor(dat_dicretized))$values
kappa(cor(dat_dicretized), exact = TRUE)

###### Logistic Regression ######
summary(model_1 <- glm(evacuation ~ social_cues,
                       data = dat_dicretized, family = binomial(link =
"logit")))
```

```
summary(logit <- glm(evacuation ~ unnecessary_evac_exp,
                     data = dat_dicretized, family = binomial(link = "logit")))

summary(model_1 <- glm(evacuation ~ expected_impacts + effect_evac_order,
                       data = dat_dicretized, family = binomial(link =
"logit")))

summary(logit <- glm(evacuation ~ .,
                     data = dat_dicretized, family = binomial(link = "logit")))

# ANOVA
anova(model_1, logit)

###### Standardized and performance ######
VIF(logit)
broom::tidy(logit, exponentiate = TRUE, conf.level = 0.95)
performance::r2_nagelkerke(logit)
performance::check_model(logit)

epiDisplay::logistic.display(logit, simplified = TRUE)[["table"]] %>%
as.data.frame() %>%
   rownames_to_column("var") %>% mutate(signif = case_when(`Pr(>|Z|)` <= 0.001
~ "Signif. 0.001",
                                                           `Pr(>|Z|)` > 0.001
& `Pr(>|Z|)` <= 0.01  ~ "Signif. 0.01",
                                                           `Pr(>|Z|)` > 0.01
& `Pr(>|Z|)` <= 0.05  ~ "Signif. 0.05",
                                                           TRUE ~ "Not
signif."),
                                        signif = factor(signif, levels = c("Not
signif.", "Signif. 0.05", "Signif. 0.01", "Signif. 0.001"))) %>%
   ggplot(aes(x = OR, y = fct_reorder(var, OR), fill = signif, color = signif))
+
   geom_point(shape = 21, size = 3) +
   geom_errorbar(aes(xmin = lower95ci, xmax = upper95ci), width = .1) +
   scale_colour_manual(values = rev(c("red", "green4", "blue", "black")),
                       breaks = c("Not signif.","Signif. 0.05", "Signif. 0.01",
"Signif. 0.001")) +
   scale_fill_manual(values   = rev(c("red", "green4", "blue", "black")),
                     breaks = c("Not signif.", "Signif. 0.05", "Signif. 0.01",
"Signif. 0.001")) +
   geom_vline(aes(xintercept = 1), size = .25, linetype = "dashed") +
   coord_trans(x = "log10") +
   scale_x_continuous(breaks = seq(0, 10, 1) ) +
   labs(title = "Logit regression predicting evacuation", x = "Odds ratio and
95% confidence intervals (log scale)",
        y = "", color = "", fill = "") + theme_bw()

effects_logit = margins::margins(logit)
summary(effects_logit)
```

```
par(new = TRUE, mfrow = c(1, 2), mar = c(7, 3, 3, 2), bg = "white", pty = "s")
for (var in c("expected_impacts", "effect_evac_order")){
   visreg::visreg(logit, var, scale = "response", partial = FALSE, rug = 2,
xlab = paste(var), ylab = "P(evacuation)")
}

###### BN ######
dag =
model2network("[age][freq_consult_info][evac_orders][gender][education][risk_ar
ea][social_cues][unnecessary_evac_exp][effect_evac_order|evac_orders][multiple_
concerns|age:gender:education][expected_impacts|age:freq_consult_info:gender:ed
ucation:risk_area:social_cues][evacuation|expected_impacts:effect_evac_order:mu
ltiple_concerns:unnecessary_evac_exp]")
par(new = TRUE, mfrow = c(1, 1), bg = "white", pty = "m")
graphviz.plot(dag, shape = "ellipse")
bn = bn.fit(dag, dat_bn_dicretized)
# coefficients(bn)
par(mfrow = c(1, 1), bg = "white", pty = "m")
graphviz.chart(bn, type = "barprob", grid = TRUE, draw.levels = TRUE, scale =
c(1, 1.2)) # c(1.2,2))
(pvalues = arc.strength(dag, data = dat_bn_dicretized))
par(new = TRUE, mfrow = c(1, 1), bg = "white", pty = "m")
strength.plot(dag, strength = pvalues, shape = "ellipse")
LL  = logLik(dag, dat_bn_dicretized)
k   = log(nrow(dat_bn_dicretized))/2
N   = nparams(dag, dat_bn_dicretized)
(BIC = LL - N * k)
score(dag, dat_bn_dicretized)

###### Exp.1 ######
dag =
model2network("[age][freq_consult_info][evac_orders][gender][education][risk_ar
ea][social_cues][unnecessary_evac_exp][effect_evac_order|evac_orders][multiple_
concerns|age:gender:education][expected_impacts|age:freq_consult_info:gender:ed
ucation:risk_area:social_cues][evacuation|expected_impacts:effect_evac_order:mu
ltiple_concerns:risk_area:unnecessary_evac_exp]")
par(new = TRUE, mfrow = c(1, 1), bg = "white")
graphviz.plot(dag, shape = "ellipse")
bn = bn.fit(dag, dat_bn_dicretized)
pvalues = arc.strength(dag, data = dat_bn_dicretized)
par(new = TRUE, mfrow = c(1, 1), bg = "white")
strength.plot(dag, strength = pvalues, shape = "ellipse")
score(dag, dat_bn_dicretized)

###### Exp.2: Adjusts ######
dag =
model2network("[age|education][freq_consult_info|age:gender:education][evac_ord
ers][gender][education|gender][risk_area][social_cues|age:gender:education][unn
ecessary_evac_exp|age:gender][effect_evac_order|age:gender:education:evac_order
```

```
s][multiple_concerns|age:gender:education][expected_impacts|age:freq_consult_in
fo:gender:education:risk_area:social_cues][evacuation|expected_impacts:effect_e
vac_order:multiple_concerns:risk_area:unnecessary_evac_exp]")
par(new = TRUE, mfrow = c(1, 1), bg = "white")
graphviz.plot(dag, shape = "ellipse")
bn = bn.fit(dag, dat_bn_dicretized)
pvalues = arc.strength(dag, data = dat_bn_dicretized)
par(new = TRUE, mfrow = c(1, 1), bg = "white")
strength.plot(dag, strength = pvalues, shape = "ellipse")
score(dag, dat_bn_dicretized)

###### Adjusted ######
dag =
model2network("[age][freq_consult_info|age:gender][evac_orders][gender][educati
on][risk_area][social_cues|age:gender][effect_evac_order|gender:evac_orders][mu
ltiple_concerns|education][expected_impacts|freq_consult_info:social_cues][evac
uation|expected_impacts:effect_evac_order:multiple_concerns:risk_area]")
par(new = TRUE, mfrow = c(1, 1), bg = "white")
graphviz.plot(dag, shape = "ellipse")
bn = bn.fit(dag, dat_bn_dicretized %>% select(-unnecessary_evac_exp))
par(new = TRUE, mfrow = c(1, 1), bg = "white")
graphviz.chart(bn, type = "barprob", grid = TRUE, draw.levels = TRUE, scale =
c(1, 1.2)) # c(1.2,2))
pvalues = arc.strength(dag, data = dat_bn_dicretized %>% select(-
unnecessary_evac_exp))
par(new = TRUE, mfrow = c(1, 1), bg = "white")
strength.plot(dag, strength = pvalues, shape = "ellipse")
score(dag, dat_bn_dicretized %>% select(-unnecessary_evac_exp))

###### Exp.3: Soft-evidence ######
dag =
model2network("[age][freq_consult_info|age:gender][evac_orders][gender][educati
on][risk_area][social_cues|age:gender][effect_evac_order|gender:evac_orders][mu
ltiple_concerns|education][expected_impacts|freq_consult_info:social_cues][evac
uation|expected_impacts:effect_evac_order:multiple_concerns:risk_area]")
par(new = TRUE, mfrow = c(1, 1), bg = "white")
graphviz.plot(dag, shape = "ellipse")
bn = bn.fit(dag, dat_bn_dicretized %>% select(-unnecessary_evac_exp))

ev <- list(multiple_concerns = "1")
# evidence vector
updated_dat <- cpdist(bn, nodes = bnlearn::nodes(bn), evidence = ev, method =
"lw", n = 1e6) # draw samples
updated_fit <- bn.fit(dag, data = updated_dat)
# refit: you"ll get warnings over missing levels
par(new = TRUE, mfrow = c(1, 1), bg = "white")
# plot
graphviz.chart(updated_fit, type = "barprob", grid = TRUE, draw.levels = TRUE,
scale = c(1, 1.2))
```

```
junction <- compile(as.grain(bn))
multiple_concerns_low <- setEvidence(junction,
                                     nodes = "multiple_concerns",
                                     states = "1")
querygrain(multiple_concerns_low)$evacuation
querygrain(multiple_concerns_low)$education

###### Exp.4: Soft-evidence ######
dag =
model2network("[age][freq_consult_info|age:gender][evac_orders][gender][educati
on][risk_area][social_cues|age:gender][effect_evac_order|gender:evac_orders][mu
ltiple_concerns|education][expected_impacts|freq_consult_info:social_cues][evac
uation|expected_impacts:effect_evac_order:multiple_concerns:risk_area]")
par(new = TRUE, mfrow = c(1, 1), bg = "white")
graphviz.plot(dag, shape = "ellipse")
bn = bn.fit(dag, dat_bn_dicretized %>% select(-unnecessary_evac_exp))

ev <- list(evacuation = "1")
updated_dat <- cpdist(bn, nodes = bnlearn::nodes(bn), evidence = ev, method =
"lw", n = 1e6)
updated_fit <- bn.fit(dag, data = updated_dat)
par(new = TRUE, mfrow = c(1, 1), bg = "white")
graphviz.chart(updated_fit, type = "barprob", grid = TRUE, draw.levels = TRUE,
scale = c(1, 1.2))

junction <- compile(as.grain(bn))
evac_yes <- setEvidence(junction,
                        nodes = "evacuation",
                        states = "1")
querygrain(evac_yes)$expected_impacts
querygrain(evac_yes)$effect_evac_order
querygrain(evac_yes)$risk_area
```