



Australian Government

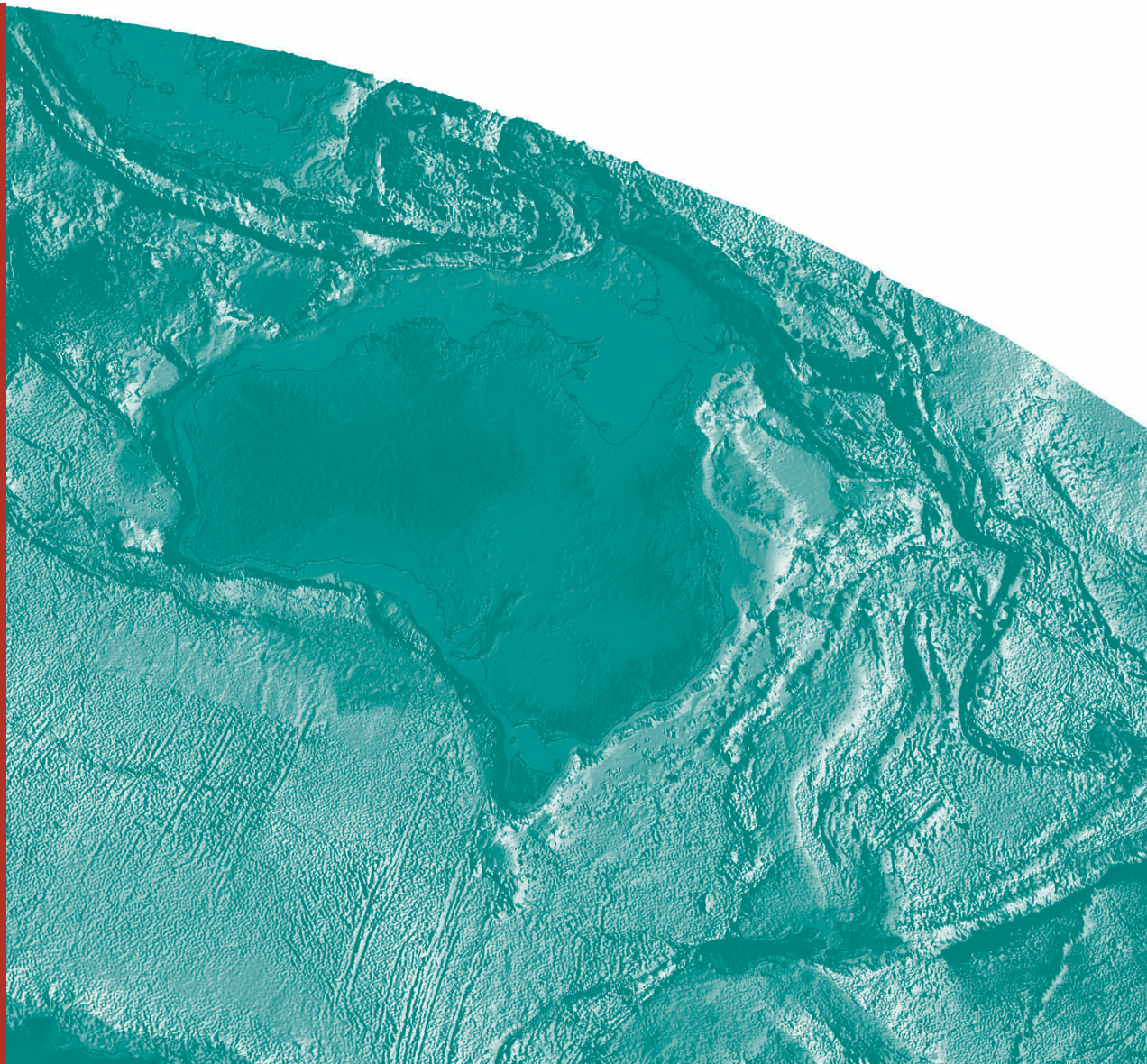
Geoscience Australia

Quantifying Social Vulnerability: A methodology for identifying those at risk to natural hazards

Dwyer, A., Zoppou, C., Nielsen, O., Day, S. & Roberts, S.

Record

2004/14





Quantifying Social Vulnerability: A methodology for identifying those at risk to natural hazards

Anita Dwyer, Christopher Zoppou, Ole Nielsen
*Risk Modelling Group, Geohazards Division,
Geoscience Australia,
Canberra, Australia*

Susan Day
*National Centre for Social and Economic Modelling (NATSEM),
University of Canberra,
Canberra, Australia*

Stephen Roberts
*Department of Mathematics, Mathematical Sciences Institute,
The Australian National University,
Canberra, Australia*



Australian Government
Geoscience Australia

Department of Industry, Tourism and Resources

Minister for Industry, Tourism and Resources: The Hon. Ian Macfarlane, MP

Parliamentary Secretary: The Hon. Warren Entsch, MP

Secretary: Mark Paterson

Geoscience Australia

Chief Executive Officer: Neil Williams

©Commonwealth of Australia 2004

This work is copyright. Apart from any fair dealings for the purposes of study, research, criticism or review, as permitted under the Copyright Act, no part may be reproduced by any process without written permission. Inquiries should be directed to the Communications Unit, Geoscience Australia, GPO Box 378, Canberra City, ACT, 2601.

ISSN: 1448-2177

ISBN: 1 920871 09 8

GeoCat No. 61168

Bibliographic reference: Dwyer, A., Zoppou, C., Nielsen, O., Day, S., Roberts, S., 2004. Quantifying Social Vulnerability: A methodology for identifying those at risk to natural hazards. *Geoscience Australia Record* 2004/14

Geoscience Australia has tried to make the information in this product as accurate as possible. However, it does not guarantee that the information is totally accurate or complete. **THEREFORE, YOU SHOULD NOT RELY SOLELY ON THIS INFORMATION WHEN MAKING A COMMERCIAL DECISION.**

Contents

Contents	iii
Foreword	v
Acknowledgments	vi
Executive Summary	vii
Introduction	1
1 Indicator Selection	8
1.1 Social Indicators	8
1.2 Social vulnerability and indicators	9
1.2.1 Study 1: Earthquake Disaster Risk Index	9
1.2.2 Study 2: Cities Project	12
1.2.3 Other vulnerability indicator studies	13
1.3 Selection Criteria	14
1.3.1 Criteria	15
1.4 Selection of vulnerability indicators	16
2 Risk Perception Questionnaire	18
2.1 Indicators and measurement	18
2.2 Perception of risk	19
2.3 Questionnaire development	20
2.3.1 Collection of national data	20
2.3.2 Generation of hypothetical individuals	28
2.3.3 Questionnaire design	28
2.3.4 Respondent demographics	29
2.4 Questionnaire results	30
3 Decision Tree Analysis	32
3.1 Decision trees and social vulnerability	32
3.2 Perception based classification	33
3.3 Decision tree analysis	35
3.4 Analysis of the risk perception questionnaire	35
3.4.1 Assigning classes of vulnerability	37
3.4.2 Developing decision rules	37

3.4.3	The relative importance of the 15 indicators	42
3.5	Decision rule findings	43
4	Synthetic Estimation	46
4.1	Perth Cities Project	47
4.1.1	Study Area	47
4.2	Synthetic estimation	48
4.3	Synthetic estimates for the Perth Case study	50
4.3.1	Matching questionnaire variables	50
4.4	Social vulnerability in the Case Study area	52
4.4.1	Estimated social vulnerability of the study area	53
4.5	Mapping social vulnerability using hazard scenarios	54
4.5.1	Mapped scenarios	54
4.6	Using the results	63
5	Discussion	66
5.1	Comparison with the Cities Project methodology	67
5.2	Limitations and future work	69
5.2.1	Indicator Selection	69
5.2.2	Risk Perception Questionnaire	69
5.2.3	Decision Tree Analysis	70
5.2.4	Synthetic estimation	70
5.3	Areas for refinement and further analysis	71
A	Risk Perception Questionnaire	72
B	Decision Tree Methodology	76
C	Microsimulation	81
D	Validation	84
	Bibliography	86
	Index	90

Foreword

The occurrence of natural hazards is not a phenomenon of recent times, however, understanding the *risk* from natural hazards is a relatively recent trend and is currently increasing at a greater rate than ever before. As population and infrastructure increases, social conditions fluctuate and the relationship between humans and their environment becomes more complex. All of these factors, and more, contribute to the wider picture of risk, including risk from natural hazards. While natural hazards will continue to occur, their capacity to become a disaster or merely a manageable event depends on many factors, including the magnitude of the hazard, the vulnerability of people and their communities, the built environment and political systems.

This report focuses on certain aspects of social vulnerability and its role in contributing to the risk from natural hazards. In particular, the study introduces a unique method of measuring the vulnerability of individuals within a household in order to contribute to the development of comprehensive natural hazard risk assessments. The research undertaken for this report has been driven by two needs: firstly, to develop a custom-made methodology of quantifying social vulnerability that can be incorporated into the risk models being developed by the Risk Research Group at Geoscience Australia. The Risk Research Group undertakes natural hazard research for the Australian Government, with the aim of developing risk models that assist decision-makers in better managing natural hazard risk to Australian communities. Secondly, the research outlined in this report has been influenced by a need to integrate social issues with hazard model development in order to investigate the greater risk to communities. Underlying this research is the need for a practical, albeit experimental, methodology of measuring elements of social vulnerability. Therefore, the report is a step-by-step account of the methodology development that aims to measure one aspect of social vulnerability, the vulnerability of an individual within a household, as a means of identifying those at risk to natural hazards.

Acknowledgments

We would like to thank Matt Hayne, Project Leader of the Risk Assessment Methodologies Project at Geoscience Australia for his support in the development of this study and his feedback on the report. Thanks also to Philip Buckle and John Handmer for their willingness to share their expertise, ideas and time with regards to social vulnerability over the past two years. Thank you also to David Robinson for his review and constructive comments.

Thank you to Dr Mihael Ankerst from The Boeing Company, Seattle, Washington for providing the the Perception Based Classification (PBC) software. And finally, thank you to all the people who completed the risk perception questionnaire.

Executive Summary

In this study, a methodology is developed to assess the vulnerability of individuals within households to risk from natural hazards. The methodology introduces a technique for measuring certain attributes of individuals living within a household that contribute to their vulnerability to a natural hazard impact. The methodology has four main steps;

Step 1: Indicator Selection

As the study is focusing specifically on *measuring* vulnerability, the indicators selected have been restricted to quantifiable indicators. Thirteen vulnerability indicators and two hazard indicators were selected for the study. The thirteen indicators are: Age, Income, Gender, Employment, Residence Type, Household Type, Tenure Type, Health Insurance, House Insurance, Car Ownership, Disability, English Language Skills and Debt/Savings. The two hazard indicators, residence damage and injuries, were included so that the following steps in the study were linked to a hazard context. The indicators are specific to people living in urbanised areas within an Australian city and were selected using selection criteria outlined in Chapter One.

Step 2: Risk Perception Questionnaire

In an attempt to identify how these indicators contribute to the vulnerability of a person within a household, a risk perception questionnaire was developed. The questionnaire was a means of collecting data on perceived vulnerability in lieu of the availability of actual vulnerability data. The questionnaire respondents were asked to rank the ability of hypothetical individuals to recover from a natural hazard impact based on their own perceptions of the situation. The hypothetical individuals were developed using the 15 indicators. The questionnaire was presented to ‘experts’ of disaster risk research and ‘non-experts’ for comparative purposes. The questionnaire results provided 1100 ranked hypothetical individuals, each with a unique set of indicator attributes.

Step 3: Decision Tree Analysis

Decision tree analysis is a classification methodology used to analyse and classify large sets of data. In this study, decision tree analysis was applied to the questionnaire data in order to sort and classify the data to find relationships between the indicator attributes. Based on data from the risk perception questionnaire, the decision tree analysis found 11 decision rules that determine high vulnerability to natural hazards. Each rule demonstrates that a combination of *two or more* indicator attributes are required

in order to predict the vulnerability of a person within a household, challenging the notion that one personal attribute can determine vulnerability. The one exception is Rule 1, which found that if a person suffers a life-threatening injury, they are automatically considered highly vulnerable.

The attributes, referred to as vulnerability indicators, of most importance relate to various levels of; injury sustained, residence damage, house insurance, income and type of house ownership. This finding suggests that individual and household finances, when combined with other specific indicators, play a significant role in determining an individual's vulnerability to a natural hazard impact.

Step 4: Synthetic Estimation

The Australian Bureau of Statistics (ABS) do not release highly detailed data relating to an individual, due to privacy laws. For example, we are permitted to know how many people in one census district (a census district is approximately 200 households in size) are over 55, how many people live alone and how many people are low income earners. However, we are not permitted to know how many people are over 55 *and* live alone *and* earn a low income, which is referred to as cross-correlated data.

As a result, synthetic estimation was undertaken for a study area of 224 census district areas in an area approximately 25 km north-east of Perth in Western Australia. In lieu of access to cross-correlated data to the real population, synthetic estimation, a technique using available data and microsimulation models, was used. Synthetic estimates of the population in the study area were developed, providing information on which census districts contain households with people identified as 'highly vulnerable' by the decision rules. The synthetic estimates can be mapped in order to provide a useful tool for representing aspects of social vulnerability to natural hazard impacts.

Future Steps

The report demonstrates that aspects of social vulnerability can be quantified in order to contribute to an understanding of risk to natural hazards. While experimental, the methodology outlines detailed processes that can be undertaken in order to capture and measure some of the complexities relating to social vulnerability.

Introduction

Australia is exposed to a wide range of natural hazards, including earthquake, cyclone, landslide, flood, storm surge, severe wind, bushfire, coastal erosion, hail storm and drought. At potential risk from these hazards are people, buildings, transport infrastructure, economies and communities, all of which are more greatly concentrated in urban centres. As one of the most urbanised nations in the world and with an ever increasing rate of urbanisation, Australia has a population that is largely concentrated within city areas [17]. The city, as articulated by Pelling, is ‘a focal point for a wider complex of economic, social, political and environmental linkages and flows of power, energy and information’ [58]. These complex linkages can often prove to be beneficial, such as providing greater accessibility to medical and educational services, greater development of social and cultural infrastructure and higher employment opportunities. While these services increase lifestyle capacity, it must also be noted that this density of infrastructure can also increase the risk from natural hazard events. Exposure to physical infrastructure is greater in urbanised areas, increasing the probability of injury or death in a natural hazard impact. The reliance in urbanised areas on complex utility networks such as water, electricity and telecommunications, which are exposed in a hazard event, may also increase the overall risk.

How each person will fare in the event of a natural hazard is influenced not just by exposure to infrastructure, but also by personal attributes, community support, access to resources and governmental management. This network of factors affecting social vulnerability to natural hazards, combined with the complex linkages found in cities and the behaviour of the hazard itself, all contribute to the development of a risk assessment.

Natural Hazard Risk Assessments

The fields of natural hazard research, risk management and emergency management use similar terms which often have different meanings. Some applications of the various terms stem from an academic or theoretical approach, while others come from a practical and managerial approach. As a result, there are many different and equally valid risk frameworks that have been developed. For the purpose of this report, the risk assessment framework adopted by the Risk Research Group has been used and is presented in Figure 1. As many of the terms that contribute to this framework are used throughout this report, they are defined below.

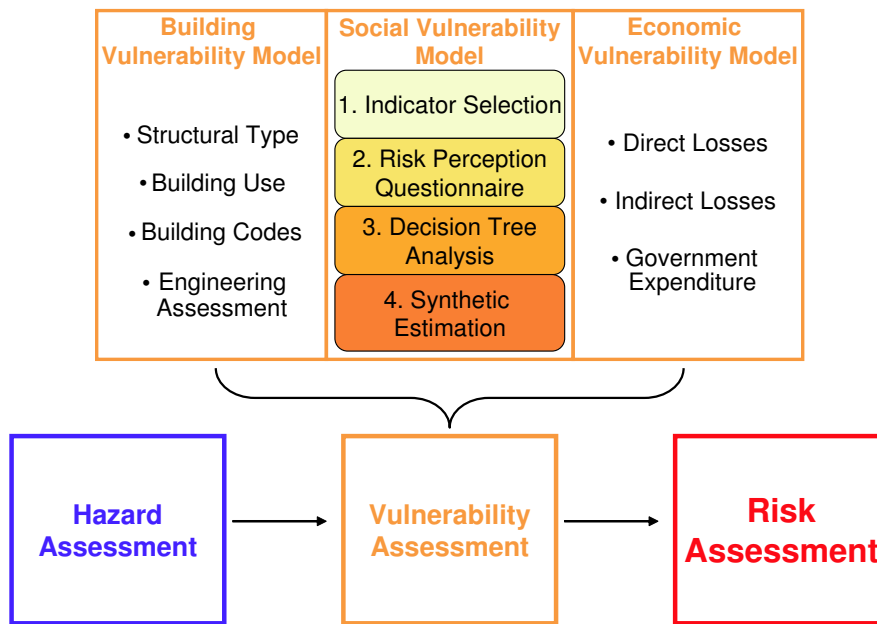


Figure 1: The steps involved in performing a Risk Assessment as followed by the Risk Research Group at Geoscience Australia. While most of the steps relate to traditional science disciplines such as geophysics, geology and engineering, the framework has now expanded to include other important factors, such as social vulnerability and economic assessments.

Risk

Risk refers to the consequences of an event. The Risk Research Group uses a simple expression to define risk: **‘Risk = Hazard * Elements Exposed * Vulnerability’**. This expression of risk is represented by the three-dimensional pyramid in Figure 2. The pyramid figure expands on the risk triangle developed by Crichton, where Crichton stated that: *”Risk’ is the probability of a loss, and this depends on three elements, hazard, vulnerability and exposure. If any of these three elements in risk increases or decreases, then risk increases or decreases respectively”* [19]. The risk pyramid aims to represent the three elements of risk in three dimensions, with the volume of the pyramid representing risk. Each edge of the pyramid is proportional to the three factors; hazard, vulnerability and elements exposed. The greater the contribution of one of the factors, the greater the volume and therefore risk.

Natural Hazard

A natural hazard is considered to be a specific natural event characterised by a certain magnitude and likelihood of occurrence. Common to all natural hazards is the uncertainty associated with both the hazards’ occurrence, the magnitude and the spatial extent of the hazard’s impact.

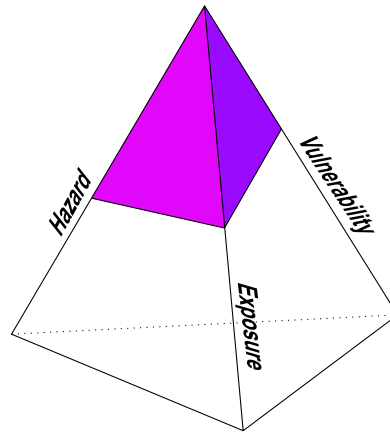


Figure 2: The risk pyramid shows the three independent factors that contribute to *risk*: *hazard*, *exposure* and *vulnerability* in a 3D format.

Elements Exposed

‘Elements exposed’ refers to the factors, such as people, buildings and networks, that are subject to the impact of specific hazard. Other elements that are not explored in this report but which are also exposed include the economy and the natural environment.

Vulnerability

Vulnerability refers to the capacity of an element exposed during the impact of a hazard event. Definitions of vulnerability to natural hazards generally refer to the characteristics of an element exposed to a hazard - road, building, person, economy - that contributes to the capacity of that element to resist, cope with and recover from the impact of a natural hazard [51].

Social Vulnerability

This study is investigating aspects of social vulnerability, an element that has been recognised as integral to understanding the risk to natural hazards [7, 59, 29]. For the purpose of this report, social vulnerability to natural hazard impacts has been simplified into four different levels;

- Individual within household (relating to personal attributes)
- Community (relating to how we interact with those around us)
- Regional/Geographical (relating to how far we are from services)
- Administrative/Institutional (relating to disaster funding and mitigation studies)

These four levels are shown schematically in Figure 3. This simplification aims to illustrate that there are diverse factors contributing to social risk from natural hazards, including those that relate to how hazards are managed by the region or nation we

live in, while others relate directly to individual attributes. This study focuses on only one aspect of social vulnerability, which is shown by the first level of vulnerability in Figure 1 and relates to an individual and their household. For the purpose of this study, this level of social vulnerability is defined as; **the ability of an individual within a household to recover from a natural hazard impact.**

Recovery

Ideally, ‘recovery’ in this situation would be determined by a person or community being able to revert to their social and economic state prior to being adversely affected by a natural hazard. However, natural hazard impacts are life-changing moments for many people. Therefore, in the context of this study, ‘recovery’ will refer to an individual within a household who attains a lifestyle state that is comparable to the one prior to the natural hazard impact.

Risk Assessment

Risk assessment refers to the analysis of various factors in order to establish the probability of a certain outcome from an uncertain event or suite of events. Such factors include the magnitude and probability of a hazard, the vulnerability of populations and the built environment and the overall loss or impact. Essentially, a natural hazard risk assessment is a study undertaken to determine the range of possible consequences from a natural hazard and is shown in Figure 1.

Risk Management

Risk management is the process of managing the possible outcomes from an uncertain event and traditionally involves four steps; “Mitigation, Preparedness, Response and Recovery”. The first two steps generally refer to the actions or measures taken prior to an impact, while the last two steps involve post-impact actions.

Risk management is a popular practice in financial and investment companies, research institutes and political studies. It involves measures of likelihood or probability and consequence. With regards to natural hazards, risk management refers to the activities undertaken to identify, control and minimise the impacts of natural hazards [51].

Mitigation

Vulnerability can be reduced by putting in place various measures that develop resiliency to a natural hazard impact. For example, if a building is vulnerable because its roof tiles are not tied down, then tying down the roof tiles can be a form of mitigation. Or, if a person is considered vulnerable because they do not understand what to do in the event of a cyclone, a simple education program will reduce their vulnerability. Essentially, these activities for reducing vulnerability, therefore risk, are called mitigation strategies, which is one step in the risk management process. It is the ultimate aim of a risk assessment to produce sufficient and accurate information that will allow a risk manager to put in place effective mitigation strategies.

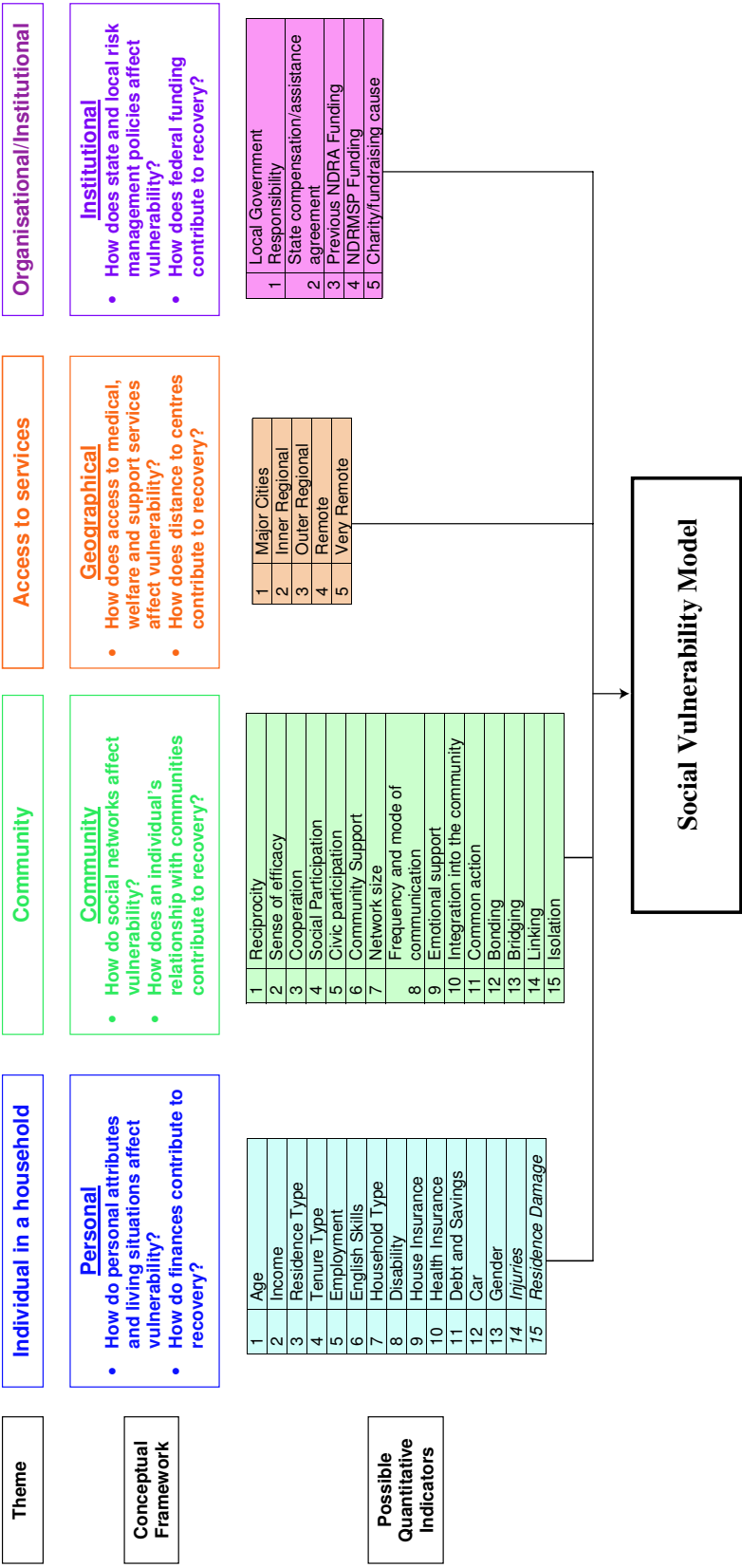


Figure 3: A schematic representation of some of the various factors contributing to social vulnerability. This study will focus on the first level of social vulnerability, which relates to the vulnerability of an individual within a household.

Objectives of the study

The research detailed in this report is the result of ongoing risk model development in the Risk Research Group at Geoscience Australia. It has been identified that the development of comprehensive risk models requires a better understanding of social vulnerability to natural hazard impacts. Therefore, this study aims to contribute to the Risk Research Group's model development by integrating a quantitative methodology of measuring aspects of social vulnerability to natural hazards.

By measuring some of the complex factors contributing to social vulnerability, we can assist risk managers in better safeguarding their communities. Some of the key questions explored in this report include:

- What factors contribute to the vulnerability of a person and their household to a natural hazard impact? Are physical attributes, such as age, disability and injury, more important than financial attributes, such as income, house ownership and debt, in contributing to vulnerability?
- Is vulnerability to a natural hazard impact perceived to be the same for different hazards?
- Can an approach be developed that is repeatable and comparable from one location to another?
- Can a single personal attribute determine vulnerability, or is vulnerability dependent on certain combinations of numerous attributes?
- Can combinations of vulnerability attributes of individuals within a household be mapped?

While these questions are not definitive of all points explored in this report, they are some of the key drivers behind the selection of methodologies, measurement processes and applications. The method is specific to urban communities in industrialised countries, however minor changes and the use of different data would make it possible for various communities in many different situations to be assessed. Therefore, one of the main hopes of this study is to contribute to the ongoing global development of vulnerability assessments that significantly value-add to total risk assessments.

Format of the report

The methodology development involves four main processes, or steps, and so Chapters 1-4 will correspond to each of the four sequential steps, while the final chapter will involve a discussion of the entire methodology and validation methods. The four significant processes involved in the methodology are represented by Figure 4 and are:

- Step 1: Indicator Selection
- Step 2: Risk Perception Questionnaire
- Step 3: Decision Tree Analysis

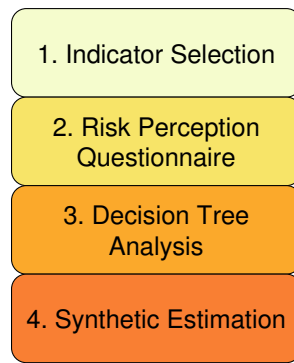


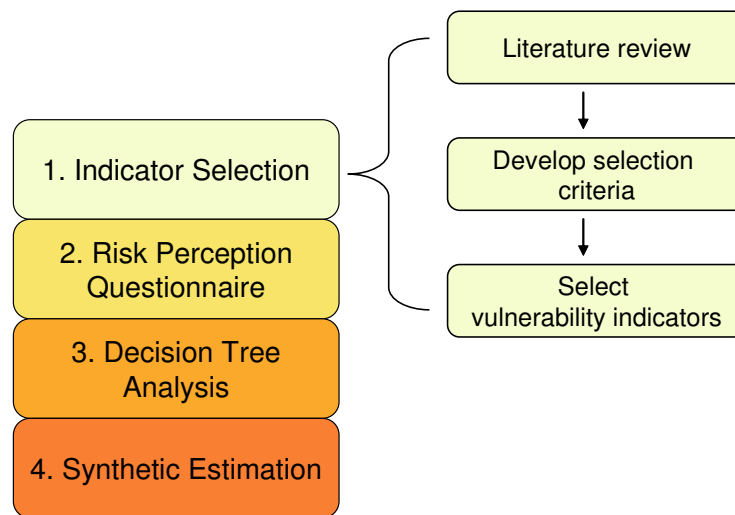
Figure 4: Four processes of investigating the vulnerability of a person within a household.

- Step 4: Synthetic Estimation

The final chapter in this report, Chapter Five, addresses some of these issues in detail and raises some points for further research and development.

Chapter 1

Indicator Selection



1.1 Social Indicators

Indicators are used to assess change over time of processes or phenomena that are difficult to directly measure [16]. The use of social indicators to monitor the change in status of people and communities has a long history in social science research and hence forms the basis of the indicators used in this methodology. Social indicators provide a means of measuring social characteristics to provide decision-makers with an effective and influential tool.

The applied use of indicators to critique social conditions originated around the 1830s, when social reformers in Europe, the UK and the USA used social statistics to improve public health conditions [16]. By the 1960s, social indicators were used by many governments around the world as a powerful component of policy development [48]. Many research and development agencies, such as the United Nations and World Bank devote substantial resources into developing social indicators and collecting relevant data on these indicators in order to develop comparative assessments of

communities across the globe. The Australian Bureau of Statistics (ABS) developed the Socio-Economic Indexes for Areas (SEIFA), which has indices of advantage, disadvantage, economic resources, education and occupation to contribute to social and welfare policy development in Australia. These indices provide some of the most comprehensive and up-to-date profiles of the Australian population.

However, it must be remembered that indicators only provide an ‘indication’ of much broader and complex social concepts. Researchers commonly note that good indicators must have a clear conceptual basis in order to measure what is intended [16]. As a result, the development of a list of indicators will inevitably exclude some features of a concept. Nonetheless, indicators can be used as a benchmark for assessing social effects and trends, and although some researchers have noted concerns with the use of social indicators, most note the importance of continuing investigations into their use [39, 38].

1.2 Social vulnerability and indicators

The concept of ‘measuring’ aspects of social vulnerability to natural hazards is one that has been explored widely in emergency and disaster literature for more than 30 years. However, research has largely focused on qualitative assessment methodologies rather than quantitative risk modelling. In part, this is due to the complex nature of people, social structures and culture, however, it is also due to the multi-disciplinary approach required to undertake such research. No single investigation into vulnerability indicators will provide a holistic and comprehensive answer, however, there are aspects of vulnerability that can be explored and represented through the development and application of quantitative vulnerability indicators. Since the 1990s, the application of indicators that ‘measure’ social vulnerability to a natural hazard has been explored. King and MacGregor [39] further explore this concept, arguing that ‘how’ and ‘why’ we measure vulnerability are just as important as ‘what’ we measure.

Most criticism targeted at incorrect application of vulnerability indicators is due largely to some of the key variables of social vulnerability being ignored or inadequately represented. There is a danger in trying to achieve a holistic answer to social vulnerability using a single method or discipline, as it is a complex, dynamic and variable aspect of disaster risk research.

The development of indicators of social vulnerability to natural hazards is a relatively small area of research, particularly within applications to industrialized nations. Some studies have investigated methodologies of vulnerability indicator development within a comprehensive risk assessment, including the Earthquake Disaster Risk Index [20] and the Cities Project [29, 28, 44]. Both of these methodologies acknowledge that social vulnerability is as much a part of risk as building damage, hazard magnitude and economic loss.

1.2.1 Study 1: Earthquake Disaster Risk Index

The Earthquake Disaster Risk Index (EDRI) was developed by Rachel Davidson [20], who describes it ‘as a composite index that allows direct comparison of the relative overall earthquake disaster risk of cities worldwide, and describes the relative con-

Indicators	Mexico City	San Francisco	Tokyo
xh1	21	46	49
xh2	12	48	44
xh3	82	41	32
xh1-xh3	27	46	45
xh4	19	33	37
xh5	20	65	65
xh6-xh7	31	39	48
xh4-xh7	23	46	50
xe1-xe4	45	43	89
xe5	54	33	93
xe6	27	65	66
xv1-xv5	43	23	24
xv6	32	56	28
xc1	36	39	97
xc2-xc3	46	26	76
xr1	68	34	17
xr2-xr3	54	23	18
xr4-xr5	68	70	-4
xr6	38	16	42
xr7-xr9	34	39	38
EDRI	38	37	54

Table 1.1: EDRI ranks for Mexico City, San Francisco and Tokyo, demonstrating that all factors, including hazard and vulnerability are significant in calculating total risk. For further information, refer to Davidson’s thesis. [20].

tributions of five factors to that overall risk’. The relationship between these factors and their associated measurable indicators used in the EDRI are shown in Figure 1.1. The EDRI is one of the earliest risk indices to incorporate both structural and physical damage indicators with social and economic indicators, providing a more holistic approach to measuring overall impact from a natural hazard. Davidson demonstrates that even in urban regions with low seismicity an earthquake could turn into a major disaster depending upon other characteristics of that city, such as population, use of adequate building codes, national GDP and housing vacancy rate. One of the strengths of the EDRI is that it has been re-evaluated over time so that trends in hazard intensity, economics and socio-demographics can be monitored.

Davidson developed an EDRI index of numerous cities around the world, providing a relative risk ranking of each city. In the calculation of the EDRI for each city, five main factors are measured and contribute to the overall EDRI rank (Figure 1.1). The relevance of each factor used to calculate the EDRI for Mexico City, San Francisco and Tokyo is shown in Table 1.1. A city such as Mexico City has a low hazard score compared with San Francisco, but a much higher vulnerability score. Therefore, the overall EDRI rank for Mexico City and San Francisco are similar, but for different reasons.

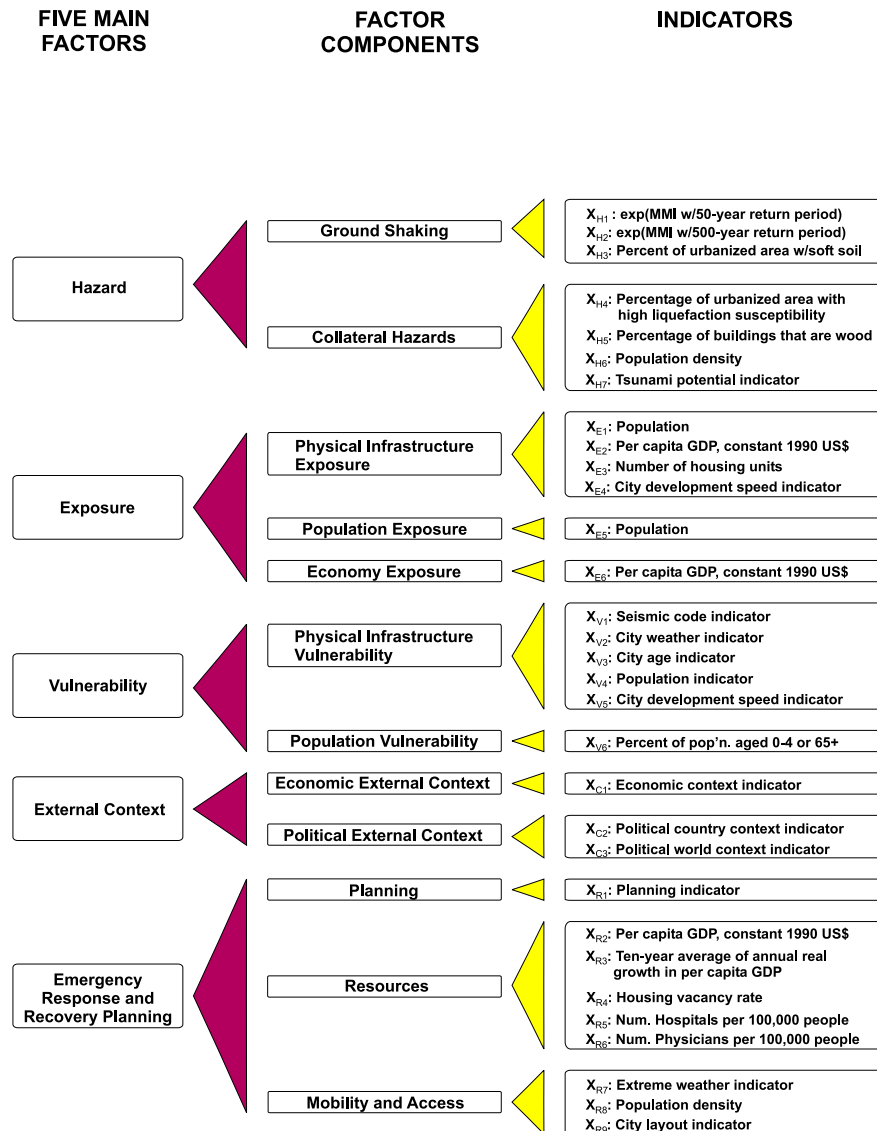


Figure 1.1: The five factors contributing to earthquake risk and their associated measurable indicators used in the Earthquake Disaster Risk Index (EDRI) [20].

1.2.2 Study 2: Cities Project

The Cities Project methodology for assessing relative community vulnerability was developed by Granger [29], who notes that a major influence in developing the methodology was Davidson's EDRI. The main difference between the two studies is that the Cities Project ranks numerous small areas, such as Census districts or local governments, within one city against each other, while the EDRI ranks numerous international cities against each other. Granger also notes that the EDRI is computationally and data intensive and for the purpose of small area comparison, the Cities Project methodology uses existing data and is designed to assist local governments [28]. In outlining the methodology, Granger and Hayne comment that there are few 'worked-through examples of a risk-index' such as the EDRI [28]. Both methodologies are two of the most comprehensive quantitative risk assessments available for natural hazard risk decision makers.

The Cities Project methodology for assessing community vulnerability involves the development of indicators that contribute to an overall 'relative risk rank'. The indicators are derived largely from the Australian Bureau of Statistics (ABS) 1996 Census [52] data and are grouped into categories referred to as the '5 esses'. These categories are *Setting*, *Society*, *Security*, *Sustenance* and *Shelter*. Within these five themes, the indicators are a collection of physical, structural, economic and lifestyle factors chosen to measure a community's vulnerability. The indicators have been chosen for their suitability for assessing risk from a range of natural hazards across a single city and provide a vulnerability estimate at the level of Census Collector's District (Census District), which is approximately 200 households. Granger notes:

"Because we are interested in showing the relative importance of each Census District to overall community vulnerability it was assumed that the most appropriate statistic to use would be the rank of the Census District in each measure. The use of rank is not without its problems. Inclusion of several variables that are highly correlated, or indeed derived from the same basic statistic, will obviously bias the outcome. Similarly, the inclusion of variables that have little, if any, bearing on community vulnerability could also distort the results. We feel, however, that with the careful selection of variables, rank is an appropriate statistic to reflect the relative significance of Census Districts." [28].

The Cities Project focuses on the vulnerability issues surrounding individuals and households. The '5 esses' indicators and their measurable variables used in the South-East Queensland Report [28] are listed in Table 1.2 and are considered representative for major urban areas. Some of the indicators are explained below:

- *Age*: Those over 65 and those under 5 were considered more vulnerable
- *Gender*: Females were considered more vulnerable than males.
- *Average House Occupancy*: Larger households were considered more vulnerable than smaller households.

The addition of the value for each indicator contributes to a final vulnerability assessment that is dimensionless. However, the final risk index calculates a relative

Setting	Shelter	Sustenance	Security	Society
Terminal facilities	Houses	Logistic Facilities	Public safety	Community facilities
Population density	Average house occupancy	Water supply facilities	Business premises	Large families
Gender	Flats	Telecommunications	Relative socio-economic disadvantage index (SEIFA/ABS)	Single parent families
	Average flat occupancy	Lifeline length	Economic Resources	Visitors
	Residential ratio		People under 5 years	Education and occupation
	Road network density		People over 65 years	New residents
	Cars		Households renting	No religious adherence
	Households with no car		Un-employment	Elderly living alone

Table 1.2: Vulnerability indicators, grouped by the 5 essences (Setting, Scene, Shelter, Security, Society), used in the Cities Project on multi-hazard risk assessment in South-East Queensland [28]

rank for each area, in this case the census district, being studied. For each of the census districts, values for each of the 31 indicators shown in Table 1.2 were added, with the final score representing a relative rank of vulnerability. The ranking methodology does not explore the relationship between indicators, that is, there is an assumption that each indicator is independent. In addition, each indicator (except for two where Granger has weighted some facilities/buildings as greater than other indicators depending upon a subjective judgement) has the same weight. However, the strength of this methodology is that it is relatively simple and consistent.

1.2.3 Other vulnerability indicator studies

There are many studies of community vulnerability to natural hazards. With the exception of EDRI and the Cities Projects, there are few comprehensive, quantitative and applied methodologies. However, there are some other important studies that, while not necessarily comprehensive in terms of assessing physical, social and economic vulnerability, do investigate quantitative assessments of vulnerability to natural hazards using indicators. Some of the more useful studies include:

- **A Hurricane Disaster Risk Index (HDRI)** [41] This study is based on the EDRI methodology [20], but explores risk to hurricanes rather than earthquakes. Four factors of risk; (i) hazard, (ii) exposure, (iii) vulnerability, (iv) emergency response and recovery are measured and analysed to give various cities an HDRI rank.
- **Creating a Safer City: A Comprehensive Risk Assessment for the City of Toronto** [24] This study is similar to the Cities Project as it assesses the risk to one city from multiple natural hazards. Using a simple cumulative approach to

ranking the different administrative areas of Toronto, rather than requiring any algorithms as employed by the EDRI and HDRI, this risk assessment also recognises the importance of including hazard factors combined with social factors. Some of the hazard factors include spatial extent of hazard and probable magnitude, while some of the social factors include housing and age distribution of people. Like the Cities Project, the Toronto assessment does not weight any of the social variables.

- **Transportation Performance, Disaster Vulnerability and Long-Term Effects of Earthquakes** [15] In reviewing the affects of the 1995 Kobe earthquake on infrastructure, Chang based her vulnerability ranking on five indices: 1) *Relative Damage Index* (dependent on damaged buildings per population), 2) *Vulnerability index* (percent of jobs found in small businesses), 3) *Employment Change Index* (change in total number of jobs in the administrative area between the 1996 and 1991 censuses), 4) *Competitiveness Change Index* (the unexplained residual from the shift-share analysis normalized to 1996 employment, or what is typically referred to as the change in competitiveness for the locality), 5) *Accessibility Index* (highway transport accessibility). These indices establish relative rankings of the various administrative areas in Kobe to provide an earthquake risk assessment following the 1995 Kobe earthquake. This study is one of the few that has integrated impacts to network systems in developing a quantitative index, which is integral to assessing how and when people recover from a natural hazard impact.
- **HAZUS Model** [4] Restricted to flood, earthquake and hurricane hazards in the USA, the HAZUS model is perhaps one of the very few integrated risk assessment models. With input from many researchers and with over ten years of development, the HAZUS model is a technical model that also employs limited quantitative measurements of social vulnerability factors, such as ethnicity and gender. However, other social factors, such as age, disability, insurance levels and the links that social vulnerability has to greater economic losses are not included.

1.3 Selection Criteria

The indicators selected for various studies depends on the purpose of the study, the research discipline being explored and the final application. Academics and organisations involved with indicator development commonly devise selection criteria to ensure that the most appropriate indicators are selected for their purpose. Common themes are listed in Table 1.3. The criteria for indicator selection for this study has borrowed largely from established reviews and literature, in particular, the Rossi and Gilmartin [61] study referred to in the development of the EDRI and are outlined in the section below.

Andrews and Withey, 1976 [2]	Davidson, 1997 [20]	Cobb and Rixford, 1998 [16]	Krumpe, 2000 [40]	King and Macgregor, 2000 [39]
Monitored over time	Validity	Clear and conceptual basis	Quantitative	Developed with a theoretical model
Disaggregated to relevant social level	Data availability	Narrow range	Reliable	Fixed set of tested indicators
Coherent	Data quality	Multiple indicators for same phenomenon	Responsive	Existing data availability
	Quantitative	Make substantive justice a priority	Sensitive	
	Objectivity	Reveal causes, not symptoms	Discriminating	
	Understandable	Have control over resources	Indicative	
	Directness		Significant	

Table 1.3: Selection criteria for the development of an indicator set. This study borrows from established criteria lists, but it is by no means exhaustive of appropriate criteria, merely representative.

1.3.1 Criteria

- **Support Concept** The most important aspect of indicator development is to ensure the indicators selected serve the needs of the research question” [39]. Indicators, therefore, must be viewed as tools used to articulate a concept.
- **Validity** Indicators need to accurately represent concepts expressed in the model or be acknowledged as valid substitutes for these concepts [20]. To ensure validity, the indicators must use credible data and be verifiable.
- **Data Availability and Quality** Most criteria lists specify that data must be available for each indicator and from a reliable source. To ensure quality, the data must be credible and reproducible.
- **Sensitivity** Indicators assume a temporal aspect, as they are designed to measure change in a system or process. Indicators must be aligned with the time-scale on which they capture change, whether days, months or years. This will ensure that the indicators are sensitive to change over time and therefore provide greater insight into the details of the factor being measured.
- **Simplicity** Indicators are used to represent concepts to a variety of people and therefore should be easily understood while also reflecting the complexity of the concept represented. Therefore, it is ideal that indicators are unambiguous and accessible.
- **Quantitativeness** Indicators must be measurable via a readily understood method. Clear methods of measurement will encourage wider indicator acceptance and limit the bias or subjectivity of the data collator. A well explained method of quantitative measurement will provide clarity and can be comprehended by decision makers.

Number	Indicator	References
1	Age	[66, 59, 12, 39, 29]
2	Income	[66, 59, 8]
3	Residence Type	[59, 8]
4	Tenure	[59, 45]
5	Employment	[14]
6	English Skills	[14, 45]
7	Household Type	[39, 29, 14, 45]
8	Disability	[59, 14]
9	House Insurance	[67]
10	Health Insurance	
11	Debt and Savings	
12	Car	[66, 29, 14]
13	Gender	[59, 26, 29]
14	Injuries	[4]
15	Residence Damage	[4]

Table 1.4: The thirteen socio-economic indicators and two hazard indicators used in this study to establish the vulnerability of a person within a household to natural hazard impacts.

- **Recognition** The concept of measuring vulnerability is not new, however, the techniques used can often vary and new developments are made over time. For the purpose of this study it is considered important to use indicators that have already been recognised by researchers as important contributors to social vulnerability.
- **Objectivity** Indicators must have classifications that can be explored further by other methods, such as surveys, statistics and data analysis, and by other researchers. Therefore, the indicator should remain the same over time, however the data for the classifications will change to reflect trends.

1.4 Selection of vulnerability indicators

The indicators chosen for this study, listed in Table 1.4, have been selected from extensive literature reviews, discussions with researchers and with the aim of exploring quantitative methods of assessing the vulnerability of an individual within a household to a natural hazard. While not exhaustive of factors that contribute to a person's vulnerability, this combination of quantifiable variables should provide an indication of which people are the most vulnerable in a community in the event of a natural hazard impact. Other factors contributing to a person's ability to recover include spiritual, emotional and psychological capacity, as well as sense of community and other less tangible factors. These factors are listed in Table 1.5 and while investigating qualitative indicators is beyond the scope of this study, it is noted that they are also important in understanding social vulnerability to natural hazards.

The indicators in Table 1.4 have been selected with a similar philosophy to the EDRI. That is, the indicators of social vulnerability are only relevant when used in combination with a hazard assessment. However, this study varies from the EDRI and other vulnerability studies listed, as it explores the relationship between the selected

Indicator	References
Sense of Community	[1, 39, 13]
Emotional Capacity	[14, 57]
Psychological Capacity	[25, 57]
Trust in Authority Figure	[26, 37, 22]
Understanding of Natural Hazard	[39, 1, 29]
Perception of Risk	[37, 63, 33, 25]
Capacity for Change	[23, 34]
Core Beliefs and Values	[39]
Preparedness and capabilities of Local Government	[33, 37]

Table 1.5: Nine qualitative indicators of a person's vulnerability to natural hazard impacts. These indicators are complex and more difficult to quantify than those shown in Table 1.4 and are not included in this study.

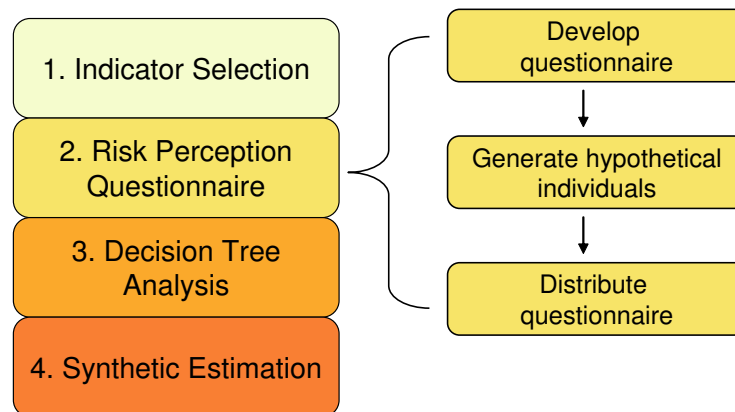
vulnerability indicators of a person within a household.

As this study is focusing on investigating aspects of social vulnerability and not hazard, two indicators relating to hazard have been included in order to provide a context for investigating vulnerability. These indicators are Injuries and Residence Damage and are listed as indicators 14 and 15 in Table 1.4. Indicators 1 to 13 are socio-economic variables that provide an insight into an individual's characteristics, while the indicators Injuries and Residence Damage relate to the impact of the hazard. The inclusion of these two indicators allows us to investigate social vulnerability in terms of recovery, as outlined in the introduction, while also investigating the relationship the other indicators have with Injuries and Residence Damage. Therefore, while it is acknowledged that the indicators Injuries and Residence Damage are not vulnerability indicators, but hazard indicators, they are included as the 15 selected indicators within this report.

Although the indicators listed in Table 1.4 are frequently noted in the literature as contributing to a person's vulnerability, they do not necessarily, in isolation, make a person vulnerable. However, a combination of these indicators, or the relationship between indicators, may render an individual highly vulnerable to a natural hazard impact. One of the main objectives of this study is to explore whether a combination of quantitative indicators, based on the likelihood of one indicator being associated with others, will provide a more accurate representation of vulnerability. To investigate this further, we must gain further insight into perceptions of vulnerability and risk. The next chapter outlines a Risk Perception Questionnaire that was developed and distributed in order to explore the relationships between the 15 selected indicators and their importance in various natural hazard scenarios.

Chapter 2

Risk Perception Questionnaire



2.1 Indicators and measurement

The use of indicators to measure the vulnerability of people and communities has been explored in various natural hazard assessment studies, as discussed in Chapter One [29, 20, 24, 4]. Most do not apply weights to the vulnerability indicators and hence the indicators are generally considered to be independent and equally important variables. That is, the effects of a combination of particular indicator values compared with other combinations are not explored. Some studies weight indicator values according to subjective perceptions of the importance of certain indicators [28]. However, the studies referred to in Chapter 1 have not explored the effects that combinations of indicators will have on a person's vulnerability. For example, an elderly person may not be vulnerable just because of age, but when combined with living alone, not having a car, having a disability and a low income, vulnerability may increase. However, if an elderly person lived with other people, had both house and health insurance and had a very high level of savings, then their vulnerability may decrease. This chapter will introduce a means of addressing this issue.

Indicators and weighting

As a condition of her application of weights to indicators, Davidson notes that ‘no amount of clever mathematical manipulation will uncover the ‘correct’ weights for the EDRI, because no single correct set of weights exists *a priori* [20]’. Davidson’s student, Lambert, further explored Davidson’s study of indicator weights to employ the analytical hierarchy process (AHP) for weighting the indicators of vulnerability in her study of the Hurricane Disaster Risk Index.

To investigate the possibility of applying weights, two approaches can be taken; one which investigates the subjective application of weights based on a researchers’ local knowledge, experience and intuition. This approach is generally qualitative and can vary greatly according to the perspective of the researcher. The second approach can employ more objective methods, focusing on a quantitative approach that minimises a researcher’s opinion or bias. This study does not dismiss the subjective approach, which is appropriate in some situations, and recognises that every approach will have some level of subjectivity. However, there is a need to further investigate if and how objectivity, through quantitative approaches, can be used to understand aspects of vulnerability.

2.2 Perception of risk

Many indicators of vulnerability to natural hazards cannot be measured by data sets due to the scarcity of comprehensive and relevant data pertinent to natural hazard events. However, the role of risk ‘perception’ can be of significance when studying social vulnerability measures. Slovic noted that if people perceive a risk to be real, then they will behave accordingly [63]. Therefore, capturing information regarding people’s perception of risk is valuable in understanding people’s behaviour. While perception of risk is not the same as actual risk, it does provide some insight into how people may behave in the event of a natural hazard in the absence of the availability of actual data-sets. Perception of risk also provides a view into what people value and what importance they place on certain factors in the event of an actual natural hazard impact. Such information is useful in determining how people will recover if these factors are affected during a hazard event

Each day people conduct their own risk management decisions, which include consciously and subconsciously reviewing the possible consequences and benefits of risk. For example, does the consequence of running across the road against a red light outweigh the benefit of arriving on the other side to catch the bus in time? People will make different decisions based on their own perception of risk which, in turn, is founded in their own education, experience, fear and emotional capacity. Studies of risk perception are vast and span many research disciplines from psychology and geography to engineering systems and sociology. It is beyond the scope of this study to engage in the epistemology of risk perception, however the basic concepts are essential to the use of the selected indicators in a quantitative assessment of vulnerability.

Early research investigating strategies for studying perceived risks involved the development of the psychometric paradigm by Fischhoff *et al.* [25]. The paradigm investigates the means by which people make quantitative decisions on their perceived risk to a situation and is widely referred to in the literature. While some researchers

have challenged the psychometric paradigm, it has demonstrated that perceived risk can be quantified and predicted in the demonstrated situations [63, 25, 37]. It must also be acknowledged that a comparison between perceived risk to natural hazards and actual risk measured after an impact has not been explored using the psychometric paradigm. However, in situations where actual risk is often unknown or untested, such as the risk from natural hazards, perceived risk may be considered a substitute for actual risk. Therefore, any investigation into quantitative assessments of vulnerability to natural hazards, where the impact of the hazard is untested, should take into account the perception of the natural hazard risk.

2.3 Questionnaire development

The development of the questionnaire is based on asking participants to rank the vulnerability of ten hypothetical individuals, where vulnerability is defined as ‘a person’s ability to recover’, to various natural hazard impacts. For the purpose of this study, this rank will be considered a measure of social vulnerability. The individuals will be ranked according to the participant’s perception of the impact of the natural hazard as per the indicators used in Table 1.4. Essentially, it is a measure of the perceived capacity of a person to recover from an impact event given the variables presented.

2.3.1 Collection of national data

In order to develop hypothetical individuals that are as realistic as possible, that is, that they resemble an urban Australian community, reliable data sources are required. At the time of the data source collection for this study, the 1996 Census undertaken by the Australian Bureau of Statistics (ABS) was the most current and comprehensive data source. While most of the data was obtained from the ABS Census, not all of the 15 indicators are represented in the Census, as noted in the section below. The data sources are used to provide a fair representation of Australians that could be exposed to a natural hazard. The data and context for 12 of the indicators are shown in the following figures. The rationale for their selection is also discussed. The two hazard indicators and one vulnerability indicator, House Insurance, do not have cross-correlated data for representation in figures, so their context and rationale is explained as follows:

Indicator 9: House Insurance

Approximately one-third of Australian houses are uninsured, for both house and/or contents insurance. Insurance becomes a significant factor in a person’s ability to finance repairs or rebuild after a natural hazard impact [8].

Indicator 14: Injuries

The HAZUS risk assessment program has an injury classification scale, which has been slightly modified for this study. Four classes have still been used, however, the class of ‘killed or mortally injured’ has been replaced with ‘no injuries’. The four classes are therefore: 1)No injury, 2)Basic medical treatment without hospitalisation, 3)Hospitalisation but expected to recover and 4)Hospitalisation with life-threatening

injuries. This study is investigating vulnerability, defined as 'ability to recover' of an individual within a household, which is not applicable to the deceased. However, as the impact of a natural hazard is greater than the physical injuries received, the class 'no injuries' has been included.

Indicator 15: Residence Damage

The HAZUS risk assessment program has a building damage state classification scale, which has been slightly modified for this study. An extra class of 'no damage' has been added to the HAZUS classes with a total of five classes. The classes are: 1)No damage, 2)Slight damage, 3)Moderate damage, 4)Extensive damage and 5)Complete damage.



Figure 2.1: **Indicator 1: Age.** Population by Age: The distribution of the Australian population by age in 1996 [52].

There is a uniform decline in the population number after 50 years of age, with few individuals over 104 years of age. The average life expectancy of non-indigenous born Australians is 75 years for males and 81 years for females [56]. Mobility, access to resources, and financial capacity are some of the recovery issues for older and younger people [14, 39, 66, 59, 29].

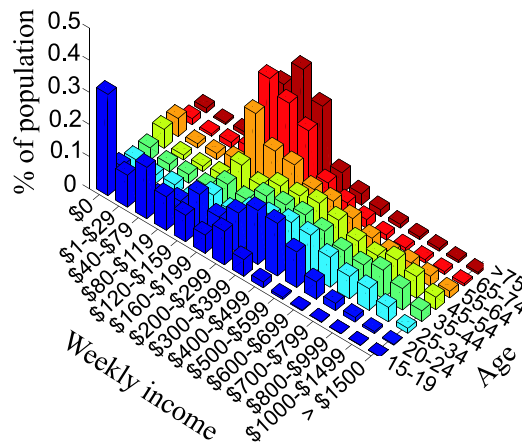


Figure 2.2: **Indicator 2: Income.** Weekly Individual Income by Age: Individual weekly income (after tax) distribution by age in 1996 [52]. The median income of 15-19 year olds is \$40-\$70. The median income rises to \$300-\$399 for 20-24 year olds, \$400-\$499 for 35-54 year olds, declining to \$160-\$199 for those over 55 years of age [56]. Income is one of the key indicators representing ability to pay for services and resources that may otherwise not be readily available after a hazard impact, such as accommodation, cars, clothing and comfort items [8, 66, 29].

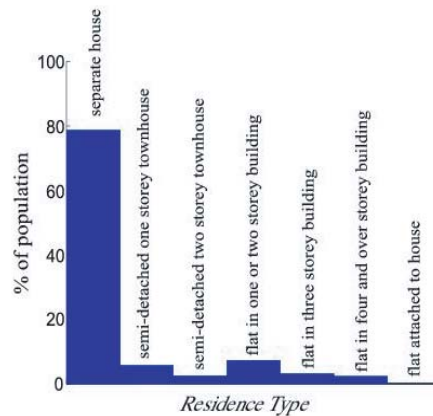


Figure 2.3: **Indicator 3: Residence Type.** Residence Type by percentage of population: The distribution of household types in 1996 [53]. Over 78% of the population live in a house. Flats are the next most common dwelling type, with 12% falling into this category. Residence type provides an insight into the physical safety of people, for example, a flood will be less of a risk to those living on the 4th floor of a multi-storey building than an earthquake.

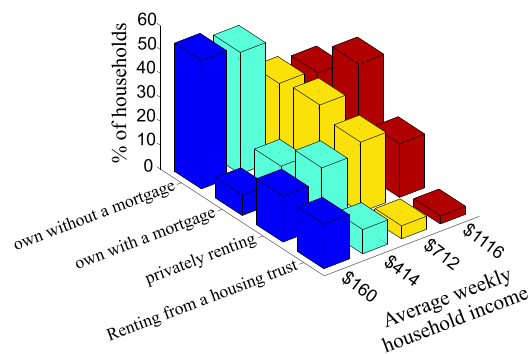


Figure 2.4: **Indicator 4: Tenure Type.** Tenure Type: The distribution of household types in 1996 by average weekly household income from the Household Expenditure Survey, Australian Bureau of Statistics [56].

Approximately 53% of home owners have an average weekly income of \$160, with 51% having \$414. The majority of 20-34 year olds have not purchased a home and 25% of them rent privately. 5% of the population rent from a housing trust. Structural damage to a residence after a natural hazard impact will create varying financial burdens according to whether the residence is rented, owned or mortgaged [59]

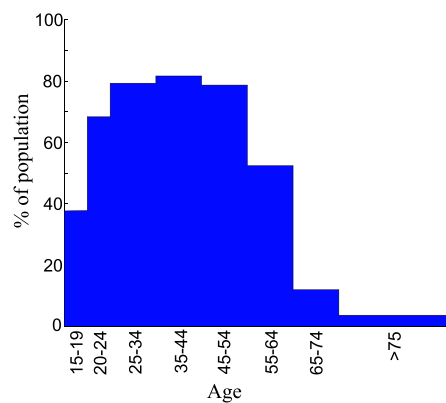


Figure 2.5: **Indicator 5: Employment** Employment by Age: The distribution of employment (working more than hour per week) by age within Australia in 1996 [52].

Over 78% of 25-54 year olds are employed in Australia, while 38% of 15-19 year olds are employed. 52% of those eligible to retire still work. Employment, regardless of income, indicates that a person is part of a network of people and has the capacity to earn money, both of which are important in the event of a natural hazard impact.

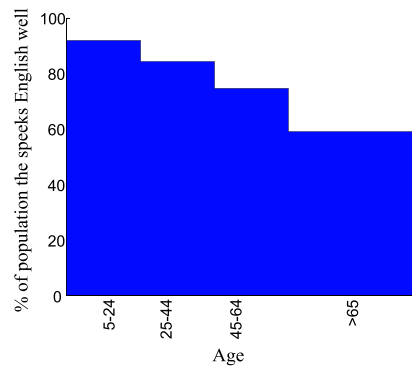


Figure 2.6: **Indicator 6: English Skills.** Proficiency in English by Age: The proportion of the Australian population, by age, that can speak English well in 1996 [53]. In 1996, 99% of Australians spoke some level of English, with 15% speaking a language other than English at home. Communication in disaster risk management is essential and can be difficult if proficiency in English is an obstacle to effective communication [14]. Migrants and refugees may have come from situations where there is a mistrust of Government, affecting their trust of authority’s advice or warnings.

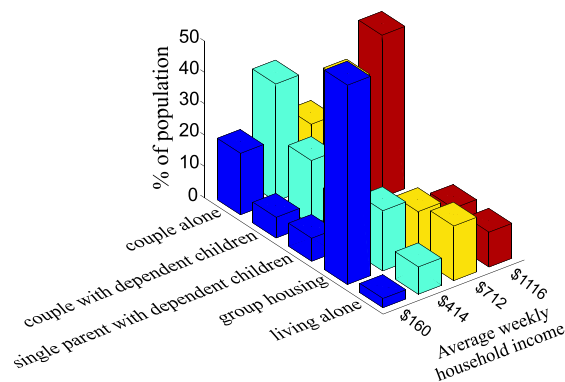


Figure 2.7: **Indicator 7: Household Type.** Household composition by Income Quintile Group: The distribution of household types within Australia [53]. Over 6% of households in Australia are single parents with dependents, 24% are living alone and 9% live in group housing. In a disaster the most immediate support network is often the immediate household and hence extended family, friends, partner or a group house implies some level of support network. Those who live alone may be less likely to have an immediate support network and may be more vulnerable if injured at home during an impact.

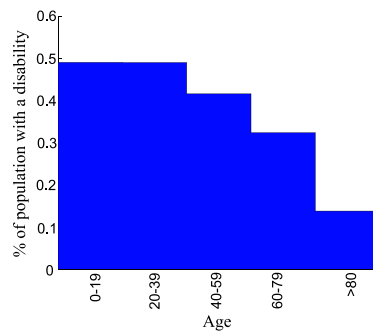


Figure 2.8: **Indicator 8: Disability.** Users of all CSDA (Commonwealth-State Disability Agreement) Services: The proportion of the Australian population with disabilities by age [50].

Approximately 0.3% of the Australian population live with a disability, according to the Australian Institute of Health and Welfare. Of these, 76% have a mental/emotional disability, 13% a physical disability, 5% a sensory disability and 6% a neurological disability. People with sensory disabilities (speech, hearing, vision) are at greater risk from not receiving information before, during and after a hazard event. It must be noted that 'people with a disability' is one of the hardest groups to measure, as many people don't identify as a person with a disability, while the elderly do not often identify as disabled but merely 'old'. For this study, data relating to the use of Government disability services is used.

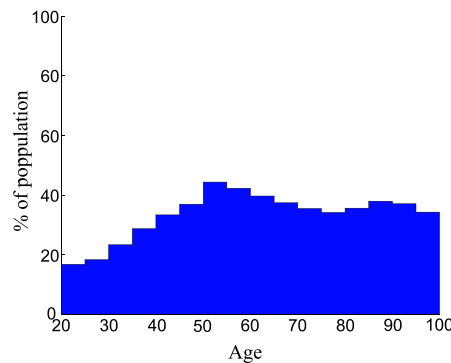


Figure 2.9: **Indicator 10: Health Insurance** Private Hospital Insurance: The proportion of the Australian population by age that has health insurance [18]. Approximately 40% of the population 55 years of age and greater has health insurance. Australia's health system ensures that medical treatment is available to everyone during times of crisis. However, long-term treatment, including elective surgery, could be at risk if a person does not have insurance.

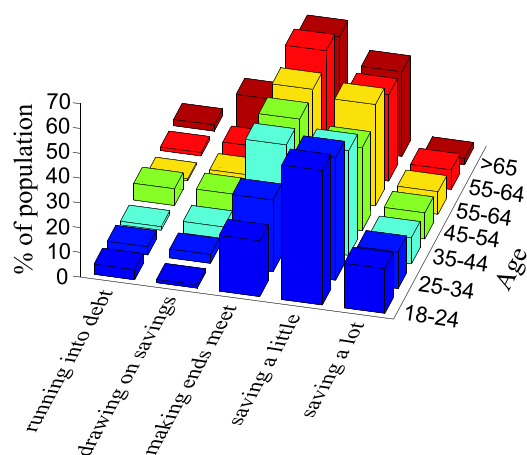


Figure 2.10: **Indicator 11: Debt and Savings.** The Financial situations of households by Age: financial situation of Australian households by age during September 2000 [31].

Over 2.5% of households are running into debt, with 0.5% in the age group 45-49 years old and 0.5% in the age group 65 years of age and greater. Approximately 6% of the population is drawing on savings. Over 36% of the population are making ends meet. This compares with 42% that are saving a little, mainly in the 18-44 year old age group and only 11% reporting that they save a lot. Households with high levels of savings suggest that they have some financial resources to utilize in times of great need.

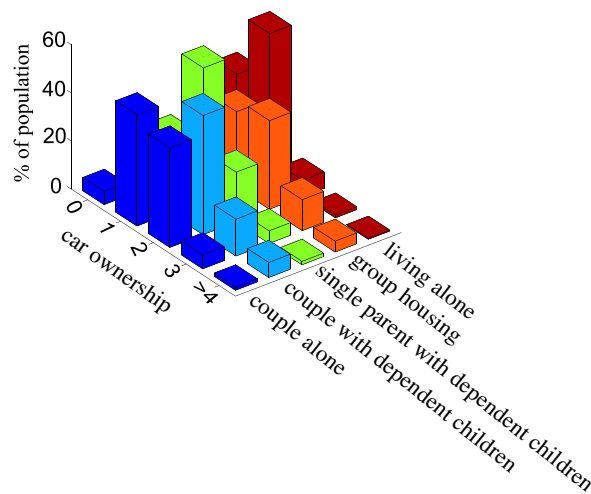


Figure 2.11: **Indicator 12: Car Ownership.** Number of motor vehicles by Household Type: the distribution by income of the number of cars owned by the Australian population [53].

12% of households do not own a car. Over 42% own a single car, 34% own two cars, 9% own three cars and 3% own four or more cars. Transport is important in both mitigation and recovery processes [14, 29]. However, in times of hazard impacts, a car is another element that is exposed and therefore susceptible to damage.

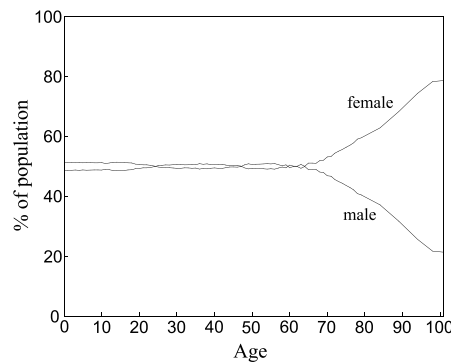


Figure 2.12: **Indicator 13: Gender.** Gender per percentage of total population: The distribution of males and females by age 1996 for the Australian population [53].

There are still some situations in Australia whereby women are more disadvantaged compared to men, for example, women's incomes are lower overall and the majority of single parents are women [56]. Australian urban centres have people from numerous cultural backgrounds, some that traditionally disempower women. Some researchers have demonstrated that women are more vulnerable to natural hazards, while others have established that women are better able to come together to support each other and recover more quickly than males [29, 59, 26].

Indicator	Classification	Restriction/Assumption
Age	< years	As the individual developed refers largely to a household situation, it is assumed that most people aged under 15 live with older people
Income	<\$280 p/w	People who are earning money from employment will receive more than \$280 per week, which relates to the Henderson Poverty Line. It is assumed that anyone with less than this amount per week will be receiving a benefit, therefore the questionnaire will refer to people 'receiving' when it is <\$280 per week, and 'earning' when it is >\$280.
Household Type	Group	Those noted in the Census as multi-family households and group households are all classified as group
Debt and Savings	Saving a lot	This classification is restricted to those earning >\$280 as it is assumed anyone below the poverty line cannot 'save a lot'.
Car Ownership	3+ cars	Those receiving below the poverty line, <\$280 p/w, have been excluded from this category.
House Insurance	Renting	Those who are renting will only have contents insurance and not house insurance

Table 2.1: Constraints imposed in the generation of hypothetical individuals

2.3.2 Generation of hypothetical individuals

Development of the questionnaire involves the generation of hypothetical individuals, with a range of values from the 15 indicators, by which participants can rank their perception of ability to recover. To explore the possible combinations of indicator attributes, a program was developed to generate hypothetical individuals. The program took into account the data distribution and represented the attributes accordingly. The aim was to run the program to create unique hypothetical individuals that resembled a diverse range of typical Australians.

As the program randomly gathered the data for each indicator, some of the data combinations did not seem realistic or feasible. For example, a person with an income of \$2 per week and owning 3 cars while saving a lot, is not likely. Therefore, to ensure that the program created hypothetical individuals that were realistic and possible, more restrictions were required to define which characteristics were admissible. For example, the Henderson Poverty Line and CentreLink allowance thresholds [30] were constraints imposed on income. Table 2.1 provides further details of the data and constraints used in the program. While these were often assumptions on the part of the authors, they ensured that the hypothetical individuals created were realistic so that respondents to the questionnaire would not be perplexed or frustrated with highly unlikely situations.

2.3.3 Questionnaire design

The program generated 10,000 hypothetical individuals using data relating to the 15 selected indicators. For the purpose of distribution ease, each questionnaire contained ten hypothetical individuals numbered in each questionnaire. The questionnaire required

Form	Experts	Non-Experts	Total
Distributed:Personally	30	347	377
Distributed:E-mail	83	0	83
Received:Personally	12	63	75
Received:E-mail	39	0	39
Received:Mail	0	38	38

Table 2.2: Distribution of the questionnaire and associated returns

each individual to be ranked on a scale of 1 to 10 for each of the four hazards listed; earthquake, flood, landslide and cyclone. The scale used for ranking the individuals was from 1 to 10, with 1 indicating ‘little time’ to recover, and 10 indicating ‘infinite time’. The median rank, 5, indicated ‘sufficient time to interrupt life’ (Refer to the Appendices for a copy of the questionnaire). The questionnaire preamble deliberately avoids placing absolutes on some definitions and also on the ranking scale, so that people were able to bring their own perception and experience to the questionnaire, while avoiding prejudiced definitions and semantic debates, such as vulnerability or resilience.

The outcome of the hazard impact is presented to the reader, hence the type of hazard should not theoretically be of importance. For example, if a person is hospitalised with minor injuries and the residence is destroyed, the hazard in question should not be of relevance to a person’s ability to recover. However, some hazards are more frequent in Australia, some hazards receive greater government assistance and some hazards are represented differently in the media. People who have experienced a particular hazard and recovered quickly, such as a cyclone, may perceive cyclones to require less recovery time than an earthquake, which they may have not been exposed to, even if the resultant situation is the same. People who perceive the type of hazard as irrelevant with regards to the outcome, have the option to rank all hazards the same for each scenario.

2.3.4 Respondent demographics

The questionnaire was distributed to people living in large urban centres of Australia between November 15th 2002 and January 5th 2003. 110 questionnaires were completed and returned, which equates to 1100 ranked hypothetical individuals and an overall response rate of 33% (Table 2.2).

460 questionnaires (equating to 4600 hypothetical individuals) were developed and distributed via personal delivery, email or postal mail. The questionnaire was emailed or personally distributed to 113 ‘experts’ with 51 completed responses returned (45% return rate). 347 questionnaires were distributed to ‘non-experts’ via mail or personal distribution, with 101 completed responses being returned (29% return rate). 152 questionnaires were returned completed, 5 were returned without completion and explaining why. The overall response rate is 33%.

To further investigate the concept of perception, the questionnaire was distributed to ‘experts’ and ‘non-experts’. The term ‘experts’ and ‘non-experts’ are used to describe researchers/practitioners of disaster risk management and those who do not work in the area of disaster risk management respectively. There is often a greater em-

Experienced a hazard?	Expert	Non-Expert	Total
Yes	26 (50%)	26 (26%)	52 (32%)
No	20 (40%)	48 (46%)	68 (47%)
Not stated	5(10%)	28 (28%)	31 (21%)
Total	51	101	152

Table 2.3: Hazard Experience. Approximately one-third of respondents indicated they had experienced a natural hazard event.

Experienced hazard and ranked each the same?	Expert	Non-expert	Total
Yes	7 (29%)	9 (31%)	16
N/A	4 (17%)	7 (24%)	11
No and ranked their hazard higher	6 (25%)	4 (14%)	10
No and ranked their hazard lower	7 (29%)	9 (41%)	16
Total	24	29	53

Table 2.4: The Hazard experienced. For the respondents who indicated they had experienced a natural hazard, the greater number either ranked all hazards the same or perceived the hazard they experienced to require less recovery time than the hazards they have not experienced.

phasis on the opinion of ‘experts’ as opposed to ‘non-experts/lay people’ in numerous research fields, including disaster risk management [37]. However, when investigating vulnerability and how people will recover from an impact event, the perception of the general public must be considered alongside experts. Slovic argues that, when risk is involved, if people perceive something to be true or impending, then they will most likely behave in a manner that supports that perception [63]. Jasanoff argues that all perception of risk, whether lay or expert, represent partial views of the situations that threaten us, and hence any policy development should reflect more than just the judgements of the experts or the non-experts [37].

2.4 Questionnaire results

Initial results from the questionnaire highlight some interesting aspects of the perception of risk to natural hazards. Selected results are tabulated in Table 2.3 and Table 2.4.

There was a difference in how experts and non-experts rank each hazard for the same outcome. Approximately two-thirds of both experts (31%) and non-experts (35%) considered that the type of hazard does make a difference to a person’s ability to recover, even when the outcome of each hazard is the same. While it cannot be deduced from this study exactly why this perception exists, it may be due to the frequency of some hazards in Australia, the type of media coverage some hazards receive and the type of support people impacted by hazards receive from the government. For example, people perceive earthquakes as requiring more recovery time than landslides, even when injury, residence damage and personal attributes are the same. This may be due to the sense that earthquakes can happen ‘anywhere, anytime’ and that the media coverage of earthquakes, both domestically and international, depicts the hazard as fatal and causing damage. It may also be due to the fact that earthquakes generally have



Figure 2.13: Destroying over 400 homes, injuring dozens and causing 4 deaths, the Canberra bushfires of January 2003 affected the entire Canberra community. Data collection, as being conducted here by Geoscience Australia, focused on structural damage to homes.

a broader impact on services and infrastructure when compared with landslides, which are generally more localised. However, as the questionnaire only investigate impacts to individuals and their household and not communities or regions, the data cannot be used to support the theory that broad-scale hazards equal longer recovery, even if it appears a possibility.

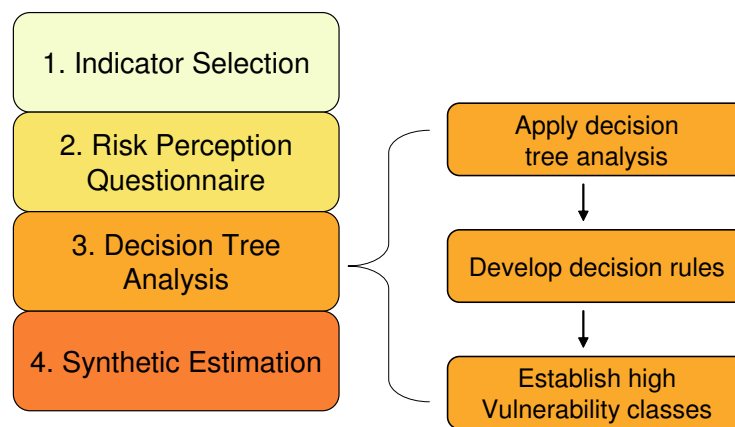
Table 2.3 shows the amount and percentage of respondents who have experienced a natural hazard. Half of the experts and only a quarter of the non-experts have experienced a natural hazard. Of those who had experienced a hazard, most have encountered a flood event (34%), while the next most popular was bushfire (27%), representing Australia's two most frequent hazards afflicting urban areas (Figure 2.4). 18% had experienced a cyclone, 8% have experienced an earthquake event and 4% experienced a landslide event.

The results shown in Table 2.4 provide some interesting points for discussion but are not based on a sufficient data to make any definitive conclusions. However, it is worth noting that respondents, both expert and non-expert, who have experienced a hazard have a tendency to rank the hazard they experienced as requiring less or the same recovery time as other hazards. On average, they did not perceive the hazard they experienced to render people more vulnerable than other hazards. This may suggest that people have a greater wariness or concern for hazards that they have not experienced.

Some of these points for discussion and issues raised will be explored in greater depth using decision tree analysis in the next chapter. The decision tree analysis will provide a clearer insight into the relationship between individual vulnerability indicators and the perceived risk of a person to a natural hazard impact.

Chapter 3

Decision Tree Analysis



This chapter examines decision tree analysis as a means of exploring one of the objectives of this study; to investigate the relative importance of personal and household attributes in contributing to vulnerability to a natural hazard. The focus of the application is to test the usefulness of applying a decision tree analysis to the data from the risk perception questionnaire. This chapter will introduce decision tree analysis and outline the process as relevant to this study, however, a more detailed account of decision tree methodology is presented in Appendix B.

3.1 Decision trees and social vulnerability

One statistical technique that can be used to explore large quantities of data for the purpose of classification and prediction is decision tree analysis. Decision tree analysis assists those involved in decision making by focusing on the essential data required to answer specific questions. Some of the more common applications of decision tree analysis have been in the insurance and banking industries where they have been used to predict the financial risks of customers. However, applications have been used in areas as diverse as medical diagnosis and landslide risk assessment [10, 32]. Application of decision tree analysis to studies of vulnerability is not new either, with a study exploring famine and chronic food insecurity in developing nations applying decision

Variable	Definition
Maize/Sheep	Retail price of maize/producer price of sheep terms of trade
Cereal	Gross production of all cereals in tons
Family Size	Average size of rural household
Health	Index of health infrastructure based on need
Market Travel	Distance to big market
Land	Average arable land owned
Females	Share of female heads in total number of household heads
Income	Average farm income per capita

Table 3.1: Vulnerability indicators used in a decision tree analysis of the *awraja* in Ethiopia, where *awraja* is an administrative district. [69]

tree analysis to predict scenarios [69]. Decision tree analysis has also been used in other studies of food insecurity conducted by the International Food Policy Research Institute [35], while the World Bank has also used decision tree analysis to explore issues of vulnerability to poverty in developing nations [5].

The study of famine and food insecurity, by Yohannes and Webb, demonstrates how decision tree analysis can be used to identify indicators that characterise different food access scenarios [69]. Using the decision tree analysis software, CART (Classification and Regression Trees), the authors of the study found that vulnerability ‘to food insecurity and famine cannot be measured by single discrete variables’ [69]. That is, vulnerability cannot be determined by a single indicator, but rather a combination of many indicators. Examples of the indicators are shown in Table 3.1. These indicators, like any study into vulnerability, are from a range of personal, geographical and financial attributes. By ‘bringing together a range of factors that reflect the behaviour and livelihood conditions of the people most affected’, the study purports that decision tree analysis can contribute to vulnerability assessments [69].

3.2 Perception based classification

One approach to knowledge discovery and the data mining process is to allow the user to interact with the data. The interaction between the user and the data is facilitated by the use of visualisation techniques developed specifically for large multidimensional data sets.

The perception based classification developed by Ankerst *et al.* [3] incorporates a pixel-oriented visualisation technique which maps each attribute value to a fixed colour map and displays attributes in different sub-windows. For example, in the circle segment technique, Figure 3.1, the whole data set is represented by a circle which is divided into d segments. Each attribute of the training data is represented in a separate segment of the circle. Within a segment, each attribute value (or class label) is visualized by a single coloured pixel as shown in part a of Figure 3.2. The colour of the pixel is established by mapping the training instances to attribute lists containing the attribute value and its class label, as shown in part b of Figure 3.2.

The circle segment technique has been applied to the survey data using all attributes and all class labels. The circle segments corresponding to the 15 attributes are shown in Figure 3.2. The rationale for the development of visualisation based classi-

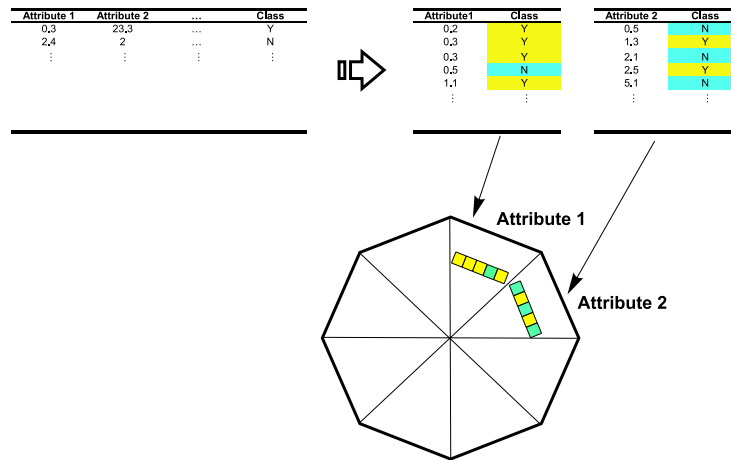


Figure 3.1: Eight attributes with binary class labels are visualized in the circle segment technique used in perception based classification.

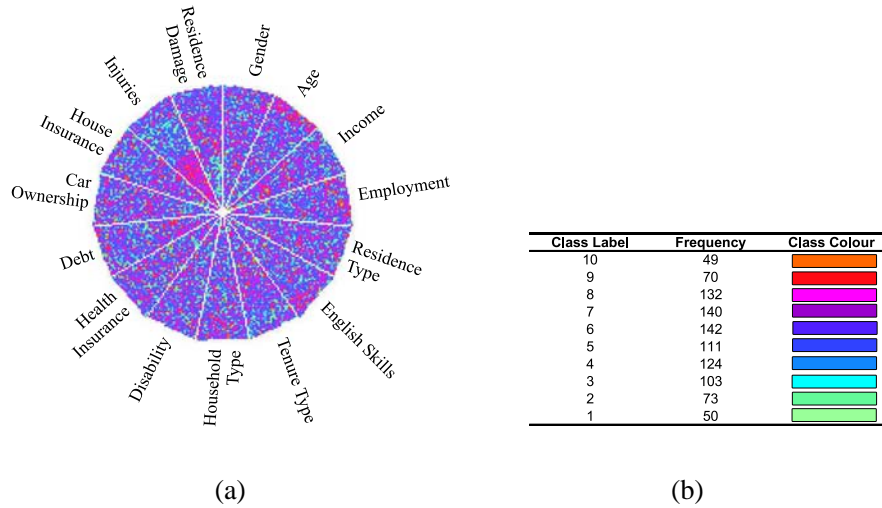


Figure 3.2: Ten colours responding to the ten attribute values (or class labels) of 'little time to recover' which is number one, to 'infinite time' which is number 10 (a), and the perception based classification representation of all the survey data as a circle (b).

fiers is that it gives the user the opportunity to interactively explore the data. However, for application to the risk perception questionnaire and other data sets, patterns in the data can be difficult to identify visually. One notable exception is the cluster of high vulnerability class labels associated with injuries. Under these circumstances automatic data mining techniques can be employed, where machine learning is performed with little human intervention, such as the CART decision tree analysis used in this chapter.

3.3 Decision tree analysis

Decision trees are a schematic way of representing alternative sequential decisions and the possible outcomes from these decisions. The analysis begins by placing all the data, in this case the results from the risk perception questionnaire, into a decision tree analysis program which, for this study is CART [9].

The program sorts through all the data, partitioning or splitting the data into subsets of increasing homogeneity. The point at which the data is split into subsets is determined by the attribute that leads to the split with the most homogenous subsets, as shown by the classes formed by splitting data, by orthogonal splitting, in part *a* of Figure 3.3. The process of splitting continues until all data has been processed to form a tree that has classified all the data, as shown in part *b* of Figure 3.3. Decision trees also provide a ranking of how important individual attributes are for predicting possible outcomes. The tree, as shown in Figure 3.4, has a structure that includes the root node, branch, decision and leaf nodes. Once built, a decision tree can be used to classify data by starting at the root node of the tree and moving through it until a terminal node is encountered. Each terminal node provides a decision rule, or outcome, that allows us to make a predictive statement about the data. A more detailed description of decision tree analysis is provided in Appendix B.

The maximal tree has a structure that includes the root node, branch, decision and leaf nodes and looks similar to that shown in Figure 3.4. Once built, a decision tree can be used to classify data by starting at the root node of the tree and moving through it until a terminal node is encountered. The class assigned to each terminal node then corresponds to a predicted outcome for the data in question.

3.4 Analysis of the risk perception questionnaire

The results of the risk perception questionnaire were applied to the CART decision tree analysis [9]. To ensure consistency was maintained from one respondent to another, the results of only one hazard scenario were selected. While it was discussed in the previous chapter that there were some differences between the results for the four different hazard scenarios, it was also highlighted that the differences, considering the small sample size, were relatively minor. Flood was chosen as a representative as the average vulnerability rank to flood (5.5) also correlates to the average rank for all hazards (5.53). It is the only hazard where by both experts and non-experts had the same average rank. These reasons are considered sufficient to support the use of the results of vulnerability to flood for the initial attempt at decision tree analysis.

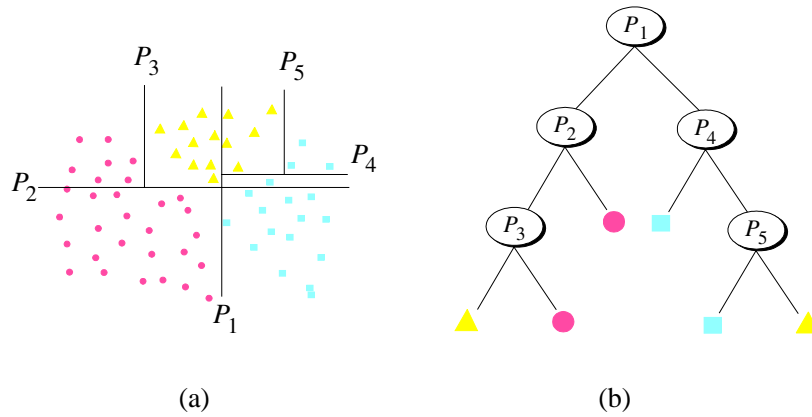


Figure 3.3: (a) Splitting of a two dimensional attribute data set with three classes using orthogonal and (b) the decision tree produced using orthogonal splitting of the two attributes oblique splits.

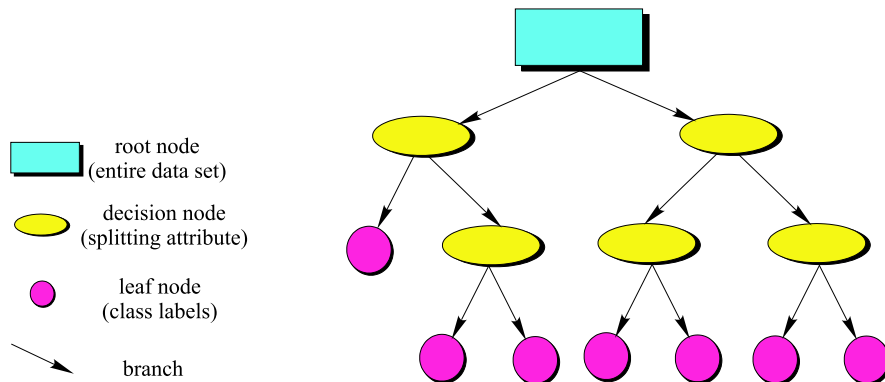


Figure 3.4: The four components that are associated with a decision tree; the root node, decision node, leaf node and the branch.

3.4.1 Assigning classes of vulnerability

The risk perception questionnaire provides 10 options for ranking a particular scenario, effectively creating 10 classes of vulnerability. Ten classes were selected in order to capture a wide variation in response to the hypothetical scenarios and to test whether the decision tree could pick up subtle differences. Different respondents may have been likely to rate similar instances slightly differently, making a precise distinction among ranks nearly impossible. For example, is there much difference between 4 and 5, or 3 and 4? One way of overcoming this problem is to organise a redistribution of the ten classes into new classes of vulnerability, which are more likely to be predictable.

The original ten classes were therefore redistributed in a systematic manner and the difference between the various vulnerability classes was investigated. Some of the redistributed groups lead to more accurate classes than others. For example, keeping all ten classes of vulnerability for the decision tree analysis results in a range of accuracies, from 43% to 79%. Hence, any decision rules developed from a completed decision tree analysis will have a fairly low predictability rate for any new data. The aim, therefore, is to get the highest predictability rate without compromising the data or class accuracy.

The other redistributions tried were for 5, 3 or 2 classes of vulnerability. Individual classes towards the middle of the spectrum are, in general, the least predictable while extreme classes and aggregated classes can be predicted with reasonable confidence. The best redistribution of the ten classes that assured a fair probability and accuracy was two classes, where levels 1–5 were mapped to *Low* and 6–10 were mapped to *High* perceived vulnerability. While two vulnerability classes are a significant shift from the original ten classes, the move towards better predictability of vulnerability addresses one of the aims of the study.

3.4.2 Developing decision rules

The decision tree was generated in CART using the GINI splitting rule and testing was done by 10-fold cross validation (as discussed in page 79). With these settings and the class aggregation described above, CART produced a tree with 21 internal and 22 terminal nodes.

The 22 terminal nodes generated by the decision tree corresponds to 22 decision rules of which 11 are rules for the affirmative case; that is, the person in question has high perceived vulnerability to a major flood. We only discuss the affirmative rules as we are investigating the socio-economic characteristics that contribute to vulnerability to natural hazards. An example of part of the decision tree highlighting rules 4 and 5 is shown in Figure 3.5. Instances where the no affirmative rule is found are necessarily classified as low vulnerability. A person is thus considered vulnerable if one of the affirmative rules stated in Table 3.2 hold true. The same (affirmative) rules are stated algorithmically in Figure 3.6. It is a matter of personal preference which representation is regarded the most comprehensible. The individual affirmative rules are listed below with commentary on why they have been formed and whether they can be considered reasonable:

Rule	Expression
1	Injuries > 3.5
2	$2.5 < \mathbf{Injuries} \leq 3.5$ & Age > 60.5
3	$2.5 < \mathbf{Injuries} \leq 3.5$ & $46.5 < \mathbf{Age} \leq 60.5$ & Residence Damage > 2.5
4	Household Type $\in \{1, 3, 5\}$ & $2.5 < \mathbf{Injuries} \leq 3.5$ & Age ≤ 46.5 & Residence Damage > 3.5
5	House Insurance = 0 & Household Type $\in \{2, 4\}$ & $2.5 < \mathbf{Injuries} \leq 3.5$ & Age ≤ 46.5 & Residence Damage > 3.5
6	Injuries ≤ 2.5 & Residence Damage > 3.5 & Age > 66.5
7	Injuries ≤ 2.5 & Residence Damage > 4.5 & $51.5 < \mathbf{Age} \leq 66.5$ & Income ≤ 216.5
8	Injuries ≤ 2.5 & Residence Damage > 4.5 & $43.5 < \mathbf{Age} \leq 51.5$
9	Tenure Type $\in \{1, 4\}$ & House Insurance = 0 & Injuries ≤ 2.5 & Residence Damage > 3.5 & Age ≤ 43.5
10	Tenure Type $\in \{2, 3\}$ & Injuries ≤ 2.5 & Residence Damage ≤ 3.5 & Age > 64.5
11	Household Type $\in \{4, 5\}$ & Tenure Type $\in \{1, 4\}$ & Injuries ≤ 2.5 & Residence Damage ≤ 3.5 & Age > 73.5

Table 3.2: Affirmative decision rules. A person is perceived to have high vulnerability if one of these logical expressions is true.

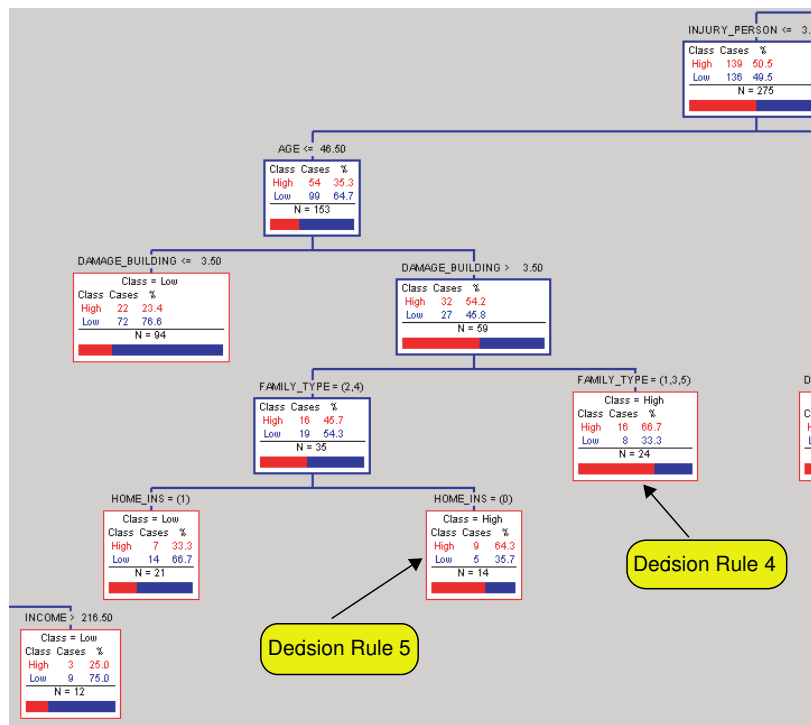


Figure 3.5: A snapshot of the decision tree developed for this study, showing rules 4 and 5.

Rule 1

- Life-threatening injury

Interpretation: A life threatening injury will take precedence over other personal attributes.

Rule 2

- Aged 60 years and over
- An injury that requires hospitalisation but expected to recover

Interpretation: Older people will take longer to recover from mild to serious injuries.

Rule 3

- Aged between 47 and 60 years
- An injury that requires hospitalisation but is expected to recover
- Residence extensively damaged or completely destroyed

Interpretation: Time off work, due to injury, could place a financial strain on those approaching retirement.

Rule 4

- An injury that requires hospitalisation but is expected to recover
- Aged 46 or less
- Residence completely destroyed
- Either a single parent or lives alone

Interpretation: Losing your home without a strong household network places you without immediate support.

Rule 5

- An injury that requires hospitalisation but is expected to recover
- Aged 46 or less
- Residence completely destroyed
- Either lives as a couple or in a group household
- No house insurance.

Interpretation: Being responsible for the wellbeing and shelter of others in the home and losing your house without insurance could create an enormous financial burden.

Rule 6

- Aged 67 years or more
- Either has an injury that requires basic medical treatment or no injury at all
- Residence extensively damaged or completely destroyed

Interpretation: People in this age demographic are generally retired and often have all their finances invested in the house they own. Hence any major damage to the home could put a strain on their finances as they are less likely to have an ongoing source of income.

Rule 7

- Aged between 52 and 66
- Either has an injury that requires basic medical treatment or no injury at all
- Residence completely destroyed
- Has a very low income

Interpretation: Losing a home when you are nearing retirement on a very low income would be a large financial burden, particularly if you also sustained an injury that affected your capacity to work.

Rule 8

- Aged between 44 and 51 years
- Either has an injury that requires basic medical treatment or no injury at all
- Residence completely destroyed

Interpretation: People of this age often have families and are in the middle to end of their working life, therefore total destruction of their home, which many people are starting to own in this age bracket, could be a financial strain. But it must be noted that, this may be an example of the sensitivity of decision trees to small training sets as mentioned on page 76. In this case, Rule 7 and 8 would collapse into a rule stating that vulnerability is considered high for individuals with moderate injuries (≤ 2.5), with the highest possible building damage and aged between 44 and 66 irrespective of their income.

Rule 9

- Aged 43 or less
- Residence extensively damaged or completely destroyed
- No house insurance
- Either owns their house or rents from the government

Interpretation: Government housing lists are long, and so waiting could take some time. For people who owned their house, extensive damage/destruction without insurance would be a large financial burden.

Rule 10

- Aged 65 years or more
- Either has an injury that requires basic medical treatment or no injury at all
- Residence either not damaged, suffered minor or moderate damage
- Either have a mortgage or rent privately

Interpretation: The perception here is having immediate financial commitments (mortgage or rent) is a factor making people vulnerable in case of a disaster. Note that the role of tenure type in Rule 9 has been reversed in this case.

Rule 11

- Aged 74 and over
- Either has an injury that requires basic medical treatment or no injury at all
- Residence slightly or moderately damaged, or no damage at all

- Either owns their house or rents from the government
- Either lives alone or in a group house

Interpretation: This decision rule has the greatest combination of indicator attributes and relates to people 74 and over. This suggests that older people have a broader range of attribute combinations and may therefore be more vulnerable to natural hazard impacts.

3.4.3 The relative importance of the 15 indicators

The relative importance of indicators in this study is listed in Table 3.3. The indicator 'Injuries' is perceived as the most important discriminator for vulnerability. Injury will affect a person's ability to recover, regardless of their personal attributes such as income or age. However, this indicator is closely connected to the context of the hazard scenario rather than the context of the individual's characteristics, and is therefore considered a hazard indicator, as mentioned in Chapter Two. It is to be expected that the hazard indicator 'Injuries' is considered the most important indicator, as it directly relates to survival.

The second most important indicator is 'Residence Damage' with a relative importance score of less than 25% of Injuries. Again, this indicator is also tied to the hazard scenario rather than the individual and, as expected, perceived as an important factor in the ability to recover from a natural disaster. The fact that the decision tree analysis found that both the hazard indicators, Injury and Residence Damage, are the most important discriminators for vulnerability to natural hazards, demonstrates that the analysis is intuitively reasonable. It also highlights the difference, outlined in Chapter Two, between hazard and vulnerability indicators and the role they play in a natural hazard impact.

The indicators House Insurance, Income and Tenure Type are the next most important indicators after Injuries and Residence Damage and therefore the most important socio-economic vulnerability indicators. All three have relative importance scores just above 10%. House insurance and Income are reasonably straightforward in the way they affect the perceived vulnerability as discussed in the outline of each decision rule. Tenure Type, however, is slightly different as it affects vulnerability in different ways depending on the other indicators. For example, Rule 9 suggests that owning a house or living in government housing contributes to vulnerability, but only when the individual is younger than 43, doesn't have house insurance and the residence suffers extreme or complete damage. However, Rule 10 states that the opposite tenure types - owning with a mortgage and renting - contribute to vulnerability but only when an individual is over 65, has a minor injury and the residence suffered minor, moderate or no damage. These two rules and their use of Tenure Type demonstrates that no one indicator variable, such as renting or owning a house, is always indicative of vulnerability.

Next is Age, with a relative score of just over 2%, while the remaining indicators have scores below 1% and only Debt, Employment, Car Ownership, and English Skills have relative scores above 0.01%. The lowest scoring indicators are Household Type, Health Insurance, Residence Type, Disability and Gender. Household Type, Disability and Gender have traditionally been regarded as important factors in assessing the vulnerability of individuals, yet are not considered of high relative importance by the

Indicator	Score
Injuries	100.00
Residence Damage	24.23
House Insurance	12.36
Income	11.59
Tenure Type	9.61
Age	2.10
Debt	0.06
Employment	0.05
Car Ownership	0.02
English Skills	0.01
Household Type	0.00
Health Insurance	0.00
Residence Type	0.00
Disability	0.00
Gender	0.00

Table 3.3: Relative indicator importance according to the decision tree analysis.

analysis [39, 29, 14]. However, Disability was only present in 10% of the survey cases and may therefore not have been sufficiently represented to make a difference. This should be noted as a point for further investigation. The indicator Health Insurance may have scored low due to the comprehensive public health system present in Australia through the government funded medicare and pharmaceutical benefits schemes. Respondents may have regarded private health insurance as unimportant in the case of a major personal injuries, as they would be treated regardless of their insurance status.

3.5 Decision rule findings

The risk perception questionnaire provided results that addressed some significant issues discussed in Chapters One and Two. These include:

- The decision tree rules demonstrated that no one personal attribute in isolation is responsible for vulnerability. Complete destruction of a residence was shown to make everyone vulnerable, regardless of other attributes, however, as noted, this is a hazard indicator rather than a vulnerability indicator.
- While it appears that, regardless of which hazard impacts a person, if the result is the same then the recovery period from the hazard should be the same. However, as residence damage and home insurance were the most significant attributes according to the decision tree, then an earthquake, which usually causes more structural damage than a flood, may take longer to recover from.
- Income, house insurance and age were three of the highest four indicators of relative importance after injury and residence damage. The other was tenure type. House insurance and income were more significant than age (by a factor of 6). House insurance, income and tenure type all relate to a person's financial situation, highlighting the significance of finances in the perception of vulnerability to natural hazards.

- Experts did rank ability to recover as lower, or quicker, than non-experts, albeit a relatively minor difference.

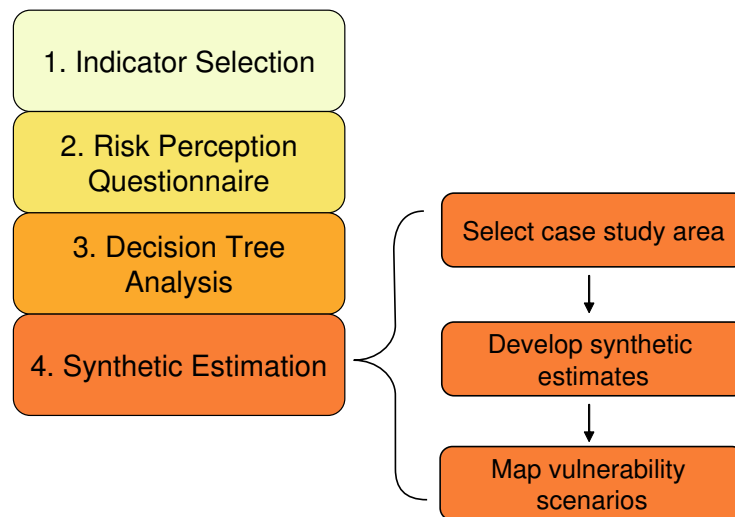
The decision rules provide an interesting insight into how we can view the vulnerability of an individual within a household to natural hazards. The 11 high vulnerability decision rules demonstrate the importance of investigating a range of personal and household characteristics when analysing vulnerability to natural hazards. The decision rules also demonstrate that using a knowledge discovery technique, such as decision tree analysis, is a feasible means of exploring complex social data. However, the methodology outlined thus far requires an application to further explore its value. The next chapter introduces a case study area in Western Australia for application of the decision rules.

Vulnerability = 'Low'	<i># Default if no affirmative rule is found</i>
if Injuries > 2.5:	
if Injuries > 3.5:	
Vulnerable = 'High'	<i># Rule 1</i>
else:	
if Age > 46.5:	
if Age > 60.5:	
Vulnerability = 'High'	<i># Rule 2</i>
else:	
if Residence Damage > 2.5:	
Vulnerability = 'High'	<i># Rule 3</i>
else:	
if Residence Damage > 3.5:	
if Household Type in [1,3,5]:	
Vulnerability = 'High'	<i># Rule 4</i>
else:	
if House Insurance == 0:	
Vulnerability = 'High'	<i># Rule 5</i>
else:	
if Residence Damage > 3.5:	
if Age > 43.5:	
if Age > 66.5:	
Vulnerability = 'High'	<i># Rule 6</i>
else:	
if Residence Damage > 4.5:	
if Age > 51.5:	
if Income ≤ 216.5:	
Vulnerability = 'High'	<i># Rule 7</i>
else:	
Vulnerability = 'High'	<i># Rule 8</i>
else:	
if House Insurance == 0:	
if Tenure Type in [1,4]:	
Vulnerability = 'High'	<i># Rule 9</i>
else:	
if Age > 48.5:	
if Age > 64.5:	
if Tenure Type in [2,3]:	
Vulnerability = 'High'	<i># Rule 10</i>
else:	
if Age > 73.5:	
if Household Type in [4,5]:	
Vulnerability = 'High'	<i># Rule 11</i>

Figure 3.6: Affirmative decision rules - algorithmic form. A person's perceived vulnerability is determined by the value of High vulnerability or Low vulnerability.

Chapter 4

Synthetic Estimation



The fourth and final step in the methodology provides the opportunity to apply the decision tree rules to a spatial area in order to analyse the vulnerability of individuals within a household in a real community. The application of the decision rules to a study area is presented as maps that identify the areas with a lower or higher proportion of households with highly vulnerable individuals. Maps are a vital part of emergency management and town planning and are also useful means of presenting spatial information.

This chapter will introduce the synthetic estimation techniques used to develop the population estimates of highly vulnerable individuals within a study area. The area selected for this study is within Perth, Western Australia, and the final part of the chapter will present and discuss the vulnerability maps that have been developed with the synthetic estimation results. For this study, an area of 224 Census Districts (approximately 220 households per Census District) in Perth, the capital of Western Australia, has been chosen. Perth contains approximately 71% of Western Australia's population, or about 1.3 million people [56]. Perth is situated beside the flood prone Swan River and approximately 150km south-west of Australia's most active earthquake zone. There are



Figure 4.1: Registering 6.9 on the Richter Scale, the Meckering earthquake of 1968 destroyed the small town located 130km east of Perth. Structural damage from the event can still be seen today in buildings in the Perth CBD.

also several limestone belts to the north and south of Perth where hazardous karst limestone systems are located, while the city's coastline suffers from coastal erosion as a result of high winds and storms.

4.1 Perth Cities Project

Perth is the location currently being researched for Geoscience Australia's multi-hazard risk assessment study referred to as the Perth Cities Project. Beginning in 2001, the Perth Cities Project¹ covers an area represented by 46 Local Government Areas (LGA), from Gingin in the north, south to Busselton and east to Northam. As the entire Perth area is too large for this vulnerability study, a smaller area has been chosen and is represented by a collection of 224 Census Districts, involving the Wanneroo and Swan local governments and is shown in Figure 4.2.

4.1.1 Study Area

The study area selected for this case study is within the Perth Cities Project area shown in Figure 4.2. The collection of Census Districts within the study area also aggregate to Statistical Local Areas (SLA), so that benchmark or comparison data are readily

¹The study has required the development of collaborations with State and Local Government and private companies in order to exchange data and align risk management needs. Due to these strong collaborations, data availability and spatial information, the Perth Cities project provides an ideal case study opportunity for the application of the social vulnerability methodology

available at the SLA level for use in any future studies. The agreed final selection, shown in Figure 4.2, totals 224 Census Districts and include all 153 Census Districts in the South-West Wanneroo SLA, all 44 Census Districts in the South-East Wanneroo SLA and 27 Census Districts in the Swan SLA.

4.2 Synthetic estimation

The aims of the case study application are to model the distribution of vulnerability in 224 Census Districts in Perth and to assess the role of synthetic estimation in a methodology aiming to quantify vulnerability to natural hazards. The techniques used to develop synthetic estimates in this case study were developed by NATSEM at the University of Canberra and are presented in greater detail in a paper presented by Day and Dwyer [21].

The main problem encountered in studying the personal attributes of people within a community is obtaining access to detailed data on the 15 indicators used in the 11 decision rules that have been used in this study. The ABS do not release such highly detailed data due to privacy laws. For example, we are permitted to access how many people in one Census District (a Census District is approximately 200 households in size) are over 55, how many people live alone and how many people are low income earners. However, we are not permitted to have access to how many people are over 55 and live alone and earn a low income, which is referred to as cross-tabulated data.

Synthetic estimation was used to develop estimates of the populations attributes for the 15 indicators. The method uses Census data for real spatial areas in order to create best fit data, which are referred to as synthetic estimates. This best fit creates the most accurate representation in the absence of access to actual data. This technique of synthetic estimation used in creating estimates of the vulnerable individuals in households for each of the Census Districts in the study area can be summarised into four steps:

- **1. Extract relevant HES household, person and expenditure data.** The majority of the variables are household reference person and household variables in the household level data. In addition, disability status of the household reference person is extracted from the person level data and insurance variables are extracted from the expenditure data.
- **2. Extract weights for sample Census Districts.** MarketInfo The data used to develop the hypothetical individuals relied on a data sources, predominantly the Census, which were not location specific. However, in this case study, we need to tie the data to specific Census Districts. To ensure the characteristics of individuals within households in each of the 224 Census Districts is the most accurate, a weighting method is needed. In this study, MarketInfo was used, which has developed specific weights linking data with location. Weights for the 224 Census Districts in the study area were extracted.
- **3. Generate synthetic data.** The extracted data in each HES record were classified and weighted in order to determine the number of people in each Census District with each characteristic.

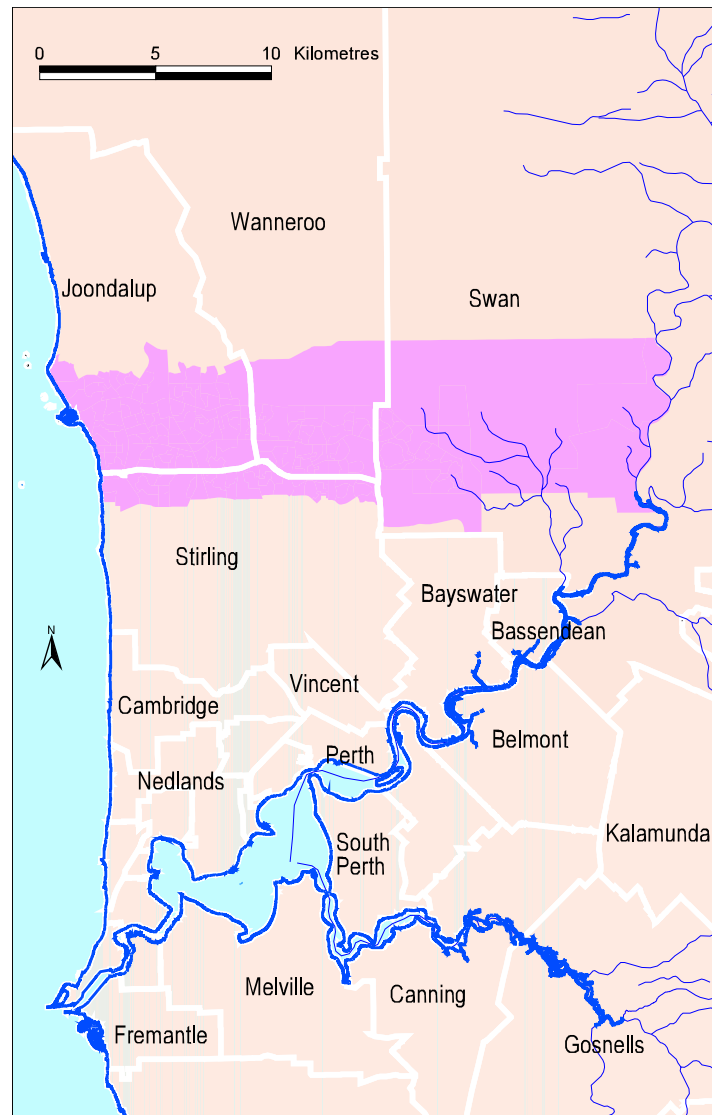


Figure 4.2: The aggregated Census Districts across 3 Statistical Local Areas used in the study. The case study area includes southern Wanneroo and Swan Local Governemnts and is part of the larger Geoscience Australia Perth Cities Project study area.

- **4. Synthetic socio-economic data** The resulting synthetic data for each Census District has over 70 fields (i.e. all data variables for each of the 15 selected indicators). Each field name indicates the variable and each field in the synthetic data contains the number of households or reference people in that variable in that Census District.

4.3 Synthetic estimates for the Perth Case study

The risk perception questionnaire describes an individual in terms of particular personal characteristics, but there is no information on the characteristics of other members of the household (apart from whether a partner or dependents exist). Hence the unit of analysis is persons, but not all persons in the household. Therefore the synthetic data must replicate these persons and those characteristics in order for the decision rules to be appropriately applied to determine the small area geographic distribution of vulnerability.

In this application, it was decided to use the HES Household Reference person (a person chosen according to certain criteria articulated by the ABS) as a surrogate for the randomly generated hypothetical individual. The household reference person is defined in the ABS HES technical paper [54] as one of the usual residents aged 15 years and over that meet one or more of the following criteria:

- partner in a registered or de facto marriage
- lone parent with dependent child(ren)
- person with the highest income
- the eldest person.

Table 4.1 shows the ratio of males to females for each family composition class. From this, we can deduce that the characteristics of household reference people are unlikely to match those of a randomly generated person. For example, the household reference person is male in over 70% of cases, while one parent families are lead by a woman in nearly 90% of cases. The sex distribution of household reference people is less skewed in group and lone person households. This choice of unit of analysis means that the synthetic results are for the number of households or household reference people in each Census District as a surrogate for the number of hypothetical individuals in each Census District.

4.3.1 Matching questionnaire variables

Often HES data has more detail than the questionnaire's classes of interest, so HES classes were merged in order to match the classes. HES surrogates for English skills, disability and debt and savings were agreed as there were no exact matches. Table 4.2 shows the variables selected from the 1998/99 HES which best match the vulnerability indicators. The third column in this table shows the classes used for each vulnerability indicator.

Sex	Couple only	Couple with child(ren)	Group	One parent	Alone
Male	73	72	56	11	45
Female	27	28	44	89	55
Total	100	100	100	100	100

Table 4.1: Percentages of family composition by sex of household reference person. (Source: ABS Household Expenditure Survey 1998/1999)

Vulnerability Indicator	HES characteristic variable	HES classes used
Age	Age of the household reference person	15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75+
Car	Number of registered cars and motorcycles in the household	0, 1, 2, 3+
Debt and Savings	Main source of household emergency money	Savings, Financial Loan, Family Loan, Other, None
Disability	Severity of the restriction of the person	None, Restriction
Employment	Labour force status and status in employment of the household reference person	Employed, Not Employed
English Skills	Year of arrival of the household reference person in Australia	Australia born, Recent(1996-1999), Prior to 1995
Gender	Sex of the household reference person	Male, Female
Health Insurance	Hospital, medical, dental insurance	Yes, No
House Insurance	House Insurance, Contents Insurance, House and contents	None, Contents, House, Both
Household Type	Household Family Composition	Lone, Couple with kids, couple no kids, single parent, group households
Income	Total gross weekly income of the household reference person	Q1: \$0 to \$221, Q2: \$221 to \$358, Q3: \$358 to \$564, Q4: \$564 to \$842, Q5: \$842 to \$4,087
Residence Type	Dwelling Structure	House, 2Storey, 3Storey, 4Storey, Flat
Tenure	Nature of Housing Occupancy	Own, Mortgage, Government rental, Private rental

Table 4.2: Vulnerability indicators paired with 1998/99 HES variables. (Source: ABS Household Expenditure Survey 1998/1999)

Classes	Residence Damage Description	Injury Description
1	not damaged	no injuries
2	slightly damaged	requires basic medical treatment without hospitalisation
3	moderately damaged	requires hospitalisation and is expected to recover
4	extensively damaged	requires hospitalisation with life threatening injuries
5	completely destroyed	

Table 4.3: Injury and residence damage scales, modified from HAZUS [4].

Injury/Damage	Damage1	Damage2	Damage3	Damage4	Damage5
Injury1	Scenario1	Scenario1	Scenario1	Scenario2	Scenario3
Injury2	Scenario1	Scenario1	Scenario1	Scenario2	Scenario3
Injury3	Scenario4	Scenario4	Scenario5	Scenario6	Scenario6
Injury4	Scenario7	Scenario7	Scenario7	Scenario7	Scenario7

Table 4.4: Of the 20 possible scenarios based on the 5 Damage and 4 Injury Classifications, 7 actual scenarios were identified by the decision tree rules. Refer to Table 4.3 for an explanation of the injury and damage classifications.

4.4 Social vulnerability in the Case Study area

The decision tree analysis of the questionnaire data revealed that Injury and Residence Damage were the most important indicators in determining an individual's ability to recover from a natural hazard. Consequently, no single spatial assessment of vulnerability based on socio-economic characteristics alone is possible. Rather, one based on different injury and damage scenarios is necessary. Therefore, it is essential to have a hazard context when assessing the relationship that socio-economic indicators have with each other and the hazard when assessing vulnerability to natural hazards. Table 4.3 shows the different injury and damage scales in Chapter 2.

If the five classes of damage were combined to explore every possible scenario, there would be 20 logical combinations, as shown by the matrix in Table 4.4. However, the decision rules produced by the decision tree have identified only seven distinct scenarios. That is, when the data is put into the decision tree, some of the possible scenarios are not differentiated. For example, when the injury level is 2, it does not matter if residence damage is 1, 2 or 3 as the decision rules show no differentiation. Table 4.5 shows the scenarios and their associated injury and damage scores, as well as the nodes in the decision tree that are populated in the scenario, and the socio-economic variables used in the rules.

A particular scenario and its associated injury and damage scores are assumed to affect the whole of every Census District in the study area whereas in an actual event, different buildings and people will be affected to differing degrees. This issue will be better resolved when sufficient hazard models are developed and can be overlain on the vulnerability maps. Unlike the injury and damage states, the five socio-economic characteristics used in the decision tree rules (house insurance, income, tenure type,

Scenario	Injury	Damage	Mean of individuals within a household	Relevant decision rules	Relevant indicator variables
1	1 or 2	1, 2 or 3	8	10, 11	age, tenure, family type
2	1 or 2	4	31	6, 9	age, insurance, tenure
3	1 or 2	5	79	6-9	age, insurance, tenure, income
4	3	1 or 2	38	2	age
5	3	3	123	2, 3	age
6	3	4 or 5	187	2-5	age, family type, insurance
7	4	1, 2, 3, 4 or 5	247	1	none (based on injury scale)

Table 4.5: Scenarios with corresponding decision rules and indicator variables. The mean of the household reference people in the high vulnerability class in each Census District is also listed.

age and household type) vary across Census Districts. In most cases, injury, damage and one demographic characteristic (age) determine the node. At most, three socioeconomic characteristics are required to distinguish between nodes, for example node 9 in scenario 3 (injury 1 or 2 and damage 5), age, home insurance and tenure are required.

4.4.1 Estimated social vulnerability of the study area

The resulting synthetic vulnerability data contains, for each Census District, the number and percent of household reference persons in the high vulnerability class for each of the seven identified scenarios. For each level of injury, the average number of household reference people in each Census District that are affected increases as residence damage increases, regardless of any other indicators 4.5. For example, for injury level 2, which corresponds to Scenario 1, 2 and 3, the average percentage of people deemed highly vulnerable increases from 3% (when damage is 1, 2 or 3) to 32% (when damage is 5). The mean count of people deemed highly vulnerable increases from 8 (when damage is 1, 2 or 3) to 79 (when damage is 5).

For injury level 3, which corresponds to Scenario 4, 5 and 6 the mean percentage of people deemed highly vulnerable increases from 16% (when damage is 1 or 2), 50% (damage state 3) to 76% (when damage is 4 or 5). The average count of people deemed highly vulnerable increases from 38 (when damage is 1 or 2) to 187 (when damage is 4) to 247 (damage state 5).

In addition, the above table shows that in most scenarios for a particular injury level, the range of the data (that is, the difference between the minimum and maximum values) increases as the damage increases. For example, for injury level 2 the range in percentage of people deemed highly vulnerable increases from 15% (when damage is 1, 2 or 3) to 65% (when damage is 5). The range in count of people deemed highly vulnerable increases from 62 (when damage is 1, 2 or 3) to 424 (when damage is 5).

For injury level 3 the range in percentage of people deemed highly vulnerable

increases from 49% (when damage is 1 or 2) to 62% (when damage is 4), then drops back down to 39% (when damage is 5). The range of people deemed highly vulnerable increases from 316 (when damage is 1 or 2) to 718 (when damage is 4) and then to 1052 (when damage is 5).

4.5 Mapping social vulnerability using hazard scenarios

Maps of the seven scenarios have been developed for the Perth study area and are shown from Figure 4.3 to Figure 4.9. The maps show the percentage of each Census District that have households in the high vulnerability class. Maps with percentages of vulnerable households have been used to demonstrate the significance of the social indicators in determining social vulnerability.

Each of the seven scenario maps has fourteen selected Census Districts that have been numbered from 1-14. While each of the seven maps allows us to compare Census Districts within a given scenario, the numbered Census Districts provides the opportunity to compare Census Districts between the scenarios. This comparison, which demonstrates that the thirteen selected indicators do contribute to recovery capacity of individual within a household, will be discussed in greater detail in the next section.

4.5.1 Mapped scenarios

The seven mapped scenarios have been produced with the same distribution so that individual Census Districts can be compared both within a single scenario and across all seven scenarios. That is, each of the 224 Census Districts are coloured according to the percentage of households in the high vulnerability group, with every 10% increment assigned a different colour. A cursory glance at the seven different scenarios demonstrates that, *regardless* of the state of personal injury and residence after an impact event, there are some socio-economic characteristics that do contribute to perceived ability to recover. The only exception is Scenario 7 (fig:Natsem20), which is the highest level of injury (life threatening). In this scenario, it does not matter what level of damage a residence receives or what personal characteristics are, as every Census District has 100% high vulnerability, as shown in Figure 4.9. This is reasonable, for it can be presumed that for a person with life threatening injuries, vulnerability will be high regardless of other socio-economic variables.

Scenario 4, which is based on an injury that requires hospitalisation but expected to recover and none to slight residence damage, shows an average of between 0% to 20% highly vulnerable people within the census districts. When compared with Scenario 5, Figure 4.7, which has the same level of injury but the next level of residence damage (moderate instead of slight), the percentage of highly vulnerable increases to an average between 40% to 70%. This increase in the percentage of people perceived to be highly vulnerable indicates that damage to residence is an extremely important factor in the perception of ability to recover.

Scenario 6, Figure 4.8, which also has the same level of injury as Scenarios 4 and 5 but much higher residence damage (from extensive damage to complete destruction), has an average of 70% to 90% highly vulnerable people. Again, this may be due to the importance of house insurance in the decision tree analysis, however it is also due to

Scenario 1

Injury 1 or 2, Residence Damage 1, 2 or 3

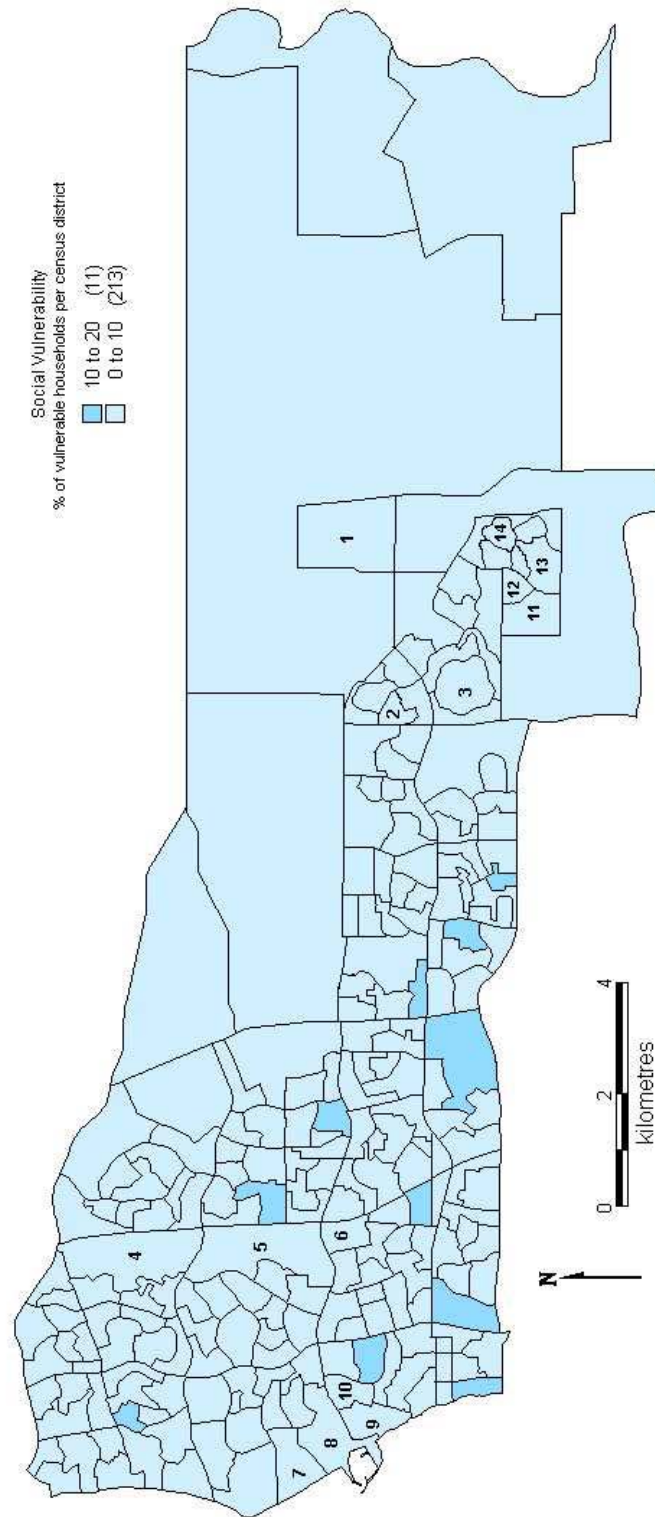


Figure 4.3: Scenario 1. This scenario visualises the estimated high vulnerability Census District areas when both injury and residence damage are relatively minor.

Scenario 2

Injury 1 or 2, Residence Damage 4

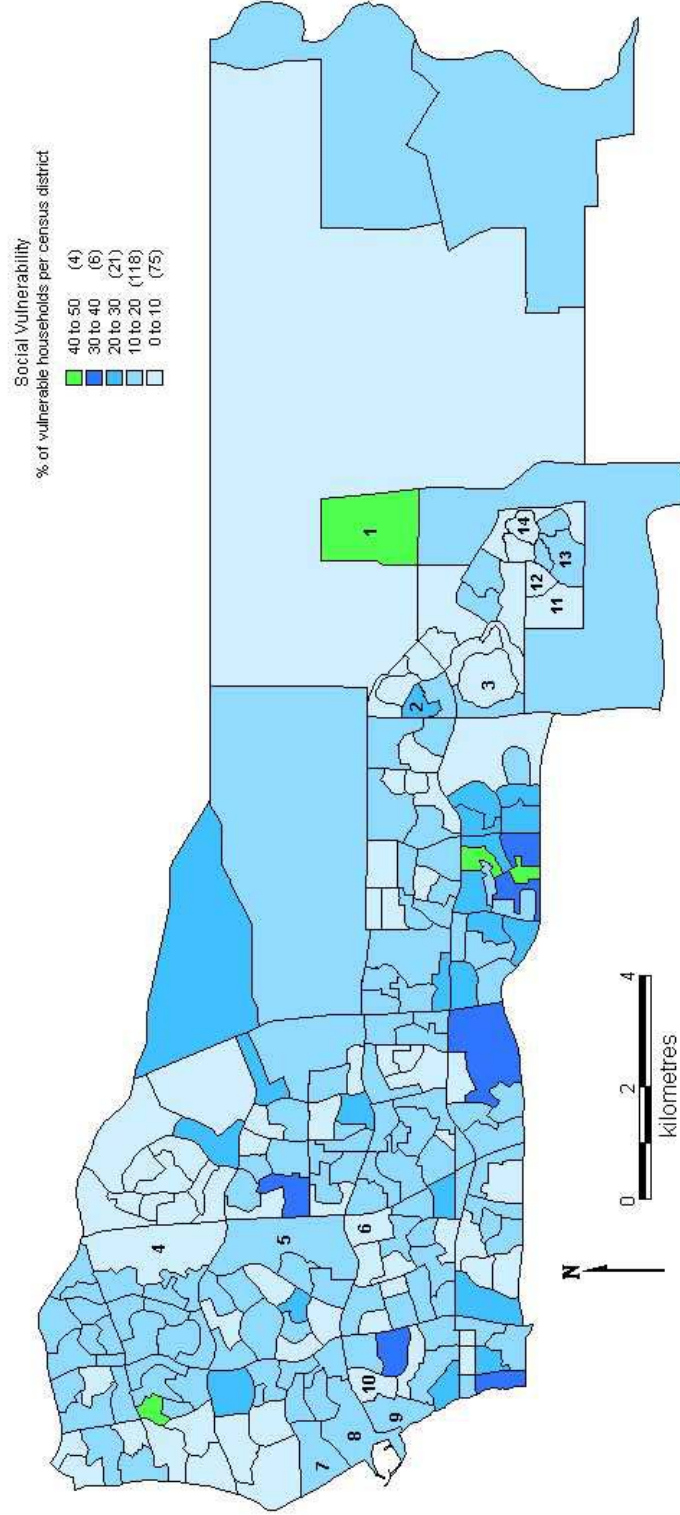


Figure 4.4: Scenario 2. This scenario visualises the estimated high vulnerability Census District areas when injury is relatively minor and residence damage is significant.

Scenario 3 Injury 1 or 2, Residence Damage 5

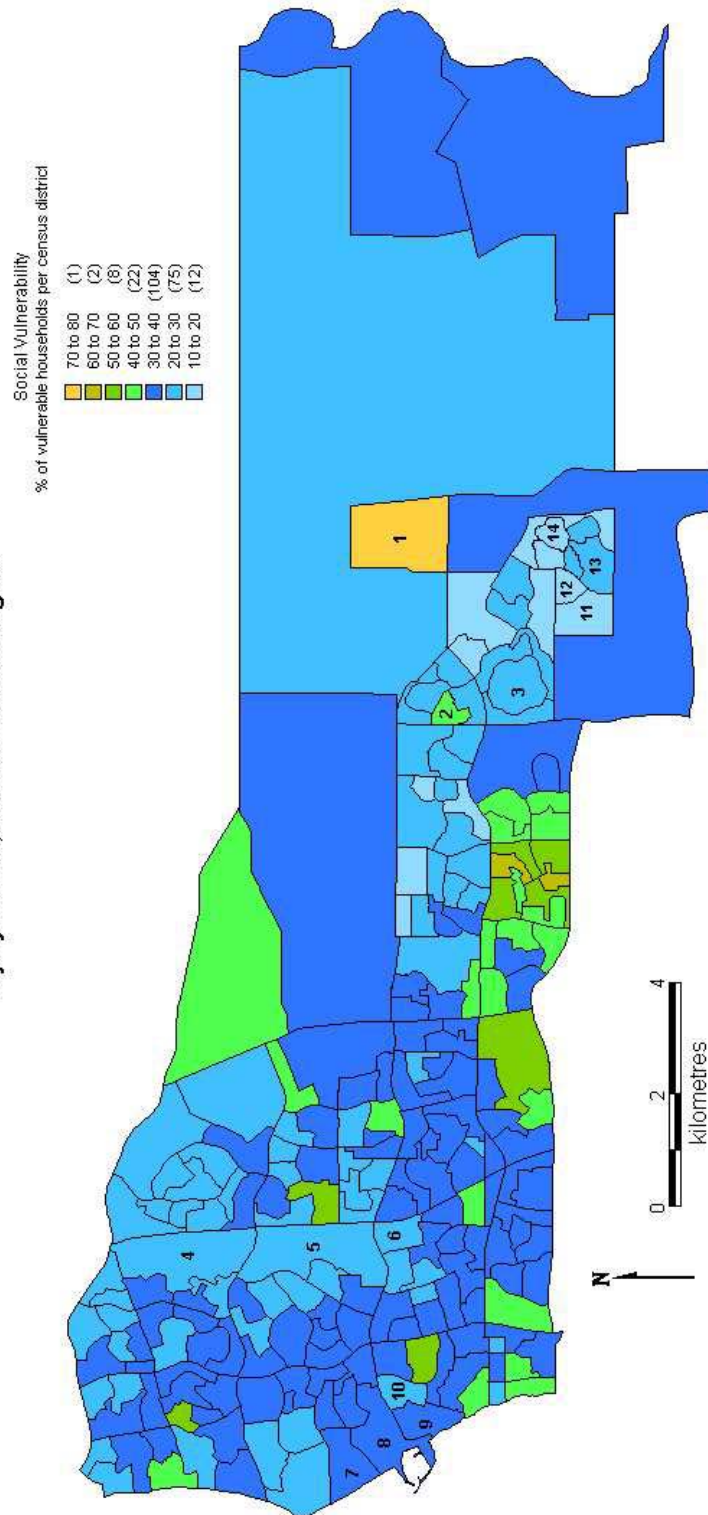


Figure 4.5: Scenario 3. This scenario visualises the estimated high vulnerability Census District areas when injury is minor and the residence is completely destroyed.

Scenario 4 Injury 3, Residence Damage 1 or 2

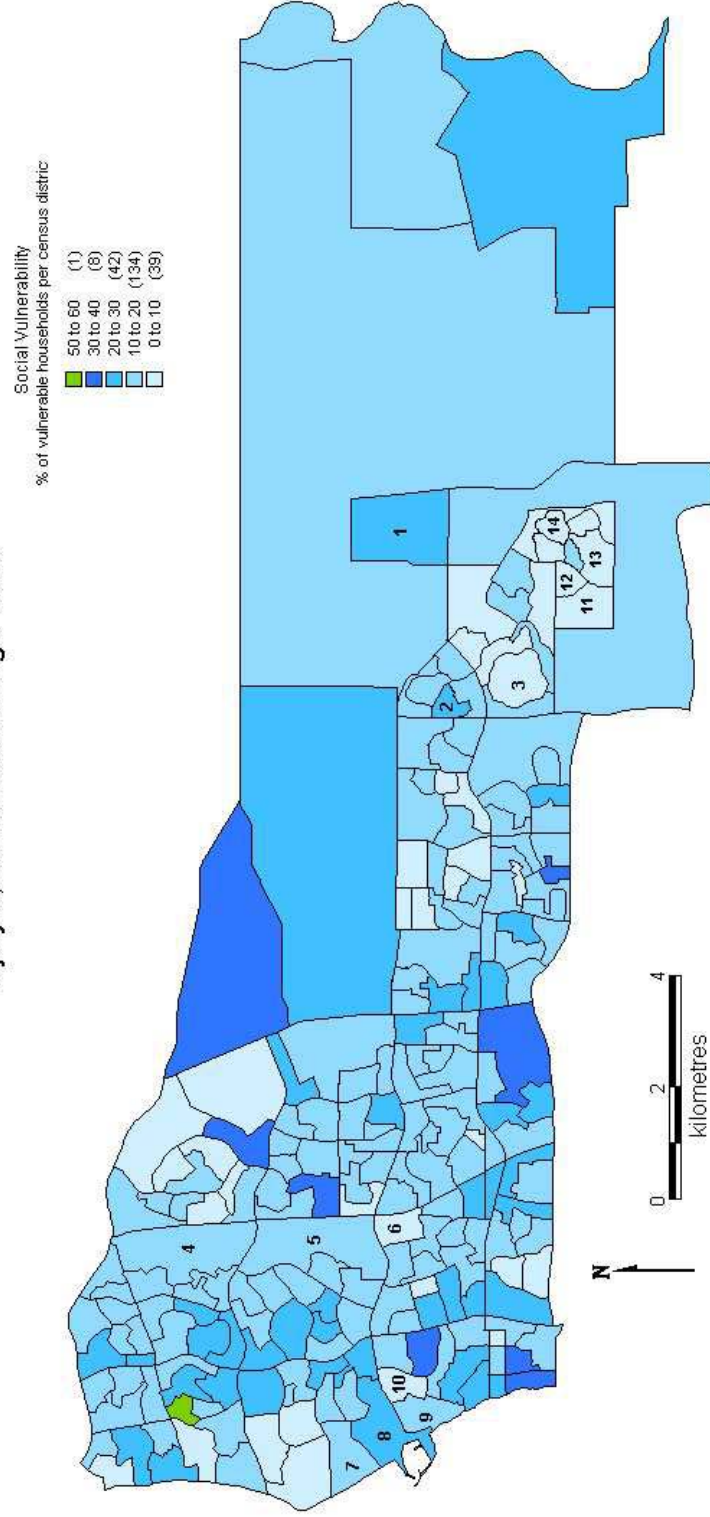


Figure 4.6: Scenario 4. This scenario visualises the estimated high vulnerability Census District areas when injury is moderate and residence damage is zero to slight.

Scenario 5 Injury 3, Residence Damage 3

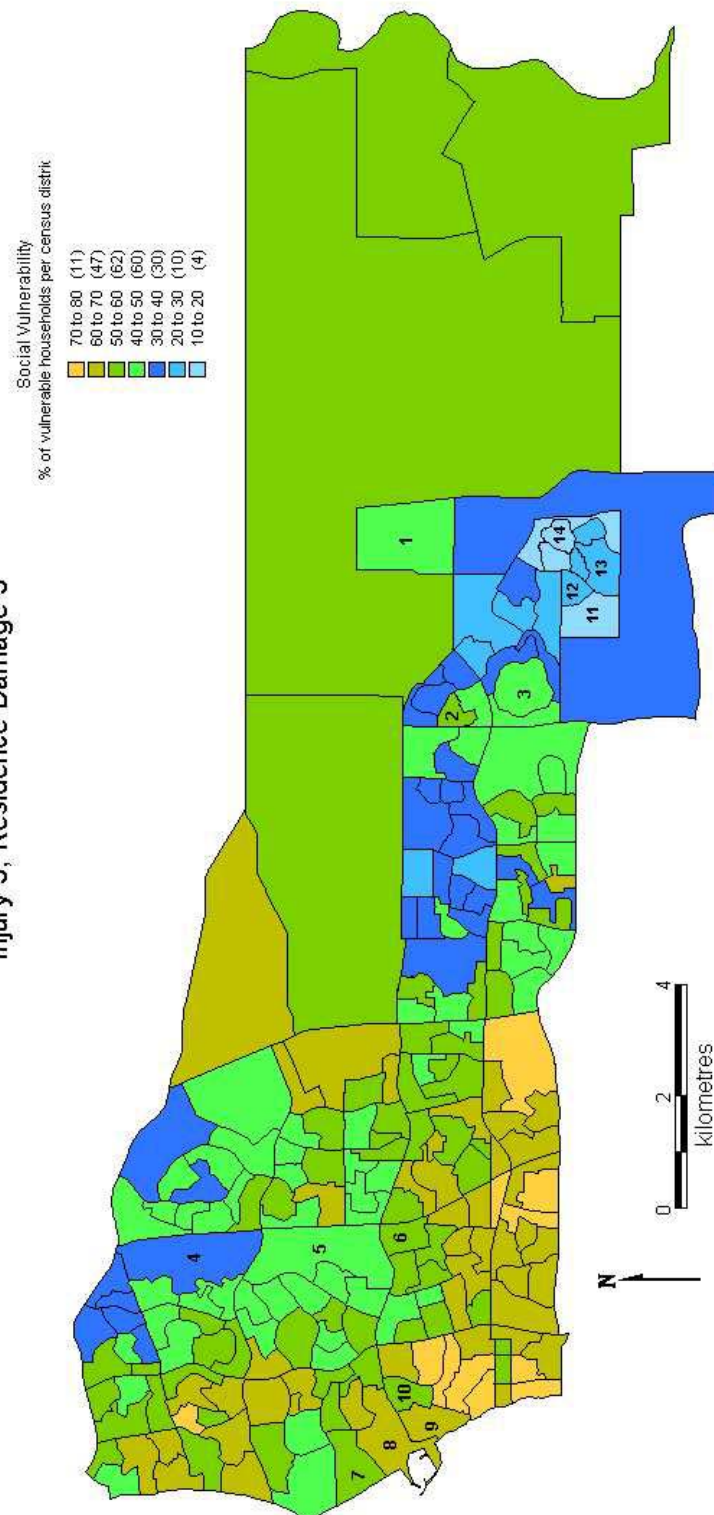


Figure 4.7: Scenario 5. This scenario visualises the estimated high vulnerability Census District areas when both injury and residence damage are moderate.

Scenario 6

Injury 3, Residence Damage 4 or 5

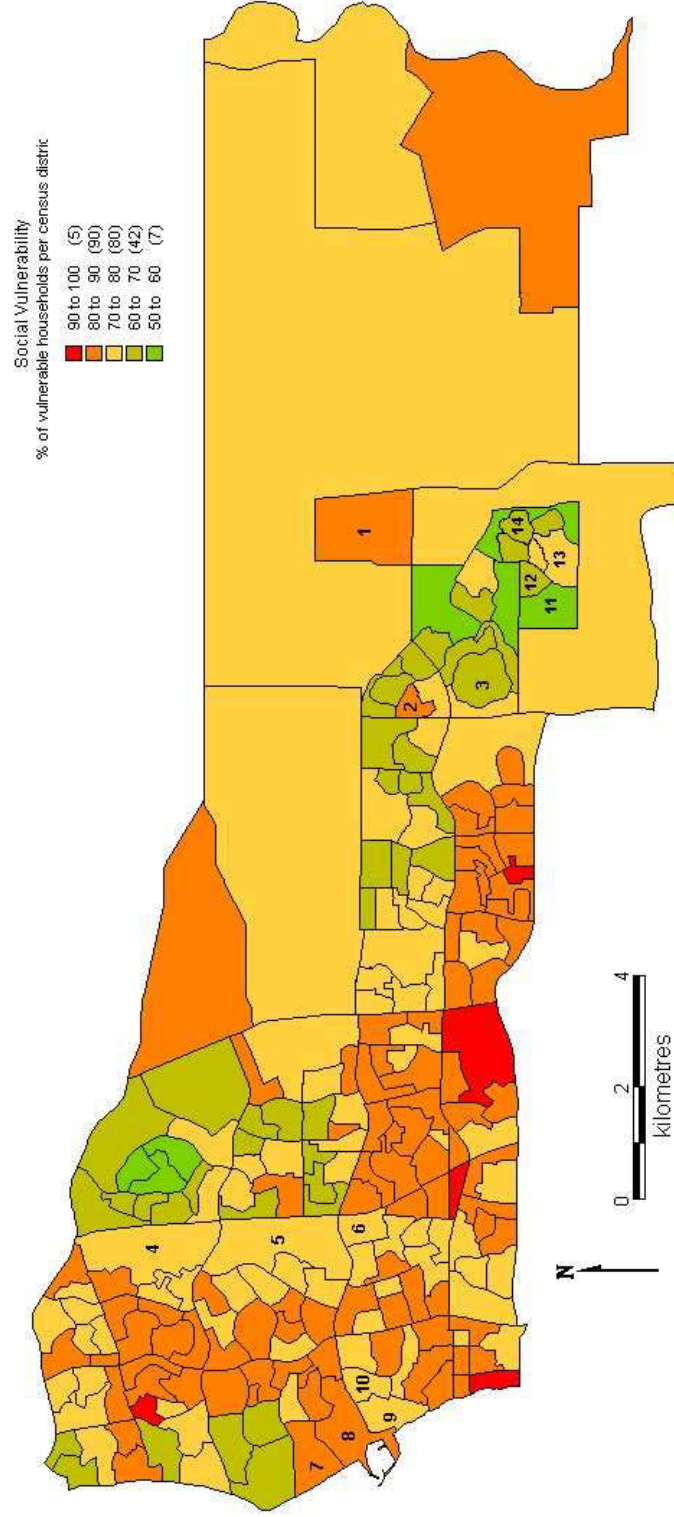


Figure 4.8: Scenario 6. This scenario visualises the estimated high vulnerability Census District areas when injury requires hospitalisation and residence damage is extensively damaged to completely destroyed.

Scenario 7

Injury 5, Residence Damage 1, 2, 3, 4 or 5

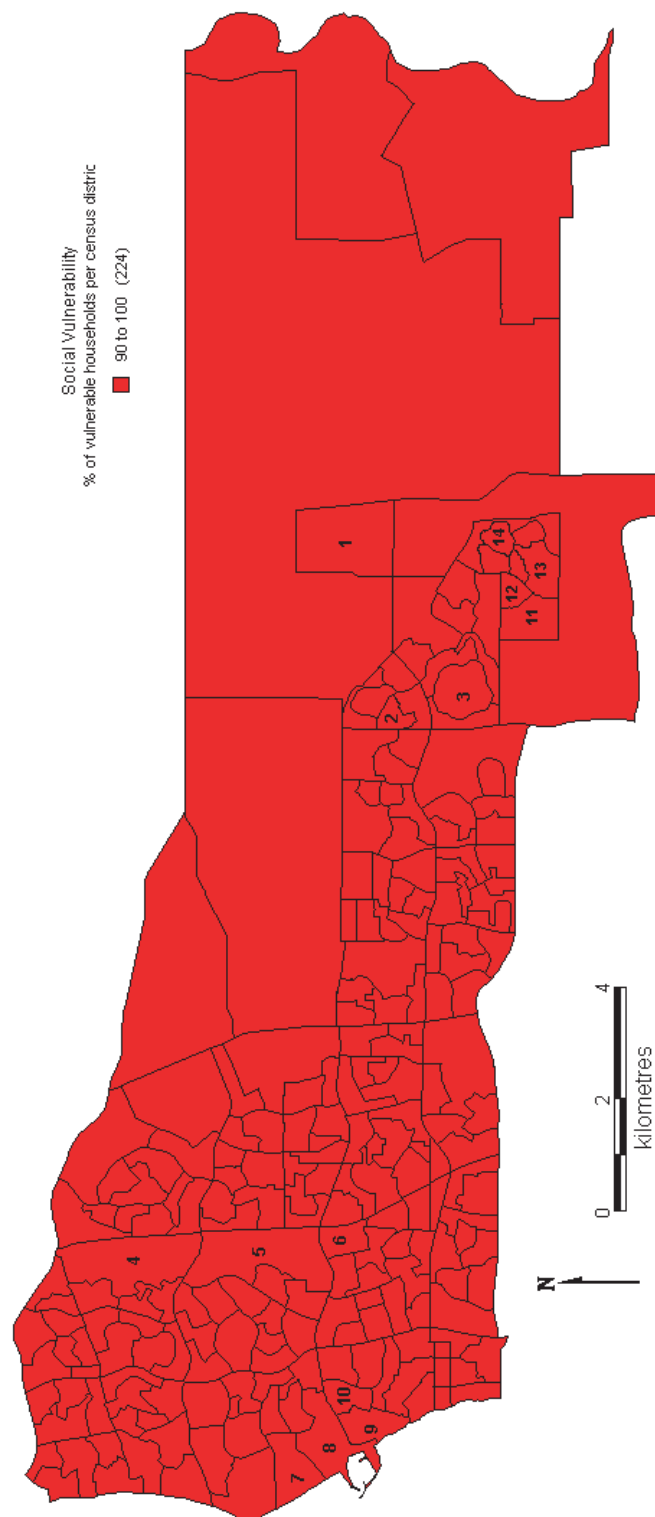


Figure 4.9: Scenario 7. This scenario visualises the estimated high vulnerability Census District areas when injury is life threatening, regardless of the condition of the residence.

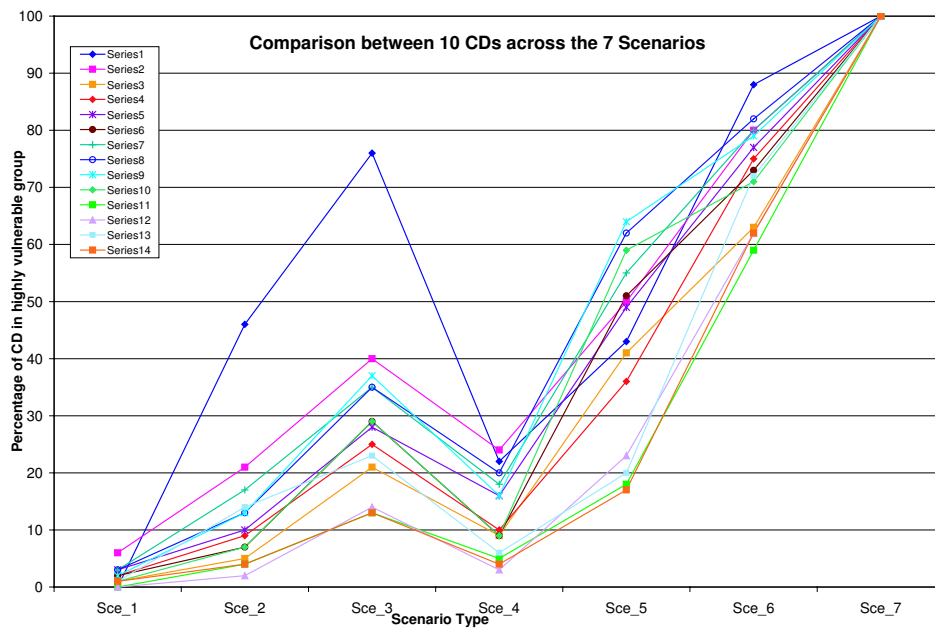


Figure 4.10: Comparative graph of 14 census districts in the study area across the 7 scenarios. The graph demonstrates that although injury and damage do provide the greatest influence on the vulnerability levels of each census district, other factors contribute to changes in vulnerability.

income, tenure type and age, which are the next most significant indicators. The combination of these indicators, according to the 11 decision rules of high vulnerability, will influence the percentage of highly vulnerable people in the mapped areas.

Example Census Districts

In order to further explore the Census Districts within and across scenarios, each of the seven scenario maps has 14 selected Census Districts that have been numbered from 1-14. These Census Districts show a cross-section of the Wanneroo area, with Census Districts 1 to 3 and 11 to 14 located in the central to central-south of the study area in some of the more recently developed suburbs. Census Districts 4, 5 and 6 are adjacent and are located along one of the main roads in the study area, while Census Districts 7 to 10 are also adjacent and located on the central coast of the study area in one of the more established parts of the Wanneroo study area. Each of the 14 Census Districts exhibit diverse vulnerabilities within each scenario and across scenarios, as is demonstrated in Figure 4.10.

The graph shows the 14 Census Districts across the seven scenarios. Census District 1 is an anomaly and can be explained, in part, by the abnormally low number of households (21) compared with the average of 200 per Census District. The graph also demonstrates that even though each scenario of fixed residence damage and injury influences vulnerability, as does the combination of socio-economic factors, the *rate* of change in vulnerability for each Census District is also different. This confirms that

some of the 11 high vulnerability decision rules are more relevant in some scenarios than others, according to certain injury and damage levels. The rate is different for Census District 2, in the newer development area, compared with Census District 8, which is in the coastal, more established area.

Damage and Injury

Census District 2 has the highest percentage of high vulnerability households (ignoring Census District 1) in Scenarios 1, 2, 3 and 4. These scenarios have lower levels of injury but high residence damage. However, Census District 9, which had lower vulnerability levels than Census District 2 in those scenarios, has a greater number of higher vulnerability households for scenarios 5 and 6. These scenarios, while also for high levels of residence damage, are also for higher injury damage. The decision tree rules, as shown in Figure 3.6 in the previous chapter, demonstrate that when injury level is high, age becomes more relevant to vulnerability. That is, people over 48 and 64, acknowledging other relevant attributes, may be more vulnerable when the hazard impact causes a moderate to high level of injury.

Figures 4.11 and 4.12 provide a comparison of the vulnerability of Census Districts when injury is relatively high (Scenarios 4, 5 and 6) and when damage is relatively high (Scenarios 2, 3 and 6). The maps exhibit that, regardless of higher levels of injury or damage, other attributes, represented by the vulnerability indicators, are important in determining social vulnerability. When injury is high, most of the Census Districts in the study area have an average of between 40% to 60% of households in the high vulnerability class. The cluster of lower household vulnerability, as indicated by the blue (20% to 40%) in Figure 4.11, is for newer suburban growth in the south-central region. This may relate to the fact that the decision tree determined when injury is significant, older age is an important attribute indicating vulnerability. The newer suburban areas in the Wanneroo area attract younger families and couples, thus reducing the average age of the Census Districts.

When residence damage is higher, most of the Census Districts in the study area have an average of between 30% to 50% of households in the high vulnerability class. Vulnerability to a natural hazard impact is, on average, less when residence damage is high compared with high injury. However, as it is not the case for every Census District as the vulnerability is dependant on the relationship of the other 13 social vulnerability indicators.

4.6 Using the results

The four-step methodology outlined in this report is intended to be integrated into a broader risk assessment, as outlined in the Introduction. However, it is also hoped that the methodology is an example of a useful quantitative vulnerability assessment that can be used to manage some of the social issues, such as the vulnerability of an individual within a household, relevant to impacts from a natural hazard. Therefore, the next chapter will explore some of the issues surrounding application, validation and usefulness of this social vulnerability assessment methodology.

Vulnerability when Injury is high (Scenarios 4, 5 and 6)

Percentage of vulnerable households in each CD

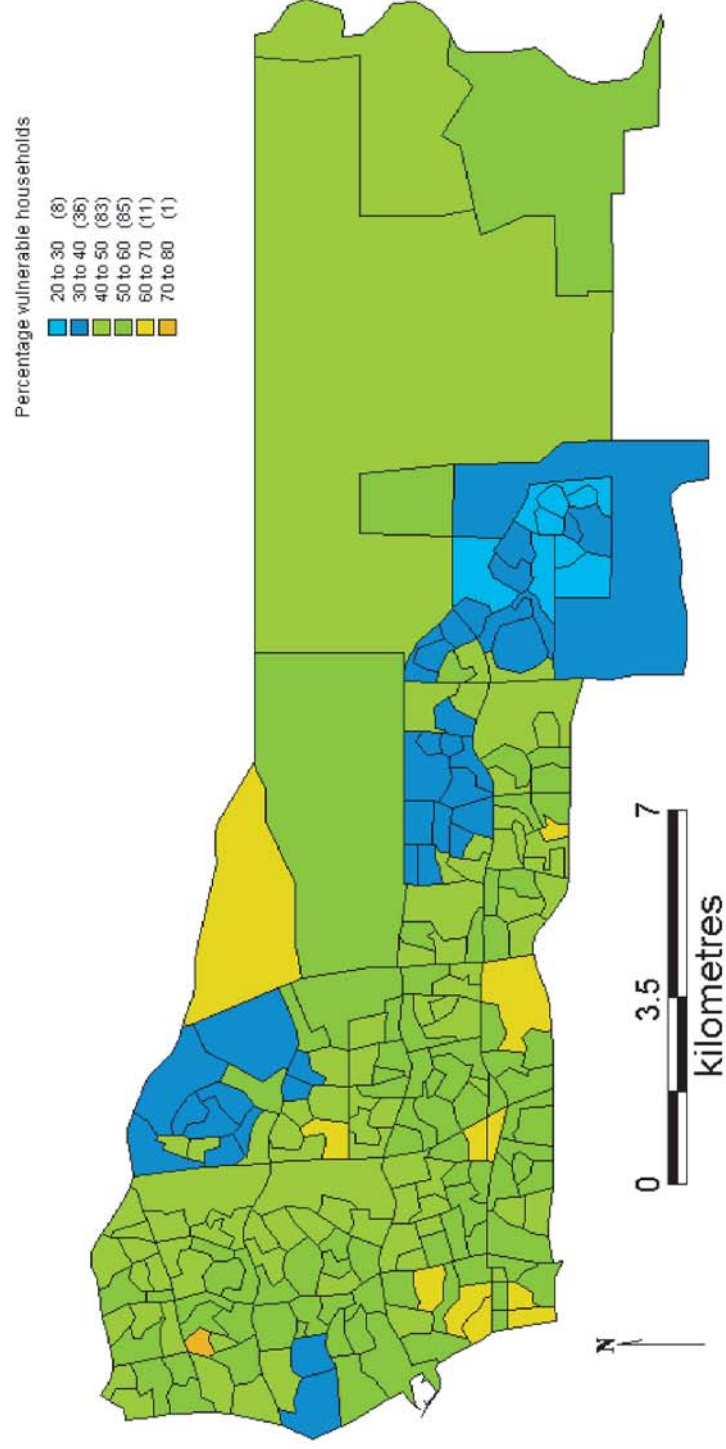


Figure 4.11: Percentage of vulnerable households in each Census District, when Injury is high (Scenarios 4, 5 and 6).

Vulnerability when Residence Damage is high (Scenarios 2, 3 and 6)

Percentage of vulnerable households in each CD

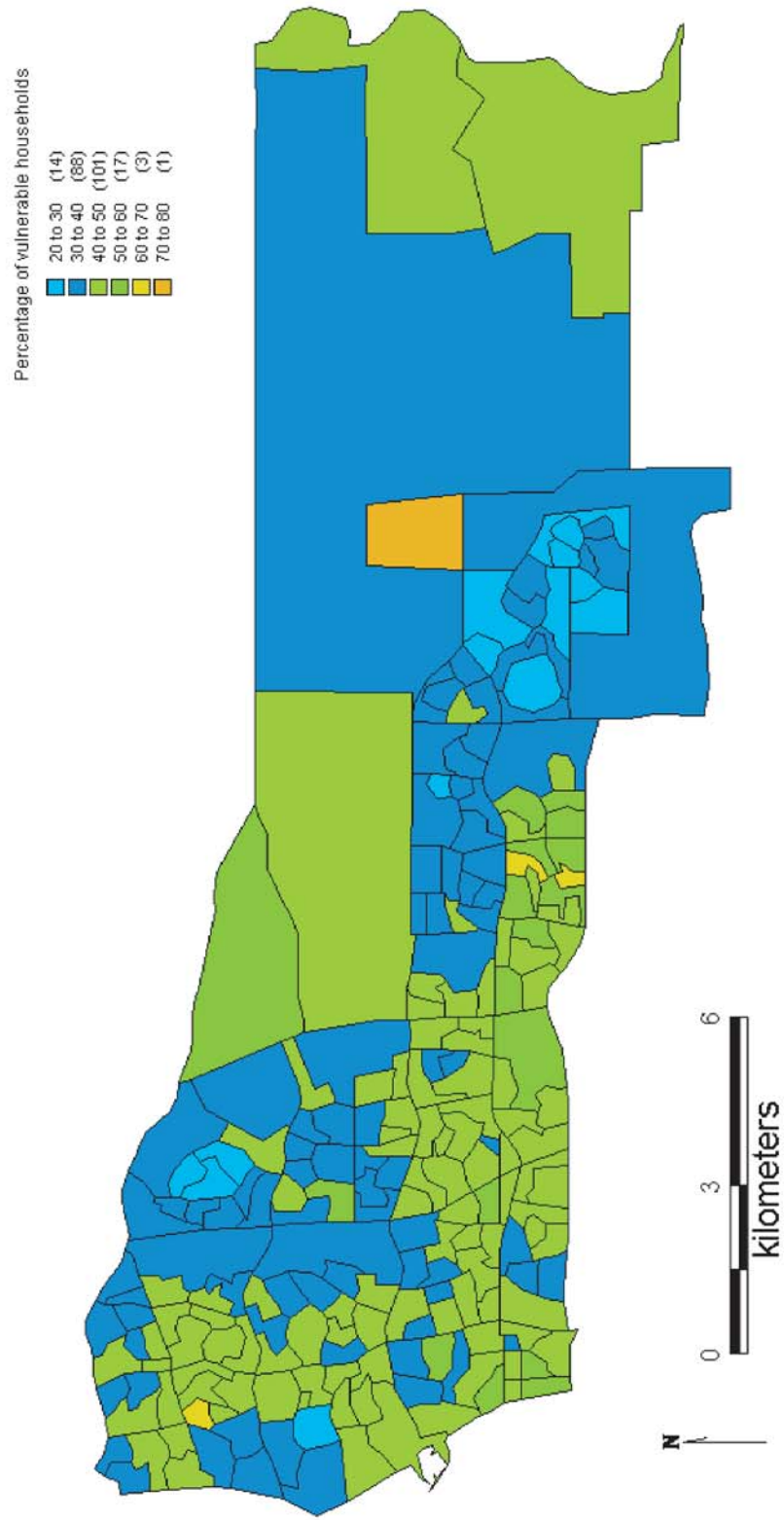


Figure 4.12: Percentage of vulnerable households in each Census District, when Residence Damage is high (Scenarios 2, 3 and 6).

Chapter 5

Discussion

This report has presented a new approach to measuring the vulnerability of an individual within a household to natural hazards. The need to explore a new approach has been driven by two significant factors; firstly, to value-add to the hazard model development undertaken at Geoscience Australia. Risk model development aims to encompass all aspects of natural hazards in order to provide a comprehensive risk assessment for government decision makers. Secondly, the study hopes to contribute to the ongoing development of vulnerability assessments that assist decision makers in safeguarding communities. Exploring methods for better understanding who is at risk in our communities will ultimately lead to better risk management by government and local decision makers.

The report has introduced a four step methodology to measuring the vulnerability of an individual within a household, involving;

- 1. Indicator Selection
- 2. Risk Perception Questionnaire
- 3. Decision Tree Analysis
- 4. Synthetic Estimation

With future refinements and developments, such an approach has the potential to contribute to policy development affecting natural hazard management at a government level.

This final chapter will provide a synopsis of the results outlined in the previous chapter and present two possible methods of validation for the Perth case study application. The chapter will then conclude with a review of each of the four steps, outlining limitations and areas for future development and refinement.

The results of the application of the quantitative methodology to the Perth case study are introduced in the previous chapter. As any study into the vulnerability of a community to natural hazards demonstrates, there is no single answer - vulnerability is dynamic and depends a range of factors. However, the results demonstrate that vulnerability can be explored and represented quantitatively. Social vulnerability to natural hazards, in this study, is represented by perception to risk, as was demonstrated

in the finding that the main factor contributing to a person's ability to recover is injury state and residence damage.

This study has challenged some of the assumptions about the personal attributes that contribute to a person's ability to recover from a natural hazard, for example, aside from a life-threatening injury, no single factor makes a person vulnerable. Financial indicators were the most significant after injury and residence damage, and yet their role in vulnerability depends on other factors. For example, having a mortgage may only make you vulnerable if you are of retirement age and have no house insurance (Decision Rule 5). Being over the age of 60 may only contribute to vulnerability if you sustain an injury that requires treatment (Decision Rule 2). Renting a home may *increase* your vulnerability if you are moderately injured and over 65 (Decision Rule 10), as you may have more immediate financial commitments. However, renting may *lessen* your vulnerability if you are younger than 43 and your house was destroyed, compared with those of the same age who owned their destroyed home but did not have insurance (Decision Rule 9).

5.1 Comparison with the Cities Project methodology

The Cities Project methodology is different from the methodology introduced in this report, especially with regards to some of the indicators used, the weighting of indicators and the classes of vulnerability. However, the objective is the same for both methodologies; producing a vulnerability map of a spatial community that can be integrated into a multi-hazard assessment and which has the potential to be used by emergency managers, policy developers and town planners.

To provide a comparison, the Cities Project methodology as outlined and detailed in Granger and Hayne [28], was applied to the same 224 Census Districts in the Wanneroo area of Perth, Western Australia. As per the methodology in this report, the 1996 census data was used to ensure consistency. The Cities Project methodology presents final results of a community's vulnerability in one final map, referred to as the 'Community Vulnerability Index'. For the Wanneroo case study area, the map of the 'Community Vulnerability Index' is presented in Figure 5.1.

The Cities Project methodology uses a different indicator list from the one used in this study. Therefore, some attributes of the community investigated are not referred to in one study or the other, making a direct comparison impossible. The Community Vulnerability Index mapped in Figure 5.1 has Census Districts ranked from 0 to 50, which is different from the *percentage* vulnerability ranking used in the methodology explored in this report.

The Community Vulnerability Index has greater variability in Census District vulnerability closer to the coast than inland, similar to the mapped Scenarios 2, Figure 4.4, and 4, Figure 4.6 in this study. However, any further comparison is difficult due to the different indicators used to arrive at the results for both the Cities methodology and the this report.

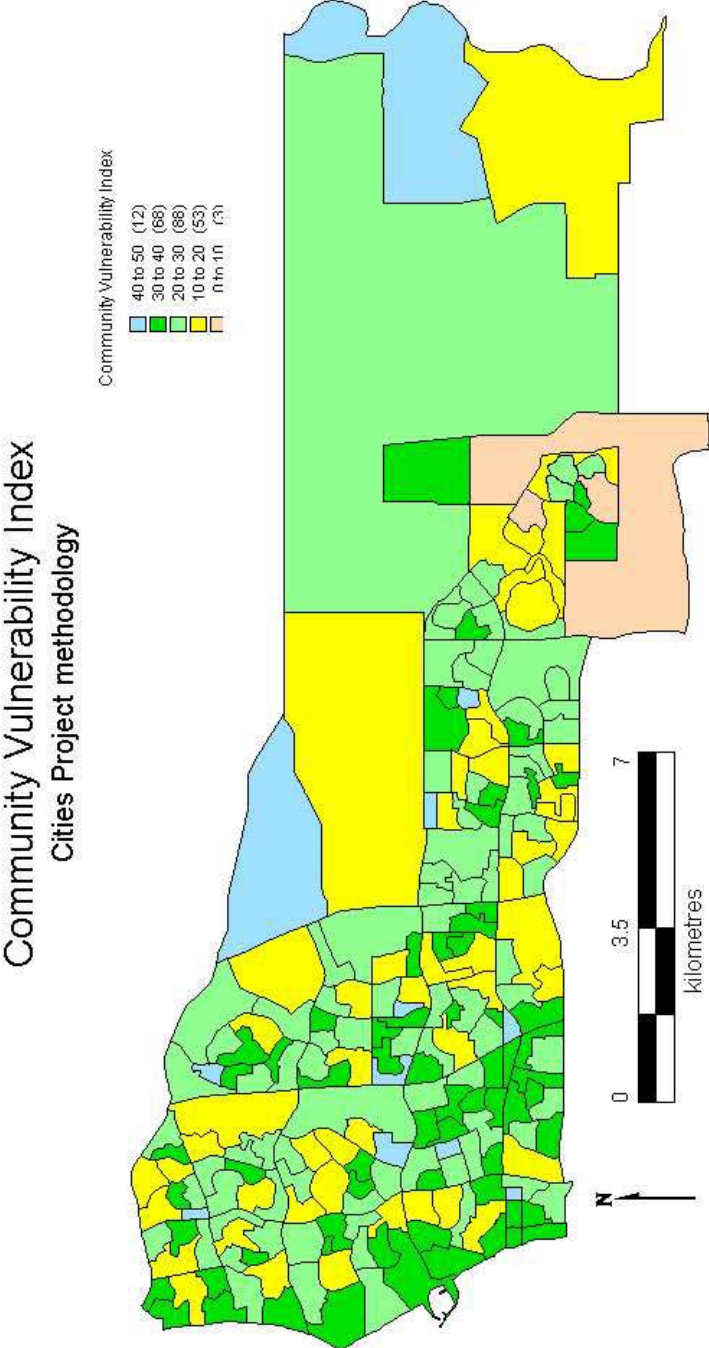


Figure 5.1: The Community Vulnerability Index for the Wanneroo study area using the Cities Project vulnerability methodology

5.2 Limitations and future work

Like any experimental methodology, this study has raised some questions that need to be explored in future research in order to refine vulnerability assessment development. Each of the four steps explored in this report have posed some questions and have highlighted areas for further work in future studies. While this vulnerability assessment methodology is intended to be layered with hazard and economic assessments in order to gain a better insight into risk from a natural hazard, the steps taken can be addressed in their own right. Some of these points for further work are outlined in the following:

5.2.1 Indicator Selection

The indicators were selected from a broad literature review and were selected according to the criteria outlined in Chapter Two. They have also been selected according to their relevancy and applicability to the Australian urban environment. Any investigation into social vulnerability to natural hazards in regional and rural Australia, or other countries, would need to incorporate other vulnerability indicators more relevant to the needs of those communities. This could include issues such as distance to services and assets vs cash-flow, especially for those employed in the agricultural industry.

The indicators used in this study explore social vulnerability at the individual and household level. However, for a broader representation of social vulnerability to natural hazard impacts, the development of community vulnerability indicators (such as resource availability, social capital measures) and regional vulnerability indicators (such as dependency on regional economy, access to service areas) would be important.

5.2.2 Risk Perception Questionnaire

The development of the questionnaire was undertaken primarily to test a methodology, that is, to test whether there is a possible technique that focuses on predominantly objective measures of investigating the relationship between vulnerability indicators. As such, the main focus of this research is to explore quantitative and analytical methods, rather than profess expertise in the discipline of sociology, in particular, questionnaire development. Points raised by Schuman and Presser in developing social surveys were taken into account and can be referred to in Schuman [62]. Therefore, this study acknowledges that if this method demonstrates to be a viable means of obtaining data for decision tree analysis and as a result, provides a fair technique for representing vulnerability, then improvements to the questionnaire that must be made include:

- Developing, with the expertise of sociologists and social survey experts, a more comprehensive and conventional questionnaire.
- Clarify the scenarios presented as per person or per household. The current scenario of hypothetical individuals utilises data that refers to both a person and to a household and may influence or confuse final results.
- Investigate further methods that could be used to capture ‘perception’ of vulnerability or risk. A realistic data set based on factual data after a natural hazard

impact (including recovery time) would be the ideal data set, but does not currently exist. In the event of a natural hazard impact, it would be highly desirable to instigate a post-disaster data collection project based on the 15 selected indicators.

The questionnaire used data that was representative of the Australian population. As a result, some attributes, such as disability and poverty, which represent a small percentage of the population, may not have been adequately represented in a small questionnaire distribution. Future questionnaire development and distribution may need to take into account overall *population* statistics as well as representative *distribution* samples. One possibility could be to have these variables ranked differently by questionnaire respondents and been assigned greater importance in the decision tree. Disability for example occurred in only 10% of the hypothetical cases and was ranked twelfth out of the thirteen socio-economic variables by the decision tree. The data are crucial to the final result.

5.2.3 Decision Tree Analysis

Decision tree analysis was used for interrogating data in this study, as the main purpose of the analysis technique is to sort and classify the data. It was also selected over other techniques, such as neural networks, as it is considered simpler, especially with regards to the application in this study.

The decision tree analysis investigated the relationships between the 15 selected indicators of vulnerability. However, as noted in Chapters One and Two, two of the indicators, Injury and Residence Damage, are more outcome-dependent than the other indicators. While these two indicators are essential in establishing the possible effects of a natural hazard, they greatly influenced the perceived importance of social vulnerability as they were ranked as the two most important indicators. When adequate and comprehensive hazard models are available, injury and damage will be established using probabilistic methods specific to certain areas. As such, these two indicators can be removed from the questionnaire and replaced with a probability of impact. This will both influence the impact on the household and their recovery and influence the decision tree analysis results. However, until such hazard models are available, it is important to provide a *context* for social vulnerability to natural hazard when investigating perceptions of risk.

5.2.4 Synthetic estimation

Synthetic estimation was used in this study to estimate the socio-economic profiles of each Census District as well as to estimate vulnerability to a natural hazard. It would be interesting to examine the synthetic estimates of selected variables, which were not used in the decision rules, to see if their relationship with the synthetic estimates of vulnerability accords with expectations. This comparison could be performed on particular Census Districts (of a reasonable size) that consistently appear in the high or low vulnerability class across scenarios. Such an assessment might provide insights into which variables should be considered in future studies.

5.3 Areas for refinement and further analysis

Stephen and Downing [64] argue that good vulnerability assessments must measure the right things, at the right scale, with suitable conceptual underpinning". The methodology explored in this report has attempted to achieve this, while also ensuring that it has a real application to a spatial community. However, of course, there are some issues that need further exploration, such as:

- What is the best scale for emergency management and effective disaster mitigation planning? Perhaps SLA (Statistical Local Area) is more effective or manageable than Census District? Should we focus on individual vulnerability or household, community, regional vulnerability?
- Which decision rules are relevant to specific Census District areas with higher percentages of highly vulnerable people? While the decision tree has noted financial indicators, outside of injury and residence damage, as the most significant attributes to vulnerability, which areas are they more important in?
- What other factors, such as political climate, local government policy, emergency service capabilities and welfare services, will contribute to a person's ability to recover from a natural hazard impact?
- Which personal attributes as listed by the vulnerability indicators, can be influenced prior to a hazard impact in order to lessen vulnerability? For example, can mitigation issues address house insurance, health insurance and income (in terms of providing greater welfare services) before the impact?

These questions need to be further addressed in order to better refine vulnerability assessments and their role in natural hazard risk management. With regards to the methodology development, there are also some areas which could be refined in future studies, including:

- Analysing the selected indicators of vulnerability for applicability and developing indicators that are community/location specific, as was done in the Yohannes and Webb [69] Ethiopian vulnerability assessment.
- Refine the risk perception questionnaire and distribute to a wider audience.
- Validation of the vulnerability map results with a wider audience, such as welfare officers, community services and managers involved in the area being assessed.

The international literature which has been referenced in this report, alongside the many other studies of vulnerability to natural hazards and other large-scale disasters, continuously highlight the complex issues surrounding methods of assessment. This report acknowledges these concerns, while also arguing the need to continue trying new approaches. It is only through ongoing investigations and experimental applications that we will be able to gain a better understanding of how to contribute to more resilient communities in the event of natural hazard impacts.

Appendix A

Risk Perception Questionnaire

Measuring community impact to natural hazards: Developing a quantitative model to contribute to disaster risk management

This survey is part of Geoscience Australia Urban Risk Group's research into understanding the risk posed to communities from natural hazards. The impact of natural hazards on people within a community has long been recognised as a difficult assessment to make. Each person has different lifestyle characteristics that contribute to how they experience and recover from an extreme event. For the purpose of this study, the impact shall be considered in terms of ability to recover, that is, how long it would take someone to return to a similar lifestyle situation prior to the natural hazard impact. It is anticipated that some people will take little time to recover while others may never return to a similar lifestyle prior to the impact and therefore have an indefinite recovery period.

The aim of this study is to try and capture your perception of a person's ability to recover from a natural hazard impact based on combinations of a small selection of lifestyle characteristics. For each of the hypothetical individuals presented overleaf, created from Australian Census data, please rank your perception of the individual's recovery ability for each of the 4 natural hazards. Assume that the individual lives in an urban community in an industrialised nation and that no essential services have been completely destroyed so that each person has access to medical, welfare and other services if required. The hypothetical individuals have been created randomly from a program so that every individual in every survey distributed is unique. The greater the number of individuals ranked, the greater our understanding of vulnerability may be, and therefore your input is very important.

At the extremes, 1-2 denotes very little disruption, 4-5 represents sufficient time to interrupt life and 9-10 represents a person's inability to recover from the impact of the natural hazard. There are no correct or incorrect answers only your perception of recovery ability. An example is below:

Person 10 is a male aged 46, is employed, earns \$328 a week and is running into debt. They live as a couple with no dependents and own with no mortgage a house with houseinsurance and has 1 car. This person does not have health insurance, does not have a disability and speaks English well. Due to the hazard impact, this person requires hospitalisation and is expected to recover and their residence is slightly damaged. According to the recovery scale, indicate how long do you think it will take for this person to recover from the following impact?

- A destructive earthquake (1-10)....7.. A major flood(1-10)....6..
- A major landslide (1-10)....5.. A major cyclone (1-10)....6..

Please indicate if you have been exposed to a:

- Major flood..... Destructive earthquake.....
- Major landslide..... Major cyclone..... Another major hazard.....

Thank you very much for your time

Person 1001 is a male aged 72, is not employed, receives \$112 a week and is drawing on savings. They are single and live in group housing and own with no mortgage a house with houseinsurance and has 0 cars. This person does not have health insurance, does not have a disability and speaks English well. Due to the hazard impact, this person sustained no injuries and their residence is not damaged. According to the recovery scale, indicate how long do you think it will take for this person to recover from the following impact?

- A destructive earthquake (1-10).....
- A major flood (1-10).....
- A major landslide (1-10).....
- A major cyclone (1-10).....

Person 1002 is a male aged 37, is not employed, receives \$275 a week and is making ends meet. They are single and live in group housing and own with no mortgage a house with houseinsurance and has 0 cars. This person does not have health insurance, does not have a disability and speaks English well. Due to the hazard impact, this person requires basic medical treatment without hospitalisation and their residence is moderately damaged. According to the recovery scale, indicate how long do you think it will take for this person to recover from the following impact?

- A destructive earthquake (1-10).....
- A major flood (1-10).....
- A major landslide (1-10).....
- A major cyclone (1-10).....

Person 1003 is a female aged 42, is not employed, receives \$1315 a week and is making ends meet. They live as a couple with dependent children and privately renting a house with contents insurance and has 3 cars. This person does not have health insurance, does not have a disability and does not speak English. Due to the hazard impact, this person requires hospitalisation and is expected to recover and their residence is slightly damaged. According to the recovery scale, indicate how long do you think it will take for this person to recover from the following impact?

- A destructive earthquake (1-10).....
- A major flood (1-10).....
- A major landslide (1-10).....
- A major cyclone (1-10).....

Person 1004 is a female aged 40, is employed, earns \$347 a week and is saving a lot. They are single and live in group housing and own with no mortgage a house with no houseinsurance and has 1 car. This person does not have health insurance, does not have a disability and speaks English well. Due to the hazard impact, this person requires hospitalisation with life threatening injuries and their residence is completely destroyed. According to the recovery scale, indicate how long do you think it will take for this person to recover from the following impact?

- A destructive earthquake (1-10).....
- A major flood (1-10).....
- A major landslide (1-10).....
- A major cyclone (1-10).....

Person 1005 is a female aged 25, is not employed, receives \$320 a week and is drawing on savings. They live as a couple with no dependents and own with no mortgage a semi-detached one storey townhouse with no houseinsurance and has 1 car. This person does have health insurance, does not have a disability and speaks English well. Due to the hazard impact, this person requires hospitalisation and is expected to recover and their residence is completely destroyed. According to the recovery scale, indicate how long do you think it will take for this person to recover from the following impact?

- A destructive earthquake (1-10).....
- A major flood (1-10).....
- A major landslide (1-10).....
- A major cyclone (1-10).....

Person 1006 is a female aged 42, is employed, earns \$811 a week and is making ends meet. They are single and live in group housing and privately renting a house with no contents insurance and has 3 cars. This person does not have health insurance, does not have a disability and speaks English well. Due to the hazard impact, this person sustained no injuries and their residence is moderately damaged. According to the recovery scale, indicate how long do you think it will take for this person to recover from the following impact?

- A destructive earthquake (1-10).....
- A major flood (1-10).....
- A major landslide (1-10).....
- A major cyclone (1-10).....

Person 1007 is a male aged 23, is employed, earns \$729 a week and is saving a little. They live as a couple with no dependents and rent from a housing trust a house with contents insurance and has 1 car. This person does not have health insurance, does not have a disability and speaks English well. Due to the hazard impact, this person requires hospitalisation and is expected to recover and their residence is slightly damaged. According to the recovery scale, indicate how long do you think it will take for this person to recover from the following impact?

- A destructive earthquake (1-10).....
- A major flood (1-10).....
- A major landslide (1-10).....
- A major cyclone (1-10).....

Person 1008 is a female aged 78, is not employed, receives \$153 a week and is running into debt. They live alone and privately renting a semi-detached one storey townhouse with contents insurance and has 0 cars. This person does have health insurance, does not have a disability and speaks English well. Due to the hazard impact, this person requires hospitalisation with life threatening injuries and their residence is moderately damaged. According to the recovery scale, indicate how long do you think it will take for this person to recover from the following impact?

- A destructive earthquake (1-10).....
- A major flood (1-10).....
- A major landslide (1-10).....
- A major cyclone (1-10).....

Person 1009 is a male aged 60, is employed, earns \$594 a week and is making ends meet. They live as a couple with no dependents and privately renting a house with contents insurance and has 0 cars. This person does not have health insurance, does not have a disability and speaks English well. Due to the hazard impact, this person requires hospitalisation with life threatening injuries and their residence is completely destroyed. According to the recovery scale, indicate how long do you think it will take for this person to recover from the following impact?

- A destructive earthquake (1-10).....
- A major flood (1-10).....
- A major landslide (1-10).....
- A major cyclone (1-10).....

Person 1010 is a female aged 39, is employed, earns \$434 a week and is making ends meet. They live as a couple with no dependents and own with no mortgage a house with house insurance and has 1 car. This person does not have health insurance, does not have a disability and speaks English well. Due to the hazard impact, this person requires hospitalisation with life threatening injuries and their residence is moderately damaged. According to the recovery scale, indicate how long do you think it will take for this person to recover from the following impact?

- A destructive earthquake (1-10).....
- A major flood (1-10).....
- A major landslide (1-10).....
- A major cyclone (1-10).....

Appendix B

Decision Tree Methodology

Identifying attributes that are perceived to make a person vulnerable is considered here to be a ‘Knowledge Discovery in Databases’ exercise. Knowledge discovery is defined as ‘the non-trivial extraction of implicit, unknown, and potentially useful information from data’ [27] and refers to techniques developed to transform large quantities of data into useful and understandable information [68]. One such technique for extracting useful knowledge from data is through data mining techniques, which refer to a class of algorithms used to identify patterns in data. The primary purpose of data mining is predictive modelling, a task which seeks to establish a functional relationship between two sets of attributes in a database in order to predict values of one of the sets based on values of the other. Usually, one variable is selected as the *target* or *response* variable and the resultant predictive model can be used to establish the outcome of the target given any combination of the predicting attribute values. When the response variable is categorical, that is there are a fixed set of outcomes, one speaks of a classification problem. When it is continuous, there are an infinite number of outcomes, usually a real number, the problem is said to be a regression problem, where one wants to predict a real number.

Decision tree induction is one of the most widely used Knowledge Discovery predictive techniques because it;

- Can handle both discrete (a value which has a finite number of distinct values, such as the number of cars owned, whether or not someone has house or health insurance) and continuous data (a value which has an infinite number of possible values within an interval, such as income, age and injuries)
- Makes no assumption as to the shape or form of the distribution of the underlying data,
- Is fast and efficient,
- Is robust with regards to missing values and noisy data,
- Is straightforward to interpret and comprehend,
- Can provide an indication of attributes that are most important for prediction or classification, and
- Is nonparametric: Decision trees do not require additional information (such as a functional form) besides that already contained in the data set.

The major weakness of decision tree induction (as well as other nonparametric methods) is that it is prone to errors in classification if many classes are involved or if there is a relatively small training data set.

		<i>continuous</i>		<i>continuous</i>		<i>categorical</i>		<i>categorical</i>	
	ID	Age	Income	.	.	.	English	Disability	Scale
instance	1	57	317				yes	no	5
	2	32	215				no	no	8
	3	24	517				yes	yes	7


	<i>n</i>	42	1057				yes	no	5
									class label

Figure B.1: Data structure in decision tree induction.

Decision tree construction

Decision tree construction, referred to as induction, attempts to identify prediction rules from a set of observations collected from a population. The individual observations are called 'instances' and a set of observations is called the 'data set'. Let $D = \{(\mathbf{x}_i, y_i), i = 1, \dots, n\}$ be a data set that contains n instances, also known as 'records' or 'examples' or 'events'. All instances must have the same number of attributes (also denoted indicators) and a 'class label', 'concept' or 'target variable' here denoted as $y_i, i = 1, \dots, n$. However, it is not essential that each attribute has a value. Each record is characterized by a vector $\mathbf{x} = (x_1, x_2, \dots, x_d)$ of d attribute values x_k and a class label y . In the case of a classification problem the label is chosen from a finite set $\mathbf{y} \in \{y_1, y_2, \dots, y_c\}$ of c classes. The space of all n instances is called the 'instance space'. The attribute space represents the dimensionality of the space, d , with each attribute supplying a single dimension.

An attribute value $x_k \in \mathbf{x}$ may be either continuous consisting of integers or real numbers, or categorical which is a set of values. Figure B.1 illustrates a typical instance space containing instances, class labels and continuous and categorical attributes. The classification problem can be stated formally as the task of finding a classification model $f : \mathbf{x} \rightarrow y$. This is simply a mapping of the set of attributes into one of the class labels.

A decision tree analysis recursively partitions, or splits, the instance space into increasingly homogeneous disjoint subsets with the aim that each subset should contain as many instances with one class label as possible. This notion is often referred to as 'purity'. Each subset is then assigned to one class based on the class of the majority of instances. Typically, more than one subset is assigned to the same class. Those subsets correspond to different ways that a class can be reached. Starting at the root node, which contains the entire data set, all attributes are tested to see which attribute is the best for partitioning, or splitting, the data into two different and more homogeneous subsets. Moreover, for each attribute the best splitting value is determined. A visual representation of this concept is shown in Figure 3.3 in Chapter Three.

The process of selecting a new attribute and splitting the data set is now repeated this time using only the data set associated with each internal node. The recursive algorithm picks the best attribute and the best split and never looks back to reconsider earlier choices. This process continues until the instances associated with each terminal node or leaf node all have the same target class label. The resulting tree is referred to as the *maximal tree*. Each branch represents a

rule, which progressively refines the classification and can also be used to classify future data.

The structure of a decision tree and its components; the root node, branch, decision and leaf nodes are shown in Figure 3.4. Once built, a decision tree can be used to classify an instance by starting at the root node of the tree and moving through it until a terminal node is encountered. The class assigned to each terminal node then corresponds to a predicted outcome for the instance in question.

Rule Based Classifiers

The purpose of a classification technique is to find the appropriate classification model that summarizes the relationship between a set of attributes and the class. The classifier is defined by a set of rules. Decision tree induction is a supervised learning classifier which can be converted into rules that represent a summary of the knowledge content of the data. The rules in a classifier are determined by the path from the root to a leaf node of a decision tree, with one rule created for each branch along the path. All paths taken can be expressed as nested 'if-statements', which can be interpreted as rules.

Most decision tree induction algorithms are based on the original work of Hunt *et al.* [36]. Other well known decision tree algorithms are CART [9], C4.5 [60] and OC1 [46]. These algorithms differ in the way they handle missing attribute values, continuous attributes, the splitting criteria and how they establish the size of the tree.

Pruning

The maximal tree will do the best possible prediction on the data set it was trained on, i.e. every instance will be correctly classified by following the generated rules. This is referred to as 100% predictive accuracy or a misclassification rate of 0.0. The misclassification rate is defined as the ratio of instances that were classified wrongly to the total number of instances n .

However, a maximal tree is unlikely to classify new data very well because it has been tailored to the idiosyncrasies of one particular data set. Data sets contain a certain amount of noise which can arise from errors in measuring the attribute and class values of an instance. There is also the uncertainty in the selection of the attribute set and not all the relevant properties of an instance are known. If a classifier fits the instances too closely, it may fit noisy instances instead of a sample from the population. This is undesirable because the decision tree may not necessarily fit other data sets from the same population equally well. This phenomenon is well known in data mining as 'over-fitting' and poses a challenge to all predictive and regression models [6]. The success of any predictive model depends on its ability to distinguish between noisy instances and patterns in the data.

The remedy for over-fitting is the use of a separate test data set together with a pruning technique where parts of the tree are eliminated. The premise is that both the training and test data sets will contain similar errors or noise. The misclassification error on the test set is then calculated for every sub-tree (sharing the same root) and the subtree with the smallest error is selected for the final model.

While the misclassification error with respect to the training data will decrease with the tree's complexity, the error with respect to the test set will find an optimum somewhere between the root of the tree and the maximal tree. When the size of the tree is too small, under-fitting occurs. An under-fitted decision tree will perform poorly on the training data set as well as an independent data set. When the size of the tree is too large, over-fitting occurs. An over-fitted decision tree will perform perfectly on the training data set but poorly on the independent data set. The pruned tree has the optimal size and is expected to perform well on both training and test data. Not only does this final pruned tree perform well as a predictive model, the number of instances supporting each subset will be larger than those of the maximal tree and the rules derived from it will be less complex than those of the maximal tree.

In cases where a test set is not available or if the number of instances is so small that one can not be obtained by sampling one can use a technique called ' k -fold cross-validation'. The data set is divided into k subsets. The decision tree is built on $k - 1$ subsets and the model error is evaluated on the remaining subset serving as a test set. This procedure is repeated until each subset has been used once as the validation set. The model error is the average error made for the k simulations.

Attribute Selection Criteria

Attribute selection is one of the basic issues in decision tree learning. The attribute selection criteria is used to establish which test condition provides the best split of a collection of instances into a smaller set. Attributes which provide poor overall performance are either ignored or are embedded deep within the tree.

One strategy for ensuring that the attributes provide the greatest information about the class label is to select a suitable attribute that will 'split' the data set. The gain in information implies a reduction in uncertainty. In general, the amount of information gained is inversely proportional to the probability of that event. For example, relative to what is expected, more is learnt from being told a relatively unlikely event has occurred.

Let c_i denote the number of instances from class y_i , n_j the number of instances in tree node j and c_{ij} the number of instances in node j from class y_i . An estimate of the probability of obtaining these values from a data set is given by

$$p_{ij} = \frac{n_{ij}}{n}, p_i = \frac{n_i}{n}, \text{ and } p_j = \frac{n_j}{n}. \quad (\text{B.1})$$

An estimate of the probability of an instance in node j belonging to class i is therefore

$$p_{i|j} = \frac{c_{ij}}{n_j}. \quad (\text{B.2})$$

If, for example, all classes are equally represented in node j then $p_{i|j} = 1/c, i = 1, \dots, c$. In general

$$\sum_{i=-\infty}^{\infty} p_{i|j} = 1 \text{ for all } j$$

The degree of heterogeneity or impurity in each node can be defined in several ways leading to a number of splitting strategies that can be employed in the search for the best splitter. An important splitting rule used in this study is the Gini Index. The Gini index used in CART [9] is simple, performs well in general and is suggested as the default splitting rule in CART. The Gini index is defined by

$$\text{Gini}(j) = 1 - \sum_{i=1}^c p_{i|j}^2 \quad (\text{B.3})$$

If all instances of \mathbf{S} belong to the same class then $p_{i|j} = 1$ for that i , the remaining class probabilities would be zero and the the Gini index would be 0 corresponding to no diversity in node j . On the other hand, if node j contained equal numbers of each class $p_{i|j} = 1/c$, where c is number of classes, the Gini index would attain the maximum value $1 - 1/c$ corresponding to the greatest possible diversity or *impurity* of that node. Let for example node 1 have $c_{11} = 60$, $c_{21} = 40$ and node 2 have $c_{12} = 7$, $c_{22} = 3$, then $\text{Gini}(1) = 1 - (0.6^2 + 0.4^2) = 0.48$ and $\text{Gini}(2) = 1 - (0.7^2 + 0.3^2) = 0.42$. Node 2 is thus 'purer' according to the Gini index.

Splitting Data

As the instances are split into smaller subsets then the impurity measures for the subsets should also reduce. This is a result of the subsets becoming purer as the split continues. Ultimately

a subset will contain instances from the same class only and further splitting is meaningless. In addition, the larger the reduction in the impurity resulting from splitting, the more discriminating the test condition. These conditions lead to a splitting function which is the difference between the weighted average impurity measure of the nodes before (parent nodes) and after (children nodes) splitting. The splitting function is given by:

$$\text{Splitting function } \Delta = I(\text{parent}) - I(\text{children}) \quad (\text{B.4})$$

in which $I(.)$ is the impurity measure given, for example by (B.3). The splitting function is called the 'information gain' if entropy is used in (B.4). The attribute that produces the largest splitting function which minimizes the impurity or maximizes the information gain should be used as the next test condition.

Missing Values

Missing attribute values are a common problem in many data sets. Although instances that contain missing attribute values could be ignored, this may significantly reduce the data set and is therefore undesirable. One solution to this problem is to replace a missing attribute with the most common value among instances from the same class. An alternative, more complex procedure is to assign a probability to each of the possible values of x_k rather than simply assigning the most common value to the missing attribute. These probabilities can be estimated from the observed frequencies of the various values of x_k among the instances at a node.

Decision trees handle missing values by keeping track of 'surrogate splitters', which are splitting rules that closely mimic the action of a primary split. In the case of a missing value the best surrogate split can be used to determine which branch to take. In cases where no surrogate is available the branch with the largest number of instances is taken.

Masking

The phenomenon of one variable hiding the significance of another is known as 'masking'. Masking can be revealed in two ways; either by rebuilding the tree without the primary splitter, or by computing an importance measure which is the sum of the improvement measures at each node for each variable in its role as a surrogate splitter. If a variable scores high in the importance measure without showing up in the tree it is an indication that it has been masked by a primary splitter but would be able to contribute in case the primary splitter wasn't. These issues become important if some variables are easy to collect while others are difficult.

Appendix C

Microsimulation

Microsimulation models, such as MarketInfo, are generally used for predicting the immediate distributional impacts of government policy change to various communities. The key use of microsimulation models has been to predict the immediate impacts of changes in tax and social security policy. The models start with microdata i.e. ‘low-level’ population data; typically the records of individuals from a national sample survey conducted by a national bureau of statistics. This is one of the most important advantages of large scale microsimulation models. Being based on unit records, it is possible to examine the effects of policy changes for narrowly defined ranges of individuals or demographic groups. Further, they can mirror the heterogeneity in the population as revealed in the large household surveys. The models can also be used to forecast the outcomes of policy changes and *what if* scenarios, that is, the results can describe what, under specified conditions, may happen to particular individuals and groups [11].

Microsimulation modelling research findings are generally based on estimated characteristics of the population. Such estimates are usually derived from the application of microsimulation modelling techniques to microdata based on sample surveys. These estimates may be different from the actual characteristics of the population because of sampling and non-sampling errors in the microdata and because of the assumptions underlying the modelling techniques. The microdata do not contain any information that enables identification of the individuals or families to which they refer.

Reweighting

As surveys are based on a sample of the population, each individual or household within the survey must be weighted to represent the total number of that type of household or person within the population (also called ‘grossing up’). In a similar manner, the same sample can be reweighted in order to be representative of a population within a small area.

Reweighting can be created by an integer or fractional approach. In the integer approach, the appropriate number and type of households from the sample survey that best represent the particular small area of interest are selected and each assigned a weight of one. Alternatively, all households within the sample can be given a small fractional weight so that the sum of all weights equals the population in the selected area and the sum of the fractional individuals or households best matches the characteristic profile of the area. The fractional method avoids the suspicion of a breach in confidentiality and may provide better data quality.

The SYNAGI and MarketInfo reweighting approach

The SYNAGI (SYNthetic Australian Geo-demographic Information) model employs the fractional reweighting method by using optimisation algorithms which iteratively generate a set of

Characteristic
Age
Total Individual Income
Marital Status
Labour force status and gender (cross-tabulated)
Country of birth
Occupation
Family Type
Student status
Level of highest qualification
Age and income market segments (cross-tabulated)
Housing type
Housing tenure
Household size
Number of motor vehicles
Level of mortgage repayments
Level of rent payments

Table C.1: Characteristics used in MarketInfo 2001 linkage variables[42, 43]

weights that ‘best-fit’ the profile of the area [11]. The method entails identifying linkage variables (as explained below) in the survey of interest and the Census, recoding these variables to be comparable with each other, and then reweighting the records in the survey in an iterative manner until they create a match for the target variables in the Census for the area of interest.

The MarketInfo model applies this method to the ABS 1998/99 Household Expenditure Survey (HES) and the 1996 ABS Census of Population and Housing Basic Community Profile (BCP). The MarketInfo techniques have an established reputation in commercial applications – such as market identification – and have also increasingly been used in socio-economic applications [42, 43].

Linkage variables

Synthetic populations can be generated using any selection of variables that are common to both the Census and the survey data. For example, age and sex could be used to reweight the survey. This would result in synthetic populations of each area, such as a Census District, having the correct age and sex distributions. However, it is unlikely that this would result in reliable distributions of the other socio-economic variables in the survey. More importantly, it is unlikely that the synthetic populations would reliably represent vulnerability in each Census District, as age and sex alone are not good predictors of vulnerability. Therefore, including more socio-economic variables in the reweighting procedure is essential in improving the representativeness of the synthetic populations. To improve the estimates further, certain linkage variables have been chosen specifically for the application. For example, of the 15 social vulnerability indicators, some provide a ‘link’ between the census and survey data, ensuring that there are sufficient cross-tabulations of particular variables that are deemed to be important drivers. Some of the linkage variables that the MarketInfo model uses, which are common to both the HES and BCP surveys and which have been drawn upon in this study, are listed in Table C.1.

There are 68 linkage variables used in MarketInfo derived from the household and personal characteristics that are common to both the HES and BCP surveys. The characteristics are mostly single variables, but include some cross-correlated ones and are listed in C.1.

Optimisation and convergence

The optimisation process used in reweighting consists of three linked convergence algorithms that marginally change the values of the weights. This subsequently evaluates the change in value of the socio-economic linkage variables compared with the Census variables. The first of the three algorithms focuses on each target sequentially with the aim of adjusting the weights and the resultant synthetic population closer to all target variables. This is the common technique known as 'iterative proportional fitting'. These weights are then used in the second algorithm that has a wider focus, evaluating the overall fit to all targets. The final algorithm involves a multi-dimensional search for convergence by changing a pair of household weights in a positive and/or negative direction.

The evaluation measure is the absolute residual between each of the reweighted survey values and the Census BCP target variables. In general terms, if the change in weights improves the fit to the Census targets the weights are accepted, otherwise the change in weights is rejected. This process is undertaken many times until the reweighted HES values converge on the Census BCP variables. The objective is to minimise the difference. When population estimates are derived from weights from a convergent run, the synthetic populations of each area have the correct distributions for each of the linkage variables i.e. the survey values resemble the Census values.

MarketInfo Census District weights

MarketInfo 2001 updates the small area data from the 1996 Australian census to 2001 and links it with the data from the ABS 1998/99 HES to produce a set of weights for the HES for each Census District for 2001. There are 34,410 Census Districts across Australia, with an average size of approximately 200 households.

The linkage variables used to derive the weights must be appropriate for the derived variable of interest. The five socio-economic variables, which appear in the decision rules that assign a person to the high vulnerability class, are all either linkage variables (income, tenure type, age, household type), as shown in Table C.1, or considered closely associated with them (house insurance). Consequently, the synthetic data population counts for these variables are considered reliable.

A set of weights (one for each of the HES survey records) has been generated by Market-Info for each Census District in the study area. The meaning of a weight w for a particular survey record for a particular Census District is that survey record represents w people or households in the Census District. Hence if that survey respondent is assigned to the high vulnerability class, w synthetic people in the Census District will be counted in that class.

Further detailed information on the microsimulation models used in this study and other applications can be found at the NATSEM website [47].

Appendix D

Validation

The decision rules and their interpretation, when applied to the Perth case study, are not without their limitations and these will be discussed later. However, the main question surrounding any application to a real spatial community such as Perth is, how do we know how valid this is? Unless an actual hazard impacts the community and all essential data is recorded, we will never know for certain how vulnerable the community is. Yet we can endeavour to compare or validate the mapped scenarios by employing other methodologies or discussions with the Local Government area in the Perth case study and these are detailed below.

Methodology validation

The vulnerability estimates have relied upon data that has been modelled and manipulated at various stages of the four-step methodology, therefore it is very difficult to assess the results against another methodology in any comparative analysis. Yet, the authors also acknowledge that validating the methodology introduced in this report is essential in methodology development. In their study of vulnerability to food access, Yohannes and Webb note that two methods have been explored for the purpose of 'validation' [69]. The first involves working closely with local experts whose comments can help improve the reliability of subjective data uses [69]. The second is to analyse the indicators of vulnerability against 'observed benchmark measures of crisis', i.e. an actual event [69].

The Perth case study area has not experienced a 'measure of crisis' for which sufficient data was collected, so the second method suggested by Yohannes and Webb [69] is not possible. However, the community vulnerability methodology employed by the Cities Project [28] has been externally reviewed, as discussed in Chapter One, and so will be used as a point of comparison for this study. Therefore, by applying the Cities Project methodology for the Wanneroo study area as a reference for discussion and obtaining the views of local 'experts' will be used in order to at least explore avenues for validating the four-step methodology introduced in this report.

1: Comparison with local experts' knowledge

The Wanneroo and Swan Local Government areas are active in their role as managers of hazards and emergencies within their communities. One of the hurdles faced by Western Australian local governments is that current state legislation does not mandate local government to develop and implement risk management strategies that incorporate natural hazards. Hence, limited funding and resources are available for developing comprehensive natural hazard and disaster risk management strategies.

For this study, maps of the Wanneroo case study area were posted to community workers in the Wanneroo Local Government. The maps asked the workers to indicate suburbs/areas

of high, medium or low vulnerability by circling the area on the map. From working with and in the community everyday, the community workers' develop a knowledge of the areas they consider have 'at risk' or vulnerable people. While this is not related directly to natural hazards, it was considered worth investigating whether or not the decision rules correlated with the local government's perception of 'at risk' people.

While some maps were returned, with comments, it was difficult to correlate the areas considered at risk, due to the 'at risk' people being quite different from the decision rules. For example, 'at risk' may refer to elderly people with low mobility, homeless teenagers and low socio-economic groups, yet these groups have not, in this study, been assessed for their vulnerability to natural hazards. Therefore, it is worth noting that groups generally considered 'at risk' in the community may be further affected by natural hazards, but have not been specifically focused on in this study. Hence, a comparison between the local perception of general risk and a study of specific hazard risk is difficult to ascertain in this situation.

Social Capital

The concept of 'at risk' within a community is explored in papers and studies on 'Social Capital'. Social capital refers to the often intangible elements that increase our everyday resilience in life, including trust in government, feeling safe in your neighbourhood, engaging with your neighbours, volunteering, tolerance and maintaining strong social networks [65]. While a discussion on social capital is beyond the scope of this study, it is strongly related to the broader and more intangible indicators of vulnerability, some of which were outlined as complex indicators in Table 1.5. Studies into the non-quantitative aspects of social vulnerability focus on some of the intangible factors, now referred to widely as social capital, that broadly encompass safety and security. Hence they are highly relevant to studies investigating social vulnerability to natural hazards and can be explored in further detail in Buckle [14], Pelling [58] and Phillips [59] for social vulnerability issues and natural hazards, or the ABS [55] and Harvard University [49] for specific social capital reading.

Bibliography

- [1] L. Anderson-Berry. Cyclone Rosita April 20 2000: Post Disaster Report. Technical report, James Cook University Centre for Disaster Studies, Queensland, Australia, 2000.
- [2] E. M. Andrews and S. B. Withery. *Social Indicators of Well Being*. Plenum Press, New York, USA, 1976.
- [3] M. Ankerst, M. Ester, and H. P. Kriegel. Towards and Effective Cooperation of the User and the Computer for Classification. In *SIGKDD, International Conference on Knowledge Discovery and Data Mining, KDD-2000, Association for Computing Machinery*, Massachusetts, USA, 2000.
- [4] Federal Emergency Management Authority. HAZUS 99 Technical Manual. Technical report, Federal Emergency Management Authority Agency, (FEMA), United States Government, Washington, USA, 1999.
- [5] World Bank. Poverty and Vulnerability and Public Action. <http://www.worldbank.org/wbi/socialprotection/africa/parisone/pdfpaper/%conceptseeng.pdf>, 2002.
- [6] M. J. A. Berry and G. S. Linoff. *Data Mining Techniques for Marketing, Sales and Customer Support*. John Wiley and Sons, London, UK, 1997.
- [7] P. Blaikie, T. Cannon, I. Davis, and B. Wisner. *At Risk: Natural Hazards, People's Vulnerability and Disasters*. Routledge, London, UK, 1994.
- [8] R. Bolin and L. Stanford. Shelter, Housing and Recovery: A Comparison of U.S. Disasters. *Disasters*, 15(1):24–34, 1991.
- [9] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and Regression Trees*. Statistics/Probability Series. Wadsworth Publishing Company, California, USA, 1984.
- [10] A. H. Briggs, A. E. Ades, and M. J. Price. Probabilistic Sensitivity Analysis for Decision Trees with Multiple Branches: Use of the Dirichlet Distribution in a Bayesian Framework. *Medical Decision Making*, 23(4):341–350, 2003.
- [11] L. Brown and A. Harding. Social Modelling and Public Policy: What is microsimulation modelling and how is it being used? In *"It's not just about the Dollars" Forum, 19 July 2002*, Sydney, Australia, 2002.

- [12] P. Buckle. A framework for Assessing Vulnerability. *The Australian Journal of Emergency Management*, 13(4):21–26, 1995.
- [13] P. Buckle. Re-defining Community and Vulnerability in the context of Emergency Management. *Australian Journal of Emergency Management*, 13(1):21–26, 1998.
- [14] P. Buckle. Assessing resilience and vulnerability in the context of emergencies: Guidelines. Technical report, Department of Human Services, Victoria, Melbourne, Australia, 2000.
- [15] S. B. Chang. Transportation Performance, Disaster Vulnerability and Long-term effects of Earthquakes. In *Second EuroConference on Global Change and Catastrophe Risk Management, Austria, July 6-9, 2000*, Laxenburg, Austria, 2000.
- [16] C. W. Cobb and C. Rixford. Lessons learned from the history of social indicators. <http://www.rprogress.org>, 1998.
- [17] Productivity Commission and Institute of Applied Economic and Social Research. Policy implications of the ageing of australia's population. Technical report, Productivity Commission, Canberra, Australia, 1999.
- [18] Private Health Insurance Administration Council. Quarterly statistics, september 1997. <http://www.phiac.gov.au/statistics>.
- [19] D. Crichton. *The Risk Triangle in Natural Disaster Management*. Tudor Rose, Leicester, UK, 1999.
- [20] R Davidson. *An Urban Earthquake Disaster Risk Index*. PhD thesis, Department of Civil Engineering, Stanford University, USA, California, USA, 1997.
- [21] *How vulnerable is your community to a natural hazard? Using synthetic estimation to produce spatial estimates of vulnerability*, 2003.
- [22] B. Drottz-Sjberg. Exposure to Risk and Trust in Information; Implications for the Credibility of Risk Communication. *The Australasian Journal of Disaster*, 2, 2000.
- [23] P. F. Drucker. The Discipline of Innovation. *Harvard Business Review*, pages 3–8, 1988.
- [24] N. B. Ferrier. Creating a Safer City: A Comprehensive Risk Assessment for the City of Toronto. Technical report, Toronto Emergency Medical Services, Toronto, Canada, 2000.
- [25] B. Fischhoff, P. Lichtenstein, S. Read, and B. Combs. How Safe is Safe Enough? A Psychometric Study of attitudes towards technological risks and benefits. *Policy Sciences*, 9:127–152, 1978.
- [26] M. Fordham. The Place of Gender in Earthquake vulnerability and mitigation. In *Second Euro Conference on Global Change and Catastrophic Risk Management - Earthquake Risks in Europe, Austria, Laxenburg, Austria, 2000*.

- [27] W. J. Frawley, G. Piatetsky-Shapiro, and C. Matheus. "*Knowledge Discovery In Databases: An Overview*" in *Knowledge Discovery In Databases (eds.) G. Piatetsky-Shapiro and W. J. Frawley*. AAAI Press and MIT Press, Massachusetts, USA, 1991.
- [28] K. Granger and M. Hayne. Natural hazards and the risk they pose to South-East Queensland. Technical report, Geoscience Australia, Commonwealth Government of Australia, Canberra, Australia, 2001.
- [29] K. Granger, T. Jones, and G. Scott. Community Risk in Cairns: a multi-hazard risk assessment. Technical report, Geoscience Australia, Commonwealth Government of Australia, Canberra, Australia, 1999.
- [30] A. Harding and A. Szukalska. Facts, figures and suggestions for the future, Income support and Poverty. Technical Report Information Sheet No. 2, Brotherhood of St. Laurence, July 2002.
- [31] D. Harding and M. Hammill. Mercantile Mutual - Melbourne Institute Household Saving Report, September Quarter 2000. Technical report, Melbourne Institute of Applied Economic and Social Research, The University of Melbourne, Victoria, Australia, 2000.
- [32] M. Hayne and D. Gordon. Regional landslide hazard estimation, a GIS/decision tree analysis: Southeast Queensland, Australia. In *Proceedings of the fourteenth southeast Asian geotechnical conference, Hong Kong, 10-14 Dec, 2001*, Hong Kong, 2001.
- [33] A. Heijmans. Vulnerability: A matter of perception. In *International Conference on Vulnerability in Disaster Theory and Practice*, pages 24–34, London, UK, 2001.
- [34] R. Hockley. Organisational change and mourning. In *Address to the Ninth Australian Institute of Training and Development Conference, Sydney, March 1998*, New South Wales, Australia, 1998. Australian Institute of Training and Development.
- [35] J. Hoddinott. Operationalizing household food security in development projects: An introduction. Technical report, International Food Policy Research Institute, Washington, USA, 1999.
- [36] E. Hunt, J. Marin, and P. Stone. *Experiments in Induction*. Academic Press, New York, USA, 1966.
- [37] S. Jasanoff. The political science of risk perception. *Reliability Engineering and System Safety*, 59:91–99, 1998.
- [38] J. Kasperson and R. Kasperson. International Workshop on Vulnerability and Global Environmental Change, Stockholm, May 17-19 2001. In *A workshop summary*, Stockholm Environment Institute, Stockholm, Sweden, 2001.

- [39] D. King and C. MacGregor. Using social indicators to measure community vulnerability to natural hazards. *Australian Journal of Emergency Management*, 15(3):52–57, 2000.
- [40] E. Krumpe. Monitoring Human Impacts: Criteria to Guide Selection of indicators. <http://www.cnr.uidaho.edu/rrt496/INDICATORS.pdf>.
- [41] K. Lambert. A Hurricane Disaster Risk Index. Master’s thesis, Department of Civil Engineering, University of North Carolina, North Carolina, USA, 2000.
- [42] R. Lloyd, J. Given, and O. Hellwig. The Digital Divide: Some Explanations. *Agenda*, 7(4):354–358, 2000.
- [43] R. Lloyd, A. Harding, and O. Hellwig. Regional Divide? A study of incomes in regional Australia. *Australasian Journal of Regional Studies*, 6(3):271–292, 2001.
- [44] M. Middelmann and K. Granger. Community Risk in Mackay: A Multi-hazard Risk Assessment. Technical report, Geoscience Australia, Commonwealth Government of Australia, Canberra, Australia, 2000.
- [45] B. H. Morrow. Identifying and Mapping Community Vulnerability. *Disasters*, 23(1):1–18, 1999.
- [46] S. Murthy, S. Kasif, and S. Salzberg. A system for induction of oblique decision trees. *Journal of Artificial Intelligence Research*, 2:1–32, 1994.
- [47] National Centre for Social NATSEM and University of Canberra Economic Modelling. <http://www.natsem.canberra.edu.au/research/marketinfo/2001full-marketinfo.htm>.
- [48] W. L. Neuman. *Social Research Methods: Qualitative and Quantitative Approaches*. Allyn and Bacon, New York, USA, 1997.
- [49] John F. Kennedy School of Government at Harvard University. <http://www.ksg.harvard.edu/saguaro/>.
- [50] Australian Institute of Health and Welfare. Disability support services. <http://www.aihw.gov.au/publications>.
- [51] Department of Regional Development, Environment Executive Secretariat for Economic, and Social Affairs Organisation of American States. Primer on Natural Hazard Management in Integrated Regional Development Planning. Technical report, Organisation of American States, Washington, USA, 1991.
- [52] Australian Bureau of Statistics. Cdata96, census 1996. Technical Report Catalogue No 2019.0.30.001, Australian Bureau of Statistics, Commonwealth of Australia, Canberra, Australia, 1996.
- [53] Australian Bureau of Statistics. Basic community profile, 1996 census of population and housing. Technical Report Catalogue No 2020.0, Australian Bureau of Statistics, Commonwealth of Australia, Canberra, Australia, 1997.

- [54] Australian Bureau of Statistics. Catalogue no.6544.0.30.001. 1998-99 household expenditure survey australia confidentialised unit record file (curf) technical paper, 2002.
- [55] Australian Bureau of Statistics. http://www.abs.gov.au/themes/Social_Capital.
- [56] Australian Bureau of Statistics. Ausstats. <http://www.abs.gov.au/AusStats>.
- [57] D Paton. Editorial; Disaster and Trauma Studies: Developing an Australasian perspective. *The Australasian Journal of Disaster and Trauma Studies*, 1, 1997.
- [58] M. Pelling. *The Vulnerability of Cities: Natural Disasters and Social Resilience*. Earthscan Publications, London, UK, 2003.
- [59] B. Phillips. *Holistic Disaster Recovery: Ideas for building local sustainability after a natural disaster*, chapter Chapter 6: Social and Intergenerational Equity. Natural Hazards Research Center, University of Colorado, Colorado, USA, 2001.
- [60] J. Quinlan. *C4.5: Programs for Machine Learning*. Morgan-Kaufmann Publishers, California, USA, 1993.
- [61] R. Rossi and K. Gilmartin. *The handbook of social indicators: Sources, characteristics and analysis*. Garland STPM, New York, USA, 1980.
- [62] H. Schuman and S. Presser. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. Sage Publications, London, 1996.
- [63] P. Slovic. *The Perception of Risk*. Earthscan Publications, London, UK, 2000.
- [64] L. Stephen and T. Downing. Getting the scale right: A comparison of analytical methods for vulnerability assessment and household-level targeting. *Disasters*, 25(2):113–135, 2001.
- [65] W. Stone. Social capital and social security: Lessons from research. *Family Matters*, 57:10–14, 2000.
- [66] G. F. White and H. J. Heinz. *The Hidden costs of Coastal Hazards*. H. John Heinz III Center for Science, Economics and the Environment. Island Press, Washington, USA, 2000.
- [67] O. Wilelmi and D. Wilhite. Assessing vulnerability to agricultural drought: A Nebraska case study. *Natural Hazards*, 25(1):37–58, 2002.
- [68] P. Wright. Knowledge discovery in databses: Tools and techniques. <http://www.acm.org/crossroads/xrds5-2/kdd.html#6>, 1998.
- [69] Y. Yohannes and P. Webb. Classification and Regression Trees, CART©: A User Manual for identifying indicators of vulnerability to famine and chronic food insecurity. Technical report, International Food Policy Research Institute, Washington, USA, 1999.

Index

- attribute, 77
 - categorical, 77
 - categorical, 77
 - continuous, 77
 - masking, 80
 - missing, 80
- attribute selection criteria, 79
- Australian Bureau of Statistics, 9
 - HES, 51, 82
 - SEIFA, 9
- branch, 35, 77, 78
- C4.5, 78
- CART, 78, 79
- circle segment technique, 33, 34
- class label, 77
- class labels, 77
- classification model, 77
- classification technique, 78
- concept, 77
- cross-validation
 - k -fold, 79
- data
 - collection, 20
 - continuous, 76
 - discrete, 76
 - missing, 76
 - noise, 76
 - training, 76
- data mining, 33
- data mining techniques, 76
- data set, 77
- decision rules, 37
- decision tree, vii, 6, 69, 77
- decision tree induction, 76
- Earthquake Disaster Risk Index, 9–13
- entropy, 80
- events, 77
- examples, 77
- experts, 29, 69
- exposure, 3
- Geoscience Australia
 - Cities Project, 9, 13, 47
- Gini index, 79
- hazard, 3
 - flood, 52
- HAZUS, 14
- Henderson Poverty Line, 28
- importance measure, 80
- impurity, 79
 - measures
 - Gini, 79
- indicator
 - selection, 15
- indicators
 - '5 esses', 13
 - age, 12, 18, 21, 50, 83
 - car ownership, 18
 - debt and savings, 26
 - disability, 18, 25
 - employment, 23
 - family status, 51
 - gender, 12, 27, 50
 - health insurance, 18, 25
 - history, 8
 - house and/or contents insurance, 83
 - household types, 24
 - householdtypes, 83
 - income, 18, 22, 28, 50, 83
 - injury classification, 52
 - language, 24
 - residence damage, 52
 - residence type, 22, 23
 - selection criteria, 14

- tenure, 83
- transportation, 27
- vulnerability, 9, 16, 51, 82
- weights, 13, 18
- individuals
 - hypothetical, 28
- information gain, 80
- instance, 77
 - space, 77
- internal node, 77
- knowledge discovery, 33
- Knowledge Discovery in Databases, 76
- linkage variables, 82
- masking, 80
- maximal tree, **77**, 78
- microsimulation, 81
- mitigation, 4
- NATSEM
 - MarketInfo, 48, 81, 83
 - SYNAGI, 81
- natural hazards
 - Australia, 29
- node
 - leaf, 35
 - root, 35
 - children, 80
 - internal, 77
 - leaf, 78
 - parent, 80
 - root, 35, 77, 78
- observations, 77
- OC1, 78
- optimisation, 83
- perception based classification, 33, 34
- psychometric paradigm, 19
- purity, 79
- records, 77
- reweighting, 81, 82
- risk
 - Earthquake Disaster Risk Index, 14
 - relative risk rank, 12
 - assessment
 - Perth, Western Australia, 47
 - South-East Queensland, 13
 - assessment methodology, 12, 13, 28
 - Earthquake Disaster Risk Index, 9
 - Hurricane Disaster Risk Index, 13, 19
 - perception, 19
 - risk assessment
 - City of Toronto, 13
 - risk index, 12
 - risk perception questionnaire, vii, 6, 20
 - design, 28
 - results, 30
- rule, 78
- scenarios, 52, 53
- splitting
 - oblique, 36
 - orthogonal, 36
- splitting function, 80
- subset, 77
- surrogate splitters, 80
- synthetic estimation, 46, 48, 53, 70
- target variable, 77
- total risk pyramid, 3
- under-fitting, 78
- United Nations, 8
- variable importance, 42
- visualisation, **33**
- vulnerability, 3
 - definition, 9
 - indicators, 13, 16, 51, 82
 - measuring, 9
- vulnerability maps, 54
- World Bank, 8