A Statistical and Geographical Analysis of Workplace Accidents in England and Wales

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

by

Linda Woods

May 2017

Acknowledgements

It has only been through the help and support of numerous people that this research has come to fruition, and to these people I am extremely grateful. First and foremost, I thank my supervisor, Professor Chris Lloyd, who has supported me throughout my studies. I have been incredibly fortunate to have such a knowledgeable, patient and encouraging supervisor who has helped guide me over the course of this research.

Many thanks also to Professor Alex Singleton who convinced me to pursue my doctoral degree and provided many opportunities over the years, including presenting at two international conferences. I also extend this thanks to Professor Chris Brunsdon for contributing invaluable technical advice and support.

Funding from the Economic and Social Research Council and support from the Health and Safety Executive, as the CASE award partner, made this research a reality, and to both I am hugely grateful.

A special thanks goes to my fellow colleagues at the University of Liverpool for their support and advice, particularly during the write up stages of the thesis.

The support from my friends and family has been immeasurable. A huge thanks in particular goes to my mother-in-law, Wendy Woods, for proof reading every word I have written. My Dad has been a pillar of support throughout my studies and I cannot thank him enough for his wisdom and guidance. My mum, whom we miss dearly, deserves an enormous thank you for providing me with endless opportunities and always encouraging me to succeed.

The final thank you has to be to my husband, John Woods. His unwavering support, even through the most stressful times, has provided me with the momentum to keep going, for which I will be eternally grateful.

Abstract

This research investigates the risk-factors associated with workplace accidents by analysing data generated by Reporting of Injuries, Diseases and Dangerous Occurrences Regulations (RIDDOR), a framework in which the Health and Safety Executive (HSE) collects reports of workplace accidents and injuries. It is reported by the HSE that work-related accidents are a significant problem facing today's workforce. Unfortunately, however, occupational health and safety have largely been under researched. Little work has previously been carried out surrounding the key determinants of workplace accidents, how these determinants might vary geographically, and whether physical conditions, such as the weather and levels of daylight, might impact levels of occupational health and safety. This research therefore seeks to address these gaps by examining the socio-economic and physical determinants of workplace accidents and injuries, and examines whether the relationships between these risk-factors and accident rates vary geographically and seasonally. Three distinct methods are utilised in analysing the RIDDOR data. These methods include: a global regression analysis based on a set of socio-economic characteristics of workers, a Geographically Weighted Regression analysis of these characteristics on three case study regions: North West England, North East England and London, and text mining, in the form of topic modelling the free-text fields of descriptions of incidents reported under RIDDOR. The key results reveal that age and socio-economic class are influencing factors of workplace accidents. Occupation type is also found to have an effect on workplace accident risk, with workers in low skilled jobs associated with an increased risk of having a work-related injury compared to workers in highly skilled occupations. The relationships between these risk-factors and accident rates have been found to vary geographically, with risk-factors appearing to have a stronger relationship with workplace accidents in particular seasons compared to others. Policy recommendations are formulated to equip the HSE with the knowledge of the key high-risk groups within the workplace population so that preventative measures can be established to reduce the rates of workplace accidents in the future.

Contents

List of Figures v

List of Tables vii

Chapter 1 1
Introduction 1

1.1 Introduction 1
1.2 Aims and Objectives 5
1.3 Outline of Thesis Structure 8

Chapter 2 11
Literature Review 11

2.1 Introduction 11
2.2 Social Factors 12
2.2.1 The Individual 12
2.2.2 The Job/Organisation 22
2.3 Environmental Factors 30
2.3.1 Weather Conditions 30
2.3.2 Daylight Hours 32
2.4 Geographic Variations 35
2.5 Background to Health and Safety in Great Britain 38
2.5.1 Health and Safety Policy and Legislation 38
2.5.2 Health and Safety Statistics 40
2.5.3 Future Outlook of Health and Safety 43
2.6 Summary 45

Chapter 3 49
Data and Methods Review 49

3.1 Introduction 49
3.2 Data Review 50
3.2.1 Data Context: Introduction to RIDDOR 50
3.2.2 The RIDDOR Sample Data 53
3.2.3 Census Data 61
3.3 Methods 64
3.3.1 Introduction 64
3.3.2 Use of Software 64
3.3.3 Global Regression Modelling- Generalised Linear Models 66
3.3.4 Local Regression Modelling- Geographically Weighted Regression 70
3.3.5 Text Mining and Topic Modelling 76
3.4 Summary 79

Chapter 4 81
Exploratory Analysis of the RIDDOR Data 81

4.1 Introduction 81
4.2 Background Statistics on RIDDOR Records 82
4.3 Characteristics of Injured Employees 83
4.4 Causes of Accidents 87
4.5 Injuries sustained to Body 89
4.6 Locations of Accidents 93
4.7 Time and Date of Accidents 99
4.8 Conclusion 109

Chapter 5 ..... **111**
A Global Regression Analysis of Workplace Accidents **111**

5.1 Introduction 111
5.2 Background and Methods 112
5.3 Comparing Regression Models 114
5.4 Refining the Global Regression Model 119
5.5 Conclusion 129

Chapter 6 ..... **131**
A Local Regression Analysis of Workplace Accidents **131**

6.1 Introduction 131
6.2 Background and Methods 132
6.3 Global Regression Model by Case Study Area 135
6.4 Geographically Weighted Regression 139
6.4.1 London GWR 139
6.4.2 North West GWR 146
6.4.3 North East GWR 150
6.4.4 Seasonal GWR 155
6.5 Conclusion 169

Chapter 7 ..... **173**
Exploration of the Free-Text Fields of the RIDDOR Data **173**

7.1 Introduction 173
7.2 Background and Methods 174
7.3 Text Mining 176
7.3.1 England and Wales 176
7.3.2 Case Study Regions 178
7.4 Topic Modelling: Latent Dirichlet Allocation (LDA) 183
7.4.1 England and Wales 183
7.4.2 Case Study Regions 184
7.4.3 Seasonal Differences 186

Chapter 8 ..... **197**
Discussion **197**

8.1 Introduction 197
8.2 Results 198
8.2.1 Socio-economic characteristics associated with workplace accidents 198
8.2.2 Physical factors associated with workplace accidents 204
8.2.3 Geographic variations in risk factors associated with workplace accidents 207
8.2.4 Spatio-temporal variations in risk factors associated with workplace accidents 211
8.3 Policy Recommendations 214
8.4 Review of Data and Methods 219
8.5 Future Work 223
8.6 Summary 224

Chapter 9 ..... **225**
Conclusions **225**

9.1 Introduction 225
9.2 Research Summary 225
9.3 Final Remarks 234
Bibliography **235**

Appendix **249**

Additional GWR Maps 249

1. London..... 249
2. North West..... 250
3. North East..... 251
4. Seasonal..... 252

List of Figures

Figure 4.1: Accidents recorded by year....82
Figure 4.2: Percentage of accidents by age out of employed working population....84
Figure 4.3: Industries (SIC 2011) with the highest 10 reported accident rates....85
Figure 4.4: Rates of workplace accidents by occupational group (SOC 2010)....87
Figure 4.5: Types of accidents recorded under RIDDOR....88
Figure 4.6: Rates of severity of injuries by age group out of working population....90
Figure 4.7: Fatalities by age group....91
Figure 4.8: Body parts injured....93
Figure 4.9: Accident rate by Local Authority District workplace population....95
Figure 4.10: Accidents per day....100
Figure 4.11: Accidents per month....102
Figure 4.12: Mosaic plot of accidents by month and region....104
Figure 4.13: Hours of daylight for shortest and longest days of 2012....105
Figure 4.14: Calendar heat map of workplace accidents....107
Figure 4.15: Calendar heat map of workplace accidents involving slips or trips....108
Figure 5.1: Accident Rate by Number of Workplace Zones....114
Figure 6.1: Global model residuals map for England and Wales by Workplace Zone....133
Figure 6.2: Location map for London and residuals and pseudo R^2 Maps for London GWR model....141
Figure 6.3: London SOC1 Managers, directors and senior officials- percentage of workers & GWR coefficients....143
Figure 6.4: London SOC8 Process, plant and machine operators- percentage of workers & GWR coefficients....144
Figure 6.5: London SOC 9 Elementary occupations- percentage of workers & GWR coefficients....145
Figure 6.6: Location map for North West England and residuals and pseudo R^2 maps for North West England GWR model....147
Figure 6.7: North West SOC 1 Managers, directors and senior officials- percentage of workers & GWR coefficients....148
Figure 6.8: North West SOC 9 Elementary occupations - percentage of workers & GWR coefficients....149
Figure 6.9: North West NS-SEC 7 Routine occupations - percentage of workers & GWR coefficients....150
Figure 6.10: Location map for the North East and residuals and pseudo R^2 maps for the North East GWR model....152
Figure 6.11: North East SOC 1 Managers, directors and senior officials- percentage of workers & GWR coefficients....153
Figure 6.12: North East SOC 8 Process, plant and machine operatives- percentage of workers & GWR coefficients....154
Figure 6.13: North East SOC9 Elementary occupations- percentage of workers & GWR coefficients....155
Figure 6.14: Location map and summer and winter residuals and pseudo R^2 maps for London....156
Figure 6.15: Summer and Winter SOC 1 Managers, directors and senior officials GWR maps for London and map of difference of GWR coefficient estimates....158
Figure 6.16: Summer and Winter SOC 9 Elementary occupations GWR maps for London and map of difference of GWR coefficient estimates....159
Figure 6.17: Location map and Summer and Winter residuals and pseudo R^2 maps for the North West....160

Figure 6.18: Summer and Winter SOC 1 Managers, directors and senior officials GWR maps for the North West and map of difference of GWR coefficient estimates 162
Figure 6.19: Summer and Winter SOC 9 Elementary occupations GWR maps for the North West and map of difference of GWR coefficient estimates.....163
Figure 6.20: Location map and Summer and Winter residuals and pseudo R^2 maps for the North East 165
Figure 6.21: Summer and Winter SOC 1 Managers, directors and senior officials GWR maps for the North East and map of difference of GWR coefficient estimates 167
Figure 6.22: Summer and Winter SOC 9 Elementary occupations GWR maps for the North East and map of difference of GWR coefficient estimates.....168
Figure 7.1: Top 10 words with highest frequencies.....177
Figure 7.2: Word Cloud of terms appearing at least 2,500 times 178
Figure 7.3: Top 10 words with highest frequencies- London.....179
Figure 7.4: Word Cloud of terms appearing at least 200 times- London.....179
Figure 7.5: Top 10 words with highest frequencies- North West.....180
Figure 7.6: Word Cloud of terms appearing at least 200 times- North West.....181
Figure 7.7: Top 10 words with highest frequencies- North East.....182
Figure 7.8: Word Cloud of terms appearing at least 200 times- North East.....182
Figure 7.9: Frequency of hot/cold terms for England and Wales 191
Figure 7.10: Frequency of hot/cold terms for London 192
Figure 7.11: Frequency of hot/cold terms for North West.....193
Figure 7.12: Frequency of hot/cold terms for North East 194

List of Tables

Table 2.1: SOC major groups and skill levels 23
Table 3.1: Variables contained within RIDDOR dataset 56
Table 3.2: 2011 Census data variables 62
Table 3.3: Kernel types in GWR4 75
Table 4.1: Status of injured person 83
Table 4.2: Industries in which men have the highest rates of accidents 86
Table 4.3: Industries in which women have the highest rates of accidents 86
Table 4.4: Percentage of accidents by material agent (Top 5) 89
Table 4.5: Severity of injuries by gender 91
Table 4.6: Nature of injury 92
Table 4.7: Accident rates by region 94
Table 4.8: The 10 LADs with the highest accident rates 95
Table 4.9: SICs with highest accident rates for North Warwickshire 96
Table 4.10: SICs with highest accident rates for Corby 97
Table 4.11: The 10 Districts with the lowest accident rates 98
Table 4.12: SICs with highest accident rates for City of London 99
Table 4.13: Percentage of accidents occurring on a Monday by SIC group out of all other days of the week 101
Table 4.14: Accidents per season 102
Table 5.1: Poisson Regression Model 1: Males 116
Table 5.2: Quasi-Poisson Regression Model 1: Males 116
Table 5.3: Negative Binomial Regression Model 1: Males 116
Table 5.4: Poisson Regression Model 2: Males + Age 16 to 25 118
Table 5.5: Quasi-Poisson Regression Model 2: Males + Age 16 to 25 118
Table 5.6: Negative Binomial Regression Model 2: Males + Age 16 to 25 118
Table 5.7: Model 3: Males + Age 16 to 25 + Age 56 to 65 120
Table 5.8: Model 4: Males + Age 16 to 25 + Age 56 to 65 + Population Density 120
Table 5.9: Model 5: Males + Age 16-25 + Age 56-65 + Population Density + SIC groups A to S 121
Table 5.10: Model 6: Males + Age 16 to 25 + Age 56 to 65 + Population Density + SOC groups 1, 2 & 4-9 123
Table 5.11: Model 7: Males + Age 16 to 25 + Age 56 to 65 + Population Density + SOC 1 Managers, Directors and Senior Officials+ SOC 8 Process, Plant and machine Operatives + SOC 9 Elementary Occupations + NS-SEC7 Routine Occupations 125
Table 5.12: Correlation Matrix 128
Table 6.1: Global regression model for England and Wales 136
Table 6.2: Global regression model for London 136
Table 6.3: Global regression model for North West England 137
Table 6.4: Global regression model for North East England 137
Table 6.5: Summary statistics for London GWR model 140
Table 6.6: Summary statistics for the North West GWR model 146
Table 6.7: Summary statistics for the North East GWR model 151
Table 7.1: Top 10 topics out of 200 topics for England and Wales 184
Table 7.2: Top 10 topics out of 50 topics for London 184
Table 7.3: Top 10 topics out of 50 topics for North West England 185
Table 7.4: Top 10 topics out of 50 topics for North East 186
Table 7.5: Top 10 topics out of 50 topics for England and Wales, April 187
Table 7.6: Top 10 topics out of 50 topics for England and Wales, August 187
Table 7.7: Top 10 topics out of 40 topics for London, April 188

Table 7.8: Top 10 topics out of 40 topics for London, August.....188
Table 7.9: Top 10 topics out of 35 topics for North West, April.....189
Table 7.10: Top 10 topics out of 35 topics for North West, August.....189
Table 7.11: Top 10 topics out of 30 topics for North East, April.....190
Table 7.12: Top 10 topics out of 30 topics for North East, August.....190
Table 8.1: Areas with the largest GWR coefficient estimates out of each of the three regions 209
Table 8.2: Areas with the largest GWR coefficient estimates out of each of the three regions during summer.....212
Table 8.3: Areas with the largest GWR coefficient estimates out of each of the three regions during winter.....213

Chapter 1

Introduction

1.1 Introduction

The focus of the research is to identify the risk-factors associated with workplace accidents and to identify whether these risk-factors vary geographically and seasonally. Occupational health and safety is largely under researched. This study is concerned with seeking to fill a gap in the literature surrounding the determinants of workplace accidents. The key results from this research will be used to provide policy recommendations to the Health and Safety Executive (HSE) with the aim of helping to reduce workplace accident rates in the future. This chapter provides a context to the research, outlining the aims and key research objectives of the study and sets out the structure of the succeeding chapters.

Recent statistics published by the HSE reveal that between 2014 to 2015 approximately 142 workers were killed at work in Great Britain, approximately 76,000 other injuries were sustained by workers and around 1.2 million workers were suffering from an illness, which they believed was caused or made worse by their job (HSE, 2015a). Despite these staggering statistics, occupational health and safety is largely under researched and this is due to a general lack of access to workplace accident and injury data. In fact, Barling et al. (2002) noted that less than 1% of organisational research in top journals looked primarily into workplace safety. Most of this research covered antecedents of work injuries such as workload, boredom, poor physical health (Frone, 1998) and stress, fatigue and distractions (Legree et al., 2003).

Occupational health and safety policy in Great Britain emerged from the 1800s, and the present system dates back to the introduction of the Health and Safety at Work etc. Act 1974 (HSW Act). The current structure provides a framework for regulating and enforcing health and safety standards so that all workers benefit from a workplace that is properly assessed and controlled. The HSE, created under the HSW Act, has a role of enforcing occupational health and safety standards and its mission is 'the prevention of death, injury and ill health to those at work and those affected by work activities' (HSE, 2015b).

To be an effective regulator of health and safety policy and to support enforcement action whereby employers fail to adequately ensure acceptable health and safety standards, the HSE keeps a record of all workplace accidents, injuries and cases of ill health caused by work. These incidents are reported to the HSE by employers under a system called Reporting of Injuries, Diseases and Dangerous Occurrences Regulations (most commonly referred to as RIDDOR) which was introduced in 1995. The RIDDOR database is one source of information that the HSE utilises to gain an understanding of current health and safety practice. Data are extracted periodically from the RIDDOR database to generate statistical reports and summaries on the number of workplace accidents and injuries occurring across Great Britain's workplace. This helps the HSE to gain an understanding as to how workplaces are performing to standards of health and safety.

Within the HSE's Science Plan, key areas are set out where the HSE aims to improve upon:

A. Leading others to improve health and safety in the workplace
B. Providing an effective regulatory framework
C. Securing compliance with the law
D. Reducing the likelihood of low-frequency, high-impact catastrophic events

(HSE, 2015c)

These four key areas of work contribute to one of the HSE's overall goals stated within its Business Plan: 'continuing to update and refresh guidance so that it is clear

and easy to understand' (HSE, 2015b). One of the ways it aims to do this is through their science and research programme where a large amount of work ranging within science and technology backgrounds is commissioned to contractors and researchers in academia to broaden the knowledge of the causes of workplace accidents and ill-health caused by work.

The HSE recognises a range of socio-economic and environmental risk factors thought to increase the risk of a worker having a workplace accident or injury. These are categorised into three broad interrelated aspects of health and safety which must be explored and addressed in helping to improve standards in the workplace. These categories include: the individual (which includes socio-economic characteristics such as age and gender), the job (which includes industry and occupation types), and the organisation (which includes factors relating to working conditions) (HSE, 2015d).

Research surrounding occupational health and safety is mixed in that contrasting evidence has been found between the associated risk-factors of workplace accidents and injuries. Examples of this include socio-economic characteristics of workers such as age, in which some studies find that the age of a worker has a positive relationship with workplace accidents (Nenonen, 2011; Xiang et al., 2000), whilst others find that a negative relationship exists (Chau et al., 2014; Root, 1981). Evidence also finds that factors relating to a worker's job, such as their occupation and industry type has an effect on a worker's level of health and safety (Chi and Wu, 1997; Sorock et al., 1993). A gap exists in the literature, however, upon the types of causes of accidents across industry types and why particular jobs carry a higher accident risk over others.

It is often the case that work carried out surrounding occupational health and safety considers individual level factors, rather than area-level factors. This is often the case due to the complexities of obtaining multi-level data, and also due to the theories surrounding an individual's health risk being solely based on their own characteristics (Diez-Roux, 1998). The use of area-level statistics to inform public health policy has grown in recent years, however in general, there has been less interest in investigating contextual effects in the occupational injury field (Neff et al., 2008). There are potential benefits of understanding geographic differences between workplace

accident and injury rates. The main one being that they can help to assist in the planning and prevention methods that are more tailored to the needs of specific groups within an area.

Geographic differences in workplace accidents can be prevalent due to areas being made up of having different socio-economic compositions. Examples of this could be that the proportion of younger workers may be larger in certain areas, whilst particular industries may be more concentrated in others. Physical constraints are also known to vary geographically. Whilst it is clear from the relevant literature that environmental factors may play an important part in understanding accident risk (Bridger, 2003; Findikyan and Sells, 1965), the effects of these factors on occupational safety is largely under researched.

The HSE is aware that there are clear gaps within the research surrounding occupational health and safety that need addressing. A requirement of the HSE is to improve health and safety across all workplaces by expanding upon the knowledge of the causes of workplace accidents and ill-health, through its Science Plan. It is this requirement that has prompted the HSE to become a case partner with the University of Liverpool for this research project. The HSE has provided a large sample from the RIDDOR database which contains a wealth of information with regards to workplace accidents, such as the characteristics of the injured workers, when and where the accidents occurred and the types of injuries sustained by workers. In addition to this, a free-text sample from the RIDDOR database was also provided containing descriptions of the incidents completed by the employer of the injured person. The following section of this chapter introduces the aims and objectives of the research project and the chapter concludes by providing an outline to the structure of this thesis.

1.2 Aims and Objectives

As it is clear that occupational health and safety is largely under researched, the overall goal of this research study is therefore to provide a current view of the risk-factors associated with workplace accidents, and to help inform health and safety policy. The existing evidence within the literature surrounding the causes of accidents and injuries in the workplace is mixed. This therefore prompts the first research question: What are the socio-economic risk factors related to workplace accidents and injuries?

Secondly, clear gaps within the literature surrounding occupational health and safety exist, this in particular is focused around area-level statistics and whether geographic differences exist between workplace accidents. A second research question is therefore: Do the relationships between these socio-economic characteristics and workplace accidents and injuries vary geographically?

Finally, the impact of environmental conditions on workplace accident risk is largely under researched. Whilst it is known that environmental conditions may impact upon workers' being able to carry out their job roles safely, the extent to which this varies between industrial groups or occupations, as well geographically, is not known. The third research question therefore is: What impact do environmental conditions have upon workplace accident and injuries and do these affect workplace accident risk-factors geographically?

These research questions form the basis of this thesis and have led to the formulation of a set of research aims and objectives. The three aims of this investigation are:

1. To understand which socio-economic risk factors relate to workplace accidents and injuries,

2. To investigate whether the relationship between these socio-economic characteristics and workplace accidents and injuries vary geographically; and
3. To explore the impact that environmental conditions have upon workplace accident and injuries; investigating how these affect the relationships between the socio-economic risk factors on accident rates geographically.

In order to achieve these aims, the following set of research objectives have been devised:

1. Review the existing body of literature surrounding the socio-economic and physical determinants of workplace accidents, paying particular attention to geographic variations.

2. Review the suitability of methods for analysing the RIDDOR data to achieve the aims of this research.

3. Explore the RIDDOR data for key themes to gain a background to reported workplace accidents and injuries.

4. Model the RIDDOR data to gain an insight into the relationships between the key socio-economic determinants of workplace accidents nationally.

5. Carry out a local regression analysis of the RIDDOR data to explore the geographic variations in accident risks.

6. Explore the free-text RIDDOR data, obtained in addition to the large RIDDOR dataset, to identify any key trends and to elaborate the findings from the previous global and local models.

7. Identify the original findings of the research, where the results fit into the wider field of literature on occupational health and safety and make policy recommendations for the HSE.

8. Review the success of the research project, making recommendations for improvements and possible future research.

The first two objectives have been formulated to review the literature surrounding occupational health and safety, and the methods for analysing the RIDDOR data. These objectives overall relate to the three core aims of this research. The third objective involves gaining a background view of the RIDDOR dataset that is obtained for the research project. This objective also relates to all three aims to help direct the research into discovering the basic key information about work-related accidents.

The fourth objective relates to the first aim of the research project. Understanding the socio-economic and physical risk-factors associated with workplace accidents is the key building block to this research. It is important to gain insight into which groups of individuals have a higher accident risk than others, to be able to target these with increased health and safety awareness to prevent and reduce workplace accidents and injuries.

The fifth objective relates to the second and third aims, in finding out how the relationships between the selected key socio-economic variables and accident rates vary geographically and by time. This objective builds upon the third objective, by exploring local variations in accident risks, rather than assuming the relationships are uniform across all geographic areas.

Objective 6 involves analysing a free-text RIDDOR dataset which contains descriptions of the accidents which occurred. This objective works towards understanding the second and third aims of the research project, in gaining an alternative perspective to workplace accident risks spatially and over time.

Finally, objectives 7 and 8 relate to the overall research project. The original findings from the analyses chapters are outlined and policy recommendations are made for the HSE, as well as possible future extensions to the research. In summary, the objectives can overall be divided into three parts: The literature review and methodologies (objectives 1 and 2), the analyses of the RIDDOR data (objectives 3 to 6), and the overall conclusions (objectives 7 and 8).

1.3 Outline of Thesis Structure

Chapter 2 introduces the literature surrounding occupational health and safety. It examines the existing literature for distinct risk factors associated with workplace accident and injuries. First socio-economic variables are explored, followed by physical factors, and finally geographic variations. Lastly the chapter provides a history and background to health and safety policy and legislation. Current statistics on health and safety are revealed, and the future outlook of health and safety practices are discussed.

The main data resources for this research study are discussed in Chapter 3. The first section focuses on providing a context to the RIDDOR data, followed by an overview of the open source public data that are used. The chapter also provides a review of the methodological approaches used within the research study, comparing between alternative techniques and where methods have been used in previous research.

Chapter 4 provides an overview of the themes found within the RIDDOR data obtained for the study. It sets the context of the research by presenting a range of results outlining the current statistics on workplace accidents. The chapter focuses on the main characteristics of the RIDDOR dataset which covers accidents that occurred across the whole of Great Britain. The chapter then moves on to explore basic trends in the data, based on summary statistics about the types of workers having accidents, and the types of injuries sustained as a result of the accidents.

Chapter 5 introduces a set of modelling techniques applied to the RIDDOR data to perform a global regression analysis. A selection of socio-economic variables is tested, based on what was found within the existing literature surrounding workplace accidents, and a final global model is constructed in a data-driven method. The data that are modelled for this chapter and the next two chapters (Chapters 6 and 7) relate to England and Wales. This is due to the lack of availability of workplace population data for Scotland at the time this research was conducted.

In Chapter 6, a local regression analysis is carried out on a set of socio-economic area characteristics across England and Wales. These characteristics are chosen based on the results from the global model found in Chapter 5. The aim of Chapter 6 is to explore the geographic variations in the accident data, as well as examining the physical environmental factors that are present.

Chapter 7 is a final analysis chapter, exploring the free-text fields of the RIDDOR data. A textual analysis based on first text mining, followed by topic modelling is carried out. This chapter provides a qualitative perspective on workplace accidents, exploring individual experiences to reflect on the previous chapters which consider quantitative methods of analysing accident rates.

Chapters 8 and 9 are the final concluding chapters. These draw together all of the results obtained from the research and compare them to the existing literature surrounding workplace health and safety. Key results are summarised and policy recommendations are made. Suggestions for possible improvements and future research are set out and the overall research study is evaluated.

Chapter 2

Literature Review

2.1 Introduction

Traditionally, the conception of workplace accident risk stems from a combination of individual and work-related factors. Characteristics of the individual found to influence workplace accident risk include: age and gender; and work-related factors include physical demands and hours worked of the individual (Wilkins and Mackenzie, 2007). In helping to reduce workplace accident rates, the Health and Safety Executive (HSE) recognises these factors and categorises them into three broad interrelated aspects of health and safety that must be considered and addressed in making working environments safer for workers. The three categories include: the individual (consisting of characteristics such as age and gender), the job (including factors relating to occupation and industry type) and the organisation (including working patterns and overall working environment) (HSE, 2015d). Working environments also include physical factors such as environmental impacts of seasonal weather conditions and time of day in which individuals work. These categories are explored within this chapter, providing a review of the literature on workplace accident risks.

In recent years, the importance of area-level factors influencing health outcomes, beyond the effects of individual-level factors, has grown in informing public health policy (Leventhal and Brooks-Gunn, 2000; Pickett and Pearl, 2001). Unfortunately, in relation to occupational health, geographic differences of accident risks have been significantly under-researched (Neff et al., 2008). Geographic variations in work

injuries can occur across regions, which may be apparent with varying socio-demographic compositions, for example certain occupational groups or young workers may be more prevalent in one region over another, affecting the national uniform structure of the labour force (Morassaei et al., 2013). The chapter provides a review of the relevant literature on geographic variations in occupational health, and examines how, by observing spatial patterns in geo-demographics, the risk of a worker having an accident might be impacted.

Lastly, the chapter provides a background to health and safety policy. Specifically, it gives an overview to the context of how health and safety policy have emerged and evolved to meet the needs of the working population, by providing a timeline of the key milestones in legislation. The chapter concludes with providing an overview of the current status of occupational health and safety, focusing on statistics published by the HSE, and explores the future outlook of health and safety practices across all workplaces.

2.2 Social Factors

2.2.1 The Individual

Age

There is an ongoing debate around the connection between the age of a worker and their risk of having a workplace accident. Within much of the literature on occupational injuries and workplace accidents, evidence has been presented to suggest that younger workers have a higher risk of workplace accidents than older workers. In the transportation industry, for example, Chau et al. (2014) finds that in analysing workplace injuries of 20,000 railway company workers, individuals up to the age of 25 are subject to a higher risk of workplace injury than any other age group. Similarly, in the fishing industry, 446 injuries are analysed in a study conducted by Bull et al. (2001) who finds that injury rates among young workers are highest compared to all other age groups. These findings reflect and support the

results of a large study conducted by Root (1981) in which a million workers' compensation records are analysed. It is again found that the injury rate is highest for younger workers than that of older workers.

Conversely, several studies suggest that there is a positive relationship between accident rate and age. The results of a study conducted by Xiang et al. (2000) on a sample of 1,500 farmers in China finds that the smallest percentage of work-related accidents is amongst workers below the age of 19 years and the highest percentage is found in age group 40-49 years. Similarly, in a study of occupational accidents in Local Authorities in Finland by Nenonen (2011), it is found that age group 45-54 has the highest accident rate out of all groups, with workers aged above 54 having an increased risk of commuting accidents.

These conflicting results on the relationship between a worker's age and the risk of an accident have been the cause of much debate. Salminen (2004) reflects on 63 occupational studies of non-fatal injuries and 45 studies on fatalities, over a time period spanning 21 years (1981 to 2002). It is concluded that 56% of studies presented findings of younger workers having more non-fatal injuries caused by workplace accidents compared to older workers. Only 17% of studies presented results of older workers experiencing a higher level of non-fatal injuries and 27% of studies identified no significant differences between age groups at all. In terms of fatalities, the majority of studies reviewed (64%) showed that older workers had a higher fatality rate than younger workers. The main limitation of the review is that it does not consider the proportions of subjects and injuries in each study when scoring and calculating percentages, and therefore in some cases the factor detailing how much higher or lower the injury rates were by age group are not included.

The evidence overall suggests that the age of a worker appears to be a social determinant of workplace accidents. There are a number of reasons to explain why younger workers might be more at risk of an accident than older workers. Most young workers begin their careers having no prior knowledge of workplace health and safety and often learn by working on the job. For some, tasks involve intensive manual labour, using tools and machinery, which can often be very dangerous

without sufficient experience of using such equipment. Generally, the younger the worker, the less experience they have of carrying out a job and therefore the more accident prone they are found to be (Hale and Glendon, 1987; Cellier et al., 1995). Young individuals are known to learn by making mistakes and understanding how to prevent them from occurring again (Yerushalmi and Polingher, 2006). The older a person gets therefore, the more knowledge they have acquired of the workplace, and of how to prevent accidents from occurring.

Different attitudes to health and safety appear to be linked to the age of an individual. The use of alcohol and cigarettes are often first experimented with by adolescents (Urberg et al., 1997; Zweig et al., 2001) and in a study conducted by van Exel et al. (2006) on youths' attitudes to health and well-being, it is found that in general, youths seem uninterested in their future health. Reasons for this include: because they feel physically fit, are generally satisfied with their health and in some cases because they simply do not care. This general attitude may be one of the factors influencing why younger workers are reported as having more accidents in the workplace compared to older workers.

Where safety is concerned, age has been found to be an influencing factor in terms of taking risks. Young individuals have been identified as being the most likely to take risks which could impact their health (Michael and Ben-Zur, 2007; Shapiro et al., 1998). Although youths often have the confidence to try new experiences with a care-free attitude, often the results are detrimental to the health and well-being of the individual. In car accidents for example, research shows that the younger the driver the more likely they are to be involved in road collisions (Fisher et al., 2006; Hasselberg and Laflamme, 2009). In fact, it is reported that in 2013/14, out of all working age adults in England, those aged 20-24 were most likely to attend an Accident and Emergency department to be treated for an injury or illness out of all other age groups (Baker, 2015).

Job satisfaction is thought to increase with age and the years worked by a person (Park et al., 2012). When job satisfaction increases, performance increases and greater attention to safety control, knowledge and compliance (Probst, 2002; Ready et al.,

1993). Younger workers who are starting out in their careers may have less job satisfaction due to limited experience and knowledge of the role. This therefore may affect the job performance and the attention span of the worker thus affecting overall safety in the workplace.

In terms of accident severity, several studies find that although younger workers have more accidents than older workers, the severity of injuries from these accidents are much higher for older workers. In particular, older workers are found to have a higher fatality rate caused by workplace accidents than younger workers (Salminen, 2004). In general, accident severity is found to increase steadily up until the age of 65, with evidence for a dramatic increase further after this age (Peek-Asa et al., 1999; Ruser, 1998).

Physical health declines as individuals grow older. The peak physical working age is around 20-35 years when workers are the most fit, healthy and strong (Christensen, 1955). Generally, beyond the age of 35, muscle strength declines and bodies weaken making older workers more susceptible to accidents in the workplace. This also might explain why the severity of injuries caused by accidents are higher for older workers compared to younger workers. Younger workers generally have a greater impact resistance than older workers (Brorsson, 1989). Therefore, potentially the same impact that could kill an older worker might only injure a younger worker.

Physical health and mental cognitive performance are known to deteriorate with increasing age. This becomes a particular problem for workers approaching retirement age whereby they feel tired more often, affecting their overall daily work performance (Beehr et al., 2000). Problems such as trips and falls become more common in older workers as subconscious behaviours reduce, affecting rational decision-making and the ability to prevent human error (Park et al., 2012). This in effect may be contributing to why older workers have more accidents than younger workers in some cases, or why levels of severity of injuries caused by workplace accidents differ between age groups.

The argument as to whether younger workers or older workers experience more accidents is still under debate. With the UK experiencing an ageing workforce, understanding the effect of a worker's age on health and safety in the workplace is particularly important. Only a small amount of research has been previously conducted on UK cases of occupational health and safety, and therefore gaining insight into the relationship between British workers and workplace accident risk will be insightful in preventing future accidents from occurring.

Gender

Previous research suggests that men tend to have more workplace accidents than women (Berecki-Gisolf et al., 2015; Salminen, 1994; Smith and Mustard, 2004). Specifically, several studies find that young men are the most at risk group of workers (Jackson, 2001; McCaig et al., 1998). Several factors connected with the type of work mostly carried out by men compared to women, as well as attitudes to health and safety, are thought to play a part in these accident rate differences, however these have relatively been under researched. There is also little focus in existing research made on how the gender divide issue might be addressed.

The simplest explanation to gender differences in workplace accidents is that exposures to workplace safety hazards often differ between men and women. This is attributed to the types of work men and women are mostly involved in. Generally, men and women tend to work in different occupations. Although substantial movement has occurred for male and female workers across job areas over the last few decades, differences in the careers that men and women choose still persist (Weeden, 1998; Wells, 1999). Women tend to work in jobs that are classed as safer than jobs mainly involving men, partly due to risky jobs being classed as 'unsuitable' for women based on the requirements of physical strength for particular job roles (DeLeire and Levy, 2001).

Two examples of industries in which large disparities exist between the levels of male and female employees are: construction (French and Strachan, 2015) and agriculture

(Quisumbing et al., 2014). In construction, it was reported that 16 per cent of women are employed within the industry in the UK, but two-thirds of these work in clerical roles (French and Strachan, 2015). The workforce therefore largely consists of male workers, where tasks involve heavy lifting of materials and the use of tools and machinery. In the agricultural sector, similarly, work tasks are labour intensive, requiring a greater physical strength and therefore more commonly suited to male workers rather than female workers. Unsurprisingly, based on the physical work involved in these job sectors, workplace accident rates are reportedly amongst the highest in both of these industries (HSE, 2015e).

Other work sectors in which a higher proportion of men work, compared to women, are the fire services (DCLG, 2010) and armed forces (Defence Statistics, 2014). A lot of the tasks involved in jobs in these sectors are also considered very dangerous. Not only is work physically demanding, but firefighters and soldiers are often faced with additional risk factors such as working in hot, wet or cold conditions due to working outdoors or in extreme temperatures (Kong et al., 2013; Ministry of Defence, 2016), therefore these factors might also increase male workers' risks of work-related accidents. Section 2.3 in this chapter discusses in further detail evidence around the link between environmental factors and workers' accident risks.

Men having a higher work related accident risk than women is therefore most likely attributed to the types of occupation and work-related tasks that they are involved in. Although in general male workers have been found to have more work-related accidents overall, gender differences have been reported in certain types of injuries sustained from accidents in the workplace. Many studies have found that men who are involved in a workplace accident suffer from injuries of higher severity than women who are involved in an accident at work. More specifically, more men than women are reported to have a higher risk of a fatality as a result of a workplace accident (Knestaut, 1996).

Other injuries more common for men than for women include back disorders, with around 870 cases per 100,000 reported for men, but 380 cases per 100,000 for women (HSE, 2015f). Men in the age range 35-44 years had an overall statistically

higher rate of back disorders than any other age group. Additionally to this, lower limb disorders which include meniscal tears of the knee or ankle, fractures of the ankle and foot and overuse injuries of, for example, the hip and hamstrings, are also more common for male workers (HSE, 2015f).

Female workers, however, are more likely than men to suffer from overall work-related musculoskeletal disorders (HSE, 2015f) which include disorders of the muscles, tendon, joints, cartilage or spinal discs. In terms of upper limb disorders, a larger number of women than men are found to suffer from injuries such as repetitive strain disorder, hand wrist tendon syndromes and carpal tunnel syndrome (HSE, 2015f). Work related musculoskeletal disorders can develop due to repetitive physical tasks, where a pace of work does not allow sufficient recovery between movements. In a study of repetitive strain disorders by Ashbury (1995) it was found that injury rates are highest for women compared to men in all occupations, with processing, machining and fabricating jobs, followed by construction reported as having the highest rates.

One of the reasons female workers tend to have accidents in manual occupations has been suggested to be because work stations and protective equipment are generally designed for the average sized man (Courville et al., 1991). The issue with this is that they are therefore ill-adapted to female workers, meaning that females exert their strength in uncomfortable positions more frequently than men, making it difficult for women to carry out work tasks safely. In a study by Gagnon and Lortie (1987), it was found that the risk of this on health workers in hospitals when needing to lift patients for care, resulted in female health workers using methods of pushing and pulling to transform difficult tasks into manageable jobs. As a result, women were reported as suffering from significant lower back pain in these occupations.

Age, together with gender, appears to be an additional risk factor in workplace accidents. Although conflicting evidence is found on the link between age and workplace accident risk, specific types of injuries tend to be reported more often for men and women in particular age groups. The highest rate of musculoskeletal disorders for example, is found in women aged 45-54 years (HSE, 2015f). Similarly,

the HSE reports also that upper limb disorders are higher for women aged over 45 years than for any other type of worker. With age, women and men suffer from different health problems. One example which may explain the higher rate of musculoskeletal injuries in women is that with age, women are at a higher risk of osteoporosis, increasing the risk of bone fractures as a result of trips and falls (Cherry et al., 2005).

In terms of overall health, women tend to involve themselves in less risky behaviours than men (Harris et al., 2006). Studies surrounding gender differences in risk perceptions have revealed that women are more likely than men to avoid drinking alcohol and driving, and men are found to be more likely of running an amber light whilst driving compared to women (Konecni et al., 1976). Waldron et al. (2005) found that whilst men are more likely to take risks, the result of such behaviour often has detrimental impacts, with evidence suggesting that men are much more likely to die from drowning or accidental poisoning throughout the Western world compared to women.

Harris et al. (2006) reported that women are much more likely than men to judge potential negative consequences from a risky event, and to judge the potential consequences to be significantly severe in terms of health as a domain. This is perhaps due to women being more likely than men to consider the responsibilities of their lives such as children and family life in weighing up a potentially risky behaviour (Grazier and Sloane, 2008). This is further highlighted by the findings of a study on gender and accident risk at work by DeLeire and Levy (2001), in which it is found that single parents are the most risk averse, followed by married women with children.

Women have a higher life expectancy than men (ONS, 2015) and this may be because women tend to care more about their health. Women are considered better at identifying their own health conditions and known to actively seek advice more often from a doctor for a health problem as compared to men (Case and Paxson, 2005). Men's care-free attitudes when it comes to health, and in general, men being less willing to discuss their health problems (Idler and Benyamini, 1997; Verbrugge,

1989) may explain the reportedly higher rates of accidents and injuries amongst male workers.

The origins of gender differences in workplace accident rates continues to be an under researched topic, particularly in the UK. In general, many studies have shown that men are reported as having more workplace accidents than women for reasons connected with the types of riskier occupations men undertake compared to women. Attitudes to health and safety could be an underlying factor in the division in gender differences also, however much of the existing literature lacks guidance as to how these factors might be addressed to prevent future workplace accidents from occurring. With the gap narrowing between gender-segregated jobs, with more women becoming involved in 'riskier' jobs, and more men working in traditionally female jobs, it is necessary to gain a thorough understanding as to how work-related risks could be better managed between gender groups.

Socio-Economic Class

Much of the existing literature surrounding health inequalities is based upon differences in socio-economic classes. Marmot et al. (1997) for example, identifies that health differences can be observed across the entire socio-economic spectrum, with the rate of mortality rising with decreasing socio-economic status. The National Statistics Socio-Economic Classification (NS-SEC) is a UK framework for grouping households into classes based on their occupation type and employment status. Social classifications are used by many different researchers and organisations including central government to analyse social and health variations, and thus direct policy and resource allocation. The NS-SEC consists of eight classes which are formed in a socio-economic hierarchical structure from highly skilled professions, to low skilled and unemployed. The classes include:

1. Higher managerial, administrative and professional occupations
2. Lower managerial, administrative and professional occupations
3. Intermediate occupations

4. Small employers and own account workers
5. Lower supervisory and technical occupations
6. Semi-routine occupations
7. Routine occupations
8. Never worked and long-term unemployed.

(ONS, 2010)

In terms of NS-SEC 8 several studies find that there is a link between unemployment and poor health (Schuring et al., 2015; Vaalavuo, 2016). In particular, evidence presented in a study by Theodossiou (1998) shows a link between unemployment and anxiety and depression. The likelihood of an unemployed person to suffer from mental-health issues is found to be significantly higher than all other working groups, even compared to individuals in low-paid employment. The study also finds that middle aged individuals are more likely to suffer than younger or older age groups.

Life expectancy is also found to be strongly correlated with social class. Rogot et al. (1992) reported that men aged 25 in the labour force, live on average 12 more years longer than those not in the labour force. Also, in a study by Marmot et al. (1978), it is found that men in lower employment grades are shorter, heavier for their height, have higher blood pressure, smoke more and report less leisure-time physical activity than men in higher grades. As a result, a link between mortality rates caused by coronary heart disease is found to be more significant for men in lower employment grades compared to men in higher employment grades.

Workers employed in low-skilled jobs are exposed to harsher and riskier work environments (Toch et al., 2014). Gillen et al. (2007) find that musculoskeletal disease is more prevalent in low-skilled jobs involving physical work. Other hazards, such as exposure to chemicals, are also more likely in low-skilled jobs (Alamgir and Yu, 2008). These are known to contribute to associated work-place ill-health and injuries (Fingerhut et al., 2005), which therefore implies a socio-economic occupational health divide between low-skilled workers and skilled-workers.

'The wealthy get healthy, the poor get poorly' (Davidson et al., 2006) is a common conception amongst today's health researchers. Some studies have found evidence confirming the existence of a health/wealth gap, with strong correlations found between low income individuals and obesity and diabetes (Pickett et al., 2005). Rogot et al. (1992) also confirm these findings, explaining that, as income increases for an individual, their life expectancy also increases. These studies imply that individuals with higher incomes are healthier and live longer. Many possible reasons for this relationship have been discussed; such as individuals with higher incomes, and therefore higher socio-economic status, can afford to eat healthier foods (Lloyd et al., 2011). Also, individuals with higher incomes have greater access to fitness centres and health facilities (Estabrooks et al., 2003) to maintain good levels of overall health and fitness.

Previous research has found links between people in lower socio-economic classes and accident risk, although in general, most studies consider cases of accidents in the home and outside of work. Children with a highly deprived background are found to be more likely to suffer unintentional injuries and accidents than those from less deprived backgrounds (Dowswell and Towner, 2002; Reading et al., 1999). Additionally, Istre et al. (2001) also find that the risk of house fires is highest for families with the lowest paid jobs and it is also found that these households did not have working smoke detectors in their homes or were aware of health and safety issues. Cubbin and Smith (2002) find that general injuries are socially graded, whereby the lower the socio-economic status, the higher the risk of injury. Severe accidents resulting in fatalities are also found to be more prevalent amongst individuals with low socio-economic status (Graham, 1985).

2.2.2 The Job/Organisation

Within the UK, a Standard Occupation Classification (SOC) was introduced in 1990 to enable the classification of occupations by type of work. The SOC system creates a uniform framework to enable researchers and policy makers to tabulate and analyse data based on a consistent structure of economic activity. The current SOC 2010

framework consists of four nested tiers: 9 major groups, 25 sub-major groups, 90 minor groups and 369 unit groups, with each level providing more detail of the occupation type. The groups are built around a hierarchy of defined skillsets required by workers to do a particular job. Skills are defined in terms of the nature of qualifications and how long it takes to accumulate these qualifications, as well as training and work experience required to become competent to perform tasks associated with the job.

The current 9 SOC major groups with their associated level of skill are contained within Table 2.1. The table shows varying levels of qualifications and experience or training required for each SOC group, with SOC 1 consisting of workers requiring the highest level skillset, usually a degree or higher in terms of educational attainment and significant work experience. SOC 9 generally consists of workers gaining the minimum level of education, with no significant training required to do tasks, therefore having the lowest skillset.

Table 2.1: SOC major groups and skill levels

Major Group	Nature of skills
1. Managers, Directors and Senior Officials	A significant amount of knowledge and experience of the production processes and requirements associated with the functioning of organisations/businesses.
2. Professional Occupations	A degree or postgraduate qualification, and/or a formal period of experience-related training.
3. Associate Professional and Technical Occupations	An associated high-level vocational qualification, often involving a substantial period of full-time training or further study. Task-related training is usually also provided through a formal period of induction.
4. Administrative and Secretarial Occupations	A good standard of general education. Certain occupations will require further additional vocational training to a well-defined standard.

5. Skilled Trades Occupations	A substantial period of training, often provided by means of a work based training programme.
6. Caring, Leisure and Other Service Occupations	A good standard of general education. Certain occupations will require further additional vocational training, often provided by means of a work-based training programme.
7. Sales and Customer Service Occupations	A general education and a programme of work-based training related to sales procedures.
8. Process Plant and Machine Operatives	The knowledge and experience necessary to operate vehicles and other mobile and stationary machinery, to operate and monitor industrial plant and equipment, to assemble products from component parts according to strict rules and procedures and subject assembled parts to routine tests. Most occupations in this major group will specify a minimum standard of competence for associated tasks and will have a related period of formal training.
9. Elementary Occupations	Minimum general level of education (ie. that which is acquired by the end of the period of compulsory education).

Researchers have found differences in health quality between occupational groups. Specifically, several studies have found that the biggest differences exist between low-skilled and high-skilled occupations, with workers in occupations requiring a high level skillset to have a significantly higher level of overall good health (Davidson et al., 2006; Townsend et al., 1988). More specifically, other studies have looked at a range of health conditions, including: risk of lung cancer, ischaemic heart disease and stroke (Drever, 1997), with the lowest skilled workers having the highest rates of poor health and the highest skilled workers having the lowest rates, in all cases. In addition, stress is found to be associated with low-skilled jobs or those unemployed (Creed and Macintyre, 2001; Sherman et al., 2012). This is thought to be due to the financial strain on workers in these occupations, impacting upon overall well-being.

Workers in high-skilled jobs have been found to experience less stress due to their greater sense of feeling in control and able to delegate and offload tasks to other workers (Sherman et al., 2012).

Statistics previously released by the ONS also report a significant difference in standardised mortality rates by occupational group. Men aged 20-64, in professional occupations scored a rate of 280, (per 100,000 cases) compared to those in low skilled jobs who scored a rate of 806 (per 100,000 cases) (Drever, 1997). This again suggests evidence to imply that health inequalities exist across occupational groups, with workers in high-skilled jobs having an overall higher level of health compared to workers in low-skilled jobs.

There is a lack of research overall, surrounding the link between occupational groups and workplace accident and injury risks. This is most likely attributed to there being a lack of access to data surrounding workplace accidents and information regarding injured workers (Feyer and Williamson, 1998). Most comparative research considers differences in health quality and social classes, whilst others consider the link between educational differences and deprivation with mortality rates (Kunst et al., 1998; Leclerc et al., 1990). In the HSE's annual statistics report for 2014 to 2015, however, it states that process plant and machine operatives (SOC 8) and elementary occupations (SOC 9) have statistically significantly higher than average rates of workplace injuries. On the other hand, it is also reported that professional occupations have a relatively low injury rate, which suggests that inequalities appear to exist between occupational groups in terms of accident and injury risk.

In a study conducted by Laberge et al. (2016) on unexpected events involving apprentices in low-skilled jobs, it is found that factors which lead to the occurrence of accidents and injuries are: working techniques, or material or product problems. The study highlights work undertook by Perrenoud (1999) which focuses on competency-based development, where it is found that the capacity to master unforeseen events is linked to a high level of competency. This may explain why reported cases of work-related injuries are lowest in the occupations requiring the highest level skillset.

Although limited research currently exists surrounding occupation types and workplace injuries, it is well known that workers in low-skilled jobs tend to be involved in more dangerous working environments than highly-skilled jobs. Tasks carried out in low-skilled jobs often involve working with material objects, which require moving and lifting and which increase the risk of occupational injuries (Ljungberg et al., 1989). Additionally, low-skilled jobs are also often involved in working environments which are exposed to toxins and chemicals, increasing risks of slips, and burns (Alamgir and Yu, 2008). Repetitive tasks are also commonly associated with workers in low-skilled jobs, such as workers on assembly line productions (Grobler, 2013), which often leads to poor postures and long term chronic joint pain.

Analysing occupational data is not always straightforward due to the amount of different industries there are across a range of sectors. The UK classifies industries under the Standard Industrial Classification (SIC) system, similarly to SOC, to group together business sectors and jobs by the type of economic activity in which they are engaged. Under the SIC system, industries are first divided into 21 broad sections, and then broken down into 88 divisions, 272 groups, 615 classes and 191 subclasses. Although there have been several versions of the SIC since 1948, there has been little change since 2003 to the most recent 2007 SIC, rather new concepts introduced at the highest level and new detail at the lower levels to reflect different forms of production. The broad group coding system is outlined below, indicating the broad sections (A to U) that industries can be classified into, under SIC 2007:

A. Agriculture, Forestry and Fishing
B. Mining and Quarrying
C. Manufacturing
D. Electricity, Gas, Steam and Air Conditioning Supply
E. Water Supply; Sewerage, Waste Management and Remediation Activities
F. Construction
G. Wholesale and Retail Trade; Repair of Motor Vehicles and Motorcycles
H. Transportation and Storage
I. Accommodation and Food Services

J. Information and Communication
K. Financial and Insurance Activities
L. Real Estate Activities
M. Professional, Scientific and Technical Activities
N. Administrative and Support Service Activities
O. Public Administration and Defence; Compulsory Social Security
P. Education
Q. Human Health and Social Work Activities
R. Arts, Entertainment and Recreation
S. Other Activities
T. Activities of Householders as Employers; Undifferentiated Goods and Services Producing Activities of Households for Own Use
U. Activities of Extraterritorial Organisation and Bodies

In the UK, all employers or managers must carry out risk assessments to identify potential hazards in their workplaces; evaluating the risks involved and identifying who might be harmed. They must then reflect on these assessments and put sensible precautions into place to control risks and make sure they stay controlled. There are various levels of health and safety measures put in place which vary between industries and which mostly depend upon the types of hazards and health issues likely to be linked with particular industrial sectors. These measures range across a variety of the following topics:

- Electrical safety,
- Fire safety,
- Gas safety,
- Harmful substances,
- Machinery, plant and equipment
- Manual handling,
- Noise,
- Personal protective equipment,
- Pressure equipment,

- Radiations,
- Slips and trips,
- Vibration,
- Working at height,
- Working in confined spaces,
- Workplace transport

(HSE, 2016)

In the annual statistics report produced by the HSE, it states that the most reported risk factor causing an accident or injury in the workplace is 'dealing with difficult customers, patients, pupils etc.' (HSE, 2015e). Whilst this is recognised as a psychosocial risk, it can also present itself as a physical risk in terms of violence towards workers. As Pouliakas and Theodossiou (2010) point out, however, it is difficult to measure risk perception correctly, as individuals consistently underestimate the probability of accidents occurring at work, and therefore often have misconceptions as to what risk factors they face in the workplace on a daily basis.

Generally, there appears to be a gap in the existing literature which compares health and safety workplace hazards across industrial sectors in the UK. This again, similarly to workplace accidents by occupational groups, may be due to the lack of access to data covering workplace accidents and cases of ill-health. It is common knowledge, however, that different industries carry different levels of risk, in that reported injury rates are likely to be higher where there is a larger proportion of manual workers. Such industries where this is found to be the case include: manufacturing and construction (Chi and Wu, 1997; Sorock et al., 1993), which are both heavily researched. These industrial sectors have the highest work-related accident and injury rates reported for Great Britain, as well as: agriculture, forestry and fishing, accommodation and food service activities and transport and storage (HSE, 2015e). Additionally, the industrial sectors which have the highest fatalities have been reported as: construction, agriculture and waste.

The industries in which the highest reported accident, injury and fatality rates occur, have several characteristics in common. Firstly, the work-related tasks involved in these industries require demanding physical labour which exposes workers to greater physical hazards than non-manual work (Robinson and Smallman, 2006). For workers employed in the construction industry for example, Sorock et al. (1993) report that in the US, the leading causes of death were falls, vehicle-related deaths and electrocutions. In the agricultural sector, Huyghebaert et al. (2015) find that the misuse of machines and agricultural equipment were the most common causes of accidents, and in manufacturing, injuries as a result of contact with machinery were most common (Chi et al., 2004). In all cases, the main causes of injuries within these sectors were found to arise from contact with foreign objects.

The second prevalent characteristic is that workers employed in the riskiest industries carry out their tasks in the harshest working conditions. Chi et al. (2005) find that the main sources of occupational injuries in the construction sector involve falls due to workers working at a height, for example on scaffolding or roofs, as well as falling through floor openings and from ladders. As well as the construction industry, employees in industries such as transportation and agriculture also work mostly outdoors and are therefore exposed to weather conditions which, during certain times of the year, may impede their ability to perform tasks safely, especially due to their work activities being mostly manual work (Edwards, 1996).

Workers employed in industries considered to have a low level accident risk, may have an increased risk of other health problems. For example, office workers are at an increased risk of musculoskeletal disorders, for example repetitive strain injuries, due to prolonged sedentary behaviour resulting in neck and shoulder pain (Tunwattanapong et al., 2016). It has also been widely reported that the rise in office and home based working, involving increasing time working at desk, increases rates of obesity, which can contribute to further health problems (Levine and Miller, 2007).

2.3 Environmental Factors

In some industries, the working environment has been found to be affected by environmental conditions. The UK has a varied and unpredictable weather system, however, in general, the UK experiences its warmest weather during: June, July and August, its summer months, and its coldest temperatures and strongest winds in December, January and February, its winter months. The levels of extreme weather fluctuate from year to year, with some winters having relatively mild weather conditions, and some summers having increased rainfall compared to the rest of the year. Regional differences in weather conditions are often observed across the British Isles between seasons (Mayes and Wheeler, 2013). The winter of 2010 for example, was reported as bringing a record level of snow, cold spells and low temperatures, and the most severe conditions were recorded in northern areas of the British Isles, particularly Highland Scotland (Prior and Kendon, 2011). This section explores the literature on environmental conditions as a cause of work-related accidents, both in terms of seasonal weather variations and particular times of the day or week.

2.3.1 Weather Conditions

Winter

It has been reported that around 1 in 8 accidents take place on rainy days (Edwards, 1996). This is because poor weather conditions impact negatively upon peoples' ability to carry out daily duties effectively. Daily activities such as leaving the house, going to work and using the car all involve being outside which increases the risk of slips, trips and falls. Also, in icy conditions, pavements become even more slippery and are known to increase an individual's risk of having a fall (Gao and Abeysekera, 2004). Falls can cause mild to serious injuries and those resulting from icy conditions mostly result in back injuries and fractures (Björnstig et al., 1997). Falls are a significant problem worldwide in general, with it being reported by the World Health Organization in 2012 that accidental falls are the second leading cause of unintentional injury death globally after road traffic injuries.

Road safety is also compromised when there is heavy rain, snow and icy conditions, as road surfaces become wet and slippery, increasing the risk of vehicle collisions and crashes. High winds can also affect drivers' ability to control the vehicle, and snow and fog can reduce drivers' visibility (Usman et al., 2012). These environmental factors increase the risk of road accidents by 2-3 times the amount that occur within dry weather conditions (Brodsky and Hakkert, 1988). Edwards (1996) also found evidence of geographic variations between accident rate and weather conditions. In the study it was found that most road accidents were occurring in the westerly extremities of Britain where rainfall was highest and the northern parts of England, which had highest winds in comparison to Central England.

Although existing literature provides evidence on the link between poor weather conditions and accidents, little work has been focused around work-related accidents. Findikyan and Sells (1965) do however note the impact to workers from carrying out tasks in low temperatures. It is found that the numbness of fingers produced by cold stress, a term describing the effect of cold temperatures on the body, leads to the inability to perform simple tasks such as manipulating screws, bolts, keys, switches and buttons. Clark (1961) also investigated hand skin temperature on the effect of knot-tying ability and found that performance was significantly worse affected when temperature reached a low of 13°C. Both these studies suggest that workers are unable to perform work tasks at their maximum level of performance in low temperatures, which may also impede their ability to work safely.

Summer

Although summer temperatures in the UK are uncommon to peak as high as in most neighbouring European countries, there have been considerable heat waves in British History, such as the heat wave in August 2003, where temperatures reached a record high of 38.5°C in Kent. Approximately 2,000 deaths were recorded in the UK as a result of this heatwave, as well as an increase in cases of heat stroke, dehydration and sunburn (Met Office, 2015). High temperatures, therefore, can be hugely detrimental to a person's health and it is therefore important to plan for such extreme seasonal effects.

There is currently no law which exists in the UK that specifies what the minimum and maximum temperature in the workplace should be. This can be hugely detrimental to the health and safety of workers, especially if workplace temperatures become significantly high. The ambient temperature of the human body is 37°C, and anything above 38°C can imbalance the core body heat. This can result in a strain on the human body to carry out physical work and mental task ability (Bridger, 2003). This is further reiterated by the results of a study on workplace accidents in Italy, by Morabito et al. (2006), where it is found that discomfort in hot weather causes a change in work behaviour by heat stress. This resulted in workers suffering from an increased risk of a work-related accident. High body temperatures can eventually lead to exhaustion, dehydration and heat stroke, which all have major implications on workplace performance. These conditions can cause many workers to be absent from work or in other cases, workers continuing their tasks in the workplace are likely to make mistakes more easily due to lethargy, becoming a hazard to themselves and others (Ramsey et al., 1983).

Climate change will most likely impact working environments further into the future. Increasing local temperatures result in higher exposures to heat, which during already hot seasons, can result in heat stress and therefore a reduced work capacity in heat-exposed jobs (Kjellstrom et al., 2009). Most indoor jobs may be unaffected with appropriate cooling systems and air conditioning, however not all business places will be so well equipped, particularly smaller firms with less income. Outdoor jobs however, will face high levels of heat exposure, which can be dangerous to health. To reduce heat production, workers potentially need to have a reduced workload so that they can work at a slower pace, generating less heat energy, and secondly by taking regular breaks from work to rest and recover (Bridger, 2003).

2.3.2 Daylight Hours

In winter months, days are considered shorter as there are less daylight hours, with evenings becoming darker much earlier than those in summer. Previous research

suggests that there is a higher risk of work-related injury and illness in the evening, night and early morning, with safety declining over successive nights and longer hours of work (Folkard and Tucker, 2003; Mustard et al., 2013). Folkard (1997) defines these times as 'black times' to illustrate the times of day that accidents mostly occur as being riskier than other times of the day. In particular, Mustard et al. (2013) finds a broad slot of times between 5pm till midnight, and 12am till 5am with increased accident risk, and Folkard (1997) notes a major peak in transport accidents at 3am. Generally, these 'black times' are thought to be due to two key points: workers suffering from poor visibility due to lack of lighting and workers being tired and sleepy during working hours. With a number of 'headlining' disasters occurring during night shift work, such as the gas leak in Bhopal in 1984, the nuclear disaster in Chernobyl in 1986, the Rhine chemical spill in 1986 and the Exxon Valdez oil spill in 1989 (Folkard and Tucker, 2003), attention must be drawn to the risk of reduced workplace safety during night and shift work.

In a study by Arditi et al. (2007), it is found that night time construction work is 5 times more hazardous than daytime construction work. This is because construction workers who are carrying out labour intensive tasks using tools and machinery need clear visibility. Even a slight mistake could result in a major injury or fatality. During the daytime there may be sufficient light to carry out work however during the evening and night, when a lot of construction work continues to be carried out for example in highway and railway construction, levels of light are reduced (Valentin et al., 2010). Workers therefore must rely on artificial lighting such as streetlights and headlights as well as high-visibility clothing to see and be seen after sunset. Similarly, Owens (1993) reports that there are more road fatalities during night-time due to poor visibility. Drivers need clear visibility to see oncoming traffic, judge distances, make safe manoeuvres and to ensure they are aware of pedestrians and cyclists. Surprisingly, although drivers are aware that their visibility is reduced during the night, there is evidence that night time average speeds do not differ to daytime speeds (Herd et al., 1980). This therefore potentially increases the risk of accidents on the road. Both studies, by Arditi et al. (2007) and Owens (1993) recommend that adequate lighting should be available to improve safety standards and reduce the amount of accidents in construction and transport.

There has overall been a lot of research into shift work and health impacts (Akerstedt, 2003; Frost et al., 2009; Kamdar et al., 2013). A key study by Dembe et al. (2006) finds that night shift workers suffering from fatigue, sleepiness and stress had a higher risk of occupational injury and illness. People need to be alert and have a high attention span during waking hours to prevent workplace accidents. Being alert allows workers to be aware of any potential causes of accidents or injuries in the workplace and to be able to prevent accidents and injuries from occurring (Heinrich, 1931). Shift workers however, report more sleep disturbances than standard daytime workers (Akerstedt, 2003). Saper et al. (2005) explains that given the importance of sleep in the functioning of the brain, sleep disturbance can hugely impact workplace injury and sleep loss can impair performance among workers. Sleep therefore is hugely influential on the ability of workers to carry out tasks safely and efficiently.

In terms of sick leave, several studies find that the majority of working days lost due to ill-health are Mondays (Bilimoria, 2009; Vahtera et al., 2001). Research also suggests that a high proportion of accidents and injuries occur more on a Monday than any other day of the week (Butler et al., 2014; Packer and Shaheen, 1993). This could mainly be attributed to workers being tired and less alert on Mondays, following the weekends. Several studies indicate that sleep is an essential activity for a human's alertness, attention span and memory (Bonnet and Arand, 1995; Dijk et al., 1992; Macquet et al., 1997; Smith et al., 2002). The reduction of sleep by as little as 1.3 to 1.5 hours for 1 night can reduce a person's alertness during the day by 32% (Bonnet and Arand, 1995).

In the UK, clocks are put forward an hour in March and returned back an hour in October to accommodate more natural light into the working day (Daylight Saving Time (DST)). Research shows contrasting evidence of DST upon the impact on workplace accidents and ill-health. Varughese and Allen (2001) find that sleep deprivation on the Monday following DST in the spring results in an increase in fatal accidents. It is suggested than an increase in late night and early morning driving is related to high fatalities due to drivers being sleepy. Barnes and Wagner (2009) also find that there is an increased rate of accidents in mining with workers suffering from injuries of a greater severity the Monday following the switch to DST than in

comparison to standard time days. In contrast however, Morassaei and Smith (2010) find no evidence of DST affecting work injury claims, and Holland and Hinze (2000) find no statistical differences between injury frequency in the construction industry before and after the shift in clock time.

2.4 Geographic Variations

The HSE reports rates of ill-health, non-fatal injuries and fatalities as a result of work-related accidents, by geographic region. The average rate of cases of ill-health per 100,000 workers is reported as 3,820 across a 3-year average from 2012-2015. The East Midlands, Wales, Yorkshire and Humber, South East and South West regions have average rates significantly higher than the average rate of Great Britain. In comparison, Scotland has a statistically significantly lower average rate compared to the national average (HSE, 2015g). In terms of non-fatal injuries, the HSE reports that the national average is 2,140 per 100,000 workers. The East Midlands and Wales regions both have rates which are statistically significantly higher than the national average, whereas London has a rate significantly lower (HSE, 2015g). In terms of fatal injuries, the highest rates in 2014/15 were reported in Scotland and the South West region, with the national average being 142 per 100,000 workers.

There are a number of reasons as to why cases of work-related injuries, fatalities and ill-health may vary across the country. Variations in environmental conditions such as the weather and levels of sunlight may impact upon differences geographically in accident rates. Additionally, areas are made up of different compositions of individuals, who work in a range of industrial sectors, and this therefore may impact upon geographic differences in rates of reported accidents.
A report prepared by Davies and Elias (2000) for the HSE, explores basic regional differences amongst reported workplace accidents and injuries. Specific characteristics that are explored as risk factors are: industrial and occupational composition, levels of qualifications held by those in employment, the age/sex profile of employees, size of businesses and the number of hours worked by

employees. In London it is found that the average employee injury rate is 46% below the average injury rate across all regions. This is found to be partly attributed to personal characteristics of there being the highest proportion of younger workers employed in London, compared to all other regions. In the North of England, average employee injury rates are estimated to be 38% above the average rate for all regions. In this case, it is found that the industrial and occupational composition of employment are the influential factors. A relatively high proportion of workers are employed in the construction industry, and in terms of occupation composition, a relatively high proportion work within craft and related occupations.

Further work carried out by Cameron et al. (2008), looks specifically into comparing accidents in the construction industry between Scotland and the rest of Great Britain. Jobs involving the use of scaffold, and bricklaying activities are found to be problematic areas in Scotland, however, in comparison, plant operators in England are also high-risk groups for having work-place accidents. Although the studies by Cameron et al. (2008) and Davies and Elias (2000) illustrate that some strong spatial variations appear to exist between workplace accident risks in the UK, and cover topics that haven't previously been researched in great depth, there appears to be some clear limitations. The study by Cameron et al. (2008) focuses on just the one sector, construction, rather than comparing several industrial sectors. It also compares only Scotland as an entire region to the rest of Great Britain, rather than identifying smaller-scale problematic areas. Another limitation of the study is that the data for the analysis is a small sample of accident data, covering a single year of reported incidents. As workplace accident data can change significantly each year, the work similarly carried out by Davies and Elias (2000) would need to be updated as it covers accident data from before the 1990s. Additionally, neither study considered the effects of physical or environmental conditions on workplace accident risks.

In a study by Doran (2004) it is found that social inequalities in health vary between regions. People in all social classes living within the North had poorer health than the national average for all social classes. The opposite is found for the South, whereby all classes were found to have a lower level of poor health in comparison to the national average. For people living in the richest areas, life expectancy is higher by

between 5-15 years than those living in the poorest areas of a developed country (Wilkinson et al., 2004). Möller et al. (2013) also finds that there are clear differences between the North and South in terms of ill health due to unemployment, with similarly the North faring worse than the South.

The use of area-level characteristics in impacting sociological outcomes, such as educational attainment and labour market opportunities has long been studied (Leventhal and Brooks-Gunn, 2000), however neighbourhood variation in health has received less attention (Pickett and Pearl, 2001). Generally, studies focus on individual risk level factors in identifying causes of ill-health and disease (Hertel and Mermelstein, 2016). This is partly due to theories surrounding an individual's health risk based on their lifestyle and own characteristics, but it is also down to the complex nature of obtaining and performing analyses on multi-level data (Diez-Roux, 1998).

An increasing interest in societal influences on health quality has in recent years emerged. It is more widely known that interactions between health risks at group-level and individual-level may provide new insight into understanding health status (Diez-Roux, 1998). Area-level characteristics often reflect the aggregated effect of individual characteristics, and therefore together with geographic information systems, have started to become essential tools in researching public health trends (Fletcher-Lartey and Caprarelli, 2016; Krieger et al., 2002).

In a study by Gilbert and Chakraborty (2011), the spatially varying relationships of demographic and socio-economic characteristics of individuals on the risk of developing cancer from exposure to hazardous pollution were examined. The key findings of the study were that relationships between demographic and socio-economic variables and cancer risk were not uniform across Florida; but that relationships were more significant in some areas and not evident in others. An example of this was observed for the Hispanic population proportion, in which the results showed a positive association with cancer risk overall, however the results also revealed a negative association with cancer risk in north Florida. Secondly, lower poverty rates were found to have a positive relationship with cancer risk overall, but

the results showed that in Tampa Bay, cancer risks were disproportionately distributed with respect to people in poverty.

There are several ways that geographic analysis can help towards informing public health policy. Identifying spatial trends in health risks increases the ability to target programs specifically to areas in need. These programs can be designed to include relevant information to tackle local risk factors, which may vary between areas (Neff et al., 2008). Unfortunately, in terms of occupational injury or illness, spatial statistics and GIS techniques have been significantly underused. There is therefore little research currently available to inform health and safety policy in targeting small-scale areas which may be experiencing high levels of work-related accidents and ill-health.

2.5 Background to Health and Safety in Great Britain

2.5.1 Health and Safety Policy and Legislation

Health and Safety policy in Great Britain first dates back to the 1800s when the Factories Act 1833 was introduced. This legislation brought about factory inspectors to regulate work within factories by helping to prevent overworking and workplace injuries (HSE, 2012a). Four inspectors initially entered textile mills and were able to formulate new regulations to ensure health and safety in the workplace, however by 1868, the number of inspectors had increased considerably and the Factories Act had been enforced to almost all workplaces. Several years after the Factories Act 1833, an investigation into working conditions in the mining industry took place, and the Mines Act 1842 was introduced. An inspectorate was formed in 1843 after it was discovered that miners suffered brutal injuries, lung disease and worked long hours in dangerous conditions (HSE, 2012a).

By 1893, women inspectors were introduced after several years of campaigning. Their early work focused on investigating women's working hours and health and safety in laundries (HSE, 2012a). By the end of the century, Great Britain saw the establishment of the Quarry Inspectorate following the introduction of the Quarries Act 1894 which was an extension of the Metalliferous Mines Regulation Act in 1872.

Inspectors were able to control and enforce health and safety in the quarrying industry through the new act enabling the reduction of accidents and injuries of quarry workers.

By the mid 1900s, further legislation was rolled out including the Agriculture (Safety, Health and Welfare Provisions) Act 1956, providing health and safety protection for agricultural workers coming into contact with agricultural machinery. The Act set out a framework for the notification of accidents and appointed inspectors to enforce health and safety requirements in the agricultural sector. Later, after an investigation into a serious incident at the Windscale nuclear site in 1957, the Nuclear Installations Act was also established setting out safety regulations for nuclear power stations.

The key milestone in health and safety policy was the introduction of the Health and Safety at Work etc. Act 1974. This became the principal legislation for tackling issues of health and safety in Great Britain's workplaces. The intention of the Act was to set out clear and concise health and safety rules and regulations to replace the previously complex legislation. The act covers the duties of the employers to their employees and of employees to their work and themselves. It establishes and defines the role of the Health and Safety Commission (HSC) and the Health and Safety Executive, both later joining in 2008 to form the now existing Health and Safety Executive (HSE). Furthermore, it sets out health and safety regulations, codes of practice and the provisions of enforcement for breaches of health and safety legislation and policy.

The HSE was introduced to enforce health and safety in Great Britain's workplaces by working with local authorities to tackle issues, which resulted in death, serious injury and illness from work. The HSE and Local Authorities work locally, regionally and nationally enforcing Parliamentary Acts, primarily the Health and Safety at Work etc Act 1974, as well as health and safety Statutory Instruments (Regulations). There are 201 Statutory Instruments providing greater detailed guidance on more specific health and safety topics. Examples of these include: Control of Substances Hazardous to Health Regulations 2002 (S.I 2002/2677), Manual Handling

Operations Regulations 1992 (S.I 1992/2793) and the Reporting of Injuries, Diseases and Dangerous Occurrences Regulations 1985 (S.I 1985/2023).

With the intention of reducing further accidents at work, the Health and Safety (Offences) Act 2008 came into force. The Act targeted law-breakers by introducing higher fines and longer prison-sentences for not complying with health and safety law. In 2010, Lord Young, adviser on health and safety law and regulations undertook a review of health and safety legislation. The review advised the need to improve the public perception of health and safety. It also advised that health and safety be taken seriously by employers and the public, and the need to reduce the burden of bureaucracy on businesses. In 2011, a report by Professor Ragnar Lofstedt was commissioned by Employment Minister Chris Grayling, titled 'Reclaiming health and safety for all: an independent review of health and safety legislation'. The report took into account views of a wide range of organisations and reviewed the available scientific literature on whether the current health and safety regulations were appropriate and consistent. As a summary, it was found that current legislation should not be radically altered, but rather simplified, combined or reduced to create a more efficient process.

2.5.2 Health and Safety Statistics

Although health and safety processes differ across Europe, available data suggests that UK health and safety performance is better than many European countries (HSE, 2015e). The UK is found to be one of the better performing EU countries in carrying out regular risk assessments in work establishments and in 2012, it was reported by the Eurostat that the rate of fatal injuries in the UK was one of the lowest out of all European countries and scores consistently low to the EU average (Eurostat, 2012). Figures of ill-health were also reported to be lower in the UK, however non-fatal injuries caused by workplace accidents differed, with these cases being reported as being at a similar level to other large EU economies in 2013 (HSE, 2015e).

Accidents, injuries and ill-health incidents have overall reduced considerably in Great Britain since the Health and Safety at Work etc. Act 1974 was rolled out (HSE, 2015h). Accidents in the workplace resulting in non-fatal injuries have been declining, and over the last 20 years in particular, the rate of fatal injuries has fallen. Since 2001 to 2002, new cases of ill-health have also fallen, and by 2011-2012 these figures reached an all-time low of 454,000. These records suggest that the health and safety system in Great Britain is performing well, however in more recent years, records of ill-health have risen, with approximately 516,000 new cases being reported in 2014/15 (similarly to the previous 2 years). Also, in the last 10 years, there are signs that the downward trends of fatal and non-fatal injuries are slowing, suggesting that there is still a lot of work needing to be done in improving health and safety standards.

The current figures for work related injuries, fatalities and ill-health are published annually by the HSE, who keeps a record of all incidents that are reported under RIDDOR. Out of all the cases recorded in 2014 and 2015, current figures show that 142 workers were killed at work, which is equivalent to 0.46 fatal injuries per 100,000 workers (HSE, 2015e). Additionally, approximately 76,000 other injuries were reported as occurring in the workplace, which equates to a rate of 293 injuries per 100,000 workers. Ill-health is also a significant problem with around 1.2 million workers in 2014 to 2015 having reported suffering from an illness they believed was caused or made worse by their job (HSE, 2015e). Out of these reported cases, 0.5 million were new conditions that had started within the year, as opposed to longer term conditions. Out of all these new work-related illnesses that were reported, a staggering 80 per cent were related to musculoskeletal disorders, stress, depression or anxiety.

The HSE also publishes results relating to workplace risks such as physical or psychosocial factors. Out of all accidents, injuries and ill-health cases reported to the HSE in 2014, the most common risk factor in the workplace reported was 'dealing with difficult customers, patients, pupils etc.' which was found present in two thirds of workplaces. The second highest reported risk factor by workers was 'lifting or

moving people or heavy loads', which resulted in accidents relating to manual handling and musculoskeletal disorders (HSE, 2015e).

Ill-health and injuries caused by accidents result in workers needing to take time off from their job to recover. It was reported that the total number of working days lost due to ill-health has generally declined since 2000, however the trend appears to be levelling off in recent years. In 2014/15 alone, 23.3 million working days were lost overall to ill-health and 4.1 million days were lost as a result of workplace injuries (HSE, 2015e). Out of all these cases, the average number of days taken off by workers was 15 (or 19 days for ill-health cases and 6.7 days for injuries from workplace accidents). The majority of working days lost were accounted for by stress, depression or anxiety and musculoskeletal disorders.

Loss of working days puts a huge strain on the UK's economy due to both financial costs, involving the loss of production of the worker, and also human costs, in terms of health care requirements of the individual to get back to full health. In 2013/14, in total, injuries at work and new cases of ill health cost society an estimated £14.3 billion, therefore burdening the country financially (HSE, 2015e). Although these costs have generally declined over the last decade, trends again appear to show that this decline is slowing down.

Sometimes breaches of health and safety are found and prosecuted in Great Britain. In 2014/15 alone, 728 of these cases occurred, of which the majority (586 cases) were prosecuted by the HSE in England and Wales. Local authorities prosecuted 70 cases in England and Wales, and 72 cases were found and prosecuted in Scotland. The number of enforcement notices has generally declined over recent years however, with just under 12,500 notices issued by the HSE and local authorities in 2014/15, a decrease of 10% from 2013/14 (HSE, 2015e).

2.5.3 Future Outlook of Health and Safety

The estimated working age population is predicted to increase by 16 per cent from 38.5 million in 2010 to 44.7 million by 2035 (ONS, 2011). This signifies a larger workforce in the future and therefore more individuals potentially being at risk of a workplace accident. As the rates of injuries caused by workplace accidents, and ill-health caused by work have begun to level off in recent year trends, problems will need to be better managed in future decades to avoid a growth in both workplace accident and ill-health statistics.

The HSE carries out research into health and safety practices to better understand health and safety problems and continually keep updated with new and fast changing technologies. The HSE is consistently striving to reduce workplace accidents and cases of ill-health, and aims to minimise future risks of poor health and safety practices into the future. The HSE has a 'Futures Team' which was created to gather and analyse data on future trends and plan for and adapt policy to emerging issues. In addition, as part of their ongoing research, the HSE works closely with the Health and Safety Laboratory (HSL). The HSL, consisting of scientific, medical and technical specialists, provides insight into health and safety in the workplace, by carrying out research into how to best manage risks of accident and ill-health of individuals at work.

The HSE also contract works with private sector companies, consultants, as well as government bodies and academic institutions. Work is carried out at both national and international levels, as well as looking into EU programmes. The HSE publishes a 3 year rolling research plan, which sets out a framework of commissioned research that takes place over the following 3-year period. The current 2015-2018 Science Plan sets out details of work of how HSE's 'use of science' will bring about improvements within the health and safety system. The Plan states the HSE's mission as: 'The prevention of death, injury and ill health to those at work and those affected by work activities' (HSE, 2015c). The key drivers of the Science Plan are:

A. Leading others to improve health and safety in the workplace;
B. Providing an effective regulatory framework;
C. Securing compliance with the law;
D. Reducing the likelihood of low-frequency, high-impact catastrophic events.

The HSE will deliver Part A: leading others to improve health and safety in the workplace, by working closely with individual companies and industry bodies to assist them in driving forward sector-led improvements. Where a problem may exist in particular industries or sectors, the HSE will step in and advise these workplaces on how to improve their health and safety practises. The results of this research project aim to fit into achieving part A by helping to advise the HSE on key problematic areas of health and safety.

To achieve Part B: providing an effective regulatory framework, the HSE aims to continually review policy and legislation to ensure it remains as effective as possible. The HSE is also keen to maintain a simplified structure of its regulatory framework to ensure it is easy to understand and implement. The outcomes of this research project is expected to help drive amendments to existing legislation, where necessary, to improve the effectiveness of the current regulatory health and safety framework. Part C: securing compliance with the law, states that a high proportion of inspections and enforcement action cases whereby health and safety practices have been breached, require particular science or engineering knowledge to identify the causes or problems in these workplaces. This research study aims to investigate and draw conclusions from work related accidents that may help to understand some of these causes and will evaluate findings to help identify solutions for the HSE.

Finally, Part D: reducing the likelihood of low-frequency, high-impact catastrophic events, focuses on the HSE being an effective regulator of major hazard and specialist industries. The HSE aims to continue providing evidence-based scientific and technical advice to industries in controlling hazards, and this research project will aim to support the HSE in this area of knowledge. Where findings indicate that certain industries or workplaces are more at risk due to certain kinds of materials or chemicals, for example, the HSE will be informed to determine how policies can be

managed or changed to focus on improving knowledge and understanding of these risk factors.

2.6 Summary

This chapter has explored the relevant literature surrounding the three aspects of health and safety in which the HSE considers to be important in helping to improve workplace health and safety standards. The categories are: the individual, the job, and the organisation/working pattern. In terms of the individual, social factors were explored as determinants of workplace accidents, specifically age, gender and deprivation level. The types of occupations and industrial sectors worked in, were considered for the next category, followed by physical and environmental factors for the organisation/working pattern category.

Having explored studies based around age and the likelihood of workplace accidents and injuries occurring, the general consensus is mixed in that some researchers find that young workers have a higher risk of work-related accidents compared to older workers (Chau et al., 2014), whereas in contrast others find older workers to have a higher risk compared to younger workers (Xiang et al., 2000). In general, however, much of the research carried out around this topic is based on accident data for European countries outside of the UK, and the US. Little research has currently been carried out to date surrounding Great Britain's labour force and how the age of the worker might influence the risk of a workplace accident.

In terms of gender, as a social determinant of workplace accidents, men are found to be more likely of having work-related accidents than women (Berecki-Gisolf et al., 2015). Overall this is mainly found to be due to the types of occupations in which men are engaged in compared to women. Men are generally found to be involved in more dangerous professions, and work in harsher conditions, therefore increasing their risk of coming into contact with workplace hazards (DeLeire and Levy, 2001). Again, however, similarly to age, as a risk factor, much of the research lacks any

context to Great Britain's workplaces and lacks information as to how problems might be addressed in terms of a gender-gap in health and safety standards.

The occupation and industrial sector of workers, were considered next, where it was found that overall, a significant amount of research is carried out into comparing social classes and health quality (Marmot et al., 1997). Research finds that low-skilled, compared to highly-skilled workers, have poorer health and higher mortality rates. This has been mainly attributed to the types of work in which low-skilled workers are engaged in. The types of sectors which are found to be the most dangerous in terms of hazards in the workplace include: manufacturing and construction (Chi and Wu, 1997; Sorock et al., 1993). Both these industries have a high level of jobs which involve intense physical labour activities, which are considered to be a significant cause of injuries in the workplace.

Environmental factors were explored in terms of understanding how seasonal effects of summer versus winter weather might affect workers' abilities to perform their jobs safely. Variations between times of the year emerged as an underlying determinant of accidents particularly due to poor weather conditions, prevalent in certain months. Winter appears to be a common season for accidents, due to conditions such as rain, snow and ice, increasing individuals' risks of slips and falls (Edwards, 1996). Alternatively, other studies found problems of heat stress, dehydration and lethargy to be a major problem during summer months. Daylight hours are found to be an influential risk factor in causing workplace accidents, with Arditi et al. (2007) finding that night time construction work is 5 times more hazardous than day time construction work. Although some work has been carried out into seasonal and daylight effects, little is known about the risk of accidents in the comparison between occupational groups or industrial sectors throughout the year.

Next, geographic variations in workplace accident risks were explored. The HSE publishes basic statistics which show that there is a variation in regional trends on workplace injuries, moralities and ill-health across Great Britain's workplaces. There is, however, a limited amount of research carried out on workplace health and safety by geographic area, even though disparities appear to exist. Studies by Cameron et al.

(2008) and Davies and Elias (2000), explore some differences between workplace accident risks across Great Britain, but are overall rather limited, either due to small sample sizes, not looking into small-scale areas, or that the results are very out-dated.

Lastly, the chapter provides a context to health and safety policy, revealing how legislation has emerged and evolved over the years. The chapter reports on the current health and safety statistics of Great Britain and the resulting estimated economic costs to society from lost working days due to accidents and ill-health. Although recently published statistics reveal that standards of workplace health and safety are improving, the problem of accidents and ill-health caused by work still remains a significant problem. The chapter concludes with providing an outlook of the future of health and safety practices.

Overall, this chapter has summarised a large range of risk-factors of health and safety in the workplace and it can be concluded that there are some obvious gaps in the literature. These have emerged around there being a lack of research focusing on the UK, and this is most likely due to the restricted access to data surrounding workplace accidents. Social factors in terms of age, gender and deprivation lack context to the UK specifically, and the comparisons across occupational and industrial groups are under researched. No large study, to date has considered seasonal effects on comparing workplace accidents across a range of occupations or socio-economic groups. Lastly, a huge gap in terms of geographic variations in workplace accidents and injuries also exists. The aim of this research project, therefore, is to attempt to broaden the knowledge upon health and safety standards, filling in the gaps that appear to exist in the current literature surrounding workplace health and safety. The HSE produces a Science Plan of how future work will be utilised to inform health and safety policy into the future to minimise accidents, injuries and cases of ill-health caused by work. This research project fits in to all four parts of the Science Plan, in supporting the HSE in its mission to prevent death, injury and ill health to those at work and those affected by work activities.

Chapter 3

Data and Methods Review

3.1 Introduction

This chapter provides a review of the data and methodological approaches used in this research study. The first section focuses on providing a context to the RIDDOR data, and gives an overview of the sample data that were obtained from the RIDDOR database for accident records spanning across Great Britain, providing a breakdown of the variables included within the data. The section also provides a brief overview of the strengths and weaknesses of the RIDDOR data, and how they can reliably be used to explore possible causes of workplace accidents and injuries.

The chapter continues by giving an overview of the open source public data that are used within this research study. It details the different geographical levels associated with data that can be obtained from the 2011 Census. In particular, it examines the availability of workplace population statistics at the new level geography of Workplace Zones. Variables chosen from the Census data to model socio-economic area characteristics for this research study are explained in detail at the end of the section.

The second part of the chapter focuses on the methods used within this research. It provides an overview of the choice of methods being used as compared to alternative methods that are available and details the strengths and limitations of the types of methods chosen. Specifically, it discusses the uses of Generalised Linear Modelling, Geographically Weighted Regression and topic modelling techniques.

These techniques will be applied to data covering England and Wales. The chapter concludes by providing an explanation as to what can be expected from the analyses in the succeeding chapters that follow.

3.2 Data Review

3.2.1 Data Context: Introduction to RIDDOR

It is a legal requirement in Great Britain for all employers and those who are in control of work premises to report accidents that occur in the workplace. This process is a requirement of 'Reporting of Injuries, Diseases and Dangerous Occurrences Regulations' (RIDDOR) 1995 that came into force on the 1st April 1996. RIDDOR was set up through obligations made under the Health and Safety at Work etc. Act 1974 to monitor levels of health and safety in workplaces across Great Britain.

An accident is defined as an unfortunate incident that happens unexpectedly and unintentionally which often results in damage or injury to a person or people. Under the requirements of RIDDOR, major injuries and deaths resulting from accidents at work must be reported to the HSE. For an employer, the key issues to determine whether the accident happens as a result of work are:

1. If the accident was related to the way in which the work was carried out,
2. If any machinery, plant, substances or equipment was used for work in relation to the accident,
3. If the conditions of the site or premises where the accident happened had an effect on the cause of the accident.

(HSE, 2012b)

Injuries fall under two categories: major injuries, and injuries resulting in absence from work for 3 or more days. Major injuries must be reported if they include:

- A fracture, other than to fingers, thumbs or toes;
- Amputation;
- Dislocation of the shoulder, hip, knee or spine;
- Loss of sight (temporary or permanent);
- Chemical or hot metal burn to the eye or any penetrating injury to the eye;
- Injury resulting from an electric shock or electrical burn leading to unconsciousness, resuscitation or admittance to hospital for more than 24 hours;
- Any other injury leading to hypothermia, heat-induced illness, unconsciousness, resuscitation or admittance to hospital for more than 24 hours;
- Unconsciousness caused by asphyxia or exposure to a harmful substance or biological agent;
- An acute illness requiring medical treatment;
- Loss of consciousness arising from absorption of any substance by inhalation, ingestion or through the skin; and/or
- Acute illness requiring medical treatment where there is reason to believe that this resulted from exposure to a biological agent, its toxins or infected material.

(HSE, 2012b, p. 11-12)

Deaths, illnesses and diseases, as a result of work, must also be reported to the HSE under RIDDOR, as well as dangerous occurrences such as near misses. There is an extensive list of dangerous occurrences within 'A guide to the Reporting of Injuries, Diseases and Dangerous Occurrences Regulations 1995' (HSE, 2012b). Such near-misses include for example: the collapse, overturning or failure of load-bearing parts of lifts and lifting equipment, and the accidental release of any substance that may damage health.

As part of RIDDOR, there are some exceptions in the reporting of incidents to the HSE. Injuries resulting from medical or dental treatment, or an examination carried out by or under the supervision of a doctor or dentist is exempt from needing to be

reported under RIDDOR. Similarly, injury caused by a moving vehicle on a road such as a car crash does need to be reported under RIDDOR. In addition to these, any accident carried out under the duties by a member of the armed forces whilst on duty also qualifies for exemption under regulation 10 of RIDDOR.

Originally, the employer or the person in charge reported an event by completing a paper copy RIDDOR form, or by telephoning the HSE. Seven different RIDDOR forms exist to choose from, with the first option being the focus of this investigation:

1. report of an injury,
2. report of a dangerous occurrence,
3. report of an injury offshore,
4. report of a dangerous occurrence offshore,
5. report of a case of disease,
6. report of a flammable gas incident,
7. report of a dangerous gas fitting.

The questions within the form covered personal details, such as name and address of the injured person, and the organisation in which they worked. The questions also related to the incident, where it happened, what caused it to happen, and the resulting injury. The HSE classifies the reported occupation and industry worked in of the injured person by the Standard Occupation Classification (SOC) and Standard Industrial Classification systems (SIC), which are continually updated to match the most recent systems (SOC 2010 and SIC 2007).

There is a large amount of information stored within the RIDDOR database, providing a high level of detail surrounding workplace accidents; there are however some limitations to the data stored within the database. Under the framework of RIDDOR not all incidents are reportable, those that are may be mistakenly classed under the wrong type of injury during the reporting process. An example of this would be a fractured arm potentially being selected as an over 3-day injury, whereas it should have been classed as a major injury. Issues in the data also could have

occurred with matching occupations and industries to the relevant SOC and SIC groups. Lastly, missing data is also an issue in that some fields are left blank by the notifiers of the accident, possibly because the answers were unknown at the time the form was completed.

The process of recording the incidents received by postal forms and telephone calls was found to be a very expensive and time-consuming process for the HSE. In 2011, an online form replaced the postal forms and telephone calls, with similar questions and formatting. There are 5 pages of the online form regarding the company, the accident and the injured person and the questions vary in terms of open and closed ended questions. Formats of the form differ in that rather than free-text fields on the paper form for employers to fill in for questions such as occupation, type of injury and body part injured, the online form contains drop down lists of multiple choice answers to choose from.

Since April 2012, the 'over-three-day' absence from work injuries was replaced by the 'over-seven-day' absence from work injury. Where an employee or self-employed person is away from work or unable to perform their normal duties for more than seven consecutive days, the incident must be reported to the HSE. Previously this procedure took place for those who were injured for more than three days. Over-three-day injuries do not now need to be reported but must be kept a record of within an accident book by the employer. One other recent change is that the 'material agent' field in the database which stated the material or substance involved with the accident has been removed from the online system.

3.2.2 The RIDDOR Sample Data

There is a wealth of unique information stored within the RIDDOR database including names, addresses, ages and information in relation to injuries sustained and the type of work that was being performed when an accident occurred. The only similar RIDDOR database that exists covers Northern Ireland where workplaces report to a separate Health and Safety Executive body. For this research, the HSE

covering Great Britain has provided access to a large sample of the RIDDOR database. The sample consists of approximately 822,000 records spanning across the years 2006-2011, with incidents recorded from months April to March consecutively for each year time period. For 2011, the data extracted from RIDDOR end at September due to the change in the format of the RIDDOR form to electronic. The dataset was transferred as six separate text-formatted files for ease of transfer across a secure network between the HSE and the University of Liverpool. Due to the sensitive nature of information in the RIDDOR database, prior to obtaining the data, the HSE omitted several fields from the sample. These data fields included details relating to personal information of the employer, workplace and injured person; details included: names, personal addresses, telephone numbers or email addresses which could pin point a particular incident or injury to a person or workplace.

When the sample was obtained, the data first needed to be cleaned and reformatted to create a uniform dataset that could be analysed easily and efficiently. The variables were extracted from the database by the HSE in a randomised order, potentially due to changes in the format of the RIDDOR form and the way in which records were stored. It was important therefore to ensure all fields matched up correctly between each of the six RIDDOR tables that were transferred across from the HSE, and had clear variable names so as to be able to distinguish exactly what data was stored within each column.

As the sample used for this investigation consists of mainly paper form notification, an issue with this was that those who inputted the information from the form at the Health and Safety Executive had to choose the most relevant answer which closely resembled the answer given on the paper RIDDOR form. Having someone else input and submit the electronic form introduces human error to the results. An employer filling out a form to be submitted straight into the database would have witnessed the incident first hand or gained knowledge of the incident immediately from a colleague or worker. Additionally, almost any incident recorded through the RIDDOR process is added to and stored in the database. The exceptions being if the incorrect form was filled out or if a report did not need to be made in the first place. Occasionally, non-reportable accidents are reported to the HSE. Although the HSE

attempts to identify non-reportable incidents such as road traffic accidents, workplace complaints and natural causes for example, there may be some that have slipped through onto the database. Any exact duplicate records were removed from the sample dataset once obtained, and entire missing rows were sifted out from the dataset. Incomplete rows were checked for mistakes and reordered where necessary.

The HSE provided lookup tables for some variables that were coded, for example injury type was categorised in a table and labelled as numbers such as '2' indicating amputations, '3' as loss of sight, etc., and body part injured similarly provided in a table coded by numbers such as '1' indicating eye and '2' indicating ear, and so on. An issue with this has been attempting to identify the appropriate lookup table corresponding to the correct field within the dataset. It was found that there were more lookup tables than there were questions and the lookup tables did not match the name of the variables. The correct tables needed to be matched with the correct relevant fields to avoid ambiguity. The coded answers therefore had to be unravelled and matched through an ad-hoc process to the lookup tables and with help from the HSE. Cleaning and formatting the data was time-consuming; however, to ensure the results obtained from any analysis work would be reliable, a systematic approach to data processing was essential.

The finalised structure of the acquired RIDDOR dataset consists of 30 variables (Table 3.1). The dataset holds a mixture of categorical and continuous variables, and includes: times, dates and address data, as well as short free text fields detailing for example the injury sustained by the worker. The variables are grouped into the following seven categories: administration records, workplace/accident location details, date/time of accident, type of workplace, details of the accident, details of the injured person, and type of injury.

Table 3.1: Variables contained within RIDDOR dataset

Section	No.	Variable	Description
A. Administration Records	1	**Incident Number**	The Incident Number is a unique number assigned to an individual incident by the HSE for reference in its database.
	2	**Date Created**	The Date Created variable is a record of the date each incident was recorded into the database (not necessarily the date of the incident).
B. Workplace/ Accident Location Details	3	**Company Name**	The Company Name is the name of the company/business of where the person involved in the incident works.
	4	**Company Address**	The Company Address is the address of the workplace of the injured person.
	5	**Company Postcode**	The Company Postcode is the postcode of the workplace of the injured person.
	6	**Whether or not the accident took place at company location**	This variable consists of 'yes' or 'no' answers about whether accident occurred at company address.
	7	**Enforcing Body**	The Enforcing Body is the HSE or Local Authority.
	8	**Accident Location Address**	The Accident Location Address is a field completed giving the address of the incident if different to the Company Address.
	9	**Accident Location Postcode**	The Accident Location Postcode is the postcode completed together with Accident Location Address, if the incident occurred at a location different to the Company Address.

C. Date/ Time of Accident	10	**Date of Accident**	The Date of Accident is the date the accident occurred in the format Day/Month/Year.
	11	**Year of Accident**	This is a separate variable of the year that the accident occurred.
	12	**Time of Accident**	The Time of Accident is the time the accident occurred recorded in 24-hour clock format.
D. Type of Workplace	13	**Enforcing Local Authority**	The Enforcing Local Authority is the Local Authority based on the area where the accident occurred.
	14	**Location on Premises**	This is a record of where on site or premises the accident/injury occurred.
	15	**SIC**	The SIC is the Standard Industrial Classification broad industry group of the type of workplace where the accident occurred.
E. Details of Accident	16	**Cause of Accident**	The Cause of Accident consists of pre-defined answers that were selected at the time of the recording of the incident. The answers consist of numeric values that can be matched to a look up table indicating causes of accidents, such as 'slip' or 'trip'.
	17	**Work Process**	The Work Process also consists of pre-defined answers which, when matched to a look up table, gives the type of work process involved in the accident. An example would be 'Manufacturing production and processes'.
	18	**Main Factor Involved**	The Main Factor Involved consists of pre-defined answers consisting of various material agents that were considered to be involved in an accident. An example is 'hand-held tools and equipment'.

	19	**Type of Work**	This variable is a description of the type of work that is carried out at the workplace.
F. Details of Injured Person	20	**Gender of Injured Person**	This variable gives an 'F' for female and 'M' for male for the gender of the injured person.
	21	**Age of Injured Person**	The Age of Injured Person is the age in years of the person involved in the accident.
	22	**SOC**	The SOC is the Standard Occupation Classification broad group of the injured person.
	23	**Job Title of Injured Person**	This variable is the job title of the injured person as stated by the notifier of the incident.
	24	**Status of Injured Person**	The status of the injured person consists of a selection of pre-defined answers such as 'the injured person was one of my employees' and 'the injured person was on a training scheme'.
G. Type of Injury	25	**Type of Injury**	This variable consists of answers from multiple choice answers such as 'burn' or 'strain'.
	26	**Injury Text**	This variable is a free-text description of the injury.
	27	**Body Part Injured**	The body part injured consists of answers from multiple choice answers such as 'hand' or 'neck'.
	28	**Body Part Injured Text**	This variable is a description of the injured body part.
	29	**Severity of Injury**	This variable consists of selections from pre-defined answers of injury severity such as 'major injury'.
	30	**Details of Severity**	This variable gives a free-text description of the details of the severity of the injury.

Clearly, there is a considerable volume of data within the RIDDOR dataset, which holds a significant amount of information about accidents and injuries that have occurred within Great Britain's workplaces. Several variables are solely for administration purposes such as incident number and date of recorded entry. However, the majority of the data within the RIDDOR dataset are useful for this investigation and approaches to the analysis of these elements will be discussed in detail in the second part of the chapter (3.3 Methods).

Variables 3-5 consist of the address of the workplace of the injured person and Variable 6 states whether the incident occurred at the location of this address. If the answer was 'No' to this, then the location of the accident is stated (Variables 8 and 9). The dataset therefore provides an address of where a particular incident took place, and gives a reliable geographic reference for spatial analysis.

Variables 10-12 relate to the time and date of the incident. These are a useful indicator for each incident, as they can pinpoint whether incidents occurred during particular seasons, and during the night or daytime. These variables open up a wide range of possible research into the physical factors that may be the cause of workplace accidents. Included within the data is information about the type of job role of the injured person, as well as the SIC and SOC broad group codes for the industry in which the injured person worked, and also the occupation of that person (Variables 15 and 22).

Also included are the gender and age of the worker (Variables 20 and 21). These variables provide details specifically about the type of worker, which help to provide some insight initially to the types of people having accidents in the workplace, forming the basis of the exploratory analysis chapter. The remaining variables provide information in relation to the incident, for example Variable 14 describes where specifically on site the incident occurred, and Variable 16 indicates the cause of the accident. Variables 25 to 30 focus on the injury sustained by the injured person, the type of injury, severity and body part injured. Various free-text fields exist within the dataset whereby the person notifying the HSE of the incident states the job title of the injured person and the location of the incident on the premises.

A second dataset was obtained from the HSE shortly after the full initial RIDDOR dataset was obtained. This dataset contained free-text fields of descriptions of accidents for approximately 43,780 incidents, that were recorded over a 6-month period in 2011 (1st April to 22nd September). These descriptions were written by the notifier of the incident, therefore providing first hand evidence of what actually happened at the scene. Details range from basic descriptions such as the worker tripping over something to including a more significant amount of detail such as how the accident occurred, whether an object was involved in the accident, how severe the injury obtained was, and details for example, of the working conditions or types of weather conditions when it happened. Also included within the dataset is the unique Incident Number, so that incidents can be matched to the first RIDDOR dataset.

The free-text RIDDOR dataset provides a vast amount of additional information to the full RIDDOR dataset, that has not previously been analysed in great detail. Because of this, the raw data needed to be cleaned and formatted correctly before any work could be carried out. One limitation of the free-text RIDDOR dataset is that it only covers accidents within a short time frame. This was what was available from the HSE at the time of the study. Nonetheless, the dataset provides an excellent source of information for investigating further details of workplace accidents.

In summary, the RIDDOR data obtained for this study provide an enormous amount of information on workplace accidents that is completely unique and has not been analysed in great detail previously. This research study will explore the type of workers having accidents, based on the information included within the data such as the worker's gender, age, occupation and the industry worked in. Having the locations of the incidents allows all the data to be easily mapped and matched to other sourced datasets. Also, having dates and times of incidents allows the data to be temporally analysed. The textual data, which provide details of the incidents in the words of the injured person or employer, also provide an additional source of information surrounding the specific details of reported incidents that may provide further insight to the analysis.

3.2.3 Census Data

Each incident within the sample RIDDOR dataset has an address record indicating where the accident occurred. Based on this, open source data can be attached to the dataset via a geographic location to facilitate greater understanding of the demographics of the area in which accidents are occurring. The census, which collects information about the population of the UK every 10 years, provides a comprehensive picture of the nation, allowing comparisons to be made across areas locally and nationally. The census is an extensive source of information about the current population of the UK, which makes it a good starting point for exploring the characteristics of areas in which people live and work.

There are many benefits of applying information collected from the census to this study. Most of the data collected from the recent census is available and easily accessible online having been published by the Office for National Statistics (ONS) (for data relating to England and Wales) and National Records of Scotland (NRS) (for data relating to Scotland). The most recent census data relates to the census survey carried out in 2011 which is useful for this study, as the RIDDOR dataset relates to accidents that occurred between the years 2006 to 2011. It would be appropriate to look at the current workplace population alongside the RIDDOR workplace accident data. The workplace population is defined by the ONS as the population whose usual place of work is in a particular area. People who work mainly at or from home or do not have a fixed place of work are included in their area of their usual residence.

The level of geography that would be considered most useful to model the data is Workplace Zones (WZ), which is a new geography output created by the ONS using data based on workplaces in England and Wales. WZs have been produced using the results from the recent 2011 Census and were designed similarly to Output Areas and Lower/Middle Super Output Areas in that they are based on grouping people into areas. Unlike Output Areas and Lower/Middle Super Output Areas which were designed to contain consistent population numbers based on where people live, WZs have been designed to contain consistent numbers of workers, based on where

people work. One of the main problems with WZs currently however, is their lack of availability with associated Census workplace population data across the country, specifically for Scotland and Northern Ireland. The ONS have released 2011 Census workplace population data tables for England and Wales by WZ relating to a variety of socio-economic area characteristics, however these are currently unavailable by WZs for Scotland by the NRS. For this research, it was therefore necessary to narrow the focus of the study to England and Wales rather than Great Britain in exploring the associated socio-economic area characteristics with workplace accident rates.

Based on the literature review (Chapter 2), a set of variables were chosen from the Census workplace population tables to assess patterns in accident rates. These variables were selected based on findings from the relevant literature on accidents, as possible causes or increased risk factors associated with having accidents in the workplace. The variables are outlined with the relevant dataset names in Table 3.2 below and include: sex, age, occupation, industry, socio-economic class and population density.

Table 3.2: 2011 Census data variables

Variable	Name of Dataset	Census Dataset Code
Gender	Sex by single year of age	WP1101EW
Age	Sex by single year of age	WP1101EW
Industry	Industry	WP605EW
Occupation	Occupation	WP606EW
Socio-Economic Class	NS-SEC	WP607EW
Population Density	Population Density	WP102EW

Gender

The Sex by single year of age (workplace population) 2011 Census dataset provides information on the sex of the workplace resident population in England and Wales

in 2011. It indicates the number of males and females in each small area (in this case WZ).

Age

The Sex by single year of age (workplace population) 2011 Census dataset also classifies the workplace population by single year of age by WZ. This dataset can be used to group people together in similar age groups, calculated as proportions out of the working age group population, to make comparisons between younger and older workers having accidents in the workplace.

Industry

The Industry (workplace population) 2011 Census dataset provides estimates that classify the workplace population by industry in which they work. A code is assigned to each industry type that an individual works in using the Standard Industrial Classification of Economic Activities (SIC). The broad groups will be used for this study to classify people into industry in which they work.

Occupation

The Occupation (workplace population) 2011 Census dataset provides estimates that classify the workplace population by occupation. A code is assigned to each occupation that an individual works in using the Standard Occupation Classification of Economic Activities (SOC). Coding systems for both major groups and minor groups are contained within this dataset however for this investigation major groups will only be used to structure the occupation types to create a more compact dataset for analysis. Examples of major groups include: 1: Managers, directors and senior officials and 2: Professional occupations.

Socio-Economic Class

The NS-SEC (workplace population) 2011 Census dataset provides estimates that classify the workplace population by the National Statistics Socio-Economic Classification. These statistics are used to help in developing policy aimed at the needs of the population according to their socio-economic class. The NS-SEC hierarchy ranges from NS-SEC 1 Higher managerial, administrative and professional

occupations to NS-SEC 7 Routine occupations and NS-SEC 8 Never worked and long-term unemployed.

Population Density

The Population Density (workplace population) 2011 Census dataset provides estimates for the density of the workplace population of England and Wales in 2011. Population density will help to provide an understanding as to how populated an area is, and whether it is likely that accidents occur more in urban areas as opposed to rural areas.

3.3 Methods

3.3.1 Introduction

The first part of this section focuses on the software used for carrying out the analyses for this research. The second part discusses the methodologies of the analyses. There are three main methods of analysis used as part of this study. These methods include: global regression analysis, local regression analysis and text mining. The global regression analysis is carried out as part of creating a national model of the socio-economic risk factors associated with workplace accidents and injuries. The local regression analysis is carried out to explore the geographic variations in the relationships between these risk-factors and workplace accident rates, and finally the text mining as well as topic modelling are explored next as methods for analysing the free-text RIDDOR dataset. The theory surrounding the use of these methods are individually examined, together with their application within software and examples of their use within the relevant literature.

3.3.2 Use of Software

Data processing based in R (R Core Team, 2015); has many advantages. It is the most broad-ranging statistical analysis software currently available, capable of

working with large and multiple datasets. It is a programming language that can be applied to most numerical work and incorporates up to the most advanced level of statistical testing and modelling. Unlike many other statistical software, with R being open source, it allows anyone to install and access it on any computer. This means that it is an inexpensive software to run and therefore can be used by anyone. Users are able to provide modifications and new packages to R, which makes the system extremely useful, up to date and comprehensive. With currently thousands of packages available ranging from techniques for plotting graphs and drawing maps to carrying out clustering analysis and statistical modelling, R has the considerable benefit of being a multi-discipline software environment capable of carrying out work in multiple research areas.

R works well with incorporating and working efficiently with different file types. For this investigation, this is particularly useful as the HSE carries out a lot of its analysis of RIDDOR on Microsoft Access and the datasets were transferred as .txt files, which R reads easily. Alternatively, other software packages such as SPSS and SAS do not match the range of capabilities of R. Although software such as SPSS provides easy-to-use methods of analysis, allowing the user to carry out simple calculations and statistical testing with the click of a few buttons, it fails to deliver the interactive features that R provides. With R, the user can modify graphics for exportation as precisely as needed. The user can implement advanced quantitative methods in R, whilst most other statistical software have limited statistical capabilities.

For carrying out GIS analysis, R has extensive capabilities. It has the flexibility to work with large amounts of spatial data and provides tools via multiple packages to model, analyse and visualise spatial data. In contrast, other GIS software such as ArcGIS and QGIS, although providing a good starting point for carrying out spatial analysis, have all found to be inflexible and without functions to implement some of the methods being used in this study. R is beneficial in many ways in research as work is reproducible using the command line, so if some form of analysis needs to be computed again and again, this is easily and quickly done within R.

3.3.3 Global Regression Modelling- Generalised Linear Models

Previous research suggests that that there may be an increased risk of accidents for people within certain socio-economic groups. The question is which socio-economic factors have an effect on workplace accidents and what effect do they have? For example, does a worker's gender, age or occupation effect their risk of accident or injury at work? The starting point to answering these questions is generally to consider a form of regression modelling whereby a selection of variables is chosen that may have an effect on the dependent variable. Regression analysis is a statistical method for investigating the relationship between a set of variables. Usually, the case involves seeking to find the effect of a set of variables on another, in this study- the effect of workplace accidents in an area based on particular socio-economic characteristics of the people who work within that area.

Regression modelling is widely used in studies of socio-economic characteristics. Examples include, the effect of the level of education upon income inequality and mortality (Muller, 2002), the effect of poverty and income inequality on crime (Hsieh and Pugh, 1993) and the effect of socio-economic area characteristics such as ethnicity, age and ACORN group (a classification system for neighbourhood characteristics) on quality of health (Pickett and Pearl, 2001). For this study, the variables have been selected from the recent 2011 Census and the dependent variable is the frequency of accidents and injuries from the RIDDOR dataset.

The 2011 Census tables and the variables selected from them for this part of the research have been described in detail in the previous section in this chapter. They include: age, gender, occupation, industry, socio-economic group and population density, weighting against the total working population. Previous research has suggested that these socio-economic area characteristics may have an effect on accident rate, and therefore this hypothesis is to be tested through the use of regression analysis.

The dependent variable, workplace accidents and injuries, is a count variable. To have an accident in the workplace is considered uncommon due to the fact that

general health and safety policy in the workplace has improved over the last few decades in England and Wales, particularly due to the establishment of the HSE. Given that this is the case, accidents resulting in workplace injuries and ill-health can be thought of as rare events, occurring infrequently and to only a small percentage of the population. To model count data such as this, Generalised Linear Models (GLMs) are used as the basic regression models (Nelder and Wedderburn, 1972). The GLM was developed to enable regression models to be fitted to univariate response data that follow an exponential distribution including for example: the normal, binomial, Poisson and negative binomial distributions.

GLMs describe the dependence of response variable y_i $(i = 1, \dots, n)$ on the variable x_i through a linear exponential family with probability density function:

$$f(y; \lambda, \phi) = e^{\left(\frac{y \cdot \lambda - b(\lambda)}{\phi} + c(y, \phi)\right)},$$

where λ and ϕ are known parameters and $b(\lambda)$ and $c(y, \phi)$ are known functions. The conditional mean is given by $\mathrm{E}[y_i | x_i] = \mu_i = b'(\lambda_i)$ and variance $\mathrm{VAR}[y_i | x_i] = \phi \cdot b''(\lambda_i)$.

Within the R package **stats** (R Core Team, 2015), the function **`glm()`** can be used for implementing GLMs, with the most useful arguments being:

```
glm(formula, family = gaussian, data, weights, subset, na.action, ,
start = NULL, offset, control = list(…), model = TRUE, x = FALSE, y
= TRUE,  ...)
```

For full details of the arguments of the **`glm()`** function see `?glm`.

The general starting point in modelling count data is a Poisson regression model. Poisson models have a long history of modelling accident data with as early as Bortkiewicz (1898)'s study applying the Poisson model to accidents involving mule-kicks in the Prussian army to more recent others including modelling airline

accidents (Rose, 1990) and road traffic accidents (Amin et al., 2014; Polus and Cohen, 2012).

Poisson regression is a special case of the GLM framework. The probability density function is given by:

$$f(y;\mu) = \frac{\mu^y e^{-\mu}}{y!},$$

with the mean and variance being identical, thus dispersion is fixed at 1. In R, again this can be computed using the **`glm()`** function, however, the family argument this time must be changed to `family = poisson`.

The mean being equal to the variance is a strict criterion often limiting the use of Poisson regression analysis in research because empirical count data typically exhibit overdispersion (where the variance is greater than the mean) and/or an excess number of zeros. To deal with the issue of overdispersion for counts, several methods have been developed which include the negative binomial model, quasi-Poisson model (Wedderburn, 1974), generalised Poisson (Consul, 1989) and zero-inflated models (Lambert, 1992).

There are differences between each of the alternative models to the Poisson regression model, which are discussed in (Joe and Zhu, 2005) and (Lord et al., 2005). However, as evaluated by Ver Hoef and Boveng (2007), despite these developments and alternative methods, the quasi-Poisson and negative binomial models remain the most widely used alternatives to the Poisson model. Both methods are more readily available in software packages for easier and efficient computation.

The quasi-Poisson model is a variant of the Poisson regression model. The mean and variance functions remain similar as in the Poisson model but the dispersion parameter ϕ is unrestricted and estimated from the data. This allows the estimates of the quasi-Poisson model to be identical to the Poisson model, however inference is adjusted for over-dispersion. Quasi-Poisson models can be fitted similarly to the

Poisson model using the **`glm()`** function in R and setting the family argument to `family = quasipoisson`.

The negative binomial regression model can also be considered as a generalisation of Poisson regression since it has the same mean structure as Poisson regression but it has an extra parameter to model the over-dispersion. The result is that if over-dispersion is present in the data, the confidence intervals for negative binomial regression models are likely to be smaller than those from a Poisson regression. To carry out negative binomial regression in R, the **MASS** package (Venables and Ripley, 2002) is required and the function is **`glm.nb()`**.

It can be a difficult task deciding the most appropriate model to use, with Amoros et al. (2003) opting for negative binomial regression in their road crashes study, and Ma et al. (2014) choosing quasi-Poisson for their accidents based study. One way of simplifying the choice of models is to draw comparisons and consider information theoretic approaches to help identify the most suitable model. The Akaike Information Criterion (AIC; Akaike, 1973) can be used to compare alternative models. Minimising the AIC allows for a trade-off between goodness-of-fit and degrees of freedom. As a general rule, the model with the lowest AIC is the preferred model.

Chapter 5 will focus on modelling the set of variables using a Poisson regression analysis in R. If over-dispersion is present within the data, having identified whether the variance exceeds the value of the mean, then quasi-Poisson and negative binomial models will be fitted to attempt to rectify. The AIC, which will be produced as part of the output in R, will be compared between models and the most suitable model will be chosen, using this as a guide. Chapter 5 will detail the specific methods used and draw comparisons between each, before selecting the final model which is considered to most suitably represent the data and model workplace accident rate.

3.3.4 Local Regression Modelling- Geographically Weighted Regression

The Poisson regression model, as outlined above, can be thought of as a global model, in that the relationships between the set of chosen coefficients on accident rate are representative of workers in England and Wales as a whole. In a typical regression analysis study, the parameter estimates contained within the output of the model are discussed, reflecting the relationship between the independent and dependent variables, based on the sign and magnitude of the parameter estimates. The parameter values imply that the relationship between variables are consistently the same across the country. The issue with global modelling, however, is that in many real-world cases, relationships between variables vary spatially, therefore meaning the global model lacks sufficient detail in many studies. Working in a particular industry may be riskier in one area than another, or different occupations might carry different risks in some areas compared to others. Local regression modelling would explore the relationship between these.

Global modelling provides single-valued statistics, representative of an entire region under study, however local modelling provides multi-valued statistics, with a different value given for parameter estimates in different locations within the study area. Local modelling adds considerable value because it opens up the ability to explore spatial patterns within the data, allowing data to be mapped using GIS software. To put it simply, global statistics highlight spatial similarities, whereas in comparison, local statistics highlight spatial variations. It is possible to identify potential clusters or 'hot spots' through fitting local models that indicate exceptions to the norm. As a result, local statistics may add value to a study that is missed in standard regression modelling. Chapter 6 explores a method of spatial modelling called Geographically Weighted Regression (GWR; Brunsdon et al., 1998; Fotheringham et al., 2002)). GWR uses a local weight matrix, which is calculated from a kernel function that places more weight on locations that are situated closer to each other. GWR essentially creates local regression models at each point in a space by using a weighting scheme that includes characteristics of neighbouring locations on the construction of each model (Fotheringham et al., 2002).

GWR has been used in many different research fields to model spatial data. For example, Cahill and Mulligan (2007) use GWR to support the idea that crime is a local occurrence and that local contexts should be considered in studying crime rates. Fotheringham et al. (2001) also explore the geography of school performances using GWR to model a range of socio-economic area characteristics across northern England. Spatial variations are found to exist between school areas and between socio-economic groups. Other applications of GWR include: investigating spatial variations in house prices in London (Lu et al., 2011), and examining the spatially varying relationships of demographic and socio-economic characteristics of individuals on the risk of developing cancer from exposure to hazardous pollution (Gilbert and Chakraborty, 2011). In all cases, GWR builds upon the existing global regression modelling techniques by exploring the local trends found within the data.

Other spatial modelling techniques exist including for example: multilevel modelling, the spatial expansion method and Bayesian spatial modelling; however, each of these techniques have their own limitations. Multilevel modelling for example, is an approach that handles clustered or grouped data, based on particular hierarchies. Multilevel modelling has become popular in spatial analyses where geographically referenced data often have hierarchies whereby people are nested into for example small-scale neighbourhoods, which are then nested into larger council district areas, which are further nested into cities and regions. Examples include using multilevel modelling to measure segregation at various levels: school-cohorts within schools within local authority areas (Leckie and Goldstein, 2015), and measuring variations in levels of health at different geographic levels: individual and regional levels (Arcaya et al., 2012). Multilevel modelling has the flexibility of not being limited to two hierarchical levels, but can be extended to however many are relevant to the analyses. Additionally, multilevel modelling could incorporate temporal or a mixture of temporal and spatial hierarchies, making it especially useful for longitudinal studies. One of the main limitations of multilevel modelling however, is that it usually assumes that the nature of the spatial process being modelled is discontinuous. Often spatial processes do not operate this way because the effects of spatial processes in many contexts are continuous.

Another alternative approach is the expansion method (Casetti, 1972, 1997; Jones and Casetti, 1992), which attempts to measure parameter 'drift'. It allows the parameter estimates to vary locally by making the parameters functions of other attributes, including location. This essentially allows trends in parameter estimates to be measured, and is a method that has proven important in highlighting the concept that relationships between variables might vary over space (Brown and Jones, 1985; Fotheringham and Pitts, 1995). In terms of limitations, the expansion method is restricted to revealing trends in relationships over space with the complexity of the measured trends being dependent upon the complexity of the expansion equations. This in effect, might lead to the spatially varying parameter estimates obscuring important local variations to the broader trends represented by the expansion equations (Fotheringham et al., 2000).

Bayesian spatially varying coefficient (SVC) models (Gelfand et al., 2003) are also an alternative method to GWR. SVC models define spatial correlation in the regression coefficients through a prior conditional specification of the coefficients using only neighbouring observations. The main advantage of SVC models is that they provide a valid probability model from which full posterior inference for all model parameters and further predictions can be made. SVC models are often challenging to fit from both an analytical and computational perspective. Users require experience with fitting hierarchical Bayesian models, and there is a general lack of available software to fit them. GWR provides an excellent alternative to this method, being easier to understand and being a simple method to implement in a range of different software.

Geographically weighted regression (GWR), the spatial modelling technique implemented in Chapter 6, is a method which builds upon traditional regression techniques by allowing local rather than global parameter estimates to be calculated. It is an ideal follow-on to the standard regression method utilised as part of creating a global model in Chapter 5. GWR can also be particularly useful in policy studies such as this research study. Different occupational health and safety preventative measures may be appropriate in different areas. GWR will help to highlight those areas where tailored health and safety policies should be targeted.

The idea of GWR is to fit a regression model of the form:

$$y_i = \sum_k \beta_k (u_i, v_i) x_{k,i} + \varepsilon_i$$

where y_i is a measured attribute at a location i, (u_i, v_i) are the coordinates of the location, $x_{k,i}$ is the kth predictor variable associated with the location, ε_i is a random normal variable with mean zero and variance σ^2 and $\beta_k(u_i, v_i)$ are varying conditionals on the location. GWR works by moving a search window from each point in a set of data points, to the next. A set of regions is defined around each point and a regression model is fitted within those regions, giving most weight to the points closest to the centre of the window. Similarly to the global regression model, a Geographically Weighted Generalised Linear Model (GWGLM) framework has been developed. In particular, for this study, a Geographically Weighted Poisson Regression (GWPR) is necessary.

A GWPR model is shown as:

$$y_i \sim \text{Poisson}\left[N_i \exp\left(\sum_k \beta_k (u_i, v_i) x_{k,i}\right)\right]$$

where y_i is a measured attribute at a location i, (u_i, v_i) are the coordinates of the location, $x_{k,i}$ is the kth predictor variable associated with the location, $\beta_k(u_i, v_i)$ are varying conditionals on the location and N_i is the offset variable at the ith location.

GWR models can be developed in different packages and software including: the GWR4 standalone Windows-based application (Charlton, Fotheringham and Brunsdon 2016), R packages such as **spgwr** (Bivand et al., 2017) and **GWmodel** (Lu et al., 2017), and ArcGIS GW regression tool in the Spatial Statistics Toolbox (ESRI, 2013). Each of the packages or applications have their own strengths and limitations. GWR4 was developed as a platform equipped to implement semi-parametric geographically weighted regression. It also offers a wider range of options related to GWGLM, including geographical variability assessment and the automated variable

selection routines (Brunsdon and Singleton, 2015). Gollini et al. (2015) provides a framework for implementing GWR methods. For example, **GWmodel** provides a more extensive set of GW modelling tools providing functions to conduct: a GWPCA (Geographically weighted principal components analysis; Fotheringham et al. (2002)), GW regression with a local ridge compensation (for addressing local collinearity), mixed GW regression, heteroskedastic GW regression, a GW discriminant analysis, robust and outlier-resistant GW modelling with a wide selection of distance metric and kernel weighting options.

The GWPR models are developed for this research study using the GWR4 software which is an open source software (available to download from: https://geodacenter.asu.edu/software/ downloads/gwr_downloads) that was developed by the GWR4 Development Team, consisting of several scholars including: Brunsdon, Fotheringham and Charlton, the creators of GWR. GWR4 is used as it is a simple tabbed interface enabling GWPR to be developed in an efficient step-by-step process. The first tab involves importing the data, the second involves defining the independent and dependent variables, the third tab involves setting the geographic kernel type and its bandwidth size (automated optimisation of bandwidth size is also available), and steps four and five involve running the model and saving the output files.

Once the data has been imported and the independent and dependent variables have been defined, the kernel type and bandwidth size must be determined. There are four kernel type options available to choose from in GWR4, outlined in Table 3.3. For this research, adaptive bi-square is used as it provides flexibility at a smaller scale to account for urban/rural trends. By selecting a fixed kernel, the geographic extent for local model fitting to estimate local coefficients is constant over space. In contrast, the adaptive kernel controls the kth nearest neighbour distance for each regression location. The two classic options out of the four are: Gaussian fixed kernel and Adaptive bi-square kernel. The Gaussian kernel weight function gradually decreases from the centre of the kernel, but does not reach zero. The bi-square kernel has a clear-cut range where the kernel weighting is non-zero and is more suitable for clarifying local extents for model fitting.

Table 3.3: Kernel types in GWR4

Fixed Gaussian	$\boldsymbol{w_{ij}} = \exp(\frac{-\boldsymbol{d_{ij}^2}}{\boldsymbol{\theta^2}})$
Fixed bi-square	$\boldsymbol{w_{ij}} = \begin{cases} (1 - d_{ij}^2/\theta^2)^2 & d_{ij} < \theta \\ 0 & d_{ij} > \theta \end{cases}$
Adaptive bi-square	$\boldsymbol{w_{ij}} = \begin{cases} (1 - d_{ij}^2/\theta_{i(k)})^2 & d_{ij} < \theta_{i(k)} \\ 0 & d_{ij} > \theta_{i(k)} \end{cases}$
Adaptive Gaussian	$\boldsymbol{w_{ij}} = \exp(\frac{-\boldsymbol{d_{ij}^2}}{\theta_{i(k)}^2})$

Where i is the regression point index; j is the locational index; w_{ij} is the weight value observation at location j for estimating the coefficient at location i. d_{ij} is the Euclidean distance between i and j; θ is a fixed bandwidth size defined by a distance metric measure and $\theta_{i(k)}$ is an adaptive bandwidth size defined as the kth the nearest neighbour distance.

(Nakaya et al., 2016, p.23)

The key controlling parameter in all kernel function types is the bandwidth. In terms of bandwidth selection, the user can set this manually in GWR4, or the 'golden selection search' can automatically determine the optimal bandwidth size. Bandwidths can be specified as either a fixed distance (eg. kilometres) or as a fixed number of local data (ie. adaptive distance, incorporating a fixed number of, in this case, WZs). The golden selection search finds the optimal bandwidth size by comparing model selection indicators with different bandwidths. These indicators include: the Akaike Information Criterion (AIC; Akaike (1973)) and AICc (small sample bias corrected AIC), and the Bayesian Information Criterion/minimum description length (MDL) and cross validation (CV) (Nakaya et al., 2016).

After executing the program in Step 5 in GWR4, the output is generated which is a csv file containing a set of location-specific parameter estimates. Also generated includes a results summary output file which contains descriptions of the model settings, optimal bandwidth size and model diagnostic information of the GWR

model. The parameter estimates can be analysed and explored by using GIS mapping techniques. Through visualisation methods, the relationships between the variables will be explored and any spatial trends can be investigated.

3.3.5 Text Mining and Topic Modelling

The final part of the analysis of this research study consists of analysing the free-text fields of the RIDDOR dataset. The free-text fields consist of raw data inputted from employers of the injured persons, consisting of descriptions of the incidents that occurred. A method was needed to gain an understanding of the patterns within the data, identifying any important information contained within the text.

Text mining is a method of gaining basic meaning from text that otherwise has no structure. Compared with quantitative data stored in databases, text is unorganised and unordered, making it difficult to deal with algorithmically. The aim of text mining is to gain an overall contextual understanding of the textual data to see if there are any obvious patterns emerging. Several text mining techniques exist, ranging from simple techniques such as summarisation and information extraction to more complex techniques involving clustering and topic modelling.

Summarising the textual data to find the key words that appear the most often within the text can often provide a good starting point to exploring the data. Key words can also be searched for within the text to draw comparisons between specific words or phrases, depending on what is to be found from the study. A basic framework for text mining is contained within the R package **tm** (Feinerer and Hornik, 2015). Word clouds are often used to visualise textual data as they can provide a weighted collection of key terms with the most frequently occurring words appearing larger in size and stronger in colour. Word clouds can be created using the package **wordcloud** (Fellows, 2014) in R.

Basic text mining techniques can provide some insightful findings initially, however researchers need tools to explore large collections of documents more efficiently and

to gain a more detailed insight into the data. In the field of Information Retrieval (IR) for a collection of texts, tf-idf (term frequency- inverse document frequency) is essentially a measure used to determine how important a word is to a document. The rank (or importance) of the word increases with the frequency of the word appearing in a document, however it is offset by the total number of times the word appears in the corpus. IR researchers proposed several other techniques including Latent Semantic Indexing (LSI; Deerwester et al., 1990), which clusters documents that share similarly occurring words in the latent semantic space. This approach can achieve significant compression in large collections, and therefore is very efficient in summarising large amounts of text.

Extending on LSI, Hoffman (1999) introduced probabilistic Latent Semantic Anlaysis (pLSA), an approach to automatic indexing and information retrieval. pLSA has become a popular method of topic modelling and has a solid statistical foundation, since it is based on the likelihood principle and defines a proper generative model of the data under study. In general, topic models are probabilistic models for uncovering the underlying structure of a group of texts, based on a hierarchical Bayesian analysis (Blei et al., 2003; Blei and Lafferty, 2009). To put it simply, a 'topic' consists of a cluster of words that occur frequently together in a set of textual documents. Although pLSA is a useful step towards probabilistic modelling of text, it fails to provide a probabilistic model at the level of documents.

Blei et al. (2003) introduced a new, semantically consistent topic modelling technique called Latent Dirichlet Allocation (LDA). In defining LDA, a *word* is a basic unit of discrete text data, a *document* is a sequence of N words denoted by $w = (w_1, \dots, w_{n)}$, where w_n is the *n*th word in the sequence. A *corpus* is a collection of M documents denoted by $D = (w_1, \dots, w_n)$. LDA assumes the following generative process for each document w in a corpus D:

1. Choose $N \sim \text{Poisson}(\xi)$.
2. Choose $\theta \sim \text{Dir}(\alpha)$.
3. For each of the N words w_n:

(a) Choose a topic $z_n \sim \text{Multinomial}(\theta)$.

(b) Choose a word w_n from $p(w_n|z_n, \beta)$, a multinomial probability conditioned on the topic z_n.

(Blei et al., 2003, p. 996)

LDA essentially considers a text document as a mixture of a defined number of topics and lets each word within the document be associated with one of these topics. Once the number of topics is initially defined, LDA will assign every word to a temporary topic (according to a Dirichlet distribution). LDA then loops through each word in every document, and for each word, its topic assignment is updated based on two criteria: 1. How prevalent is that word across topics? and 2. How prevalent are topics in the document? The process of checking topic assignment is repeated for each word in every document, cycling through the entire collection of documents. This iterative process is the key feature of LDA that generates a set of coherent topics that most represent the document or corpus.

Rather than finding documents through keyword search alone, topics can be uncovered and phrases or documents relating to that theme can be explored. LDA has attracted a considerable interest from statisticians and has become one of the most popularly used topic modelling techniques in research. The impressive capabilities of LDA, indicate that the technique is effective in summarising large textual documents and identifying key topics within textual data. Newman et al. (2006), for example, used LDA to extract the main themes of the 330,000 NY Times news articles. Topics surrounded a range of themes including: "basketball", "Tour de France" and "Harry Potter". Blei (2012) also extracted topics from 17,000 articles from the journal, Science. Topics included for example: "genetics" and "evolution", which helped to identify the main themes within the text.

The textual data acquired for this research study is geographically referenced based on the exact postcode location of where the event occurred, opening up the possibilities of using this resource to understand geographic differences in workplace accident rates. LDA in general is being increasingly used in geographic information retrieval. This is due to the growing sources of online user-generated content from

social media platforms which act as rich sources of information that can be used for capturing geographic data about a variety of different topics of interest. The most popular definition of such content that possesses geographically referenced data is Volunteered Geographic Information (VGI), which was first presented by Goodchild (2007). VGI has been used in a variety of different fields. Hollenstein and Purves (2010), for example, explored the use of Flickr tags in determining the types of terms used to name city centres, such as 'Downtown' across the USA, whilst Dredze et al. (2013) explored the application of Twitter geo-located text data to public health, where they developed a system for improving influenza surveillance. Other applications of text mining VGI are found in fields such as: tourism (Girardin et al., 2008), environmental monitoring (Connors et al., 2012), and land use (Perger et al., 2012).

The application of LDA can be carried out efficiently within R using either the **lda** package (Chang, 2015) or **topicmodels** package (Grun and Hornik, 2011). The **topicmodels** package builds on the **tm** package, whereby **tm** provides an infrastructure for constructing a corpus, and transforming it to a document-matrix, ready for topic modelling. As **tm** is initially used for basic text mining, for this study, to explore the textual documents, the **topicmodels** package will be used to carry out LDA.

3.4 Summary

The aim of this chapter was to provide an overview of the data and methodological approaches used in this research study. The first section of this chapter focused on giving an overview of the RIDDOR data. It continued with detailing the RIDDOR sample dataset that was obtained for the research, giving an account of the structure of the dataset, describing the variables within it, the timeframe the data covers and the number of accidents that are stored within it. The free-text RIDDOR dataset was also described in detail, explaining what the data covers, and how it matches to the full RIDDOR dataset with the unique incident number. The strengths and limitations were outlined for the dataset and how it can be potentially utilised to investigate the risks of workplace accidents.

The second part of the data review looked at the open source government data that will also be used for this research study. Specifically, it details variables chosen from the 2011 Census tables provided by the ONS, based on hypotheses constructed following from previous studies outlined within the literature review (Chapter 2). The chapter continued to review the available methods that could be applied in this investigation. Techniques including regression analysis, specifically Generalised Linear Modelling, were discussed to model workplace accidents using pre-selected variables from the 2011 Census tables. This creates a global model, which although providing an informative approach to understanding workplace accidents, misses the underlying spatial dimension. Geographically Weighted Regression was proposed as a method to investigate local trends in the data. The method was outlined and its usefulness for identifying the relationship between variables varying over space was considered.

Finally, to analyse the textual RIDDOR dataset, text mining and topic modelling were methods proposed. In particular, Latent Dirichlet Allocation (a type of topic modelling) was discussed as a method for finding topics within the data. These topics represent core emerging themes that exist within the data, with the aim of providing a sense of meaning to the otherwise unorganised text. The results from this analysis will then be used to elaborate on the findings from the global and local modelling analyses.

In the following chapters, exploratory analysis will be carried out as a background to the study, providing a statistical summary of the RIDDOR data (Chapter 4). The global regression (Chapter 5) follows, providing a model of socio-economic characteristics that relate to accidents and providing an insight into the general risks of workplace accidents. The following analysis chapter (Chapter 6), will move towards exploring the spatial element of modelling, specifically looking at spatial trends between variables across England and Wales and how relationships change geographically. The final part of the analyses chapters will be the textual analysis (Chapter 7), which will draw upon pulling out topics that are hidden within the free-text fields.

Chapter 4

Exploratory Analysis of the RIDDOR Data

4.1 Introduction

The aim of this chapter is to provide an overview of the RIDDOR dataset, setting the context of the study. To begin with, the chapter focuses on the main key characteristics of the dataset including how many accidents were recorded each year over the six years and also the status of those involved in the recorded accidents, such as whether they were employees, on a training course or work experience, or a member of the public.

The chapter moves on to explore the basic trends in the data based on summary statistics that identify the characteristics of the types of workers having accidents. Such characteristics include: the workers' genders, ages and types of occupation and the industrial sectors worked in based on their Standard Occupation Classification (SOC) and Standard Industrial Classification (SIC) groups. Details are discussed on the types of accidents that occur, and the types of injuries sustained from these accidents, as well as the severity of the injuries, discussing whether workers had a major injury, fatality or were absent from work for more than three days.

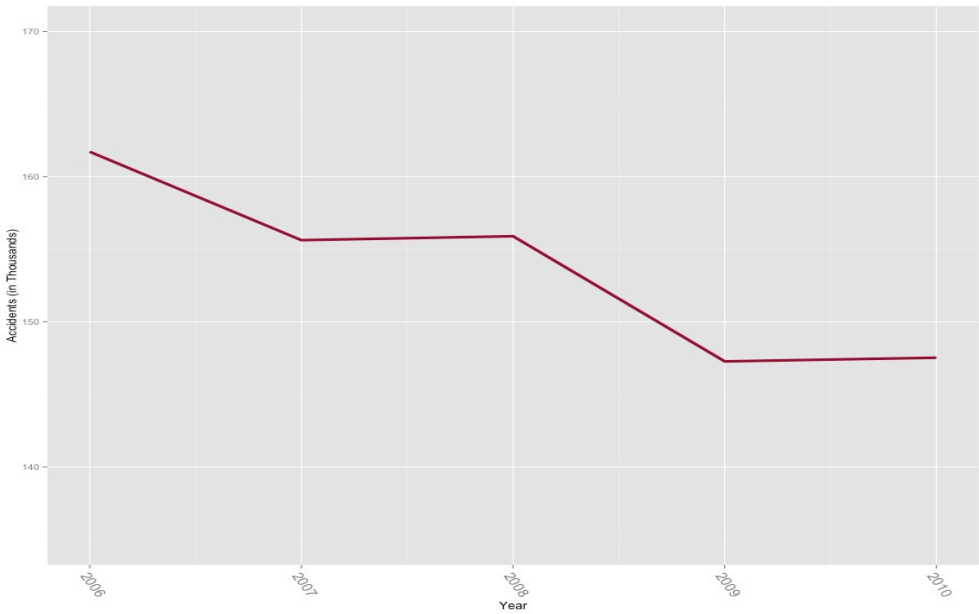
The end of the chapter focuses more in depth on the details of the accidents, looking at the location of the incidents and also when they occurred. Specifically, the RIDDOR data are mapped to identify spatial trends across Great Britain, and the

Local Authority Districts which experience the highest and lowest rates of accidents out of the working population are identified. The number of accidents by day and month is detailed next, evaluating the differences between geographic areas. This chapter provides the initial background analysis of the research study, setting the scene for the analyses chapters (Chapters 5, 6 and 7) that follow.

4.2 Background Statistics on RIDDOR Records

The total number of recorded accidents in the RIDDOR dataset between 2006 and 2011 is 822,318. Figure 4.1 shows the number of accidents per year. In 2006, around 161,700 accidents were recorded, which is the maximum number of accidents recorded for any year in this period. The lowest number of recorded accidents (excluding the half year obtained for 2011) was found in the year 2010, at just under 147,300. Generally, the number of accidents recorded annually declined from 2006 to 2011. There were slight increases in the recorded accidents between 2007 and 2008 and between 2009 and 2010, however only by around 200-250 accidents in each case.

Figure 4.1: Accidents recorded by year



The RIDDOR dataset contains records not solely about workers who were involved in an accident or had an injury or illness in the workplace, but also on members of the public. The percentage of accidents by the status of the injured person (whether they were employed or a member of the public) is shown in Table 4.1. Unsurprisingly, employees had the most accidents, with approximately 84.7% of injured people being recorded as an employee in the workplace. Small percentages of people were found to be either self-employed (1.48%), carrying out work experience (0.06%) or on a training scheme (0.12%), when being involved in a workplace accident. Members of the public were the second highest group of people (13.6%) who had an accident in the workplace, and approximately 0.04% of people were described as 'other'.

Table 4.1: Status of injured person

Status	**Percentage (%)**
Employee	84.70
Self-employed	1.48
Work experience	0.06
On a training scheme	0.12
Member of the public	13.60
Other	0.04

As this study is concerned with identifying the risks of workers having accidents in the workplace, it was necessary to remove members of the public from the study. The characteristics outlined in the following section are therefore based solely on workers who have had accidents in the workplace.

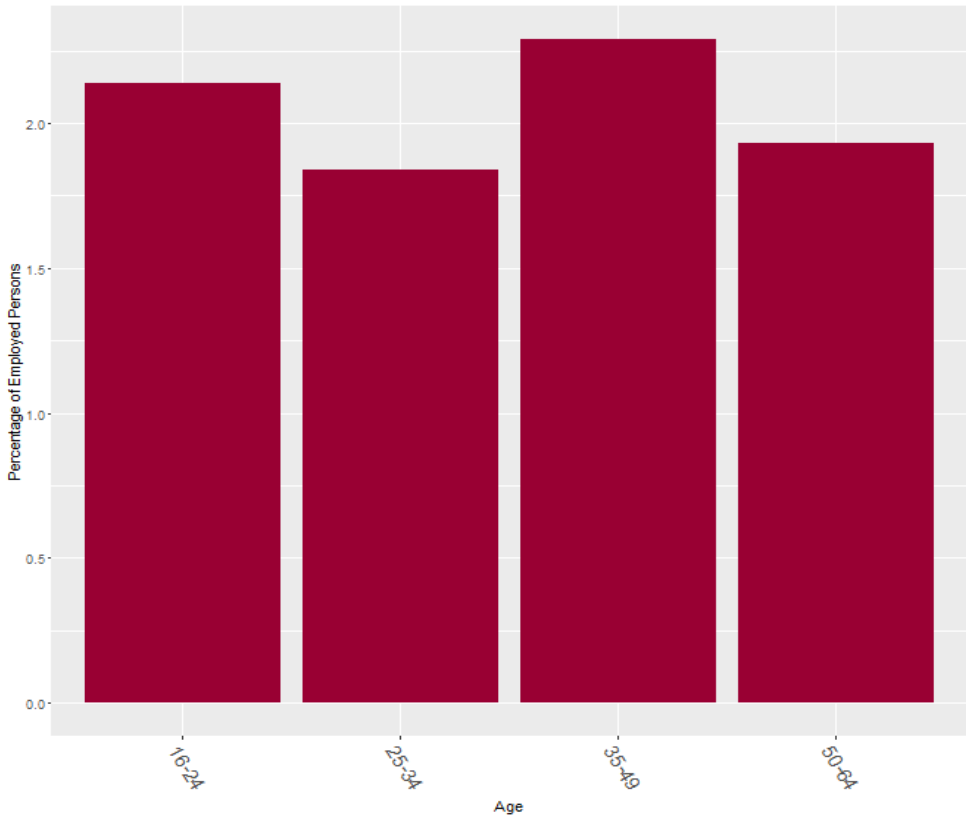
4.3 Characteristics of Injured Employees

There are many variables contained within the RIDDOR dataset that reveal details about the type of worker having accidents. This section focuses on extracting

summary statistics based on the type of worker such as their age, gender and occupation. The aim of this section is to gain a background understanding of the data based on the workers having accidents.

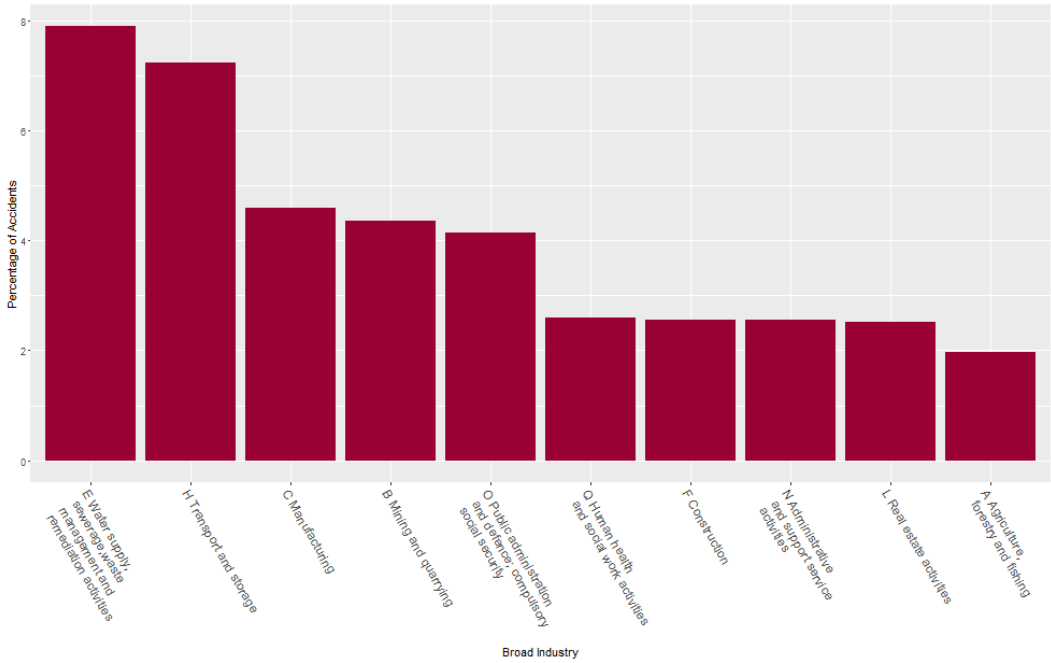
Out of the workplace population in Great Britain, around 2.37% of men were reported as having a workplace accident, whilst only 0.83% of women were reported. This suggests that men have a higher risk of having a workplace accident than women. In terms of age, Figure 4.2 shows a bar chart of the percentage of reported accidents in the RIDDOR dataset by age group out of the total number of employed people in that age group. As a proportion, workers aged 35 to 49 have reportedly the highest rate of recorded accidents, followed closely by age 16 to 24 year olds, with the lowest rates found for 50 to 64 years and 25 to 34 years, respectively.

Figure 4.2: Percentage of accidents by age out of employed working population



The incidents recorded in the RIDDOR dataset are coded by SIC 2011 Subclass. A lookup table was used to find the Broad Industry group (A to U) where workers experienced the most accidents. Figure 4.3 shows the 10 industries with the highest rates of reported accidents out of the working population. Out of all the accidents recorded, workers in SIC E Water supply, sewerage, waste management and remediation activities were found to have the highest rate of accidents (7.89%). A rate of approximately 7.22% of workers out of the working population reportedly had an accident in SIC H Transportation and storage and 4.59% of workers in SIC C Manufacturing were reported as having a workplace accident. Industries where the lowest rates of workplace accidents occurred included SIC J Information and communication (0.59%), SIC K Financial and insurance activities (0.44%) and SIC M Professional, scientific and technical activities (0.37%).

Figure 4.3: Industries (SIC 2011) with the highest 10 reported accident rates



Tables 4.2 and 4.3 below show the industries in which men and women have the highest accident rates. For men, the highest accident rate is found within SIC E Water supply, sewerage, waste management and remediation activities (10.43%). Following this, SIC H Transportation and storage has the second highest rate of accidents for men, and SIC B Mining and quarrying has the third. For women, most

accidents were reported as occurring in SIC H Transportation and storage (3.93%) followed by SIC O Public administration and defence; compulsory social security and SIC Q Human health and social work.

Table 4.2: Industries in which men have the highest rates of accidents

Industry	**Percentage (%)**
SIC E Water supply, sewerage, waste management and remediation activities	10.43
SIC H Transportation and storage	8.01
SIC B Mining and quarrying	6.83

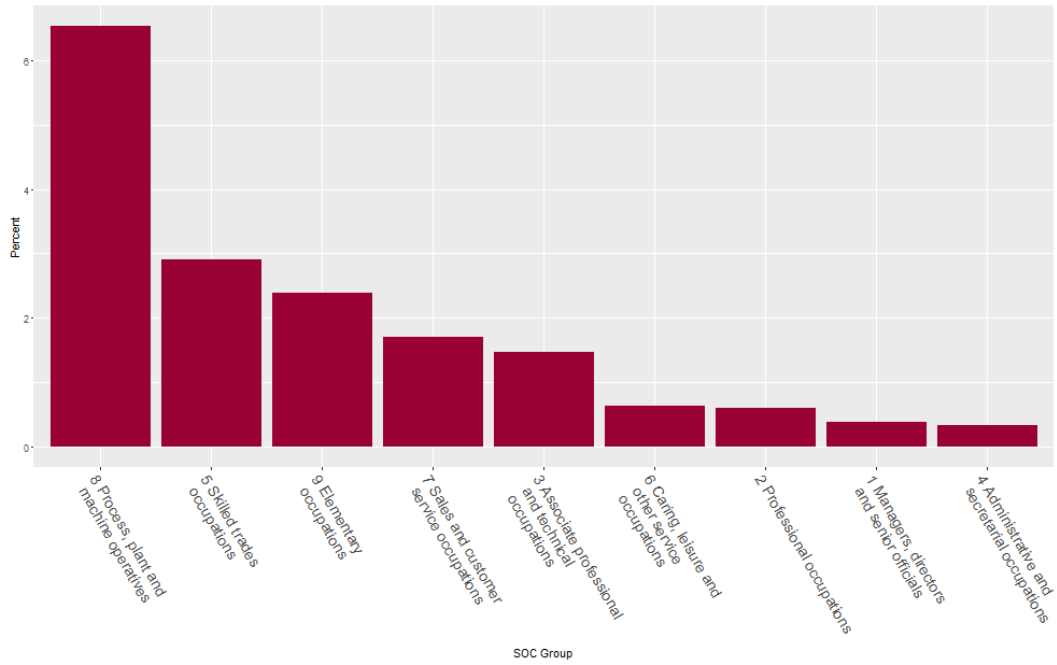
Table 4.3: Industries in which women have the highest rates of accidents

Industry	**Percentage (%)**
SIC H Transportation and storage	3.93
SIC O Public administration and defence; compulsory social security	2.96
SIC Q Human health and social work	2.39

The incidents recorded in the RIDDOR dataset are also coded by SOC 2010 Unit groups. This provides a similar level of detail to the SIC subclasses. Similarly, to the SIC framework, a lookup table was used to determine the SOC Major Group 1 to 9 from the provided unit group codes. Figure 4.4 shows the rates of accidents recorded under RIDDOR by the 9 SOC Major Groups out of the working population. Workers in occupational group SOC 8 Process plant and machine operatives had the highest rate of reported workplace accidents (6.53%). This was followed by SOC 5 Skilled trades and SOC 9 Elementary occupations, with accident rates recorded as 2.90% and 2.39% respectively. The occupations with the lowest accident rates were

SOC 2 professional occupations (0.59%), SOC 1 managers, directors and senior officials (0.38%), and SOC 4 administrative and secretarial occupations (0.32%).

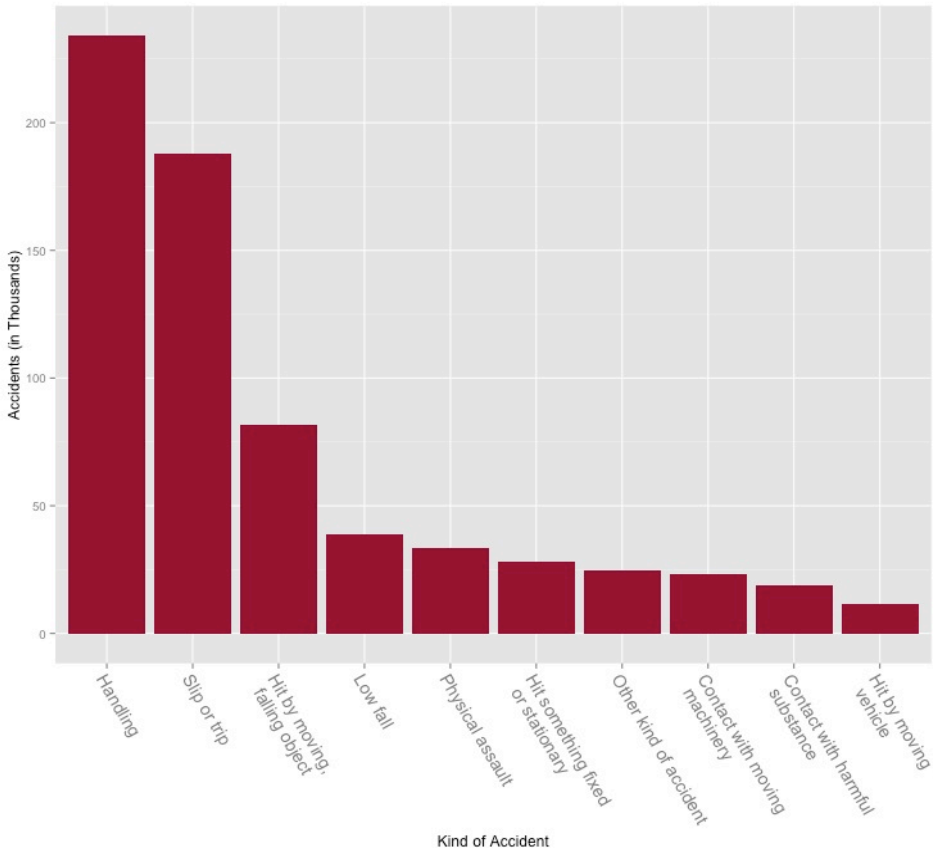
Figure 4.4: Rates of workplace accidents by occupational group (SOC 2010)



4.4 Causes of Accidents

The types of accidents with the highest records in the RIDDOR dataset are shown in Figure 4.5. The type of accident with the highest accident rate is handling, with approximately 234,000 accidents recorded. The second type of accident that was recorded most commonly in the dataset was slips or trips with approximately 188,000 accidents recorded. These two types of accidents in total accounted for almost 60% of incidents over the 6 years. Other types of accidents recorded included: workers being hit by a moving or falling object, a low fall, physical assault, hit by something fixed or stationary, contact with moving machinery or moving vehicle, or contact with a harmful substance.

Figure 4.5: Types of accidents recorded under RIDDOR



The incidents recorded in the RIDDOR dataset also include information with regards to the material agents involved in the causes of the accidents. Table 4.4 shows the material agents involved in the highest amount of accidents. Surfaces, structures and building access equipment is reported most often (22.66% of incidents). This includes for example floor, pavements or roads. The second most common material agent reported is lifting and storage systems (16.71% of accidents). The third most common was reported as people, with approximately 15.7% of accidents reporting this as the material agent involved in the accident. Other material agents, in order of those appearing the most often in the RIDDOR dataset include: machine components or materials, substances and radiation, vehicles, plant and moving equipment, office equipment and furniture, and hand held tools, amongst several others.

Table 4.4: Percentage of accidents by material agent (Top 5)

Material Agent	Percentage
Surfaces, structures and building access equipment	22.66%
Conveying, lifting, storage systems and hand held pushed/pulled transport equipment	16.71%
People	15.70%
Materials, objects, products, machine components	14.59%
Substances and radiation	5.66%

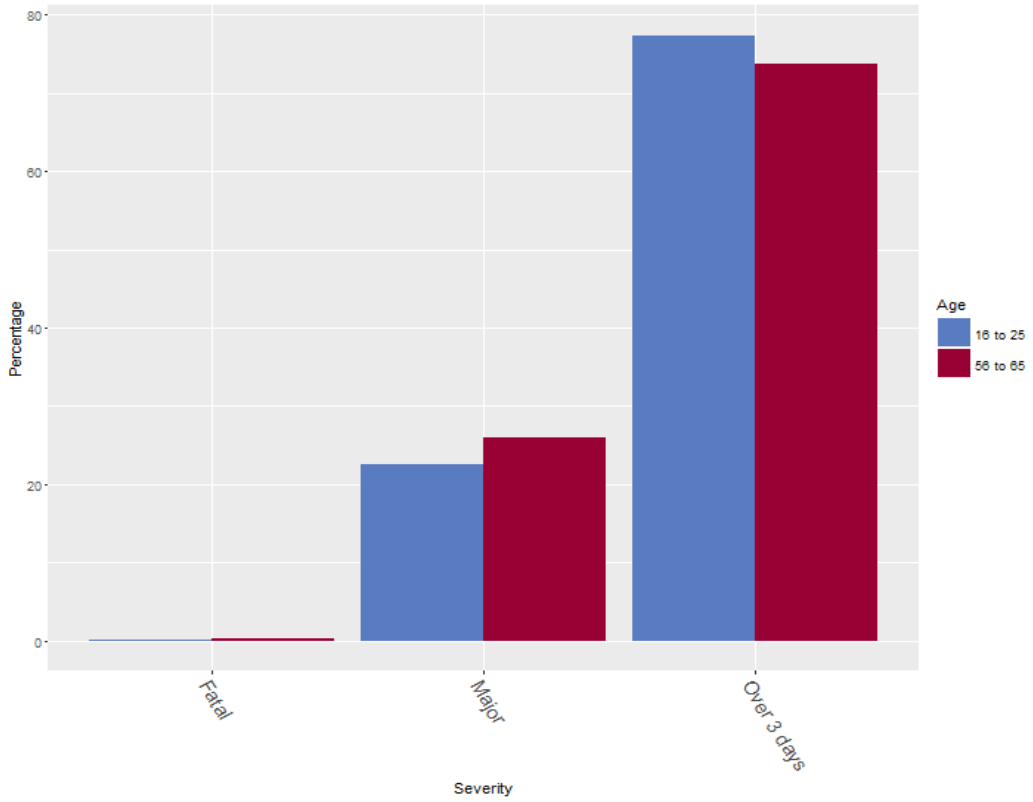
Handling and slips and trips are reportedly the most common causes of workplace accidents for workers across all industrial sectors. In terms of other causes, the third most common cause of accident in SIC C Manufacturing and SIC F Construction was being hit by a moving or falling object. This accounted for approximately 9,000 accidents in manufacturing and 17,000 in construction. Physical assault was the third most common cause of accident for industries: SIC Q Human health and social work activities (accounting for around 16,000 accidents) and SIC O Public administration and defence; compulsory social security (accounting for around 8,000 accidents). Additionally, approximately 1,300 incidents were reported as being involved in an injury caused by an animal in SIC R Arts, entertainment and recreation (the third most common cause of accident in this industrial sector).

4.5 Injuries sustained to Body

Within the RIDDOR dataset, the severity of the injury sustained by the accident in the workplace is recorded. Out of all of the accidents recorded under RIDDOR, the majority (78%) resulted in an absence from work of greater than 3 days. The second highest rate of accidents (21%) were reported as 'major', which accounts for, for example: fractures (other than to fingers, thumbs and toes), amputations, injuries likely to lead to loss of sight, any crush injury to the head or torso, serious burns or unconsciousness. Over the 6 years of reported RIDDOR data, approximately 1,121

(0.2%) accidents resulted in a fatality. Figure 4.6 shows the rates of injury severity by two age groups: age 16 to 25 and age 56 to 65. The 56 to 65 age group had a slightly higher percentage of reported fatalities than the 16 to 25 age group (0.25% compared to 0.12% respectively). A higher percentage of major injuries were reported for older workers, whilst a higher percentage of injuries resulting in an absence from work for greater than 3 days was found amongst age 16 to 25 year olds.

Figure 4.6: Rates of severity of injuries by age group out of working population



In terms of fatalities, Figure 4.7 shows the rates by all working age groups (age 16 to 65). Generally, the percentage of fatalities increases with age, with age 56 to 65 year olds having the overall highest rate of fatalities and age 26 to 35 having the lowest rate. The age 16 to 25 year olds have the third highest rate of fatalities overall. The rates of major injuries and those resulting in the worker being absent from work for greater than 3 days are similar for both men and women (Table 4.5). The rates of fatalities are higher for men than women with percentages of approximately 0.22% for men and 0.016% for women.

Figure 4.7: Fatalities by age group

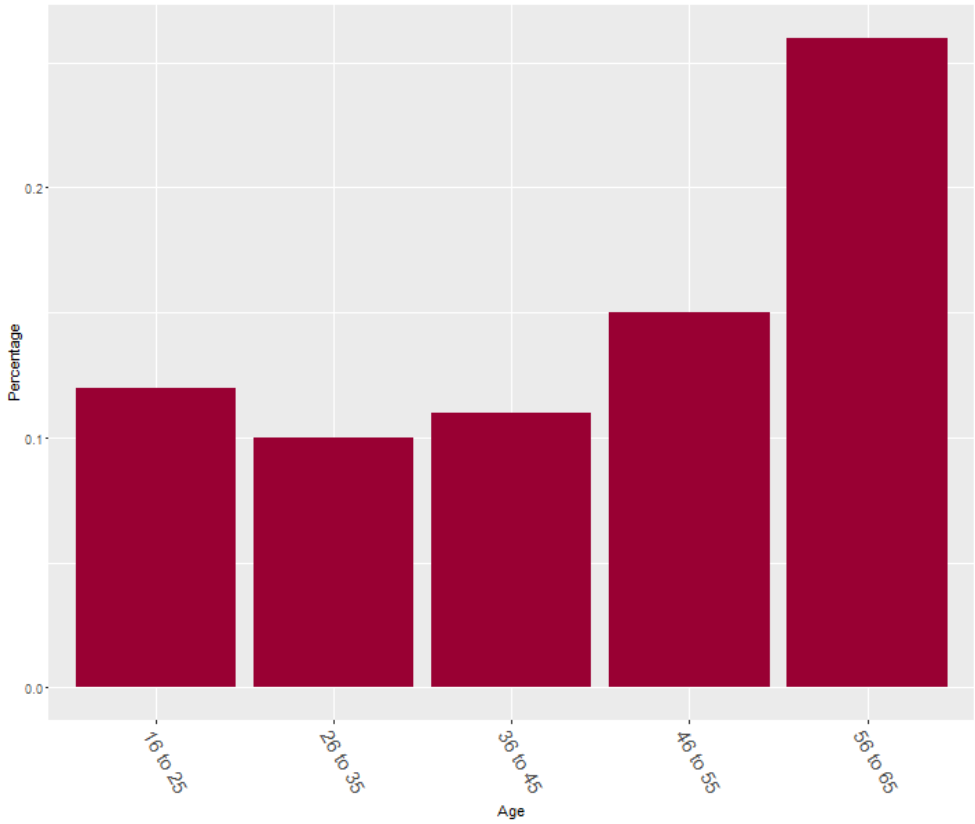


Table 4.5: Severity of injuries by gender

Severity	Females	Males
Fatal	0.016%	0.22%
Major	21.28%	21.46%
Over 3 days	78.71%	78.32%

The nature of the injuries sustained by workers is detailed in Table 4.6. Strains are the most common form of injury reported under RIDDOR as a result of a workplace accident, with over a third of cases reporting this as the nature of the injury sustained to the worker. The second most common injury is a fracture, with just under 20% of accidents reporting this under RIDDOR. The third most common injury is a contusion (bruise), with around 16% of workers experiencing this as a

result of an accident in the workplace. Other injuries include: lacerations (cuts), burns and dislocations, injuries of natural causes, as well as superficial (on the surface of the skin) and workers having multiple injuries. More severe, less common injuries include: concussion/internal injuries (0.65%), amputation (0.52%), asphyxiation or poison (0.22%), electric injuries (0.22%) and loss of sight (0.07%).

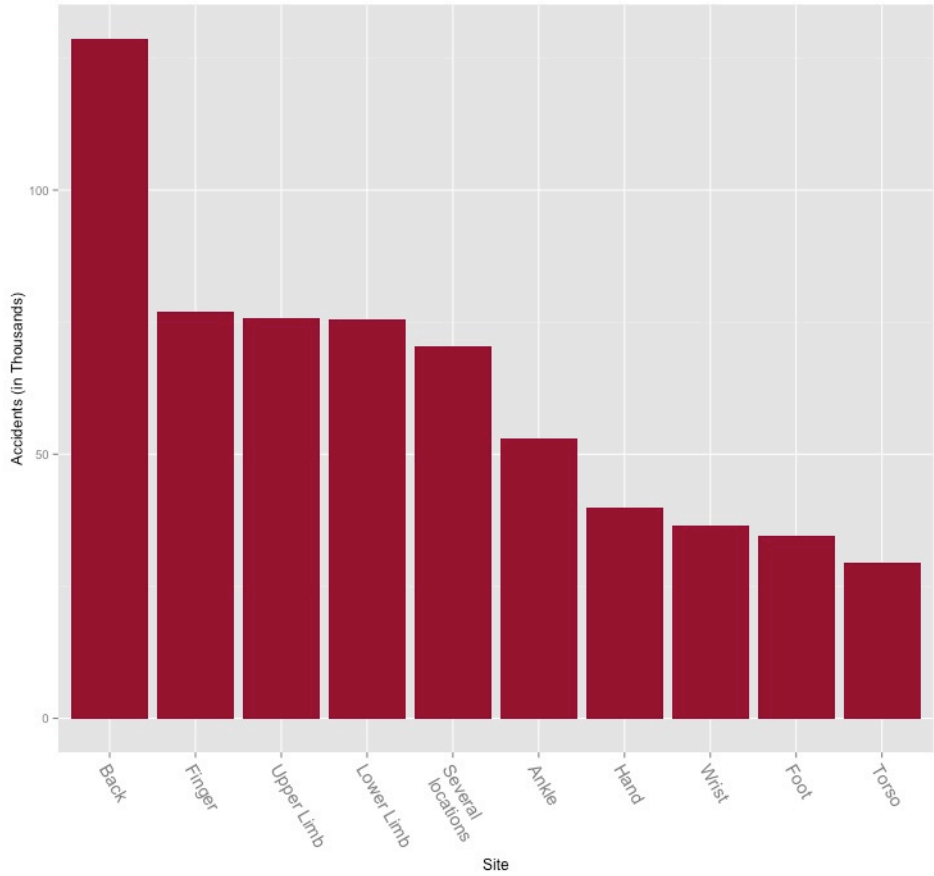
Table 4.6: Nature of injury

Description	**Percentage (%)**
Strain	37.52
Fracture	19.67
Contusion	15.78
Laceration	8.84
Multiple	5.08
Superficial	4.59
Burn	2.51
Other known	1.58
Dislocation	1.47
Other not known	1.20
Concussion/internal injury	0.65
Amputation	0.52
Asphyxiation/poison	0.22
Electricity	0.22
Loss of sight	0.07
Natural Causes	0.07

The site on the body that was injured as a result of the workplace accident is shown in Figure 4.8. Specifically, it shows the 10 body parts that are most commonly injured, with the number of accidents, in thousands, that resulted in an injury to the body part. Injuries to workers' backs are the most common form of injury found in the RIDDOR data as a result of a workplace accident. Back injuries occurred in almost 130,000 cases reported to RIDDOR, which equates to 20% out of all the

accidents recorded with a reported injured body part specified. Injuries to the finger, upper and lower limbs followed with between 75,000 and 77,000 incidents being accounted for, for each of the body parts. Around 71,000 accidents overall were to multiple parts of the body. Around 53,000 accidents resulted in injuries to workers' ankles, followed by between 30,000 and 40,000 accidents each resulting in injuries to hand, wrist, foot and torso.

Figure 4.8: Body parts injured



4.6 Locations of Accidents

The incidents stored within the RIDDOR dataset contain a postcode reference of where each workplace accident occurred. Using the ONS Postcode Directory, postcodes were first aggregated into regional areas. The rates of accidents by regional workplace population are shown in Table 4.7. The East Midlands experiences the

highest accident rate (2.76%), followed by Yorkshire and The Humber (2.72%) and Wales (2.70%). The lowest accident rates were found in the South West (2.15%), South East (1.98%) and London (1.48%) regions.

Table 4.7: Accident rates by region

Description	Percentage (%)
East Midlands	2.76
Yorkshire and The Humber	2.72
Wales	2.70
North East	2.68
West Midlands	2.67
North West	2.53
East	2.26
Scotland	2.21
South West	2.15
South East	1.98
London	1.48

Figure 4.9 shows the distribution of accident rates by Local Authority District area, out of the total population working in each area. The map shows that there is a spatially varying distribution of accidents across Great Britain, with the percentage of accidents out of the working population ranging from under 1% to over 6%. The higher rates of accidents appear to be in East and North of England, whilst the lowest rates are found in the South East and London areas. Table 4.8 shows the Local Authority District areas where the highest rates of accidents are found out of the working population. North Warwickshire has the highest rate of accidents with 6.2% found over the six years of reported incidents, followed by Corby (Northamptonshire) with 4.76% and Bassetlaw (Nottinghamshire) with 4.42%. Accident rates in Wakefield and North West Leicestershire reach 4.02%, followed by South Holland in Lincolnshire (3.93%), Hounslow in West London (3.92%) and Flintshire in North Wales (3.86%).

Figure 4.9: Accident rate by Local Authority District workplace population

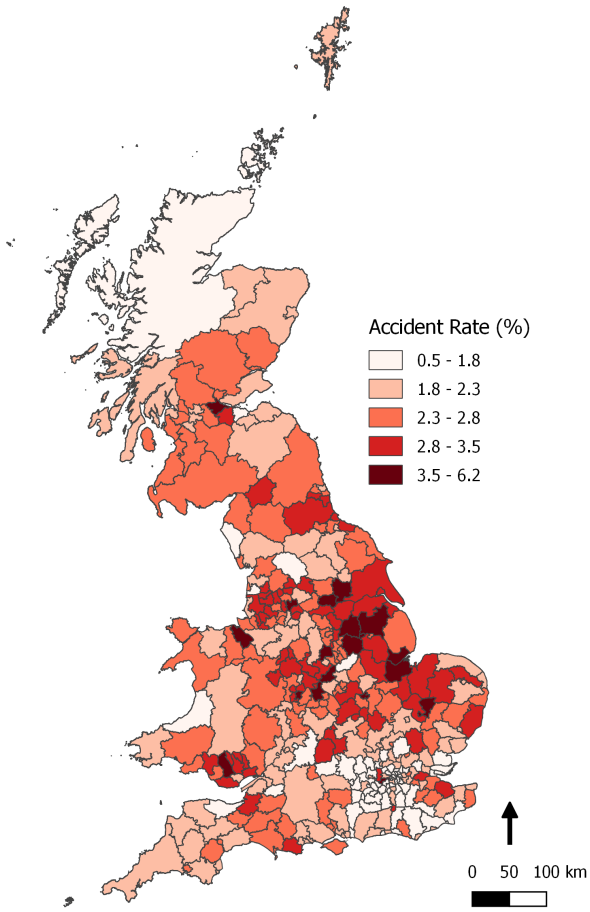


Table 4.8: The 10 LADs with the highest accident rates

Local Authority District and Region	**Accident Rate**
North Warwickshire, West Midlands	6.20%
Corby, East Midlands	4.76%
Bassetlaw, East Midlands	4.42%
Wakefield, Yorkshire and Humber	4.02%
North West Leicestershire, East Midlands	4.02%
South Holland, East Midlands	3.93%
Hounslow, London	3.92%
Flintshire, Wales	3.86%
Amber Valley, East Midlands	3.84%
Forest Heath, East of England	3.82%

The industry with the highest number of reported accidents in North Warwickshire is SIC H Transportation and Storage. Approximately 1,104 accidents in the North Warwickshire district area were reported as occurring in SIC H. However, as a proportion of the SIC H workplace population, this industry experienced the second highest rate of workplace accidents, with the highest accident rate being found for SIC B Mining and Quarrying (31.7%), as shown in Table 4.9.

Table 4.9: SICs with highest accident rates for North Warwickshire

Industry	**Number of Accidents**	**Workplace Population**	**Accident Rate**
SIC B Mining and quarrying	225	710	31.69%
SIC H Transportation and storage	1104	6567	16.81%
SIC N Administrative and support service activities	185	1923	9.62%
SIC T Activities of households as employers	1	11	9.09%
SIC E Water supply, sewerage, waste management and remediation activities	36	409	8.80%

For accidents which occurred within SIC B Mining and quarrying in North Warwickshire, 91.1% resulted in a worker being absent from work for over 3 days. The most common type of injury within SIC B in North Warwickshire was a strain (34.7% of accidents) and the most common site of injury was to lower limbs (20.8% of accidents). In comparison, the highest accident rate occurring in Corby was found in SIC H Transportation and storage (17.4%), as shown in Table 4.10. This was followed by SIC N Administrative and support services and SIC C Manufacturing. Within SIC H, most workers suffered from injuries which resulted in them being

absent from work for 3 or more days, and most commonly suffered from a strain as a result of an accident. The most common site for an injury was reportedly to workers' backs (29.4%).

Table 4.10: SICs with highest accident rates for Corby

Industry	**Number of Accidents**	**Workplace Population**	**Accident Rate**
SIC H Transportation and storage	445	2551	17.44%
SIC N Administrative and support service activities	104	1513	6.87%
SIC C Manufacturing	544	8381	6.49%
SIC E Water supply, sewerage, waste management and remediation activities	11	243	4.53%
SIC F Construction	61	1668	3.66%

The Local Authority District which experiences the least number of accidents per head of the working population is the City of London with an accident rate of approximately 0.50% (Table 4.11). Approximately 0.85% of workers in the Isles of Scilly had a workplace accident, and 0.93% in Tower Hamlets, London. The most noticeable difference between the districts with the highest rates of accidents and the districts with the lowest rates of accidents is that less accidents as a proportion of the working population appear to be reported in the South of England, particularly London. The districts with overall more accidents, as shown in Table 4.8 appear to be located in the Midlands or North of England.

Table 4.11: The 10 Districts with the lowest accident rates

District	Accidents Rate (%)
City of London	0.50
Isles of Scilly	0.85
Tower Hamlets	0.93
Camden	0.98
Westminster	0.99
Hackney	1.03
Islington	1.21
Richmond upon Thames	1.23
Hart	1.26
Harrow	1.26

The industries with the highest accident rates in the City of London district are shown in Table 4.12. The highest accident rate is found in SIC E Water supply, sewerage, waste management and remediation activities with 8.02% of workers having accidents out of the working population in that industrial sector. SIC F Construction has the second highest rate (5.05%), followed by SIC I Accommodation and food services (3.53%).

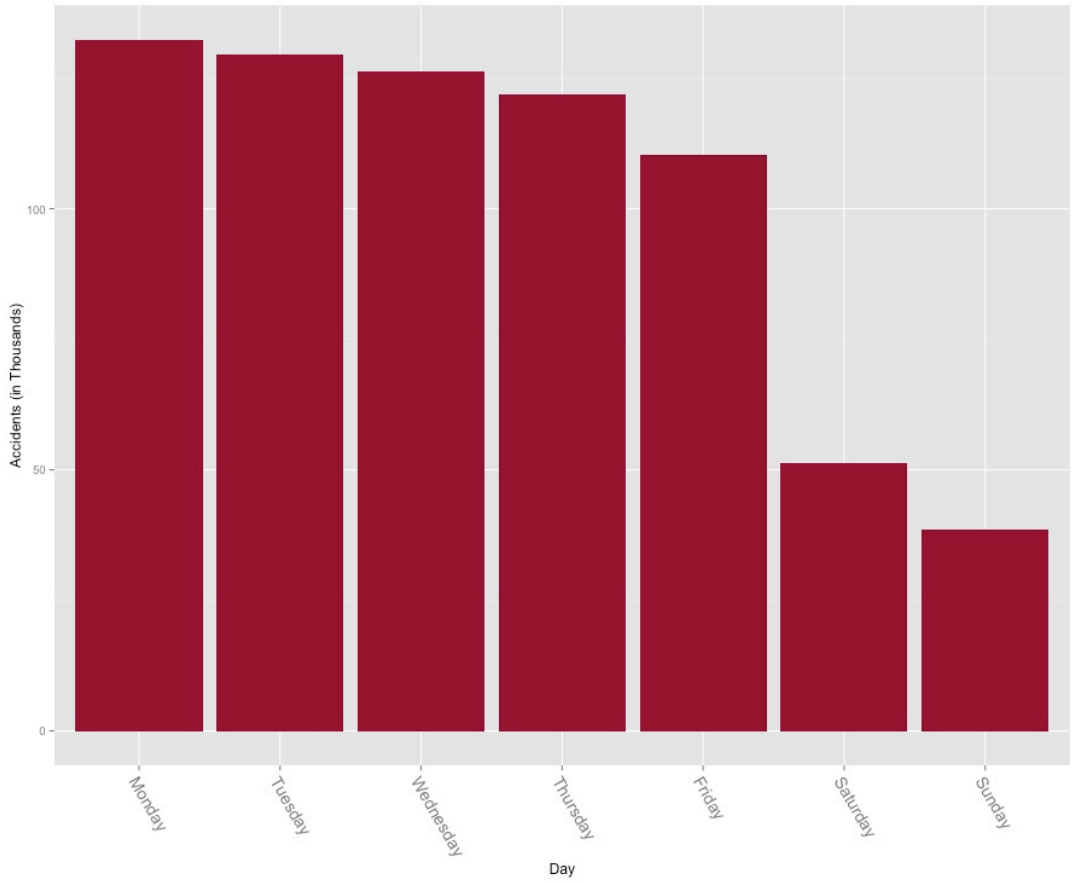
Table 4.12: SICs with highest accident rates for City of London

Industry	Number of Accidents	Workplace Population	Accident Rate
SIC E Water supply, sewerage, waste management and remediation activities	17	212	8.02%
SIC F Construction	291	5723	5.05%
SIC I Accommodation and food service activities	388	10980	3.53%
SIC H Transportation and storage	96	4832	1.99%
SIC C Manufacturing	46	2864	1.61%

4.7 Time and Date of Accidents

The RIDDOR dataset contains fields in relation to the time and date of each of the incidents recorded. These can be used to determine whether there are any particular times when workers are more at risk of having accidents over others and provide further insight into the existing RIDDOR records. Figure 4.10 shows the number of accidents occurring by day over the six years of recorded data. Around 19% of accidents occur on a Monday, which equates to around 132,300 accidents over the recorded time period. The number of accidents then gradually descends through the weekdays to the weekend with more accidents occurring on a Saturday than Sunday. The weekend trends could be explained by less workers working on Saturday or Sunday than a standard weekday and in general, working hours are shorter on a Sunday in Great Britain, which may explain why this day has the least amount of recorded accidents.

Figure 4.10: Accidents per day



Daylight Saving Time (forwarding the clocks by an hour in spring and putting them back by an hour in autumn) is thought to have an effect on our bodies by making them tired and less responsive or alert to danger (Barnes and Wagner, 2009; Varughese and Allen, 2001). The RIDDOR dataset was split into just the Mondays following the clock change in March to see whether, on average, there were more accidents occurring on those particular Mondays in comparison to a standard Monday. The average number of accidents on a standard Monday was found to be 464, however the average number of accidents on a Monday following the clocks being put forward (loss of an hours sleep) was actually less, at just 446. Interestingly, the average number of accidents following the Monday in October when clocks are put back, and in essence an hour of sleep is gained through the night, is higher than the average number of accidents recorded on standard Mondays. Approximately 483

accidents are reported on average on the Monday following the October clock change which is 19 more accidents than on a standard Monday.

Table 4.13 shows the 5 industries with the highest proportions of accidents reported as occurring on a standard Monday. SIC U Activities of Extraterritorial Organisations and Bodies has proportionally more accidents on a Monday out of all other days compared with any other SIC group with approximately 28% of accidents occurring on this day. Following this, 22% of accidents in SIC D Electricity, gas, steam and air conditioning supply occurred, the same percentage of accidents in SIC E Water supply, sewerage, waste management and remediation activities also occurred on a Monday.

Table 4.13: Percentage of accidents occurring on a Monday by SIC group out of all other days of the week

Industry	**Percentage of Accidents**
SIC U Activities of extraterritorial organisations and bodies	28%
SIC D Electricity, gas, steam and air conditioning supply	22%
SIC E Water supply, sewerage, waste management, and remediation activities	22%
SIC F Construction	21%
SIC M Professional, scientific and technical activities	21%

Figure 4.11 shows the reported number of accidents that occurred by month of the year. These numbers represent accidents reported between 1st April 2006 and 31st March 2011. The highest number of recorded accidents occurred in October (approximately 59,400). This is followed by January (58,700) and June (58,600). The month in which the lowest number of accidents occurred is December, with approximately 47,000 accidents being reported. In terms of seasons (Table 4.14), accidents mostly occurred in Spring (29%), followed by Summer (27%), Autumn (25%) and Winter (19%).

Figure 4.11: Accidents per month

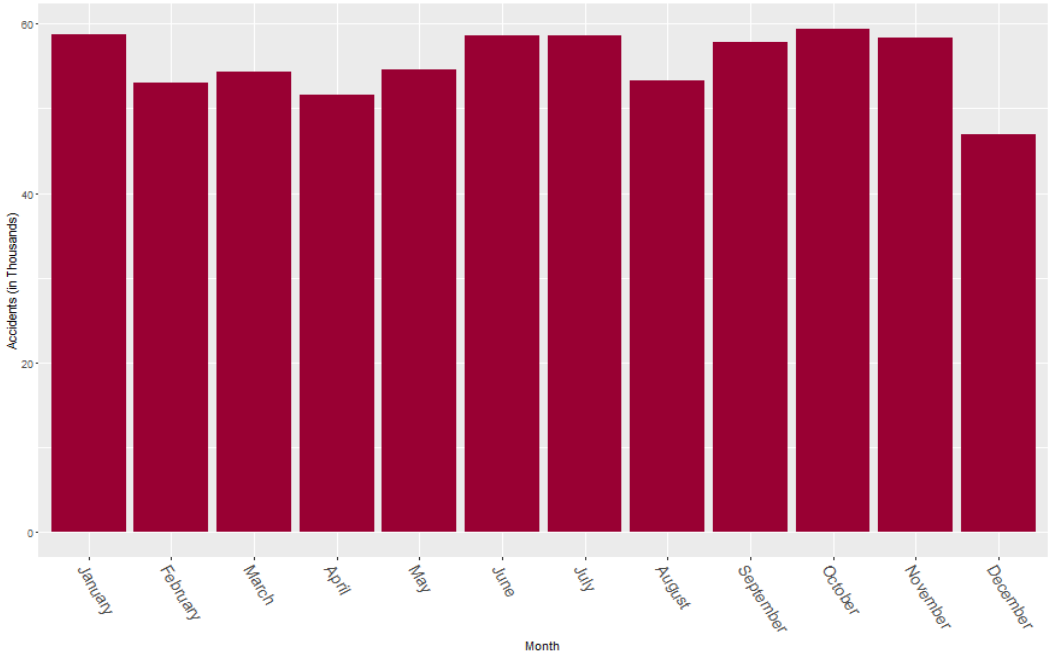


Table 4.14: Accidents per season

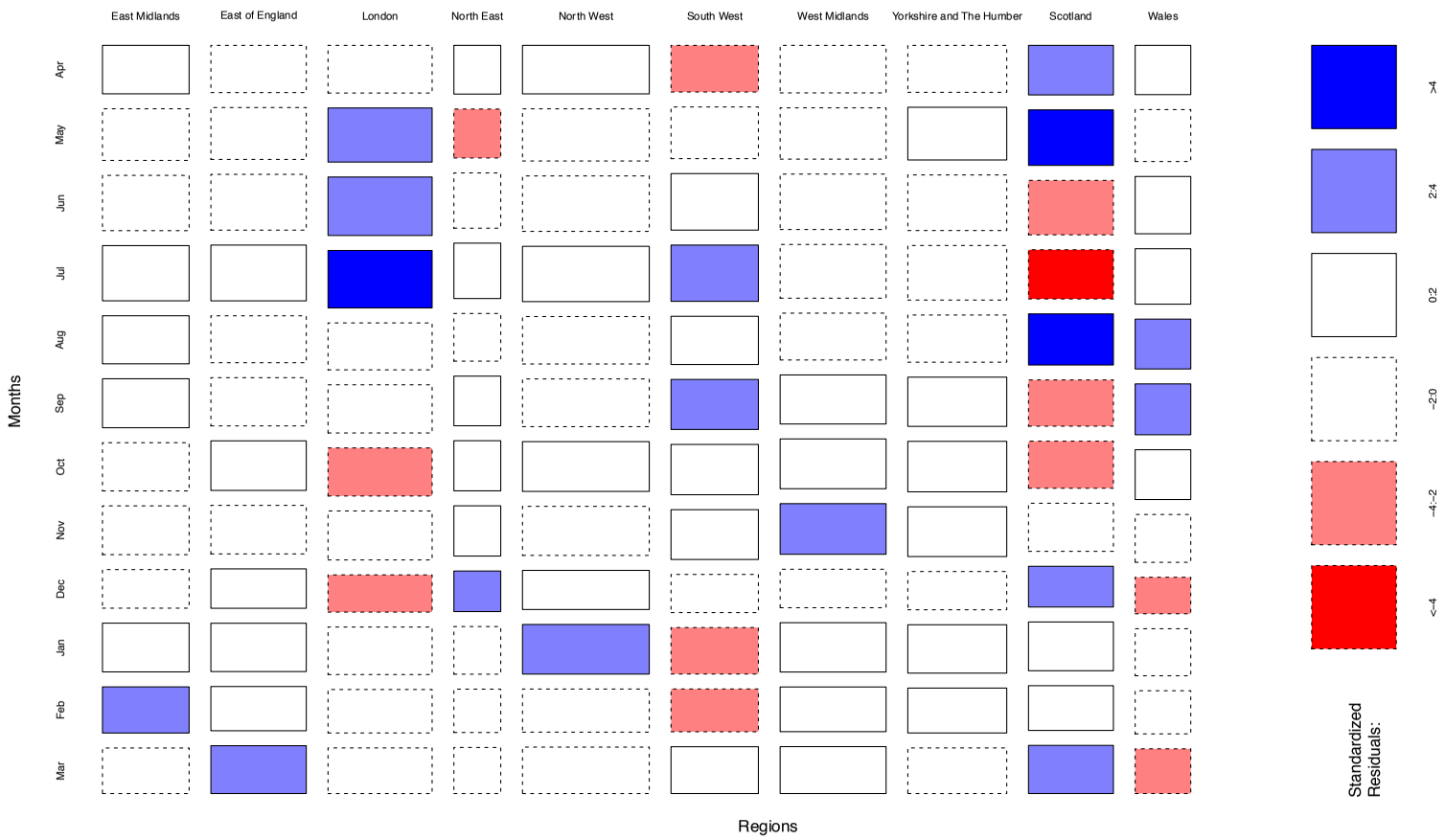
Season	Percentage of Accidents
Spring	29%
Summer	27%
Autumn	25%
Winter	19%

Mosaic plots provide a simple and effective technique for visualising contingency tables. Figure 4.12 shows a mosaic plot of a cross tabulation of the counts of standardised residuals of accidents by region and month of the year. The areas of the boxes proportionally represent the number of workers having accidents. Horizontally across the mosaic plot, is the amount of accidents by region and vertically represented is the amount of accidents by month of the year. To make inferences about the population, measures of statistical significance are provided inspired by the

chi-square test. Pearson residuals are defined which measure the departure of each cell from independence. The colours therefore represent residual shading with the solid black outlined boxes and blue coloured boxes representing more observations than would be expected under the null model (independence of variables). The dashed outlined boxes and the red coloured boxes represent a scale of fewer observations than would be expected in those particular cells. More specifically, the dark blue and dark red boxes represent the largest deviance from the null model. The main purpose of the shading however, is not to visualize significance but the pattern of deviation from independence (Friendly 2000, p. 109).

From the mosaic plot, it is evident that there is little variation in the amount of recorded accidents across the months for regions including Midlands, Yorkshire and Humber or the East of England. In comparison, the results for London and the South of England, Scotland and the North of England and Wales show that some variations exist between accidents by month of the year. There are several key findings drawn from examination of the mosaic plot. The dark blue coloured boxes (the positive standardised residuals) indicate cells whose observed frequency is substantially greater than would be found under independence. The dark red coloured boxes (the negative standardised residuals) indicate cells that occur less often than under independence. More accidents are found to have occurred for the North East, North West and Scotland in the winter months of December and January (dark blue boxes) whilst in contrast, less accidents in London, South West and Wales (dark red boxes). In contrast, there are more accidents occurring in London and the South West of England in the summer months of June and July (blue boxes), when less accidents are reported in Scotland (red boxes).

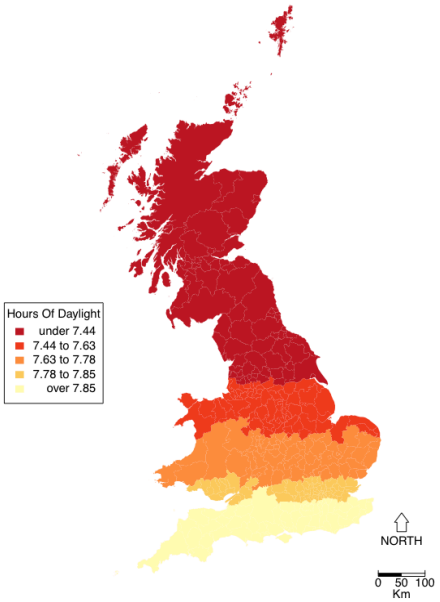
Figure 4.12: Mosaic plot of accidents by month and region



Physical factors such as weather conditions or daylight hours may have an effect on workplace accident risk (Edwards, 1996; Folkard, 1997). They may also help explain why variations in accident rates occur between areas, during particular times of the year. Figure 4.13 illustrates the differences in daylight hours between winter and summer months. The first map on the left shows the hours of daylight experienced across the country on the shortest day of the year (21st December 2012), which represents the winter months. The second map, on the right shows the hours of daylight experienced across the country on the longest day of the year (21st June 2012), which represents the summer months. As can be seen, there are more hours of daylight in the South of England in the winter, and less in Scotland (where more accidents occur during winter), and there are more hours of daylight in Scotland and the North of England in the summer, and less in the South of England (where more accidents occur during summer).

Figure 4.13: Hours of daylight for shortest and longest days of 2012

Shortest Day- 21st December



Longest Day- 21st June

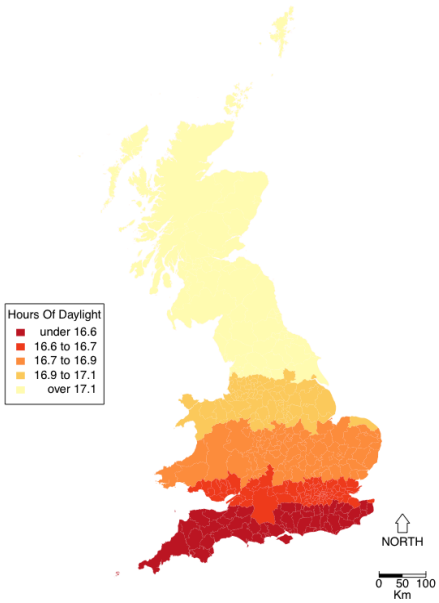


Figure 4.14 shows a calendar heat map of the recorded accidents by day. The light yellow and green colours indicate that a higher number of accidents occurred during a particular day, versus the dark orange and red colours when fewer accidents were recorded as occurring. The calendar heat map shows that a higher number of accidents occurred during the first week of January and the first two weeks of February in 2009, as can be seen by the green and light yellow squares. The Met Office reported that the weather in the UK during these two months in particular was mostly unsettled, with the first half of January being very cold with rainfall across most areas (Met Office, 2013a). It was also reported that during the first week of February it was very cold with snowfall in many areas. In fact, it was reportedly the most widespread snowfall since February 1991 (Met Office, 2013b).

The first two weeks of January in 2010, as well as early December 2010 show that a higher number of accidents occurred, again highlighted by the light and dark green cluster of colours in the calendar heat map. The Met Office reported that January 2010 was the coldest January over the UK since 1987. The beginning of the month faced widespread snowfalls and some sharp frosts (Met Office, 2013c). In December 2010, it was also exceptionally cold across the UK, being the coldest December in over 100 years. Snowfalls occurred in almost all areas, especially in the first week of the month (Met Office, 2013d).

Figure 4.15 shows a calendar heat map of workplace accidents that resulted only in a slip or trip. This heat map clearly shows the differences between dates when accident rates were at their highest versus lowest. The days when the number of accidents were highest appear to coincide with those found for overall accident rates pictured in the first calendar heat map. These are during January and February 2009, and January and December 2010. There is also a higher number of accidents found in December 2008 and 2009. In general, these are the UK's winter months, when weather is particularly unsettled, and the number of daylight hours in working days are shorter. These heat maps suggest seasonal differences between workplace accidents, that may be attributed to adverse weather conditions.

Figure 4.14: Calendar heat map of workplace accidents

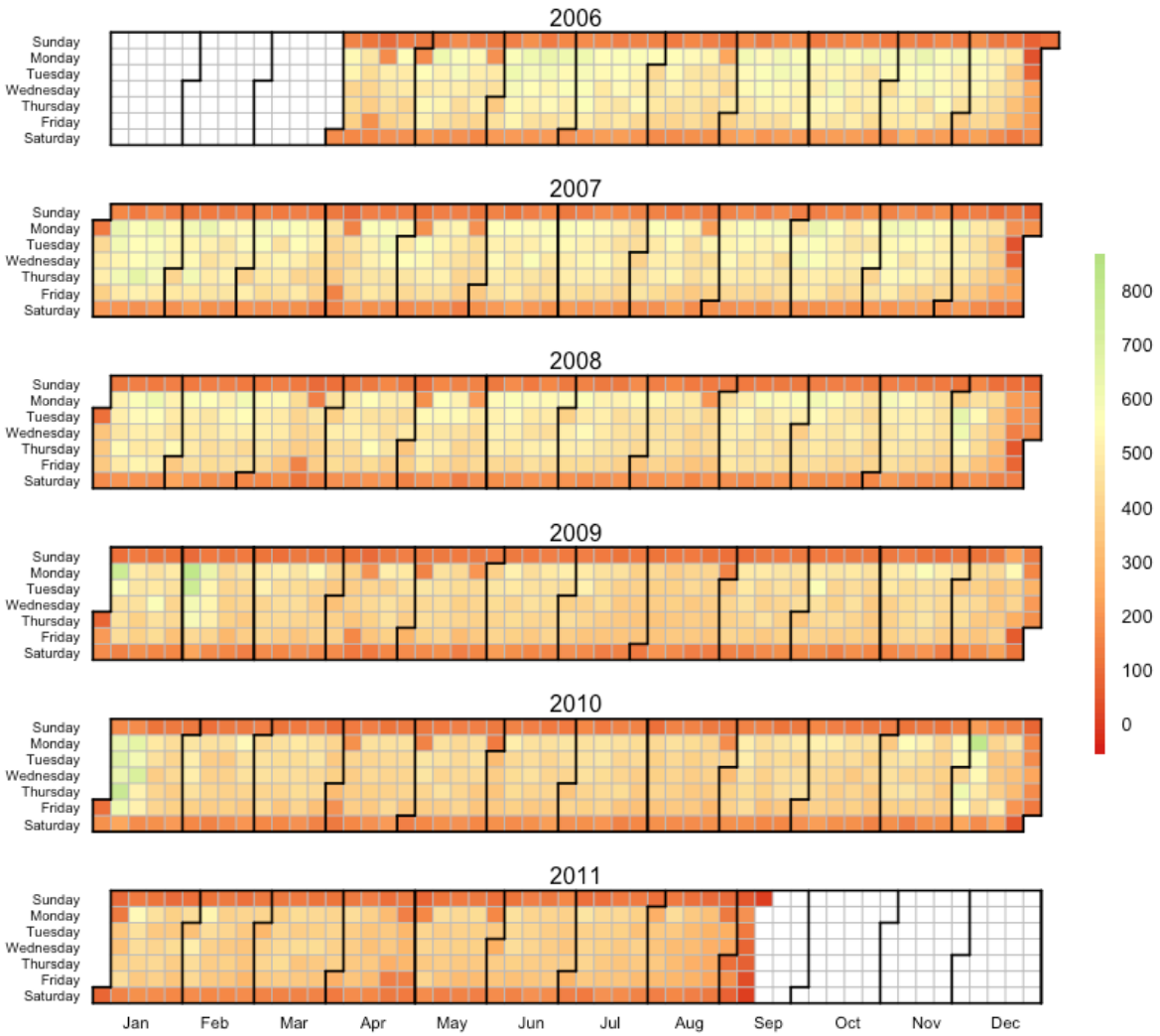
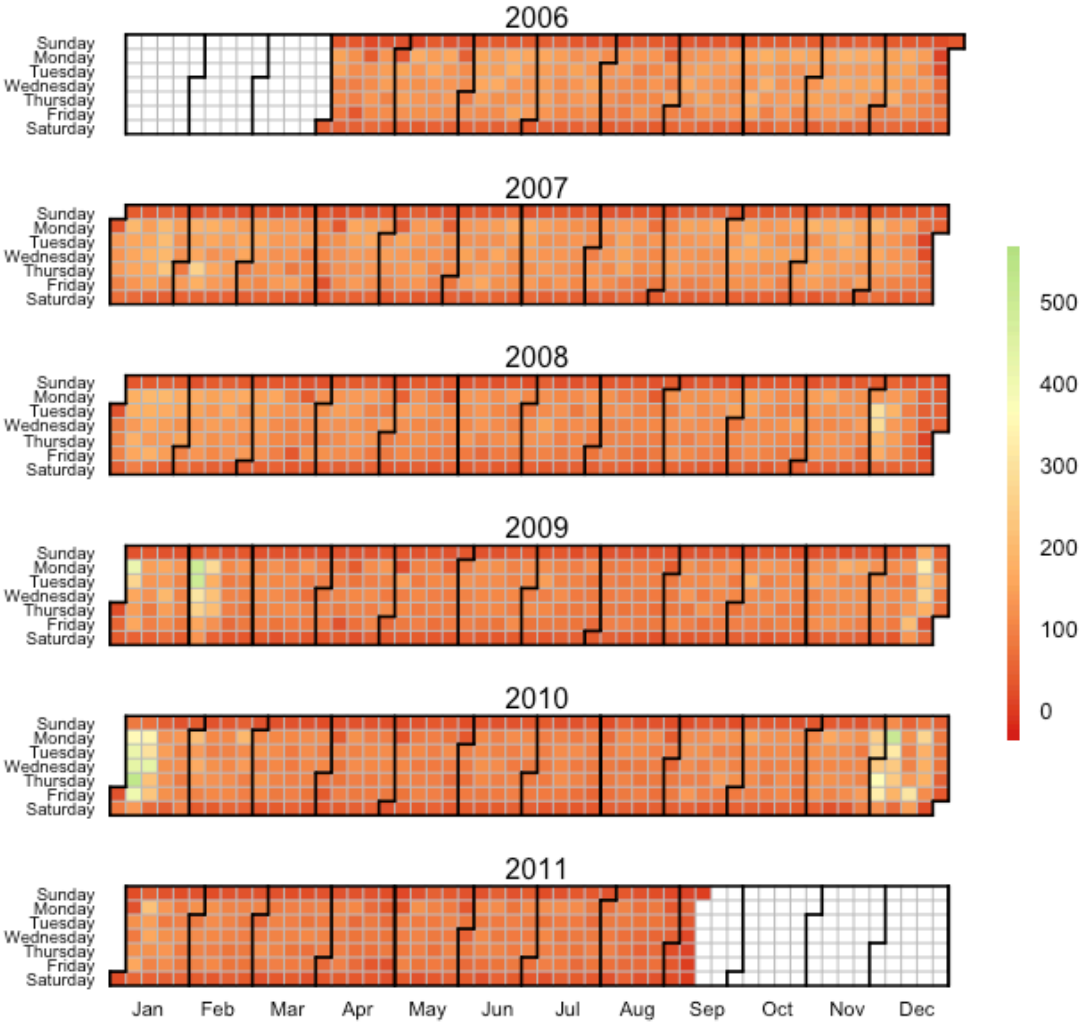


Figure 4.15: Calendar heat map of workplace accidents involving slips or trips



4.8 Conclusion

This chapter has explored the background statistics of the RIDDOR dataset, providing a context for the chapters that follow. The chapter begins with providing an overview of the number of accidents recorded in the dataset by year, and breaks down the number of accidents by status of person, which includes: employees, self-employed, those on work experience or a training scheme, or members of the public.

The chapter continues with focusing solely on the employees having workplace accidents, and explores the socio-economic characteristics of these employees. Specifically, the frequency of accidents is found by gender, age group, occupation and industry to help understand more about the types of workers having workplace accidents. It is found that more men than women have accidents in the workplace, and workers who are aged 35-49 have the highest rate of accidents out of any other age group, with age 25-34 having the least amount of accidents recorded under RIDDOR.

In terms of industry and occupational groups, SIC groups and SOC groups are compared within the chapter. It is found that most accidents occur within SIC E Water supply, sewerage and waste management industry followed by SIC H Transportation and storage. It is also found that workers who work in SOC 8 Process, plant and machine operatives have the highest rate of accidents overall out of all occupational groups.

The chapter continues by giving an overview of the types of accidents that commonly occur. It provides details on the processes involved in the accidents, whether there were any material agents or objects that caused the accidents as well as information on the types of injuries sustained to the worker. The severity of the injury is considered, as well as the nature of the injury for example whether the worker sustained a cut, bruise or a fracture etc., and details about the types of body parts injured as a result of the accident.

A brief overview of the location of accidents is also provided. Specifically, the number of accidents by the workplace population per Local Authority District Area was assessed and the industries within the areas with the most accidents were listed. Finally, accidents were grouped by the days and months in which they occurred to identify any particular patterns in terms of when accidents where occurring and why. Most accidents were reported as occurring on a Monday and the month when most accidents occurred was October. There appeared to be some spatio-temporal differences in the results, whereby more accidents were being reported in the summer months in the South of England, as opposed to the winter months in the North of England and Scotland. The calendar heat maps also highlighted some seasonal differences in workplace accident rates, which were compared with weather conditions as summarised by the Met Office reports.

Overall, this chapter has provided a summary of the basic trends of the accidents reported in the RIDDOR dataset, opening up the research study for the chapters which follow. Some interesting findings have been discussed in this chapter and these are explored further in Chapters 5, 6 and 7, looking further into the differences of the characteristics of the workers having accidents, as well as the locations in which accidents occur.

Chapter 5

A Global Regression Analysis of Workplace Accidents

5.1 Introduction

The previous chapter explored the basic statistics of the RIDDOR dataset, looking at the general trends found in the data such as the rates of men and women having accidents, the industries which experience the highest rates of accidents, and the different types of injuries that were reported to the HSE. This chapter is concerned with exploring in further detail what factors influence accident rates and therefore provides a context for understanding the risks involved in workplace accidents.

This chapter focuses on gaining a global perspective on accidents in the workplace based on socio-economic area characteristics across England and Wales. As detailed in Chapter 3, this is carried out via global regression modelling and this chapter considers alternative forms of Generalised Linear Models most suited to the characteristics of the data. The regression modelling techniques that are fit to the data include Poisson, quasi-Poisson and negative binomial regression models and the most appropriate method is selected.

A range of socio-economic variables are included within the global model to determine what factors have an effect on workplace accidents. These variables include: age, gender, occupation, industry, social classification and population density. The socio-economic variables were selected on the basis of findings from

previous studies (see Chapter 2, Literature Review), which suggested that they may have an impact on a worker's risk of having an accident in the workplace. The chapter concludes by identifying the most appropriate model which represents workers' accident risk across England and Wales' workplaces.

5.2 Background and Methods

To carry out a regression analysis, which attempts to predict a relationship between a set of variables, a sample is first needed from a large population. In this investigation, the population was the current working population of England and Wales, and the sample was the RIDDOR dataset obtained from the HSE consisting of workers who have had a reported accident in the workplace. Due to the lack of availability of workplace population data for Scotland, the sample was restricted to include only England and Wales. The model created in this case aimed to identify the types of workers who are most likely to have a workplace accident based on the characteristics of the workers in the sample.

In the regression model the dependent variable was the count of the number of people having accidents in the workplace. The independent variables were the socio-economic area characteristics chosen based on the relevant literature suggesting these impact the risk of workplace accidents. The results from the recent 2011 Census contain an enormous amount of information with regards to the current population of England and Wales. The variables chosen were extracted from the 2011 Census results and include: the workplace population of England and Wales by age, gender, socio-economic classification group, occupation group and industry group. Population density (persons per hectare) was also obtained.

Each record in the RIDDOR dataset contains a postcode of where an accident occurred in England and Wales. Postcode level reveals little about the type of worker having an accident, and therefore it was necessary to aggregate postcode data into a suitably larger geographic area, in this case, Workplace Zones (WZ), containing more workers. This was achieved via the ONS Postcode Directory lookup table. The

resulting dependent variable for the model therefore consisted of a count of accidents by WZ.

The independent variables were also obtained by WZ. Ages were grouped, and the two age groups that were used within the model were ages 16 to 25 and 56 to 65. This was to enable the comparison of accident risk between younger and older workers. In terms of gender, male workers were included in the model and counts of workers by SOC, SIC and NS-SEC groups were included within the model, as well as population densities by WZ area. An offset variable was included to account for the differing population sizes. Each model had an offset of the log of the total workplace population, ages 16 to 74, by WZ.

In terms of variable selection, two main methods were considered. The first being selecting variables to include within the model based on the relevant literature. The second method was a data-driven variable selection procedure. Automated statistical algorithms such as stepwise regression are commonly used to test a large number of combinations of variables to identify a single model with the best fit. It is widely recognised however that these techniques often lead to incorrect statistical inference and frequently exploit random variations in sample data that result in models that statistically fit well to the sample but do not generalise well to the population under investigation (Derksen and Keselman, 1992; Hurvich and Tsai, 1990; Johnson et al., 2004). Instead, alternative information theoretic approaches are often used (Anderson et al., 2000), in particular the Akaike Information Criterion (AIC) (Akaike, 1973).

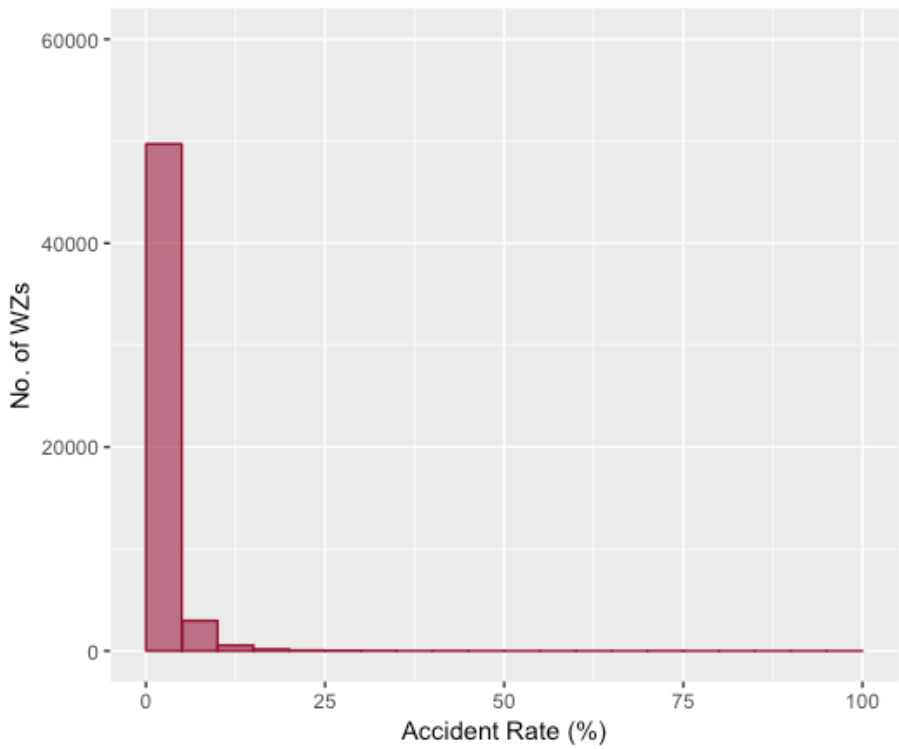
To maintain a balance between theory and data-driven based variable selection, a range of global variables was first chosen based on the relevant literature surrounding workplace accidents. This was carried out by identifying within the relevant literature, the potential risk-factors associated with having a workplace accident, and choosing variables to include within the model to test hypotheses on particular risk-factors. Specifically, these included: gender, age, occupation, industry type, socio-economic classification group and population density of an area worked in (outlined in Chapter 3). At each stage of fitting a regression model, variables were removed and added

until the most suitable model was found. Suitability was based upon the statistical significance of the addition of independent variables to the model, utilising the AIC as a measure of suitability. The model therefore with the smallest AIC was chosen as the most appropriate model overall.

5.3 Comparing Regression Models

To gain an understanding of the data, it was important to produce a model using the most appropriate distribution that best represents the data. The histogram (Figure 5.1) illustrates the frequency of accidents by WZ and reveals that accident counts are highly skewed towards zero. Workplace accidents are rare events, in that they are an uncommon occurrence. It is therefore not surprising that the RIDDOR data does not follow a normal distribution. To model this type of data, it is usual in these circumstances to consider a Poisson distribution.

Figure 5.1: Accident Rate by Number of Workplace Zones



One of the main properties of a Poisson regression model is that the mean should be equal to the variance. It is found however, that the variance in this case was much greater than the mean. By WZ, the approximate value of the mean was 2.13, and the approximate value of the variance was 9.82. As the variance was considerably larger than the mean in this case, it suggested that the Poisson distribution may not be suitable as the strict condition was not met, and overdispersion exists within the data.

The problem of overdispersion is common when carrying out Poisson regression and there are therefore several methods to explore to overcome this, which have been discussed in Chapter 3. The two most common methods used to deal with this are quasi-Poisson regression and negative binomial regression, which are tested and compared with the Poisson regression models in this chapter. The most appropriate model to represent the data is examined based on the level of statistical significance for the coefficient estimates and on the models' Akaike Information Criterion value as detailed above.

The first model considers gender as a cause of accidents in the workplace, specifically exploring whether the proportion of male workers within an area has an effect on workplace accident rate. Previous research into gender differences in workplace accidents find that men have a higher risk of accident than women (Berecki-Gisolf et al., 2015; Smith and Mustard, 2004). The gender variable is selected first as it is considered to be the most basic indicator for revealing differences between workers' accident risk. Below are the results of the first model fitted by a Poisson regression (Table 5.1) and also quasi-Poisson regression (Table 5.2) and negative binomial regression (Table 5.3). The coefficients (i.e. the independent variables) are listed for each model, together with the parameter estimates for each independent variable. Also indicated are the standard errors and the p values (PR(>|z|)), indicating the level of statistical significance of each of the coefficients. Below each of the models is a pseudo R^2 statistic and an AIC figure, with also a key illustrating where the p value is less than 0.001.

AICs depend on a distributional form and a likelihood; however quasi models are only characterised by their mean and variance, and do not necessarily have a

distributional form. This means that for the quasi-Poisson models, AIC values are unavailable. Alternative information theoretic approaches have been developed such as the quasi-AIC (QAIC; Burnham and Anderson, 2002), however these can only be used to compare between quasi-models. In terms of comparing results of the quasi-Poisson model with those of the Poisson and negative binomial models, the QAIC would not be appropriate either, and therefore has been excluded from the results.

Table 5.1: Poisson Regression Model 1: Males

Coefficients	**Estimate**	**Standard Error**	**PR(>|z|)**
(Intercept)	-3.82e+00	1.55e-03	<2e-16 ***
Males	6.72e-05	1.68e-06	<2e-16 ***

Pseudo R^2: 0.0025

AIC: 770799

*** < 0.001

Table 5.2: Quasi-Poisson Regression Model 1: Males

Coefficients	**Estimate**	**Standard Error**	**PR(>|z|)**
(Intercept)	-3.82e+00	7.47e-03	<2e-16 ***
Males	6.72e-05	8.12e-06	<2e-16 ***

Pseudo R^2: 0.0025

AIC: N/A

*** < 0.001

Table 5.3: Negative Binomial Regression Model 1: Males

Coefficients	**Estimate**	**Standard Error**	**PR(>|z|)**
(Intercept)	-4.04e+00	5.53e-03	<2e-16 ***
Males	6.96e-04	1.43e-05	<2e-16 ***

Pseudo R^2: 0.026

AIC: 344212

*** < 0.001

There are several similarities found between each of the three models above (Tables 5.1, 5.2 and 5.3). Firstly, all of the estimates are statistically highly significant as the p values are less than 0.001 for all coefficients. Each of the three models indicate that there is a positive relationship between the number of males working in an area and the number of accidents in the workplace. The coefficient estimate is 6.72e-05 for the males coefficient for the Poisson and quasi-Poisson models and 6.96e-04 for the negative binomial model. Interpreting this directly, for the Poisson and quasi-Poisson models for example, this means that for an increase of one in the number of males, the number of workplace accidents will increase by exp(6.72e-05) which is approximately 1.000067. This is equivalent to a rate of 0.0067% ((1.000067 – 1) * 100) more accidents. Similarly, under the negative binomial model, the number of accidents will increase by a rate of approximately 0.070%. The output for both the Poisson and negative binomial models include an AIC with the negative binomial model having the smallest AIC and the Poisson model having the largest. The pseudo R^2 statistic is 0.0025 for both the Poisson and quasi-Poisson models and 0.026 for the negative binomial model.

For the second model, to gain an understanding of the relationship of other demographic factors on accident risk, an age group variable was next added to the model. Specifically, the youngest workers, ages 16 to 25 was included. Tables 5.4, 5.5 and 5.6 show the results of the second model, which reveal the relationship between both males and young workers (ages 16 to 25 years) on accident rate. Again, Poisson, quasi-Poisson and negative binomial models have been fitted to the data to draw comparisons between outputs for all three types of regression models and to help identify the most appropriate model.

Table 5.4: Poisson Regression Model 2: Males + Age 16 to 25

Coefficients	**Estimate**	**Standard Error**	**PR(>|z|)**
(Intercept)	-3.81e+00	1.59e-03	<2e-16 ***
Males	9.76e-05	2.22e-06	<2e-16 ***
Age 16 to 25	-1.69e-04	8.57e-06	<2e-16 ***

AIC: 770371

*** < 0.001

Table 5.5: Quasi-Poisson Regression Model 2: Males + Age 16 to 25

Coefficients	**Estimate**	**Standard Error**	**PR(>|z|)**
(Intercept)	-3.81e+00	7.65e-03	<2e-16 ***
Males	9.76e-05	1.07e-05	<2e-16 ***
Age 16 to 25	-1.69e-04	4.13e-5	<2e-16 ***

AIC: N/A

*** < 0.001

Table 5.6: Negative Binomial Regression Model 2: Males + Age 16 to 25

Coefficients	**Estimate**	**Standard Error**	**PR(>|z|)**
(Intercept)	-4.03e+00	5.64e-03	<2e-16 ***
Males	8.98e-04	1.84e-05	<2e-16 ***
Age 16 to 25	-9.21e-04	5.64e-05	<2e-16 ***

AIC: 343930

*** < 0.001

Similarly to the results of the first model, all coefficients included within the second model are statistically highly significant with p-values less than 0.001. The inclusion of the age 16 to 25 variable in the model, changes the estimates for the intercept and males coefficients, but only slightly. The parameter estimate for the age 16 to 25 variable is negative for all three model types. Specifically, for the Poisson model for example, if the number of age 16 to 25 year old workers increases by 1, the

difference in the logs of expected counts would be expected to decrease by 1.69e-04, while holding the other variables in the model constant.

The AIC is smallest again for the negative binomial regression model and largest for the Poisson model. Although the smallest AIC in each model so far has been for the negative binomial models, the parameter estimates remain similar for all models. Also, as the Poisson regression is the simplest and easiest to implement in R and for the ease of transition into the local regression analysis in the next chapter that follows, the Poisson model has been selected as the most appropriate type of model for the global regression analysis. The following models below are therefore Poisson models only which are compared against each other based on AIC values.

5.4 Refining the Global Regression Model

The next step involved refining the model to find the variables which best explain variation in the numbers of workplace accidents. The models which follow test a variety of independent variables, such as the inclusion of extra age groups and occupational groups. Model 3 (summarised in Table 5.7) shows the results where an extra age group variable was added to the model. An on-going debate currently exists within the literature surrounding the effect of age as a risk-factor of workplace accidents (Bull et al., 2001; Nenonen, 2011). The addition of the age 56 to 65 group variable helps to contrast with the age 16 to 25 group variable. The results show that the age 16 to 25 estimate is negative, whilst the age 56 to 65 estimate is positive; this suggests that older workers have a higher risk of having a workplace accident. Note, in this model, the males coefficient has switched signs, and is now negative. Whilst this may be counterintuitive, based on the relevant literature surrounding male workers and workplace accidents, it is assumed that additional accident risk associated with being male is now being explained by the other variables within the model. All estimates are again statistically highly significant with p-values less than 0.001 and it is also worth noting that the AIC is decreasing from Model 1 to Model 3 with the inclusion of additional variables, suggesting that Model 3 is the most appropriate model tested so far.

Table 5.7: Model 3: Males + Age 16 to 25 + Age 56 to 65

Coefficients	**Estimate**	**Standard Error**	**PR(>|z|)**
(Intercept)	-3.85e+00	1.67e-03	<2e-16 ***
Males	-3.50e-05	2.84e-06	<2e-16 ***
Age 16 to 25	-3.80e-04	1.03e-05	<2e-16 ***
Age 56 to 65	1.19e-03	1.38e-05	<2e-16 ***

AIC: 763089

*** < 0.001

Continuing with incorporating additional variables into the model, the fourth model considers the same variables as in Model 3: males, age 16 to 25 and age 56 to 65, with population density added. The results from Model 4 are shown in Table 5.8. Again, all coefficient estimates are statistically significant and the AIC is the smallest out of all the models so far. In this model the males and age 56 to 65 variables have a positive relationship with accidents in the workplace, whilst the age 16 to 25 variable has a negative relationship. The population density estimate is also negative, suggesting that the more workers there are in an area, the less likely an accident is to occur.

Table 5.8: Model 4: Males + Age 16 to 25 + Age 56 to 65 + Population Density

Coefficients	**Estimate**	**Standard Error**	**PR(>|z|)**
(Intercept)	-3.82e+00	1.67e-03	<2e-16 ***
Males	6.97e-05	2.93e-06	<2e-16 ***
Age 16 to 25	-2.05e-04	9.10e-06	<2e-16 ***
Age 56 to 65	8.00e-04	1.47e-05	<2e-16 ***
Population Density	-4.28e-04	4.34e-06	<2e-16 ***

AIC: 747192

*** < 0.001

It would be interesting to consider the type of industry that workers are employed in on the risk of accidents in the workplace to identify whether a particular industry

may have an effect on accident risk more than another. Model 5 tests the variables from Model 4 together with selected industry groups. The Standard Industrial Classification groups included are SICs A to S which include a wide range of industries such as agriculture, manufacturing, transport, IT, education and health. SICs T and U, the remaining groups, are excluded from the model to account for interrelations between the SIC variables. These industrial groups represent 'Activities of households as employers- producing activities of households for own use' and 'Activities of extraterritorial organisations and bodies'. These SIC groups have a small percentage of workers employed within these sectors by WZ.

Table 5.9 shows a summary of the fifth model. The AIC is the smallest out of all the models so far, with a value of 694159. The signs of the coefficient estimates for males, both age groups and population density all remain the same as in the previous model. Almost all of the estimates for the SIC variables are negative, apart from SICs A, B and E. An issue with this model is that not all of the coefficient estimates are statistically significant, with the age 16 to 25 and SIC H variables having p-values greater than 0.05. This model has been refined multiple times with different subsets of the SIC variables, however in all cases some coefficient estimates were not significant, suggesting perhaps industrial sector may not be a good indicator for determining workplace accident risk.

Table 5.9: Model 5: Males + Age 16-25 + Age 56-65 + Population Density + SIC groups A to S

Coefficients	**Estimate**	**Standard Error**	**PR(>|z|)**
(Intercept)	-3.70e+00	2.28e-03	<2e-16 ***
Males	6.27e-04	1.03e-05	<2e-16 ***
Age 16 to 25	-2.73e-05	1.59e-05	0.086 .
Age 56 to 65	1.04e-03	3.43e-05	<2e-16 ***
Population Density	-1.16e-04	4.11e-06	<2e-16 ***
SIC A Agriculture, forestry and fishing	1.66e-03	1.03e-04	<2e-16 ***
SIC B Mining and quarrying	1.59e-03	4.20e-05	<2e-16 ***
SIC C Manufacturing	-4.03e-04	1.16e-05	<2e-16 ***

SIC D Electricity, gas, steam and air conditioning supply	-9.90e04	3.19e-05	<2e-16 ***
SIC E Water supply, sewerage, waste management and remediation activities	2.67e-04	2.42e-05	<2e-16 ***
SIC F Construction	-8.53e-04	3.06e-05	<2e-16 ***
SIC G Wholesale and retail trade, repair of motor vehicles and motorcycles	-1.46e-04	1.05e-05	<2e-16 ***
SIC H Transport and storage	-3.52e-05	1.02e-05	0.0584 .
SIC I Accommodation and food service	-4.16e-04	3.33e-05	<2e-16 ***
SIC J Information and communication	-1.30e-03	2.24e-05	<2e-16 ***
SIC K Financial and insurance activities	-7.39e-04	1.21e-05	<2e-16 ***
SIC L Real estate activities	-1.86e-03	7.16e-05	<2e-16 ***
SIC M Professional, scientific and technical activities	-2.55e-03	2.65e-05	<2e-16 ***
SIC N Administrative and support service activities	-1.07e-03	3.54e-05	<2e-16 ***
SIC O Public administration and defence compulsory social security	-5.28e-04	1.02e-05	<2e-16 ***
SIC P Education	-8.51e-04	1.28e-05	<2e-16 ***
SIC Q Human health and social work activities	-1.13e-04	5.81e-06	<2e-16 ***
SIC R, S Arts, entertainment and recreation and other service activities	-1.10e-03	3.90e-05	<2e-16 ***

AIC: 694159

*** < 0.001,

0.001 < ** < 0.01

0.01 < * < 0.05

0.05 < . < 0.1

It may be more insightful to consider occupation type on accidents in the workplace to identify how the type of job role or work carried out has an effect on accidents in the workplace. The variables included in Model 6 are similar to Model 5, but instead of the inclusion of Standard Industrial Classification groups, it includes Standard

Occupation Classification groups. For the purposes of this model, SOC Broad Group 3 has been removed to account for interrelations between variables. SOC 3 represents associate professional and technical occupations, which is similar to SOC 1 and SOC 2 in terms of the types of job roles and levels of education and training required to perform those roles.

Table 5.10: Model 6: Males + Age 16 to 25 + Age 56 to 65 + Population Density + SOC groups 1, 2 & 4-9

Coefficients	**Estimate**	**Standard Error**	**PR(>|z|)**
(Intercept)	-3.81e+00	2.02e-03	<2e-16 ***
Males	-2.68e-04	9.44e-06	<2e-16 ***
Age 16 to 25	-1.20e-04	2.03e-05	4.07e-09 ***
Age 56 to 25	1.39e-03	3.44e-05	<2e-16 ***
Population Density	-1.74e-04	4.34e-06	<2e-16 ***
SOC 1 Managers, Directors	-1.14e-03	3.40e-05	<2e-16 ***
SOC 2 Professional Occupations	-1.26e-04	6.84e-06	<2e-16 ***
SOC 4 Administrative and Secretarial Occupations	-1.94e-04	1.49e-05	<2e-16 ***
SOC 5 Skilled trades	1.70e-04	2.48e-05	6.02e-12 ***
SOC 6 Caring, leisure and other service occupations	1.60e-04	1.39e05	<2e-16 ***
SOC 7 Sales and customer service occupations	1.40e-04	1.58e05	<2e-16 ***
SOC 8 Process plant and machine operatives	7.97e-04	1.30e-05	<2e-16 ***
SOC 9 Elementary Occupations	8.38e-04	1.08e-05	<2e-16 ***

AIC: 711065

*** < 0.001

The AIC is slightly larger than that for the previous model which considered industrial sector, however the AIC is smaller than Models 1-4. All estimates are

statistically significant, with p-values < 0.001. The estimates are all similar to previous models, but with the males estimate being negative, whilst the age groups remain the same sign, as well as the population density variable. The SOC estimates vary between positive and negative values, with SOCs 1, 2 and 4 being negative and SOCs 5 to 9 being positive.

After testing different subsets of Model 6, the results of Model 7, the final model are shown in Table 5.11. The variables include: males, age 16 to 25, age 56 to 65 and population density, as well as a subset of the SOC variables and also included is NS-SEC 7 Routine occupations. The NS-SEC 7 routine occupations represents workers in positions with a basic labour contract. Examples of such occupations include: routine sales and service occupations, routine production occupations and routine technical operations. The NS-SEC measures the employment relations and conditions of occupations, which in effect reveal the structure of socio-economic positions. NS-SEC 7 is the lowest category within the hierarchical structure of NS-SEC groups, other than NS-SEC 8 which represents groups who have never worked or are long-term unemployed.

The key SOC broad groups chosen for this model are: SOC 1 Managers, directors and senior officials, SOC 8 Process, plant and machine operatives and SOC 9 Elementary occupations. These variables allow for a concise comparison between different types of occupational groups. SOC 1 Managers, directors and senior officials covers occupations whose tasks involve, planning, directing and coordinating resources to achieve the efficient functioning of a business or organisation. Most jobs within this occupational group usually require extensive business knowledge and high-level qualifications or training.

SOC 8 Process, Plant and Machine Operatives include occupations that involve workers operating industrial plant and equipment, assembling products from component parts and driving mobile machinery. Specific occupation titles include plant and machine operatives and road transport drivers. There are no formal academic qualifications required for these types of roles, however training is typically provided on-the-job. SOC 9 Elementary Occupations covers occupations that

require the knowledge and experience necessary to perform mostly routine tasks often involving the use of simple hand-held tools. The jobs involved in this broad group again like those in SOC 8 do not require formal academic qualifications but will usually have a short period of training before starting the job. Examples of occupations that typically fall into this broad group are window cleaners, lunchtime school supervisors and grocery assistants.

Table 5.11: Model 7: Males + Age 16 to 25 + Age 56 to 65 + Population Density + SOC 1 Managers, Directors and Senior Officials+ SOC 8 Process, Plant and machine Operatives + SOC 9 Elementary Occupations + NS-SEC7 Routine Occupations

Coefficients	**Estimate**	**exp(Est)**	**Standard Error**	**PR(>|z|)**
(Intercept)	-3.79e+00	0.0226	1.85e-03	<2e-16 ***
Males	-3.08e-04	0.9997	7.09e-06	<2e-16 ***
Age 16 to 25	4.57e-05	1.00005	9.95e-06	0.00012 ***
Age 56 to 65	8.76e-04	1.0009	1.64e-05	<2e-16 ***
Population Density	-2.09e-04	0.9998	4.32e-06	<2e-16 ***
SOC 1 Managers, directors and senior officials	-1.03e-03	0.9990	3.10e-05	<2e-16 ***
SOC 8 Process, plant and machine operatives	1.00e-03	1.0010	1.19e-05	<2e-16 ***
SOC 9 Elementary Occupations	8.42e-04	1.0009	1.69e-06	<2e-16 ***
NS-SEC 7 Routine Occupations	5.25e-05	1.00005	1.74e-05	0.00025 ***

Pseudo R^2: 0.20

AIC: 710540

*** < 0.001

Table 5.11 shows a summary of the results from Model 7. This model has the smallest AIC out of all the suitable models and has the largest pseudo R squared value. Additionally, all of the coefficient estimates are highly significant with p-values less than 0.001. A mix of positive and negative values exist for the coefficient estimates. Both the coefficient estimates for age groups 16 to 25 and 56 to 65 are positive, with values of 4.57e-05 and 8.76e-04 respectively. These values indicate that for an increase of one in the number of workers aged 16 to 25, the number of accidents will increase by exp(4.57e-05) = 1.00005. This is equivalent to a rate of 0.005% ((1.00005 – 1) * 100) more accidents. Similarly, for an increase of one in the number of workers aged 56 to 65, the number of accidents will increase by exp(8.76e-04) = 1.0009, ie. 0.09% more accidents.

Note again, as with several previous models, the estimate for the males coefficient is negative (-3.09e-04). This means that each male worker essentially reduces the rate of workplace accidents by a factor of exp(-3.09e-04) = 0.9997. This is equivalent to a reduction rate of 0.03% ((0.9997 – 1) * 100) in workplace accidents. As previously mentioned, this may appear counterintuitive, given the research surrounding male workers and occupational health and safety, however, it is likely that the risks associated with workplace accidents are most likely being explained by other variables within the model. The SOC 8 Process, plant and machine operatives and SOC 9 Elementary occupations coefficient estimates are positive, suggesting a positive relationship with the number of workplace accidents with these occupational groups. Every SOC 8 worker for example increases the rate of workplace accidents by approximately 1.0010 (0.1% more accidents), and for every SOC 9 worker, the rate of workplace accidents increases by 1.0009 (0.009% more accidents).

The estimate for NS-SEC 7 Routine occupations is positive (5.25e-05), suggesting that a positive relationship exists between the number of NS-SEC 7 workers and the number of workplace accidents. This means that each NS-SEC 7 worker essentially increases the rate of workplace accidents by a factor of exp(5.25e-05) = 1.00005. This is equivalent to a rate of 0.005% more workplace accidents. Finally, population density has a negative coefficient estimate, suggesting a negative relationship between the number of workers who work in an area and the number of workplace accidents.

Each worker essentially reduces the rate of workplace accidents by exp(-2.09e-04) = 0.9998, which approximately results in 0.02% less accidents.

Now that a final model has been constructed, it is appropriate to assess the relationship between the independent variables to be confident that no correlations exist between them. The correlation matrix (Table 5.12), reveals that there is no evidence of strong correlation between any of the variables included within the model. As a result, this suggests that there is no adverse impact upon the results of the model, as the coefficients are considered independent of each other.

Table 5.12: Correlation Matrix

	Males	**Age 16-25**	**Age 56-65**	**Population Density**	**SOC 1**	**SOC 8**	**SOC 9**	**NS-SEC7**
Males	1.00	0.63	0.70	0.30	0.69	0.57	0.61	0.59
Age 16-25	0.63	1.00	0.48	0.25	0.55	0.22	0.55	0.37
Age 56-65	0.70	0.48	1.00	0.06	0.44	0.42	0.58	0.54
Population Density	0.30	0.25	0.06	1.00	0.42	-0.04	0.05	-0.02
SOC 1	0.69	0.55	0.44	0.42	1.00	0.23	0.34	0.28
SOC 8	0.57	0.22	0.42	-0.04	0.23	1.00	0.51	0.70
SOC 9	0.61	0.55	0.58	0.05	0.34	0.51	1.00	0.69
NS-SEC7	0.59	0.37	0.54	-0.02	0.28	0.70	0.69	1.00

5.5 Conclusion

This chapter has explored the relationships between a set of socio-economic characteristics and workplace accidents. Based on the RIDDOR accident data provided by the HSE, a Poisson regression was first considered to model the data. As overdispersion was found to exist within the data, two alternative methods were considered, informed by relevant literature. These were quasi-Poisson regression and negative binomial regression models. The chapter explored all three methods and considered the strengths and weaknesses of each. Taking into consideration the similarity between results for the Poisson, quasi-Poisson and negative binomial regression models, and due to the flexibility and availability of the use of the Poisson regression in local regression methods within software packages, it was decided that the Poisson distribution best suited the data.

Several Poisson regression models were then tested in this chapter for the basis of representing a global model for the accident data. Variables including gender, age, industry group, occupation group, population density and socio-economic classification group were all considered and in a variable selection process of finding the most suitable model, variables were added to and dropped from the models based upon the structure of the results and the AIC values. The most suitable model chosen to best represent the data had a larger pseudo R^2 statistic than initial models considered and had the smallest AIC value. All of the estimates of the coefficients included within the final model were statistically highly significant with p-values less than 0.001.

The final model included the following coefficients: males, ages 16 to 25 year olds, ages 56 to 65 year olds, population density, SOC 1 Managers, directors and senior officials, SOC 8 Process, plant and machine operatives, SOC 9 Elementary occupations and NS-SEC 7 Routine occupations. The coefficient estimates within the final model identified characteristics which were positively associated with workplace accidents. Age 16 to 25 years olds, and age 56 to 65 year olds were found to have a positive relationship with accidents, with the age 56 to 65 coefficient having a larger positive estimate, and therefore suggesting a higher accident risk. The

males coefficient had a negative relationship with workplace accidents, as did population density, suggesting that the more workers there are in an area, and specifically the more men there are working in an area, the less accidents occur.

Out of the occupational groups, the SOC 1 Managers, directors and senior officials variable had a negative estimate suggesting a negative relationship exists with SOC 1 workers and workplace accidents. In contrast, the coefficient estimates for SOC 8 and SOC 9 were positive, suggesting a positive link between these occupational groups and workplace accidents. Finally, the estimate for NS-SEC 7 Routine occupations was also positive, which suggests that people grouped within the NS-SEC 7 socio-economic class have an increased risk of having a workplace accident.

The model overall provides a significant amount of information to provide an overview on the likely socio-economic area characteristics that have an effect on workplace accidents. This can help inform the Health and Safety Executive on the types of workers most at risk of having an accident in the workplace. This model will be the focus of the next chapter, looking at Geographically Weighted Regression as a localised regression analysis of workplace accidents. Currently, the overall model summarised in this chapter illustrates the relationship between a set of socio-economic area characteristics and accident rate across workers in England and Wales as a whole area. The chapter that follows will help to provide an understanding of how the relationship between these variables might vary geographically across different regions.

Chapter 6

A Local Regression Analysis of Workplace Accidents

6.1 Introduction

The previous chapter explored the relationship between socio-economic variables and accidents in the workplace, at a global scale, encompassing England and Wales as a whole area. This is a useful approach in terms of understanding the basic links between what causes workplace accidents and pinpointing specific high-risk socio-economic groups within the population. It is often the case, however, that relationships between variables vary geographically, and what socio-economic groups may be of a high accident risk in one area, may not necessarily be a high accident risk group within another area. This could be due to a number of reasons including environmental factors, such as weather conditions and daylight hours.

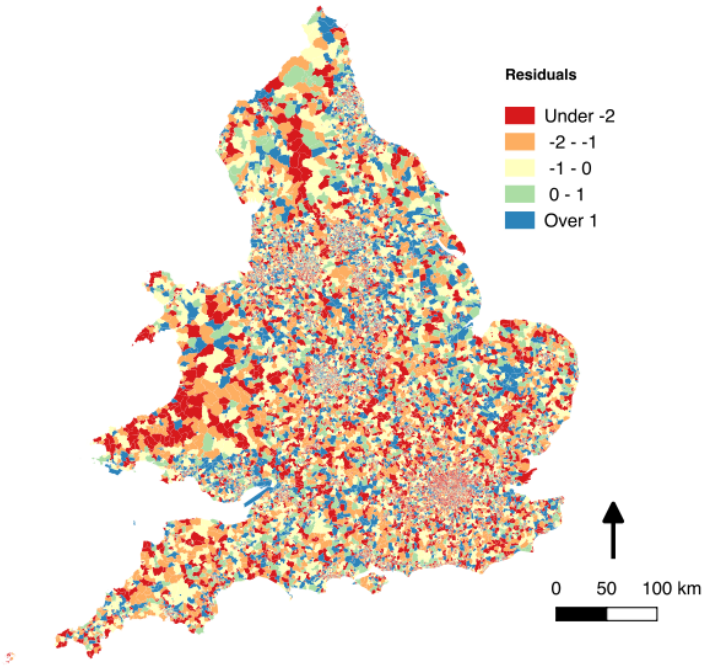
This chapter therefore explores the spatial element of workplace accidents that is missing from the 'global' Poisson regression model found in Chapter 5. The process aims to gain a local perspective on accidents in the workplace based on socio-economic area characteristics. The basis of this local regression analysis is a Geographically Weighted Regression (GWR: Brunsdon et al., 1998; Fotheringham et al., 2002) analysis of the RIDDOR dataset. GWR, in basic terms, is a method for analysing the spatially varying relationships between a set of variables and is being increasingly used in research which involves spatial data.

The RIDDOR dataset is split into three case study areas: London, North West England and North East England, and GWR is carried out individually on these areas. This selection is partly for computational reasons and partly to provide a focused comparison of regional areas. Further GWR analysis is carried out on the RIDDOR data by two seasons: summer and winter. Several key socio-economic variables are plotted to show the geographical variation of their relationship with workplace accident rate. Urban versus rural differences are highlighted across the three regional case studies, as well as season differences, and key results are discussed and summarised in the sections that follow.

6.2 Background and Methods

The global model (Chapter 5) essentially predicts the number of accidents likely to occur across England and Wales, dependent upon the number of workers in certain socio-economic groups. Figure 6.1 shows a map of the residuals of the global model across England and Wales. As can be seen from the map, the residuals vary geographically, with the model essentially predicting more accurately the number of accidents in some WZs than in others. This suggests that the model fits better in some areas than others and that therefore the relationship between the independent variables and the count of the number of accidents is not constant over England and Wales as a whole area.

Figure 6.1: Global model residuals map for England and Wales by Workplace Zone



Geographically Weighted Regression (GWR) is introduced as a method for exploring the local trends found within the data. Chapter 3 outlines the theoretical background of GWR and this part of the chapter details the stages of a GWR analysis that was carried out on the RIDDOR and census data. The first step considered was to carry out a GWR analysis for the RIDDOR data covering England and Wales, similarly to the global model. The aim of this was to identify trends, comparing patterns regionally, across a national scale as well as identifying urban to rural and coastal trends. As the data had already been organised into WZs, this next step was considered as a logical follow on from the previous work.

The GWR analysis was carried out using the standalone Microsoft Windows-based application, GWR4. GWR4 is an open source software that was developed by the GWR4 Development Team, consisting of several scholars including: Brunsdon,

Fotheringham and Charlton, the creators of GWR. The team were similarly involved in the creation of GWR R packages **spgwr** and **GWmodel**. Unfortunately, running the model on all WZs across England and Wales failed, and similarly failed attempts were made using R packages **spgwr** and **GWmodel**. It was considered that an alternative method of selecting several case study areas was required.

The RIDDOR dataset was split into three case study regions: North West England, North East England and London. These areas were chosen on the basis of them being representative of the diverse working populations across England and Wales, providing a cross section of geographical areas from the North, South, East and West of England. The North West is a region consisting of coastal and rural areas (Cumbria and Cheshire), as well as built up urban areas, including two major cities: Liverpool and Manchester. Workers are generally employed across a range of industrial sectors within the region. Similarly, the North East region contains three large cities: Newcastle upon Tyne, Sunderland and Durham. The region also consists of large rural and coastal areas including Northumberland, located within the North of the region, bordering Scotland. London was chosen as the third case study region because it is the largest urban area of the UK.

For each of the case study regions, a 5km buffer was first created around the boundaries of the regions. Any WZs that were overlapping this buffer were included within the GWR analysis for that particular area. This was to allow for a more continuous model, which considered relationships between neighbouring WZs outside of the region area, rather than creating an 'abrupt' boundary. The first step in GWR4 was importing the RIDDOR data and defining the GWR model. The X and Y coordinates were defined as the centroids of the WZs and the model type was selected as Poisson. The dependent variable was added as the frequency of accidents by WZ and an offset was added to the model, in this case (as for the global model) the log of the total working population (aged 16 to 75). This was to account for the differing population sizes within each WZ. The independent variables were added next to the model which included: males, aged 16 to 25 and 56 to 65, population density, SOC 1 Managers, directors and senior officials, SOC 8 Process, plant and machine operatives, SOC 9 Elementary Occupations, and NS-SEC 7 Routine

occupations, matching the variables included in the final global Poisson model (Chapter 5).

Also within the process, included setting the kernel function type to adaptive bi-square and selecting the golden selection search for the automatic bandwidth calculation. This was to determine the optimal number of nearest neighbours around each point to carry out GWR analysis. Regression diagnostics were used to judge the goodness of fit of the models, and the variance inflation factor (VIF) was used as an indicator of multicollinearity. Typically, a VIF value of greater than 10 suggests severe collinearity (Kennedy, 2003; Neter et al., 1989).

The output consisted of a summary text file, giving details of the GWR model such as summary statistics of the selected variables and the optimal bandwidth size. Also included within the output was a separate GWR csv file listing the estimated local statistics, including the geographically varying coefficient values. The sections that follow within this chapter outline the results obtained from running the GWR analysis and the variations across each of the regions are discussed in detail.

6.3 Global Regression Model by Case Study Area

The first table shown below is the finalised global model for England and Wales that was summarised in Chapter 5. As the RIDDOR data were split into case study areas to carry out a local regression analysis, the global model has been applied to the data for London, North West England and North East England. The results for these three global models are shown in Tables 6.2, 6.3 and 6.4, below the results of the global model for England and Wales (Table 6.1).

Table 6.1: Global regression model for England and Wales

	Estimate	Standard Error	PR(>|z|)
Intercept	-3.79e+00	1.85e-03	<2e-16 ***
Males	-3.08e-04	7.09e-06	<2e-16 ***
Age 16 to 25	4.57e-05	9.95e-06	0.00012 ***
Age 56 to 65	8.76e-04	1.64e-05	<2e-16 ***
Population Density	-2.09e-04	4.32e-06	<2e-16 ***
SOC 1 Managers, directors & senior officials	-1.03e-03	3.10e-05	<2e-16 ***
SOC 8 Process, plant and machine operatives	1.00e-03	1.19e-05	<2e-16 ***
SOC 9 Elementary occupations	8.42e-04	1.69e-05	<2e-16 ***
NS-SEC 7 Routine occupations	5.25e-05	1.74e-05	0.00025 ***
			AIC: 712540
			*** < 0.001

Table 6.2: Global regression model for London

Coefficients	Estimate	Standard Error	PR(>|z|)
Intercept	-4.16e+00	4.78e-03	<2e-16 ***
Males	-4.43e-04	2.05e-05	<2e-16 ***
Age 16 to 25	8.80e-05	2.46e-05	0.00039 ***
Age 56 to 65	5.41e-04	5.19e-05	<2e-16 ***
Population Density	-1.85e-04	6.23e-06	<2e-16 ***
SOC 1 Managers, directors & senior officials	-3.24-e04	6.03e-05	7.53e08 ***
SOC 8 Process, plant and machine operatives	-2.60e-04	6.67e-05	9.42e05 ***
SOC 9 Elementary occupations	1.54e-03	3.29e-05	<2e-16 ***
NS-SEC 7 Routine occupations	2.12e-03	7.03e-05	0.0025 **
			AIC: 114208
			*** < 0.001
			** < 0.005

Table 6.3: Global regression model for North West England

Coefficients	Estimate	Standard Error	PR(>|z|)
Intercept	-3.74e+00	4.90e-03	<2e-16 ***
Males	-1.69e-04	1.70e-05	<2e-16 ***
Age 16 to 25	-7.17e-05	2.56e-05	0.0042 **
Age 56 to 65	2.43e-05	4.63e-05	<2e-16 ***
Population Density	-7.84e-05	1.03e-05	2.98e-14 ***
SOC 1 Managers, directors & senior officials	-2.31e-03	1.03e-04	<2e-16 ***
SOC 8 Process, plant and machine operatives	8.88e-04	3.79e-05	<2e-16 ***
SOC 9 Elementary occupations	9.92e-04	3.52e-05	<2e-16 ***
NS-SEC 7 Routine occupations	1.63e-03	4.31e-05	<2e-16 ***
			AIC: 92623
			*** < 0.001
			** < 0.005

Table 6.4: Global regression model for North East England

Coefficients	Estimate	Standard Error	PR(>|z|)
Intercept	-3.82e+00	9.44e-03	<2e-16 ***
Males	-2.29e-04	3.84e-05	2.56e-09 ***
Age 16 to 25	-1.24e-03	5.09e-05	<2e-16 ***
Age 56 to 65	-6.55e-04	9.22e-05	1.21e-12 ***
Population Density	-3.72e-04	4.44e-05	<2e-16 ***
SOC 1 Managers, directors & senior officials	2.27e-03	2.02e-04	<2e-16 ***
SOC 8 Process, plant and machine operatives	7.84e-05	8.63e-05	<2e-16 ***
SOC 9 Elementary occupations	2.11e-03	1.15e-04	<2e-16 ***
NS-SEC 7 Routine occupations	1.87e-03	1.25e-04	<2e-16 ***
			AIC: 28608
			*** < 0.001

As can be seen from the tables, all of the coefficient estimates are statistically highly significant, with p-values less than 0.005. There are some noticeable differences between the coefficient estimates in each of the regions compared to the national

global model. Some coefficients in the regional models have stronger relationships with the counts of accidents compared to the coefficients in the model for England and Wales and there are some variations between the estimates in the regional models.

The coefficient estimates for the number of males variable are negative for all regions and for England and Wales as a whole. The largest negative value, indicating the strongest negative relationship with accident rate is found in London. For the number of young workers (ages 16 to 25), the coefficient estimate is positive in England and Wales and the London region, however the coefficient estimate is negative in the North West and North East regions. Similarly, for older workers (age 56 to 65), the estimates are positive for England and Wales, and the London region, as well as the North West, however is negative for the North East.

The estimates for population density are negative across all areas, with the North East having the largest negative value of -0.000372. In terms of the occupation coefficients, the estimates for SOC 1 Managers, directors and senior officials are negative across all areas other than the North East, where a positive value of 0.00227 is found. Alternatively, estimates for SOC 8 Process, plant and machine operatives are positive for all areas, but negative for London. This suggests that in London a negative relationship exists between SOC 8 workers and workplace accidents, whilst for the North West, North East, and England and Wales, a positive relationship between SOC 8 workers and workplace accidents exists. For SOC 9 Elementary occupations, all estimates are positive, with the largest value found for the North East (0.00211). This suggests that the strongest relationship between the number of SOC 9 workers and accident rate exists in the North East, compared to London, North West, and England and Wales as a whole. The coefficient estimates for NS-SEC 7 Routine occupations are all positive, with the largest estimate found for London (0.00212). All estimates are larger than in the global model covering England and Wales.

Overall, it is evident that the coefficient estimates vary between the geographic areas modelled. For some coefficients, one or several case study areas have a stronger

relationship with workplace accidents, than England and Wales. In other cases, the variables within England and Wales have a stronger relationship with workplace accidents. As a result, there is evidence shown that it may be unreliable to consider only one global model to represent the workplace population of every part of the country. The following sections in this chapter explore the local variations in the relationship between the socio-economic coefficients and accidents in the workplace, drawing upon the results found from the global models.

6.4 Geographically Weighted Regression

Geographically Weighted Regression was carried out on the RIDDOR data split by case study area. First the results are shown for London, followed by the North West, and then for the North East. A location map for each of the case study areas are presented as well as residuals and pseudo R squared maps indicating how well the GWR model fits the data. A selection of coefficient estimate maps then follow which show some key results illustrating variations in the relationship between the selected variables and accident rates by WZ. The main key coefficient maps included are for the occupation coefficients: SOC 1 Managers, directors and senior officials and SOC 9 Elementary occupations. These were selected since they provide contrasting cases which correspond to different degrees of spatial variation. Also included for some regions are NS-SEC 7 Routine occupations and SOC 8 Process, plant and machine operatives, where additional key contrasting results are highlighted. The maps of the remaining GWR coefficients are included within the Appendix.

6.4.1 London GWR

A GWR model was first fitted to data covering the London region. The optimum bandwidth size was found to be 1,269 nearest neighbours, computed using the golden selection search in the GWR4 application (as explained in Chapter 3). This indicates that the GWR analysis had taken account of the 1,269 nearest neighbouring WZs in calculating the coefficient estimates for each WZ. The model diagnostic

indicators are also produced within the text file output of the GWR results, in which the AIC is shown to be 70282, which is considerably smaller than that of the global regression model for London (114208), indicating a better fit. Summary statistics for the local coefficients are provided in Table 6.5. These include the minimum, median, maximum and interquartile range. Comparing Table 6.5 with the global model results for London (Table 6.2), there are some key differences. The median estimates in the local model differ to the global model for all coefficients and the IQR values show the spread of values around the average value for the coefficient estimates, showing spatial variation exists between WZs. In particular, the median estimate for the males coefficient is smaller than the global model estimate. Also, both median values for the local model estimates for the age groups are higher than the global model for London.

Table 6.5: Summary statistics for London GWR model

Coefficient	**Min**	**Median**	**Max**	**IQR**
Intercept	-0.10	0.53	1.89	0.30
Males	-0.002	-0.000047	0.0064	0.0012
Age 16 to 25	-0.0062	0.00075	0.0047	0.0013
Age 56 to 65	-0.0075	0.0031	0.0081	0.0029
Population Density	-0.0027	-0.00012	0.0027	0.00046
SOC 1	-0.029	-0.00026	0.0061	0.0035
SOC 8	-0.012	-0.0011	0.01	0.0046
SOC 9	-0.0041	0.0031	0.023	0.0031
NS-SEC 7	-0.026	0.0016	0.015	0.0052

Below is a location map of London (Figure 6.2), illustrating where some of the districts within the London region are located. Also shown is the residuals map of the GWR model. The map shows that the residual values, which indicate how well the GWR model fits the data, are consistently small across the areas of the WZ-level map of London. Values generally fall around zero, suggesting that the predicted values of accidents match closely to the observed values of accidents. The model

performs slightly less well in some WZs, represented by the extremities of the colour ranges- the red and blue areas. Overall, however, consistent mid-range values around zero are found across the regional map, confirming the predictive power of the GWR model. The pseudo R^2 map indicates that the model provides a moderately good fit across London, particularly in the North and West parts of the region, represented by the red and orange areas. The green and blue areas around the East and South of the region suggest a poorer fit of the model.

Figure 6.2: Location map for London and residuals and pseudo R^2 Maps for London GWR model

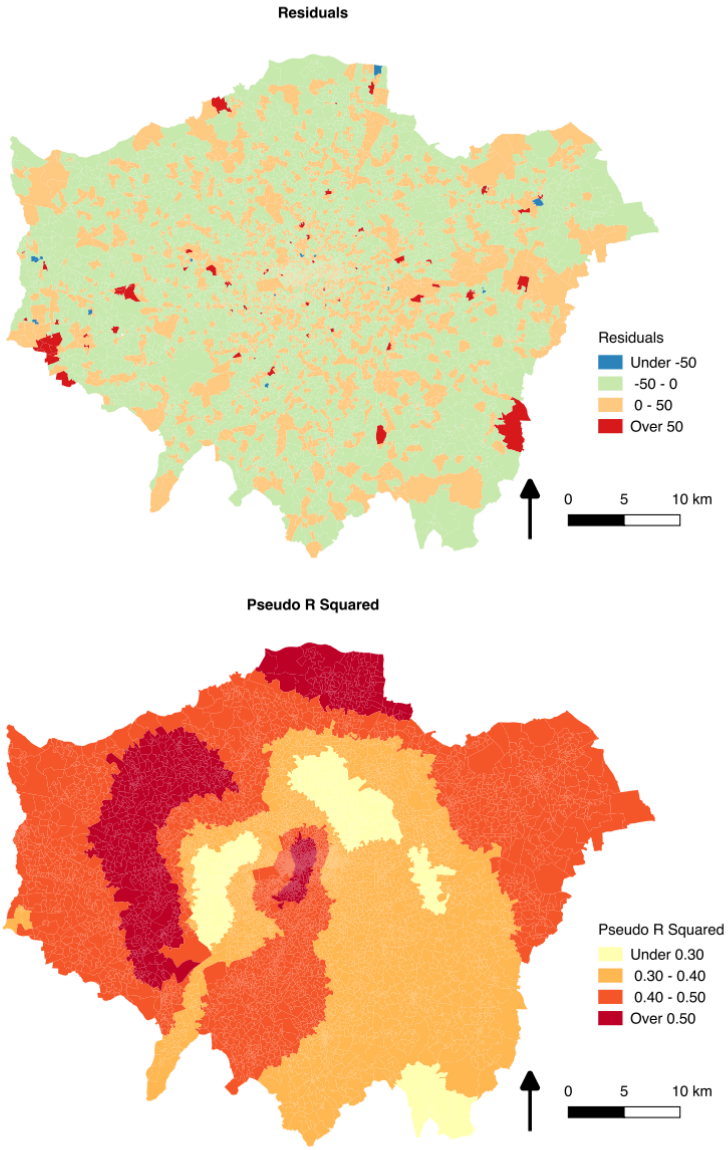


Figure 6.3 below shows the GWR coefficient results for the SOC 1 Managers, directors and senior officials variable. The first map provides a background to the percentage of SOC 1 workers there are located by WZ across the region and the second map shows the coefficient estimates from the local model. The higher percentages of SOC 1 workers appear to be located within the central areas of London, particularly the West (Westminster, Hammersmith and Fulham and Camden). Less SOC 1 workers are located within the East of London and parts of the South (Newham, Lewisham and Greenwich). The map of the GWR coefficients show a range of values (positive and negative) across the London region. The positive GWR values are located in the East, parts of the North and South West of

London. These appear to contrast with the levels of SOC 1 workers in these areas. In general, a small percentage of SOC 1 workers are employed where the GWR coefficient is highest. The negative GWR coefficients are found in the West of the region around Hounslow and Richmond upon Thames.

Figure 6.3: London SOC1 Managers, directors and senior officials- percentage of workers & GWR coefficients

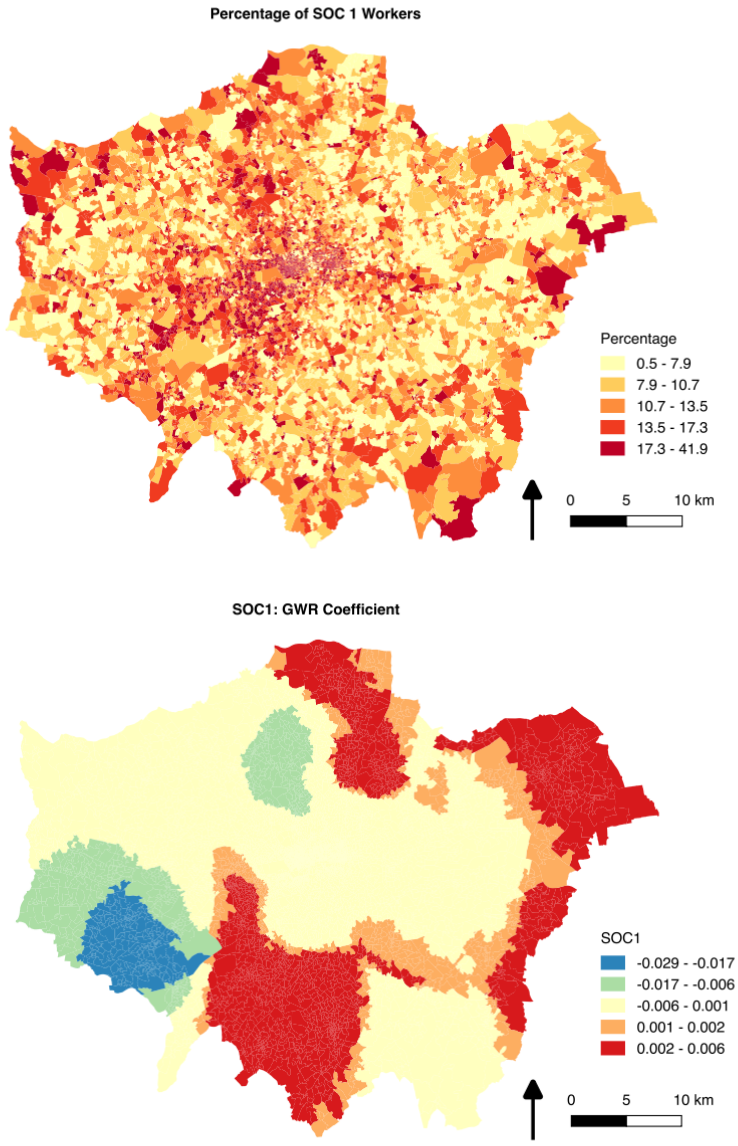


Figure 6.4 shows the percentage of workers and GWR coefficient maps for SOC 8 Process, plant and machine operatives. The largest percentages of SOC 8 workers are

located in the West and East areas of London, represented by the dark orange and red areas. The GWR coefficient map shows a range of estimate values from negative to positive. The positive coefficient estimates are mostly located in the South of the region (where a smaller percentage of SOC 8 workers are located), and negative estimates are located in the West of the region (where larger percentages of SOC 8 workers are located).

Figure 6.4: London SOC8 Process, plant and machine operators- percentage of workers & GWR coefficients

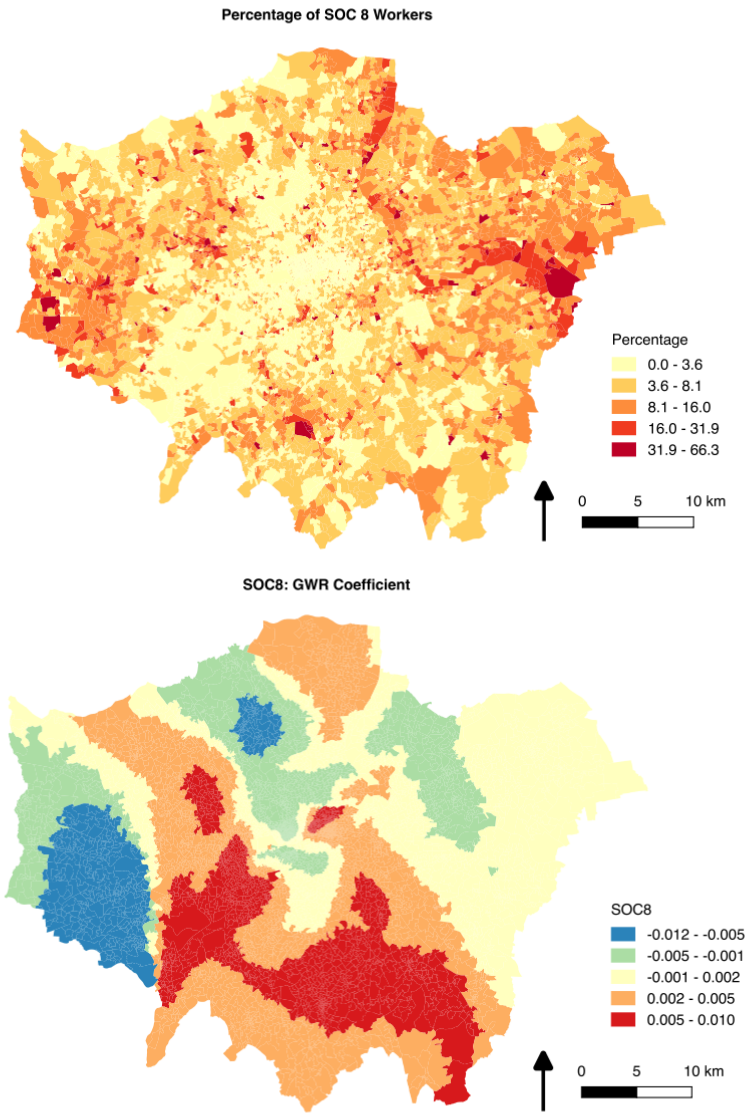
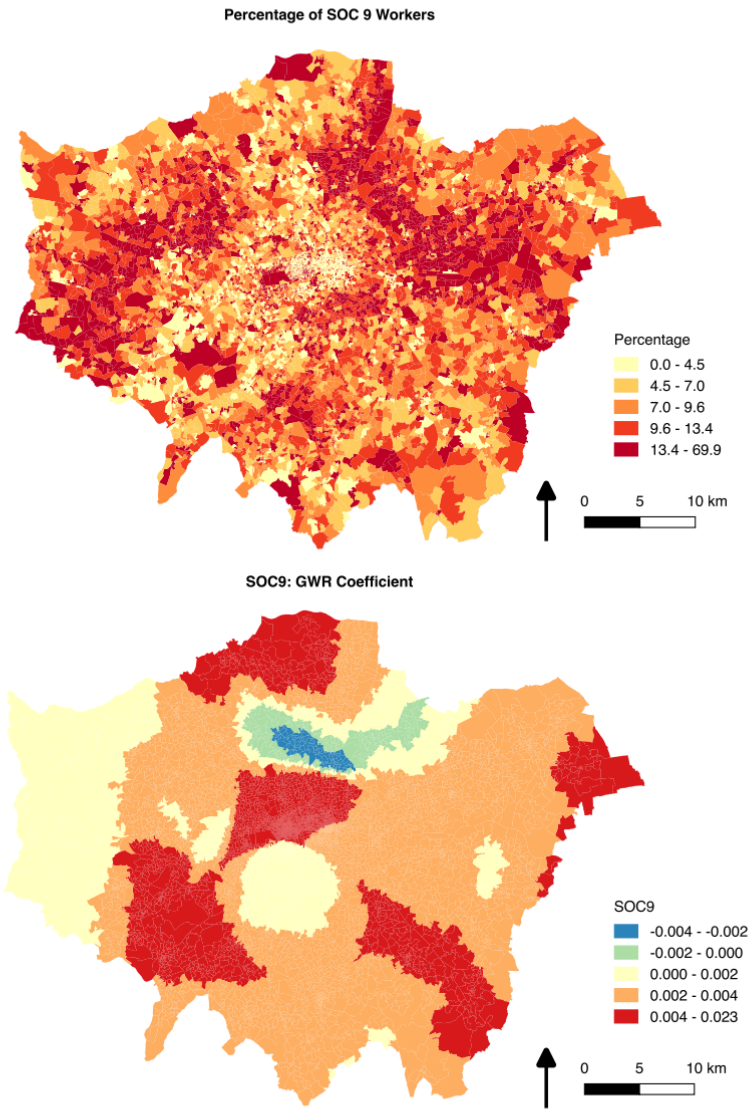


Figure 6.5 shows the maps for the percentage of SOC 9 Elementary occupations in London and the coefficient estimates for the GWR model. The percentage of SOC 9 workers is highest in the outer areas of London, particularly in the East (Newham and Redbridge) and West (Hounslow and Ealing). The smallest percentages of SOC 9 workers appear to be in central London and parts of the West such as Wandsworth and Hammersmith and Fulham. The GWR estimates vary across the region, with the highest positive values in some areas associated with higher percentages of SOC 9 workers (parts of East and West London), however also in areas where the percentages of SOC 9 workers are small. These areas include parts of central London and north London (Camden and Westminster, and Barnet and Enfield).

Figure 6.5: London SOC 9 Elementary occupations- percentage of workers & GWR coefficients



6.4.2 North West GWR

A GWR model was next fitted to data covering the North West region. The optimum bandwidth size was found to be 325 computed using the golden selection search in the GWR4 application. This means that the GWR analysis carried out for the North West considered the 325 nearest neighbouring WZs to derive the coefficient estimates for each WZ. The AIC was 53933, which is clearly smaller than that of the global regression model for the North West (92623), indicating a better fit. The summary statistics for the local coefficients are shown in Table 6.6. Most of the local model median estimates are larger than the global model estimates for the North West. The IQR again shows the dispersion of values around the median estimate, suggesting spatial variation exists between WZs.

Table 6.6: Summary statistics for the North West GWR model

Coefficient	Min	Median	Max	IQR
Intercept	-0.15	0.77	1.50	0.36
Males	-0.0039	0.00064	0.011	0.0022
Age 16 to 25	-0.012	-0.00052	0.013	0.0028
Age 56 to 65	-0.021	0.0016	0.012	0.0036
Population Density	-0.0078	-0.00012	0.0044	0.00093
SOC 1	-0.018	-0.0012	0.016	0.0083
SOC 8	-0.015	0.0019	0.016	0.0042
SOC 9	-0.015	0.0037	0.021	0.0047
NS-SEC 7	-0.026	0.0016	0.015	0.0052

Figure 6.6 shows a location map of the North West region, illustrating where some of the districts within the region are located. Also included is the residuals map of the GWR model and pseudo R squared map. The residuals map shows that the residual values for each WZ are consistently small across all parts of the region. Values generally fall around the zero value, similarly to the GWR residuals for London, suggesting that the predicted values of accidents match closely to the observed values of accidents. The model performs slightly less well in some WZs than in others,

represented by the extremities of the colour ranges- small areas of red across the region. The pseudo R^2 map also shown in Figure 6.6 similarly reveals a moderate to good fit of the model in the study area. The dark red areas suggest that the model fits well- particularly in Merseyside, parts of East Manchester and Carlisle. The yellow and light orange colours suggest areas where the model does not fit as well, with R squared values less than 0.5. These are located within the centre of the region around Lancaster as well as parts of Cheshire.

Figure 6.6: Location map for North West England and residuals and pseudo R^2 maps for North West England GWR model

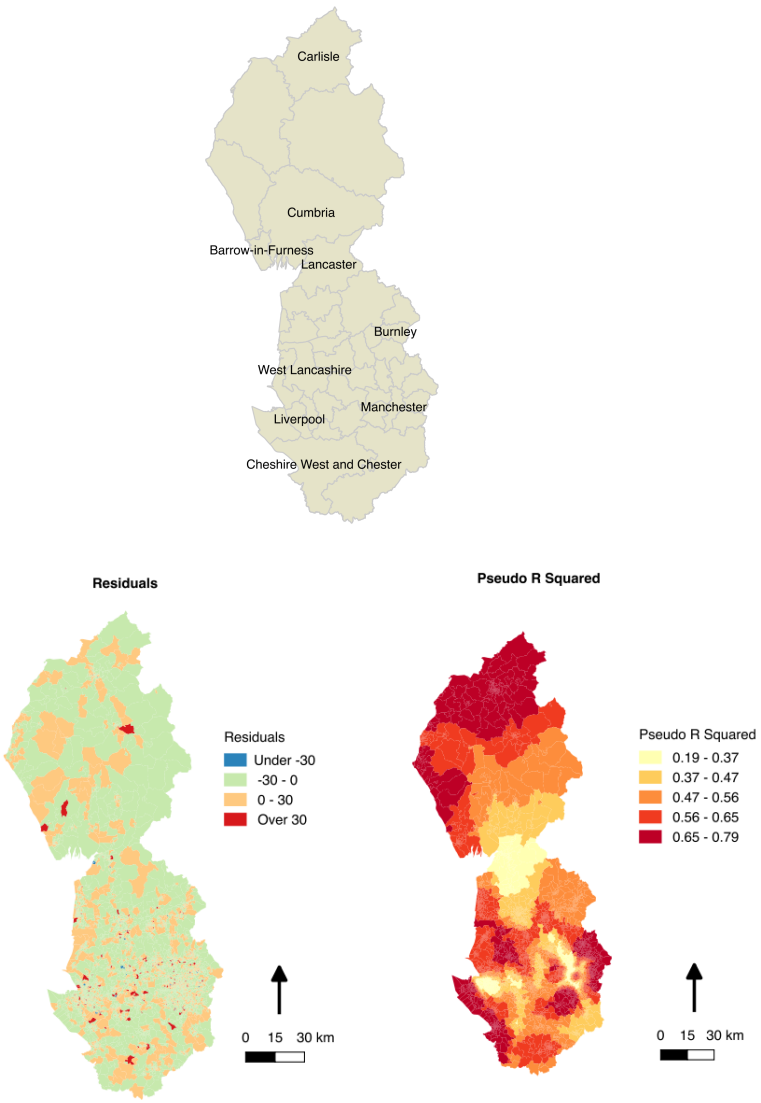
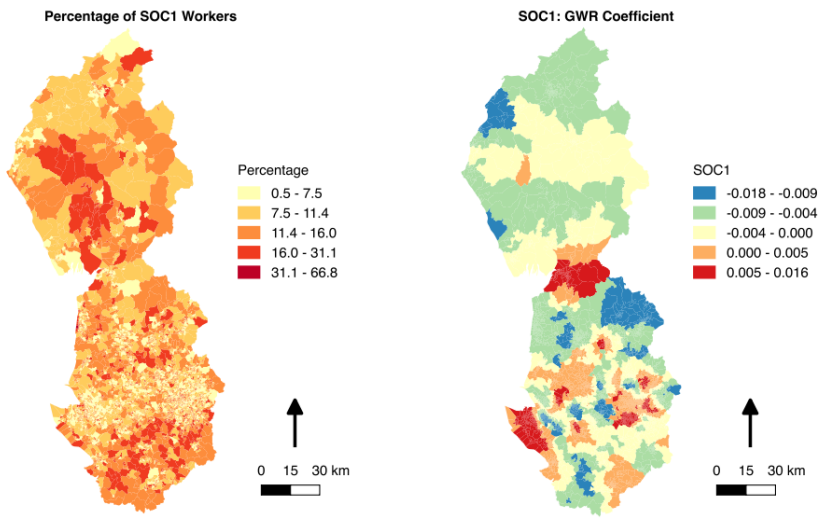


Figure 6.7 shows a map of the percentages of workers in North West England who work in SOC 1 Managers, directors and senior officials. The higher percentages are located in Cheshire, as well as parts of Cumbria. The smaller percentage of SOC 1 workers are located in Merseyside and parts of Manchester and Lancaster. The map of the GWR coefficient estimates are also shown in Figure 6.7. The red coloured areas show where the relationship between SOC 1 workers and workplace accidents is strongly positive, whereas the blue areas show where the relationship is strongly negative. The largest positive GWR estimates are found in Merseyside and Lancaster, as well as parts of Manchester. These areas are where the percentages of SOC 1 workers are small. Where the percentages of SOC 1 workers are high (for example parts of Cheshire and Cumbria), a negative relationship with accidents is found.

Figure 6.7: North West SOC 1 Managers, directors and senior officials- percentage of workers & GWR coefficients



The results for the SOC 9 Elementary occupations coefficient estimates are shown in Figure 6.8, together with the percentage of workers in this occupational group. The percentages of SOC 9 workers varies across the region, with higher percentages found in parts of West Cumbria, Lancashire and South Cheshire. The GWR coefficient estimates are mixed across the region, with some higher values associated

with areas that have a larger percentage of SOC 9 workers (particularly West Cumbria and South Cheshire). Negative GWR estimates are found in the South of Manchester and parts of Lancaster.

Figure 6.8: North West SOC 9 Elementary occupations - percentage of workers & GWR coefficients

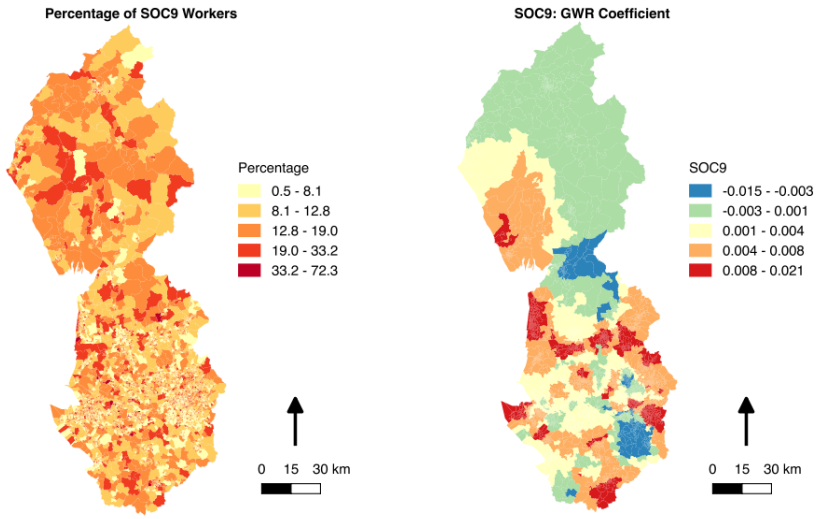
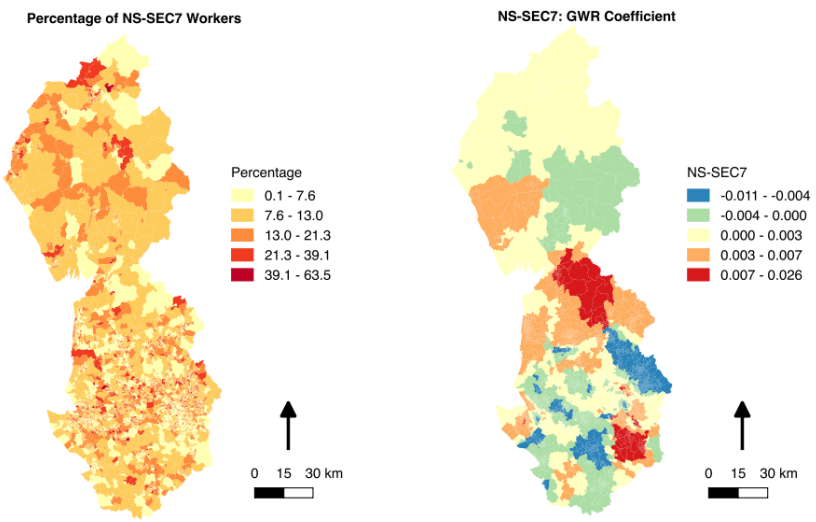


Figure 6.9 shows the maps of the percentage of workers across the North West classified as NS-SEC 7 Routine Occupations, and the GWR coefficient for NS-SEC 7. The percentages of NS-SEC 7 workers by WZ working population varies across the region. The GWR coefficient estimates also vary across the region, but with patches of positive values and negative values. The red areas around Lancashire and South of Manchester show that the strongest positive relationship between NS-SEC 7 workers and accident rate exists. These are located in areas where the percentages of NS-SEC 7 are small. Similarly, across the middle of the region and parts of Cumbria, mid-range values are found. Areas with a negative relationship between NS-SEC 7 and accident rates exists in central Manchester and parts of Merseyside and Cheshire.

Figure 6.9: North West NS-SEC 7 Routine occupations - percentage of workers & GWR coefficients



6.4.3 North East GWR

A GWR model was next fitted to data covering the North East region. The optimum bandwidth size was found to be 58, which again was computed using the golden selection search in GWR4. The AIC for the GWR model was 9208, which is smaller than that of the global regression model for the North East (28608), therefore indicating a better fit than the global model for the region. The summary statistics for the local coefficients are given in Table 6.7, these again are the minimum, maximum, median and interquartile range. Comparing these statistics to the North East global model, there are clear differences that exist between the coefficient estimates. The median local estimates are larger for the males, age 56 to 65, SOC 8 and SOC 9 variables. A switch in sign between the males and SOC 1 estimates exists between the local and global estimates. The minimum and maximum and IQR shows the spread of values for the local model, in comparison to the global model, revealing the spatial variation in results.

Table 6.7: Summary statistics for the North East GWR model

Coefficient	**Min**	**Median**	**Max**	**IQR**
Intercept	-0.92	0.67	2.36	0.66
Males	-0.014	0.0014	0.19	0.0045
Age 16 to 25	-0.030	-0.00015	0.027	0.0072
Age 56 to 65	-0.025	0.0017	0.044	0.012
Population Density	-0.064	-0.00040	0.037	0.0032
SOC 1	-0.062	-0.0032	0.050	0.018
SOC 8	-0.063	0.0036	0.053	0.0099
SOC 9	-0.047	0.0039	0.034	0.010
NS-SEC 7	-0.054	-0.00078	0.069	0.012

A location map of the North East region illustrating where some of the districts within the region are located is shown below in Figure 6.10. Also included are the GWR residuals and pseudo R^2 maps. The residuals map shows that the residual values for each WZ area within the region are consistently low across all parts of the region. Residual values generally fall around the zero mark, similarly to the GWR residuals for London and the North West. Small areas however, have slightly higher residuals, illustrated by the red colours within the map, suggesting that the model may not fit as well in these areas compared to others. The pseudo R^2 map generally reveals a good fit of the model, with most values being over 0.5 across the region.

Figure 6.10: Location map for the North East and residuals and pseudo R^2 maps for the North East GWR model

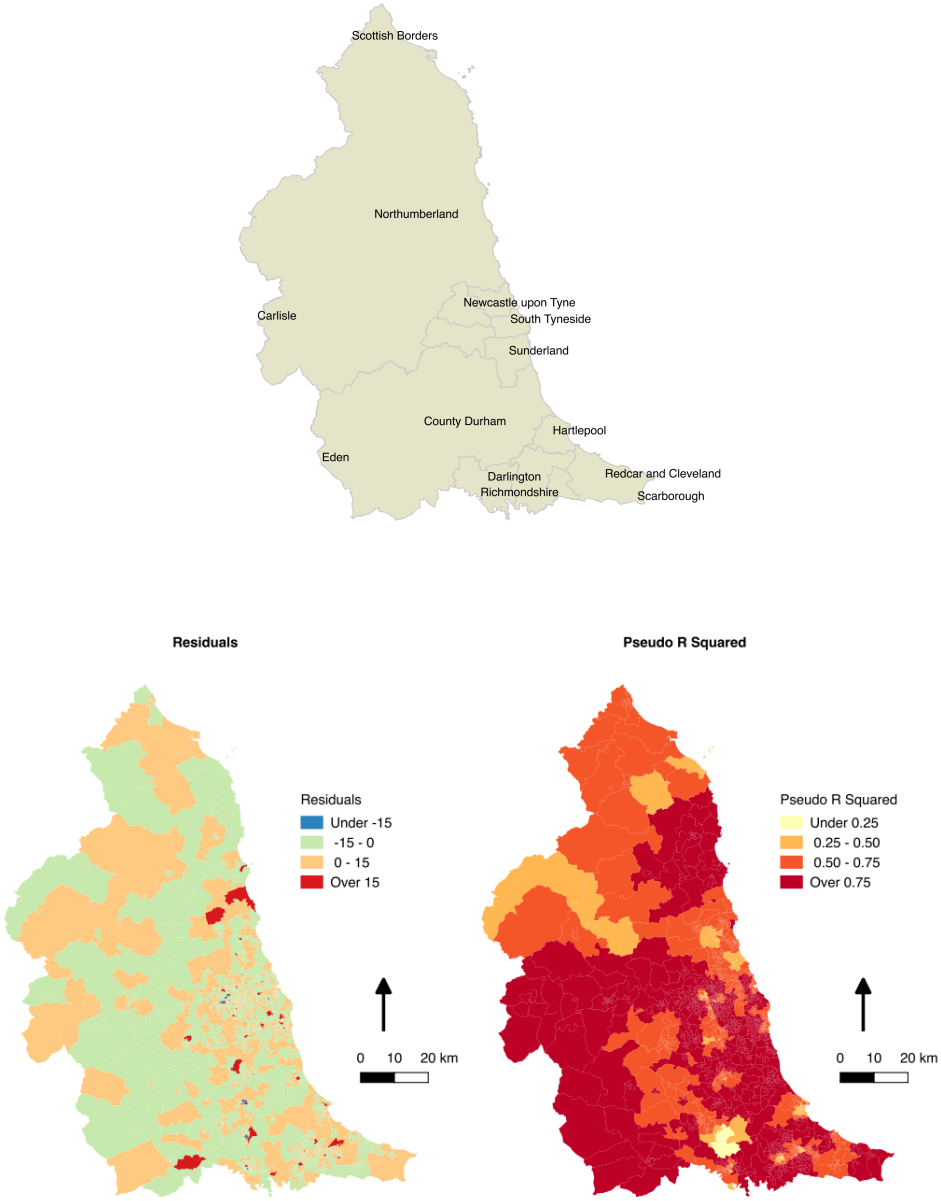


Figure 6.11 below shows two maps, the first of the percentage of SOC 1 Managers, directors and senior officials working across the North East, and the second of the GWR coefficient estimates for the SOC 1 variable. The percentage map shows that

a larger proportion of SOC 1 workers are located in the West of the region (parts of County Durham and Northumberland), whilst a smaller percentage of SOC 1 workers are located in the East of the region. The GWR coefficient estimates map shows a range of estimate values across the North East, with negative values mostly located in the North of the region, and parts of the West, whereas positive values are mostly located in the East and parts of the South. The strongest positive estimates are found in parts of County Durham and Hartlepool where slightly higher percentages of SOC 1 workers are located.

Figure 6.11: North East SOC 1 Managers, directors and senior officials- percentage of workers & GWR coefficients

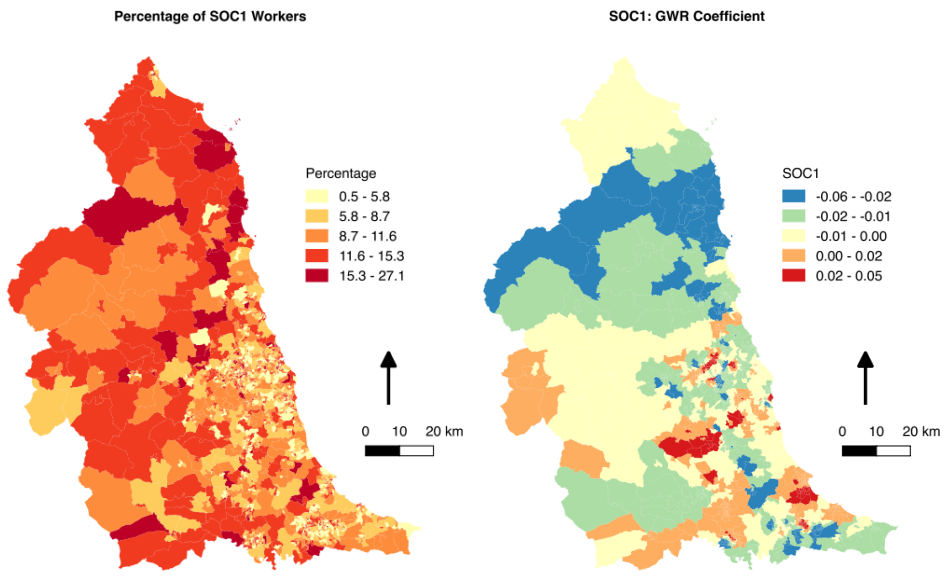
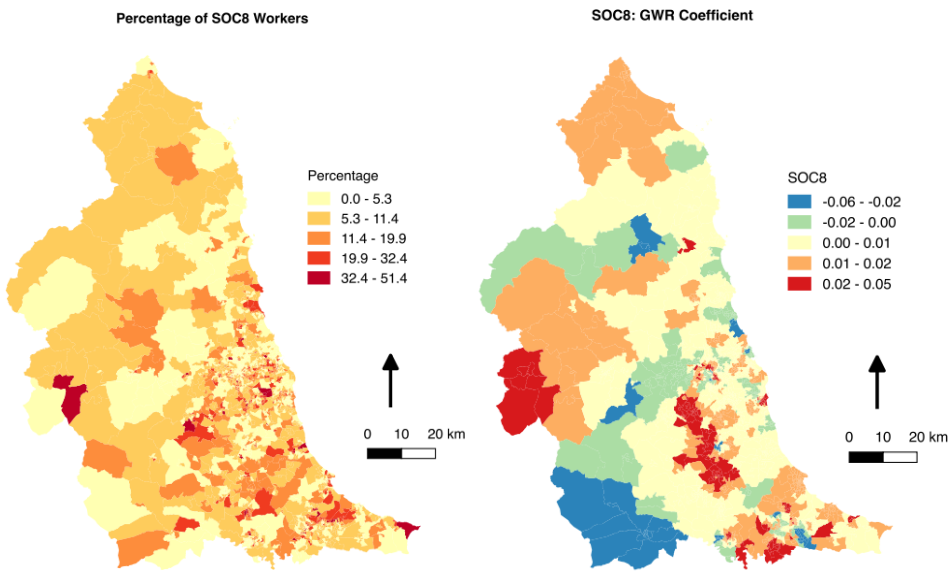


Figure 6.12 shows the maps of the percentage of SOC 8 Process, plant and machine operatives and the associated GWR coefficient estimates. The GWR coefficient map shows that along the West of the region and towards the South (particularly County Durham and Richmondshire), a positive relationship is found between SOC 8 workers and accident rate. Similarly, an area in Carlisle in the West of the region, a

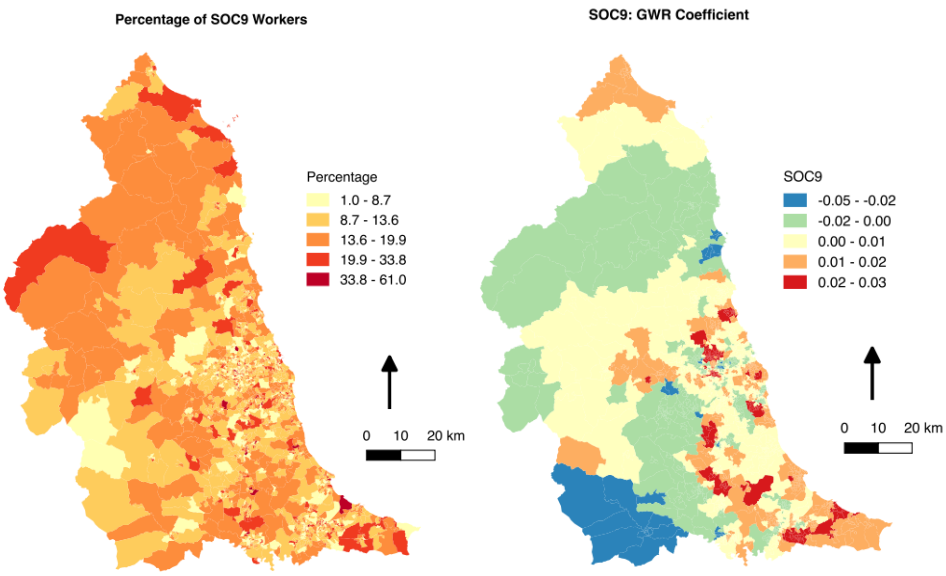
strong positive relationship also exists. In these areas, generally, the percentages of SOC 8 workers are small.

Figure 6.12: North East SOC 8 Process, plant and machine operatives- percentage of workers & GWR coefficients



The percentage maps of SOC 9 Elementary workers and the SOC 9 GWR coefficient estimates for the North East are shown in Figure 6.13. The highest percentage of SOC 9 workers are found in the North of the region, particularly Northumberland, as well as parts of Newcastle and Scarborough. The highest GWR values are found in the East and South of the region, in areas such as Scarborough, County Durham and parts of Newcastle, where the percentages of SOC 9 workers are higher. The most negative GWR estimates are found in the South West of the region and parts of central Northumbria.

Figure 6.13: North East SOC9 Elementary occupations- percentage of workers & GWR coefficients



6.4.4 Seasonal GWR

Data were next split into seasons and a GWR analysis was carried out again on each region. The aim of this was to understand whether local coefficient estimates vary through the year, and therefore whether the relationship between these socio-economic variables and accident rates strengthen or weaken during particular months of the year and in certain regions. This section presents the results for the GWR local coefficient estimates for the three case study regions, looking at summer and winter seasons only. Although not directly comparable, due to differing bandwidth sizes and sample data, contrasts will be discussed on the overall trend and spatial distribution of each, picking out any key findings or differences. For each model, a residuals map and a pseudo R squared map are shown to illustrate how well the model fits the data and short summaries are presented discussing the contrasts between the results.

London

Approximately 18,752 accidents were reported as occurring during the summer months (June, July and August) in London, and approximately 16,992 accidents were reported in the winter months (December, January and February). The automatic bandwidths calculated and utilised for the London seasonal GWR analysis were both 1,269. The AIC for the summer GWR model was computed as 25618, whereas the AIC for the winter GWR model was 29610. Figure 6.14 shows the residuals maps for the summer and winter GWR models for London. Both maps show a uniform trend of residuals across all WZs, with values ranging around the zero value. This suggests that the predicted values and the observed number of accidents do not greatly differ or vary across the region in both of the seasonal models. The pseudo R squared maps are also similar for both seasons, with the models fitting well in the west, parts of the north and east of the region.

Figure 6.14: Location map and summer and winter residuals and pseudo R^2 maps for London

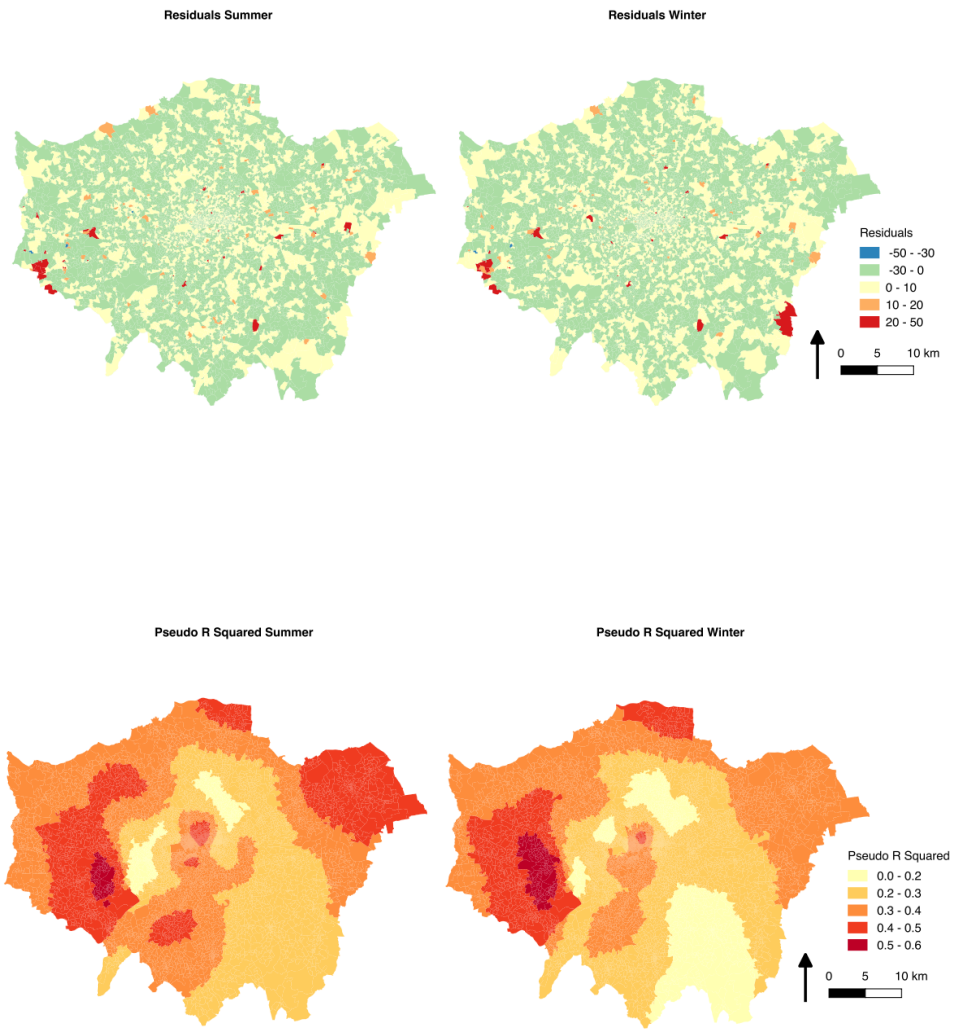
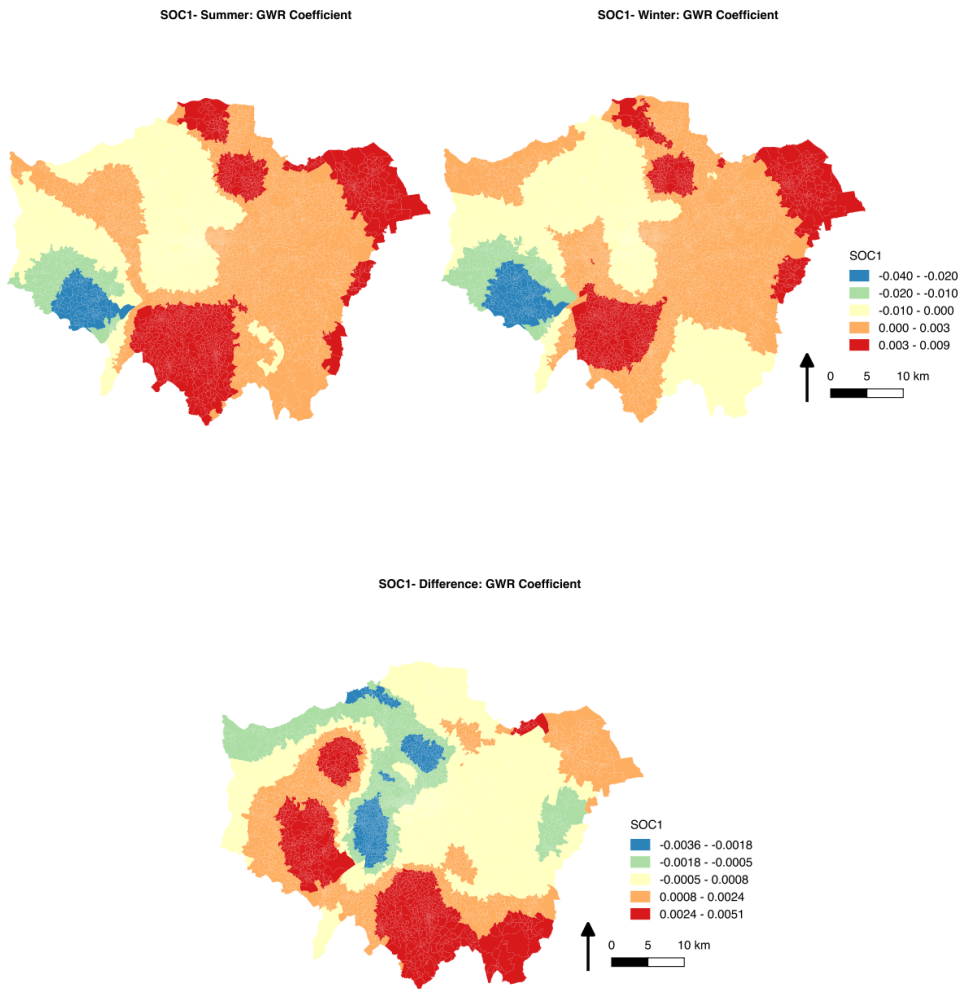


Figure 6.15 below shows the local coefficient estimates for SOC 1 Managers, directors and senior officials for the summer and winter months. Also included is a map of the difference between the summer and winter estimates, with the red and orange areas representing where the summer GWR coefficients are larger, and the green and blue areas representing where the winter GWR coefficients are larger. The summer and winter GWR coefficient maps for SOC 1 have a similar spatial structure, where the estimates are more positive in the south west area and also in the north east. The map showing the differences between the two seasons indicates that in the summer, the strongest relationship exists between the number of SOC 1

workers and the number of accidents in the south of the region and west areas including: Hounslow, Ealing and Richmond upon Thames.

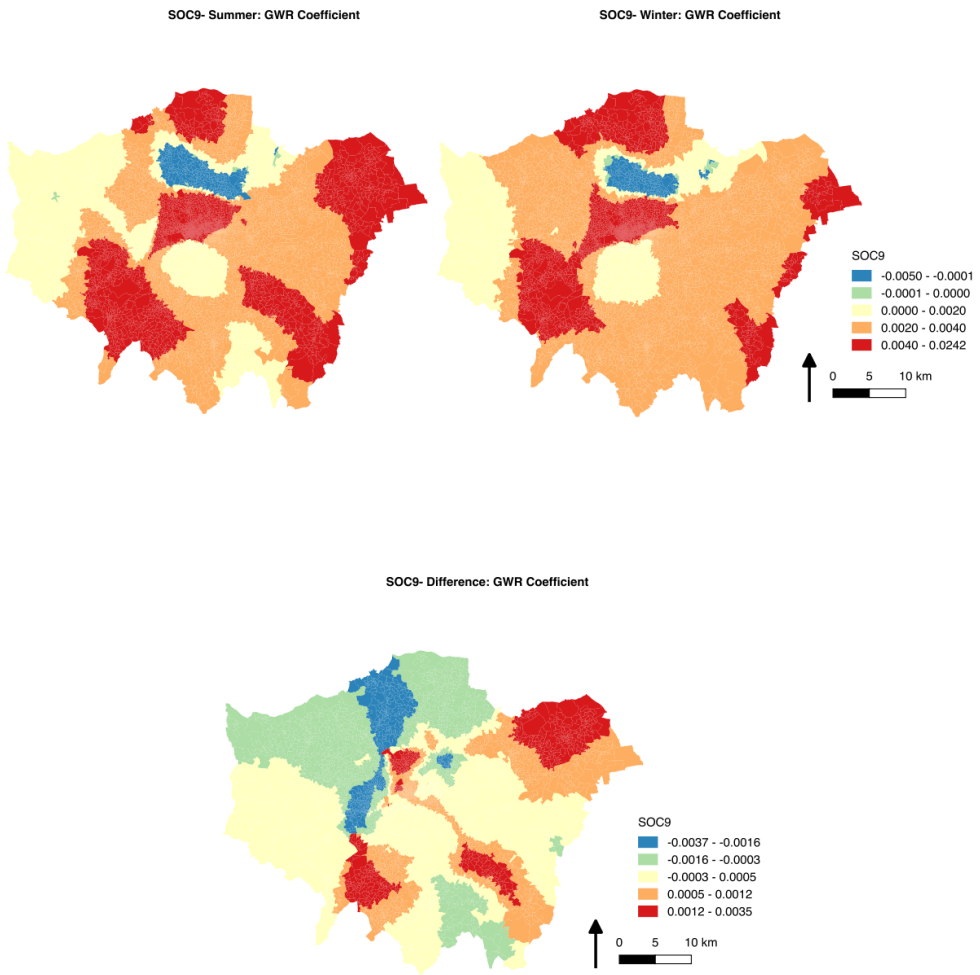
Figure 6.15: Summer and Winter SOC 1 Managers, directors and senior officials GWR maps for London and map of difference of GWR coefficient estimates



The results for the seasonal model for SOC 9 Elementary occupations are shown in Figure 6.16. A similar structure in the summer and winter maps can be seen, with negative coefficient estimates found in the northern part of the region in both seasons, whilst positive estimates are found in most other areas across London. The map showing the difference between the summer and winter GWR estimates shows

that for summer, central parts of London, as well as the North East of London have a higher GWR coefficient and therefore a stronger relationship with accident rates exists. In contrast, in the North West parts of London, in the districts of Barnet, Harrow and Brent, the coefficient estimates are larger in winter, suggesting a higher risk in these areas during the winter months.

Figure 6.16: Summer and Winter SOC 9 Elementary occupations GWR maps for London and map of difference of GWR coefficient estimates



North West England

Approximately 20,947 accidents were reported as occurring during the summer months in North West England, and approximately 19,597 accidents were reported in the winter months. The automatic bandwidths for the GWR seasonal analysis models for the region were both 468, therefore each GWR model is based on the nearest 468 neighbouring WZs. Figure 6.17 shows the location map, residuals maps and pseudo R squared maps for the summer and winter GWR models for the North West. The residuals map shows a uniform trend of values across all WZs, with a few patches of higher residual values. The pseudo R squared maps show that the model fits slightly better in Merseyside, Manchester and parts of Cumbria, represented by the red and dark orange areas. The AIC values are 20947 and 19597 for the summer and winter models respectively.

Figure 6.17: Location map and Summer and Winter residuals and pseudo R^2 maps for the North West

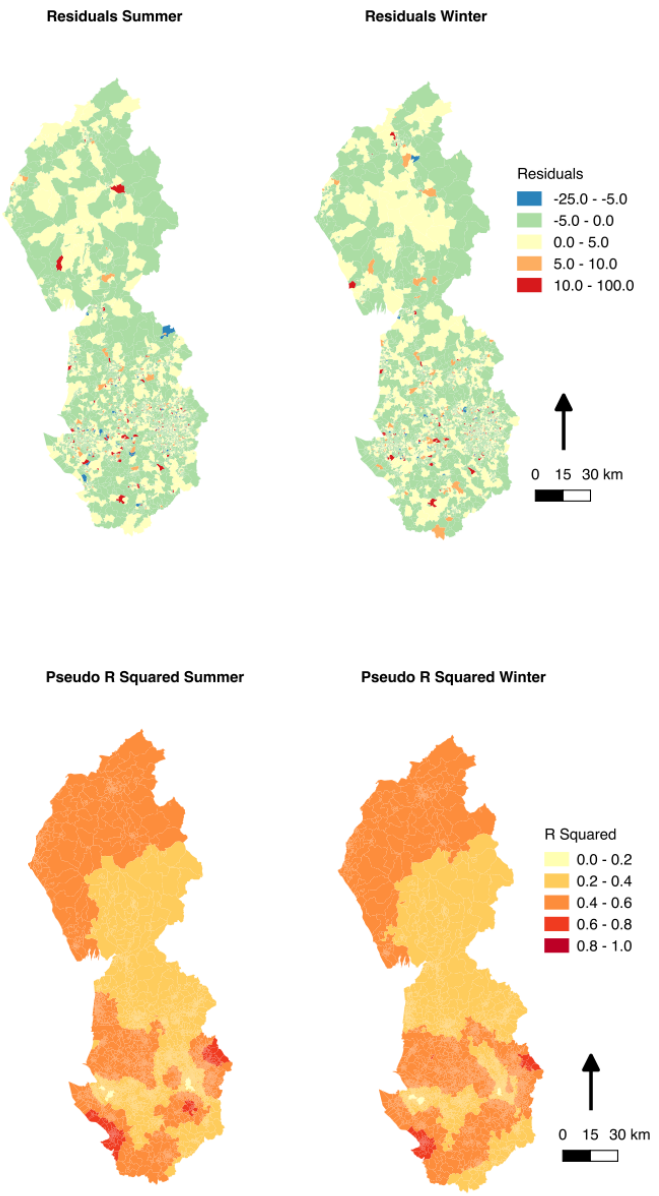


Figure 6.18 shows the GWR results for SOC 1 Managers, directors and senior officials for summer and winter in the North West of England. There is a variation of coefficient estimates across the regions in each season, with generally negative estimates found in the North of the region and most positive estimates found in the South. The map showing the differences between the coefficient estimates by season indicate that during the summer, the coefficient estimate is larger in Merseyside, as

well as parts of South Lancashire. Around central Manchester, larger coefficient estimates are found in winter, suggesting accident risk being stronger in this area during winter.

Figure 6.18: Summer and Winter SOC 1 Managers, directors and senior officials GWR maps for the North West and map of difference of GWR coefficient estimates

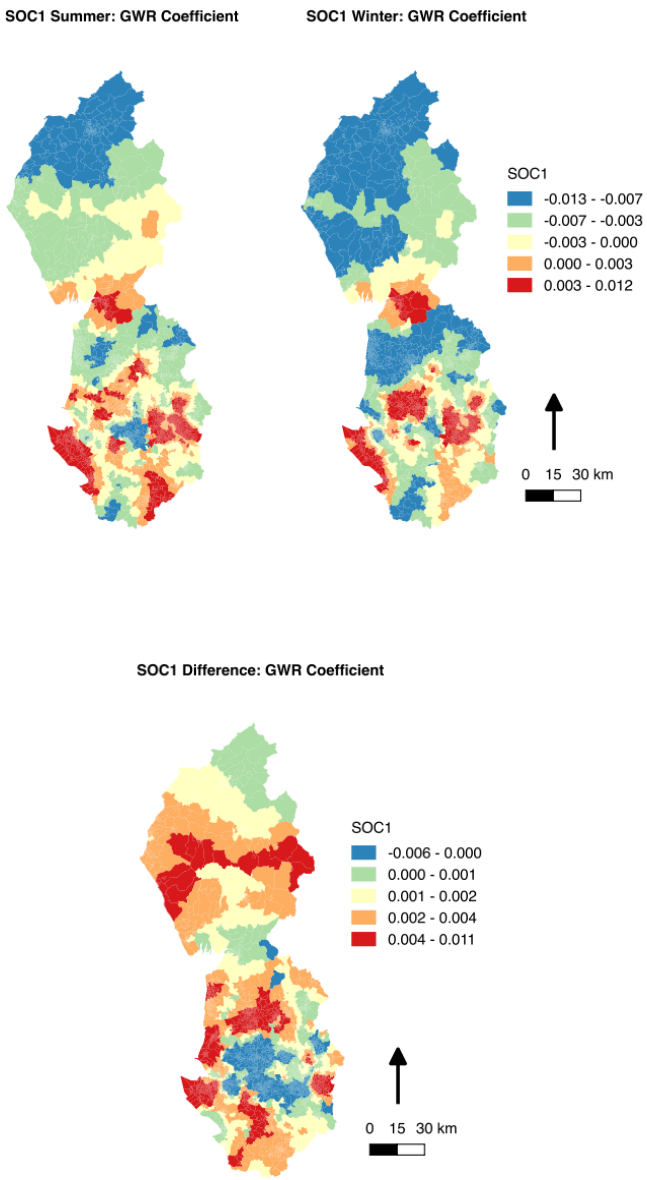
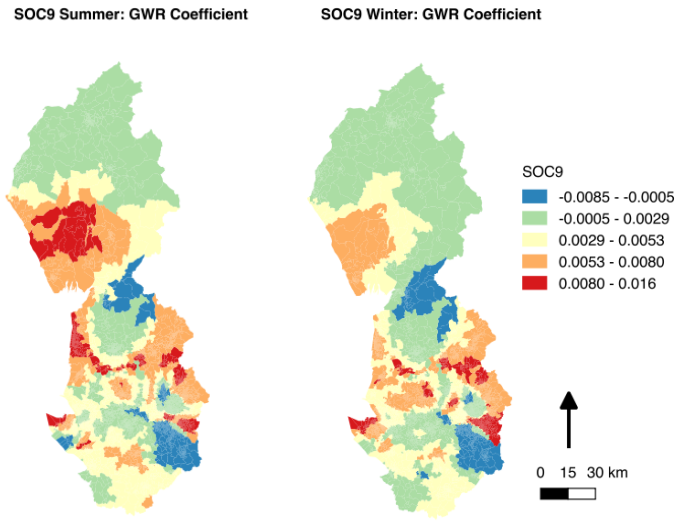
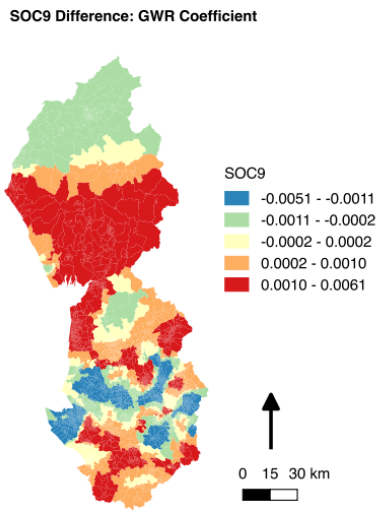


Figure 6.19: Summer and Winter SOC 9 Elementary occupations GWR maps for the North West and map of difference of GWR coefficient estimates



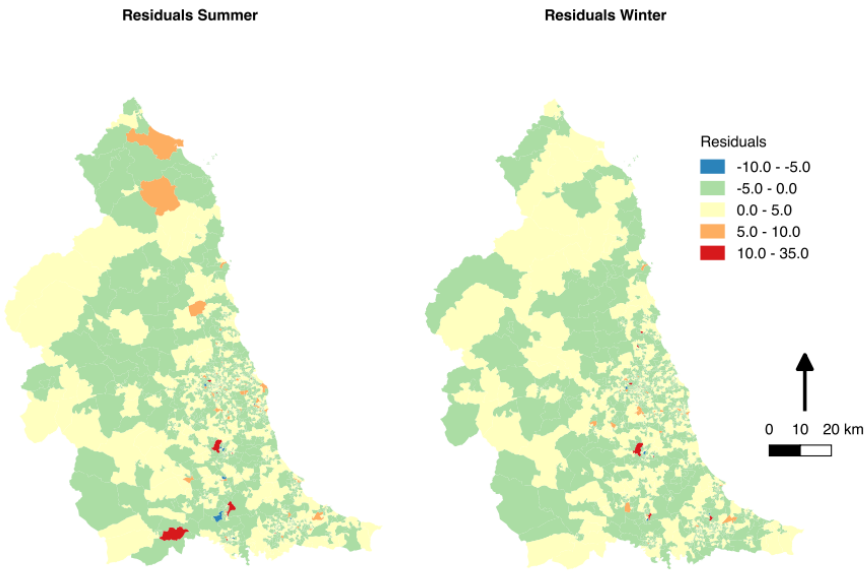


The results for the seasonal models for the SOC 9 Elementary occupations coefficient are shown in Figure 6.19. In the summer, the coefficient estimates are strongly positive in parts of the North West of Cumbria and West Lancashire. The coefficient estimates in these areas in summer are larger than those found in winter. In general, the estimates found for both summer and winter in Cheshire are negative. The map showing the difference in coefficient values indicates where the estimates are larger in the summer versus the winter. Overall, coefficient estimates for SOC 1 workers are larger in the summer in the North of the region, whilst further South of the region, in Manchester and Merseyside, the estimates are larger in Winter, suggesting some variations between seasons in different parts of the region.

North East

Approximately 7,464 accidents were reported as occurring during the summer months in the North East, and approximately 6,917 accidents were reported in the winter months. The automatic bandwidths for the GWR seasonal analysis models for the region were 83 for summer and 76 for winter. The AIC values were 7,811 and 7,270 respectively. Figure 6.20 shows the location map and residuals and R squared maps for the summer and winter GWR models for the North East. Overall, both residual maps show a uniform spatial trend across the region with some patches of darker areas, representing high residual values, in both maps. The pseudo R squared maps show a range of values across the region. Darker areas of red and orange are more prevalent indicating higher pseudo R squared values across the majority of the region in both seasonal maps. Overall these maps suggest that the seasonal GWR models would provide fairly accurate predictions for the majority of areas across the region for both seasons.

Figure 6.20: Location map and Summer and Winter residuals and pseudo R^2 maps for the North East



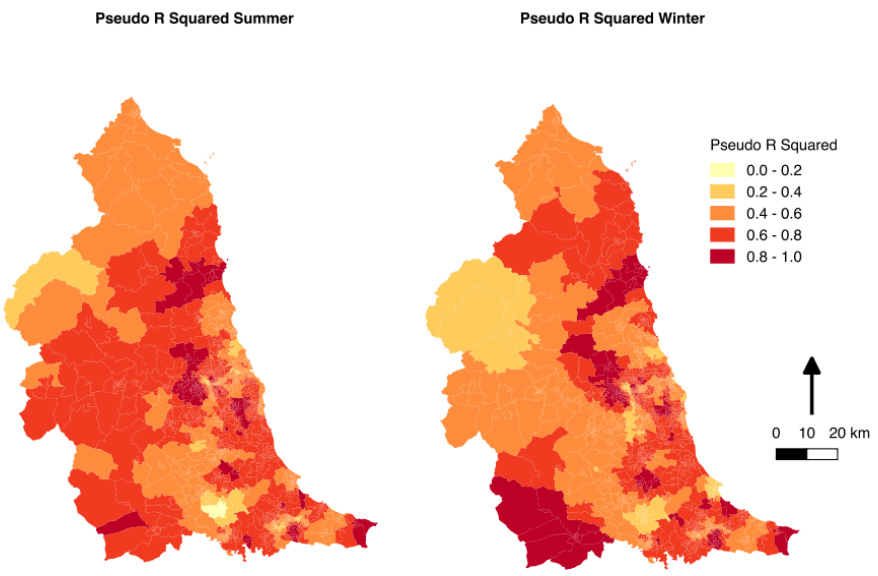
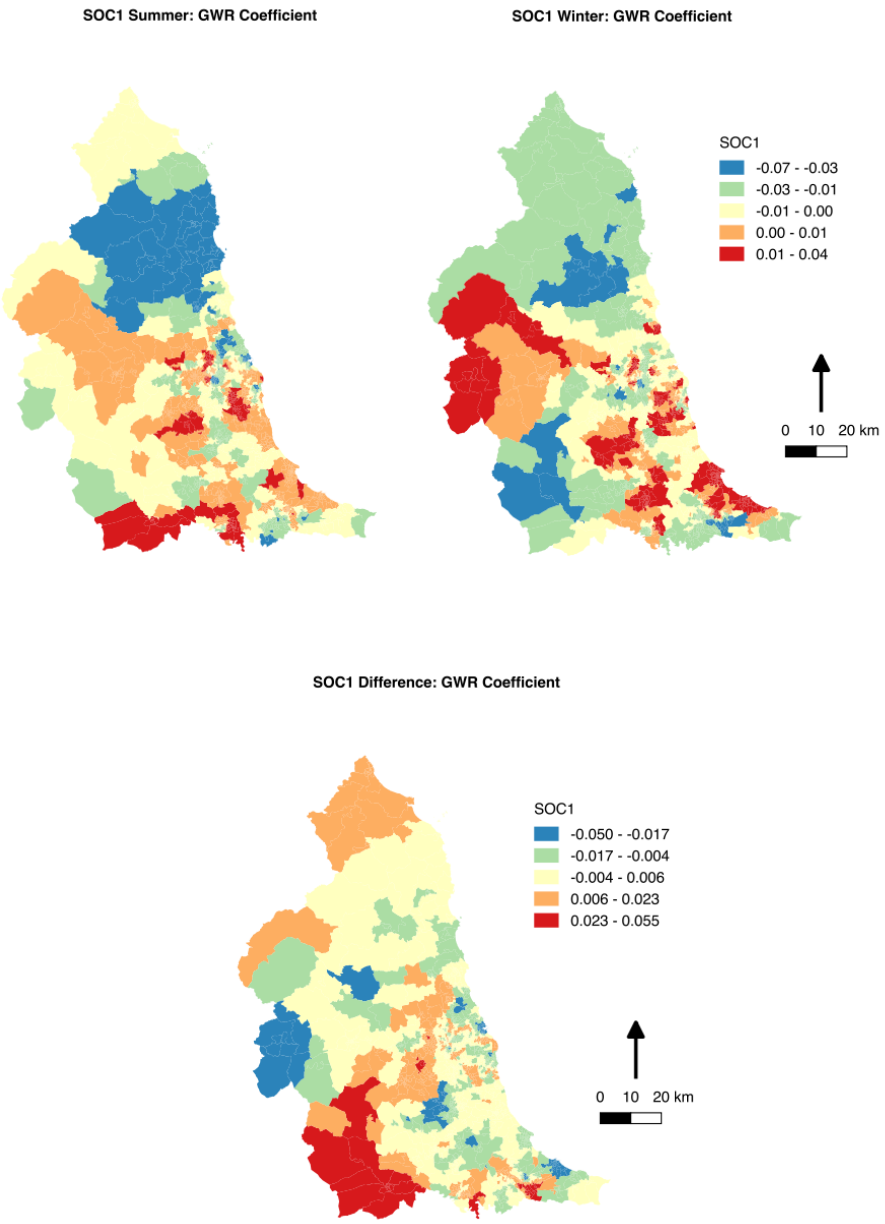


Figure 6.21 shows the results of the GWR local coefficient estimates for SOC 1 Managers, directors and senior officials for the North East for both summer and winter. Both maps show that the coefficient estimates vary across the region for the two different seasons. In winter, a stronger relationship between SOC 1 workers and workplace accident rate is found in the West of the region (Carlisle and parts of West Cumbria) compared to the summer. Alternatively, in the summer, higher coefficient estimates are found in the South of the region (in the district of Eden). These findings are clearer in the map showing the differences between the summer and winter GWR coefficients where the red and orange indicates higher estimates in summer, and the green and blue areas indicate where there are higher estimates in winter.

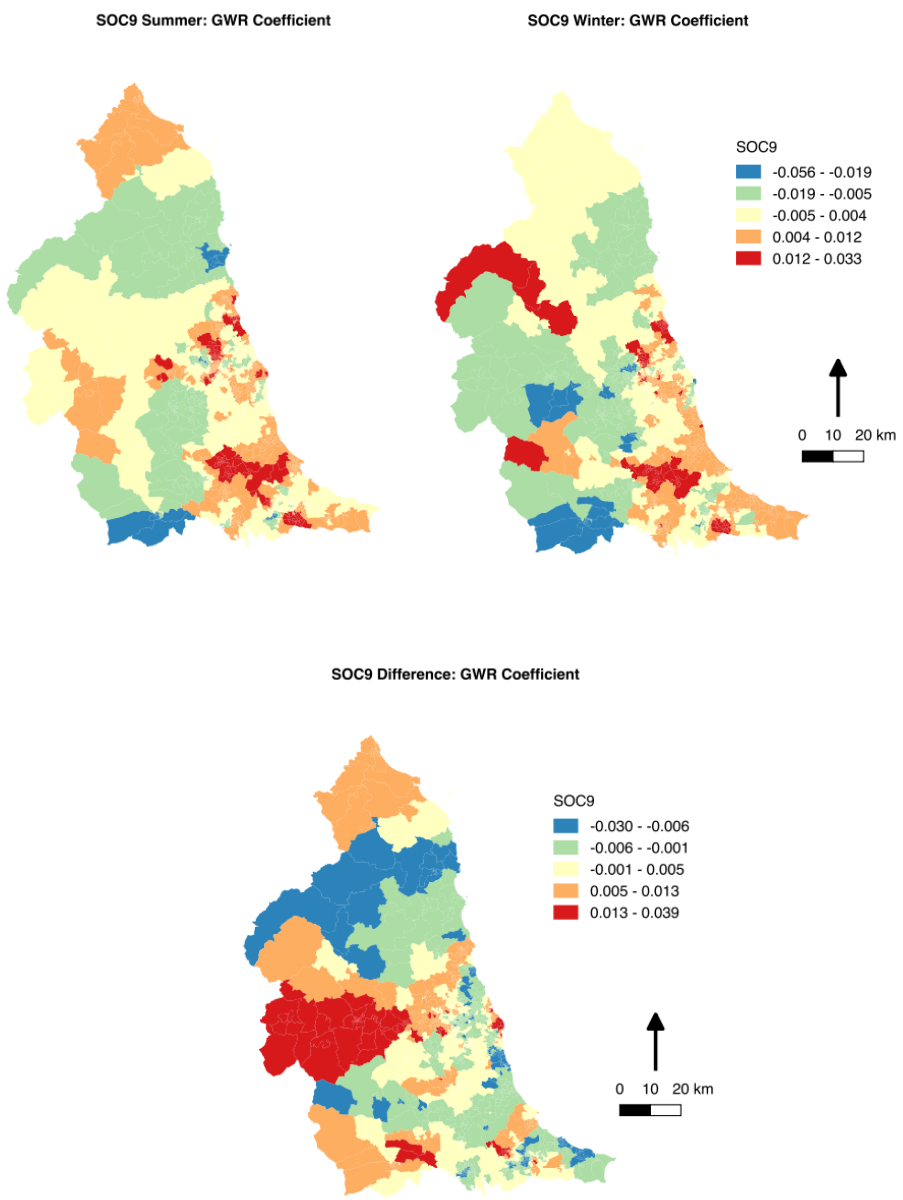
Figure 6.21: Summer and Winter SOC 1 Managers, directors and senior officials GWR maps for the North East and map of difference of GWR coefficient estimates



Lastly, Figure 6.22 shows the SOC 9 Elementary occupations local coefficient estimates for the seasonal models, summer and winter. A range of values are spread across the region for both seasons, with patches of blue and green areas representing negative estimates in the west of the region. In the winter, a strong positive relationship exists in the western area of Northumberland, represented by the red area on the map. The map showing the differences between the two seasons shows

where the coefficient estimates are larger in summer (represented by orange and red) versus winter (represented by green and blue). Larger coefficient values are seen in the North and West of Northumbria and West of the Eden district in winter, whilst a larger estimate is seen in summer in lower Northumberland and parts of lower Darlington.

Figure 6.22: Summer and Winter SOC 9 Elementary occupations GWR maps for the North East and map of difference of GWR coefficient estimates



6.5 Conclusion

This chapter has explored the relationship between a set of socio-economic area characteristics of workers and the number of workplace accidents in an area through a method of local regression analysis called Geographically Weighted Regression. The aim of this was to attempt to provide a more accurate model to explain variation in accident rates using the RIDDOR data, than was possible using a global Poisson regression model (Chapter 5). GWR was performed on three separate case study regions: London, North West England and North East England, to identify the differences in the spatial variation of these coefficients in each of these areas. The GWR local coefficient estimates were mapped for each of the case study regions, together with the percentage of workers within each of the socio-economic groups. A selection of the coefficients was presented within the chapter and overall, some strong findings emerged from the GWR analysis which identified where particular types of workers were more at risk of having a workplace accident, compared to others.

For the GWR analysis performed on the London data, it was found that the number of SOC 1 Managers, directors and senior officials in areas around North, East and South West of the region had a strong positive relationship with the number of workplace accidents. Similarly, for SOC 8 Process, plant and machine operatives, areas in the South of the region were found to have a stronger relationship with workplace accidents compared to the rest of the region. In these particular parts of London, however, the proportions of SOC 1 and SOC 8 workers were small. On the other hand, a positive relationship was found between the number of SOC 9 Elementary occupations and accident rate in parts of East and West London, where the percentages of SOC 9 workers were large.

For the North West GWR model, a positive relationship was found in urban areas such as Liverpool, Lancaster and parts of Manchester between SOC 1 workers and workplace accidents. In comparison a negative relationship was found in parts of Cheshire and Cumbria. A positive relationship was found between the number of SOC 9 workers in areas such as West Cumbria and South Cheshire (where a large

percentage of SOC 9 workers are employed). Also, for NS-SEC 7 Routine occupations, a positive relationship was found in areas with low proportions of workers classed as NS-SEC 7, particularly in parts of Lancashire and South Manchester.

For SOC 1 workers in the North East, the strongest negative local coefficient estimates were found in the North of the region and parts of the West. Positive estimates were found for the number of SOC 8 workers in the West and South of the region (particularly County Durham and Richmondshire). Generally, these were areas with small percentages of SOC 8 workers. Finally, for SOC 9 workers, positive GWR coefficient estimates were found in the East and South (for example- Scarborough, County Durham and parts of Newcastle), where larger percentages of SOC 9 workers work. Negative relationships between the number of SOC 9 workers and accident counts were found in the South West of the region and central parts of Northumbria.

The local residuals were mapped for each of the socio-economic groups and it was found from the maps that the GWR models would have predicted the number of accidents reliably across each of the regions. The GWR AIC values were considerably lower than those found for the global case study models. These measures both indicate that the GWR models provide a more accurate representation of workplace accidents than the overall global model.

The final part of the chapter discussed the GWR results for the models which were fitted to the case study regional data split into the summer and winter seasons. This was done to observe any differences between the socio-economic characteristics and workplace accidents based on the time of year in particular areas across England and Wales. It was found that there were some key contrasts between the two seasons considered, for example, in the summer months, a stronger positive relationship was found between the number of SOC 1 workers in the South and West of London (for example- Hounslow, Ealing and Richmond upon Thames). For SOC 9 workers in the North West of England, strong positive estimates were found in North West Cumbria and West Lancashire in winter, however larger positive estimates were

found in Merseyside and Manchester. In winter in the North East, larger coefficient estimates were found in the West of Northumbria, but in summer larger estimates were found in lower Northumbria and parts of lower Darlington.

This chapter has illustrated that some strong spatial variations between socio-economic area characteristics and workplace accidents exist. There is evidence provided to suggest that the global model does not adequately represent variations in all parts of England and Wales, and that there are local differences between small-scale areas. This chapter has also demonstrated that relationships between socio-economic variables and the number of workplace accidents within an area can vary between seasons, which also suggests that the level of accident risk not only varies by geographic location, but can also vary throughout the year, with some places having a higher risk of workplace accidents during particular times of the year.

Chapter 7

Exploration of the Free-Text Fields of the RIDDOR Data

7.1 Introduction

This chapter explores the RIDDOR dataset consisting of free-text fields containing descriptions of workplace accidents. The aim is to identify key themes that exist within the data, providing an alternative perspective to analysing the workplace accident data and delivering further insight into the risk factors associated with accidents in the workplace. The beginning of the chapter focuses on text mining, by exploring the most frequently occurring words that appear within the RIDDOR data. This helps to understand the structure of the data, gaining insight into what words tend to be stated the most often when accidents are reported.

The second part of the chapter focuses on topic modelling. As outlined in Chapter 3, topic modelling is a method whereby algorithms are applied to a large collection of textual data to discover main topics within an otherwise unstructured collection of documents. In this case, the text consists of descriptions of incidents reported to the HSE, which involved an accident or injury within the workplace. The overall aim of this chapter is to explore the information that is held within the free-text fields, which may provide another perspective on the findings of the previous two chapters. In particular, data will not only be split into the case study regions but also by month to explore seasonality differences, looking into whether particular weather constraints have an effect on accident risk in particular parts of the country.

7.2 Background and Methods

The RIDDOR dataset that was obtained for this part of the investigation consists of approximately 43,780 free-text fields containing descriptions of accidents reported to the HSE by employers. This data was subset into incidents reported in England and Wales, to match the study area of the global model (Chapter 5), where approximately 37,240 incident records were identified. The data covers a smaller timeframe than the original full RIDDOR dataset, covering just less than 6 months of reported incidents (from 1st April 2011 to 22nd September 2011), however it is sufficient to be able to perform some basic text mining and explore topic modelling of accident data.

The descriptions of the accidents vary in length, with some descriptions providing significant detail of the incident: how it occurred, the result of the injury and how it could have been prevented. Others can be rather short, providing a very simple description of the incident or even just the injury sustained by the worker. The aim of this chapter is to explore these free-text fields to identify any hidden core themes that exist within the data, which may have been missed from the analysis in the previous chapters. Exploration of these fields will be carried out via basic text mining, attempting to discover what exactly is being reported often within the text. The terms that have the highest frequencies will be extracted and plotted to visualise the key themes of the data, revealing which words are used the most often to describe an accident. It should be noted here that many words have several meanings depending on the context within which the word is being used. In some cases, it is therefore difficult to fully interpret the results with confidence. An example of this is the word 'hospital', which could mean a member of staff having an accident within the hospital workplace, but could also be meant in a context of the injured person being taken to hospital. Care is taken to provide an appropriate interpretation of the results.

The text mining analysis was carried out in R. The packages that were utilised were **tm**, **SnowballCC**, **RColorBrewer**, **ggplot2**, **wordcloud**, **RTextTools** and **topicmodels**. Firstly, the data were cleaned by removing punctuation and numbers from the text. Secondly, all words were transformed into lower case so that words

which were the same appeared in exactly the same format. Stop words were also removed from the textual data. These are words that are common and have no analytical value such as 'or', 'is' and 'what'. Finally, because of removing all of the above and cleaning the data, there was a lot of empty white space within the data which needed to be stripped from the dataset. Once this was all completed, the next stage was to create a document term matrix to be used for text mining and topic modelling.

The words with the highest frequencies were found and the top 10 highest frequency words were identified. Word clouds were also created to visually identify a wider range of high frequency words within the data. To further understand the data, topic modelling, specifically Latent Dirichlet Allocation (LDA) was carried out to see what core themes emerged from the data. LDA considers a text document as a mixture of a defined number of topics and lets each word within the document be associated with one of those topics. It is therefore a method of discovering sets of words associated with particular topics, which can reveal a number of themes within the text.

To carry out LDA, the number of topics to fit the data had to first be specified. As Blei and Lafferty (2009) explain, choosing the most appropriate number of topic models is not an easy task, and there are many different methods for doing this. They further state that generally choosing the number of topics that provides the best language model may be a preferred method over statistical approaches. Grun and Hornik (2011) state that models with different numbers of topics are fitted to the data and the optimal number of topics is determined based on the output of the results, in a data-driven way. For this analysis, a range of topic numbers were tested and the most appropriate topic number was chosen based on the reliability and readability of the output. Typically, a range of 50 to 300 topics is commonly used in topic modelling, with 50 topics used for small textual datasets and 300 for large datasets (Wei and Croft, 2006), therefore essentially a positive linear relationship exists between the optimum number of topics and text entries. For the topic modelling analysis in this chapter, a range of topic model numbers are tested and the best is chosen for the LDA, based on the output. Terms generated for each topic

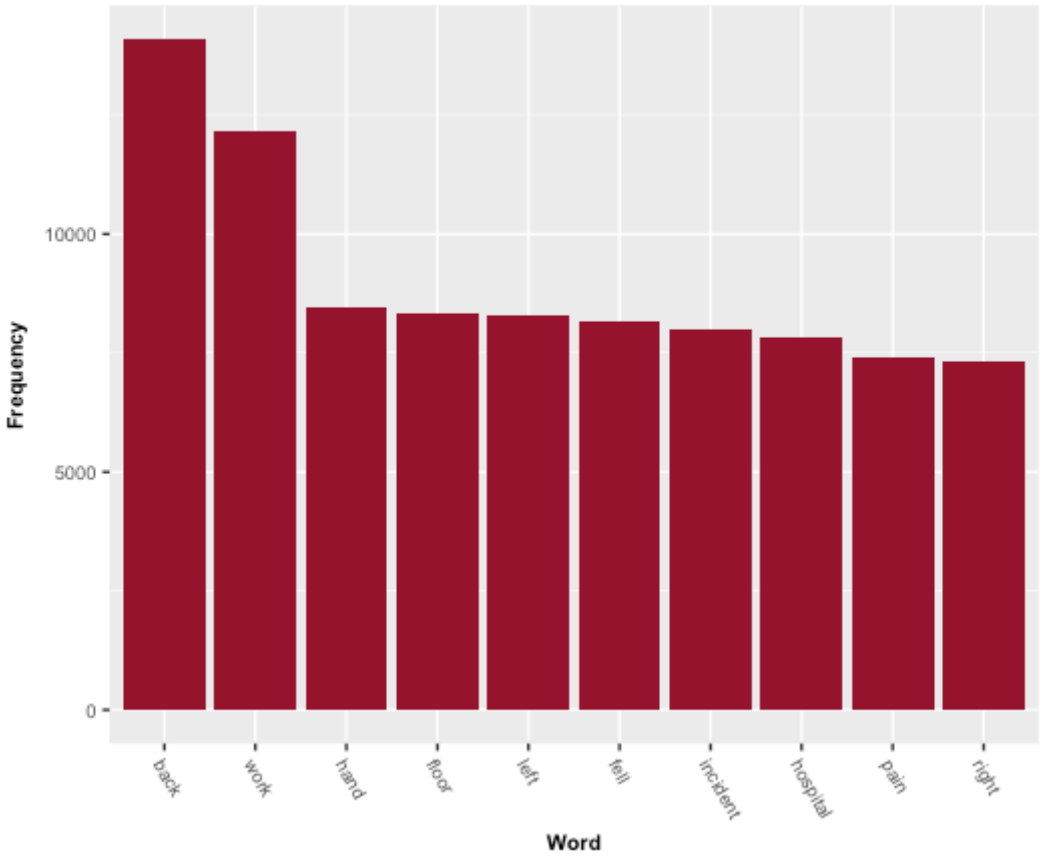
generally indicate an overall theme. The top 5 terms for the top 10 topics are shown for the data to give an overview of what each topic represents, with the most likely topics coming out top, and the most likely themes per each topic coming out top.

7.3 Text Mining

7.3.1 England and Wales

To gain a brief overview of the free-text fields within the RIDDOR dataset, some basic text mining was carried out. Figure 7.1 below illustrates the words with the top highest frequencies that have appeared in the RIDDOR dataset. The word that appears the most number of times is 'back', followed by 'work', 'hand', 'floor', 'left', 'fell', 'incident', 'hospital', 'pain' and 'right' respectively. The top word 'back' suggests that many incidents that occurred could have been due to workers injuring their back for example, but it could also indicate workers who have returned 'back' to work. The second word that appears the most, 'work', is unsurprising, given that the RIDDOR dataset details accidents that have occurred within the workplace. The third word, 'hand', indicates possible injuries involving the hands of workers- possibly broken wrists and fingers, sprains, cuts or burns. 'Floor' and 'fell' indicate accidents involving falling over, slips and tripping over the floor. The word 'incident' describes the event in which an accident occurred, and the terms 'left' and 'right' describe possibly the 'right' or 'left' arm injured for example. Finally, the terms 'hospital' and 'pain' indicate the result of the accident. 'Hospital' could be in reference to the injured person having to go to hospital, however it could indicate an accident occurring in a hospital as a workplace; for a doctor or nurse, for example.

Figure 7.1: Top 10 words with highest frequencies



To visualise the top frequency words further, Figure 7.2 shows a word cloud of the terms that appear more than 2,500 times in the RIDDOR dataset. The darker, larger text within the word cloud illustrates the words appearing more often in the RIDDOR text in contrast to the smaller, lighter coloured words in the Word Cloud. Again, as shown in the bar chart, the words that appear the most within the text are 'back' and 'work'. The word cloud, however, gives the opportunity to visualise more words than just the top 10, so in Figure 7.2, words such as 'truck', 'machine' and 'vehicle' become visible. These words all indicate that most accidents are reported as having been involved with tools and machinery.

Figure 7.2: Word Cloud of terms appearing at least 2,500 times

7.3.2 Case Study Regions

The previous chapter split the RIDDOR dataset into three case study areas to carry out GWR analysis. The text mining and topic modelling carried out in this chapter, will similarly not be based solely upon England and Wales alone, but will also be split into case study areas so that overall comparisons for this investigation can be made to the GWR analysis results. The following bar charts and word clouds are therefore based upon the RIDDOR textual data being split into the three case study regions: London, North West and North East. The data relating to London consists of approximately 3,951 reported accidents. Figure 7.3 shows a bar chart of the most frequent words of the RIDDOR dataset for London. As for the full dataset covering England and Wales, the most frequent word is 'back' followed by 'work'. The remaining most frequently occurring words are all exactly the same as the full RIDDOR set, except for the word 'staff', which replaced the word 'right'. The word cloud shown in Figure 7.4 shows the words which appear more than 200 times in the London RIDDOR dataset. There are no obvious differences between the highest frequency words found within the London dataset, as compared to the England and Wales RIDDOR dataset, however differences may become clearer in the topic modelling section of this chapter.

Figure 7.3: Top 10 words with highest frequencies- London

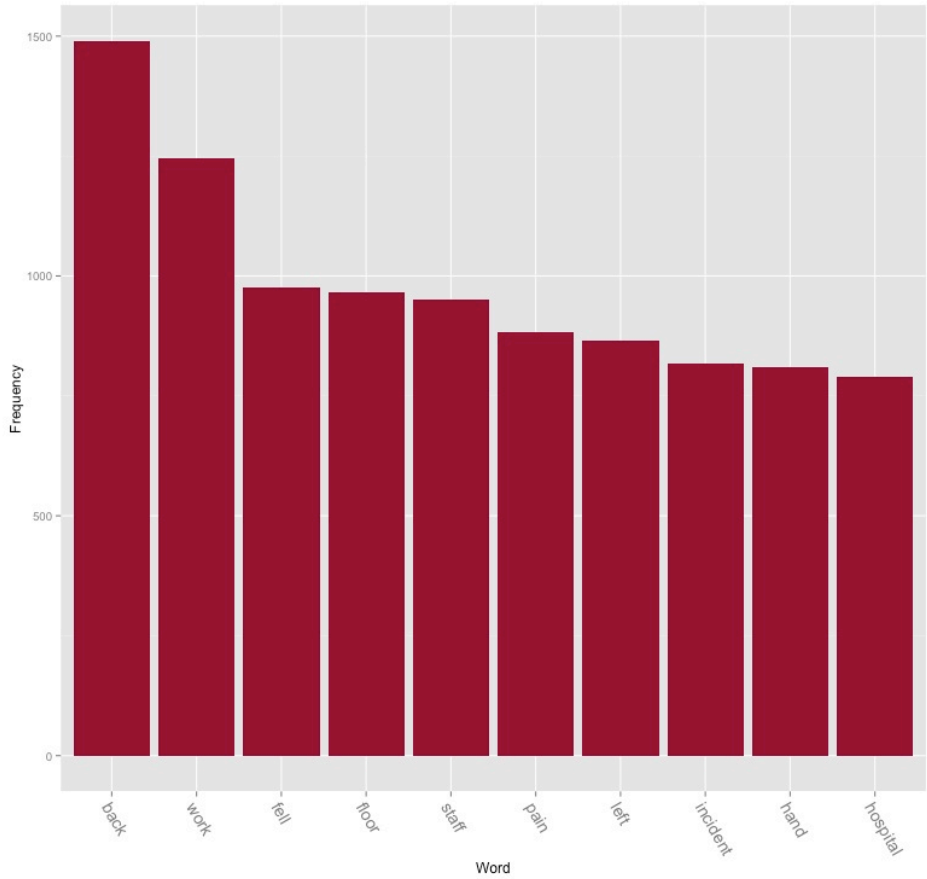


Figure 7.4: Word Cloud of terms appearing at least 200 times- London

The North West RIDDOR dataset has approximately 2,451 reported incidents. Figures 7.5 and 7.6 show the bar chart and word cloud of the most frequently occurring words for the dataset. The top 2 words, again similarly to the full RIDDOR dataset covering England and Wales and for the data covering London, are 'back' and 'work'. The rest of the top 10 terms are also similar in comparison to the previous results for London and England and Wales. The data covering the North West is smaller and therefore the word cloud showing the words that appear more than 200 times, has slightly less terms shown within it. Again, similarly to the previous results for London and England and Wales, the results from the word cloud show the most frequent terms including description words of accidents such as 'slipped', 'floor' and also possible results of accidents such as 'injury' and 'hospital'.

Figure 7.5: Top 10 words with highest frequencies- North West

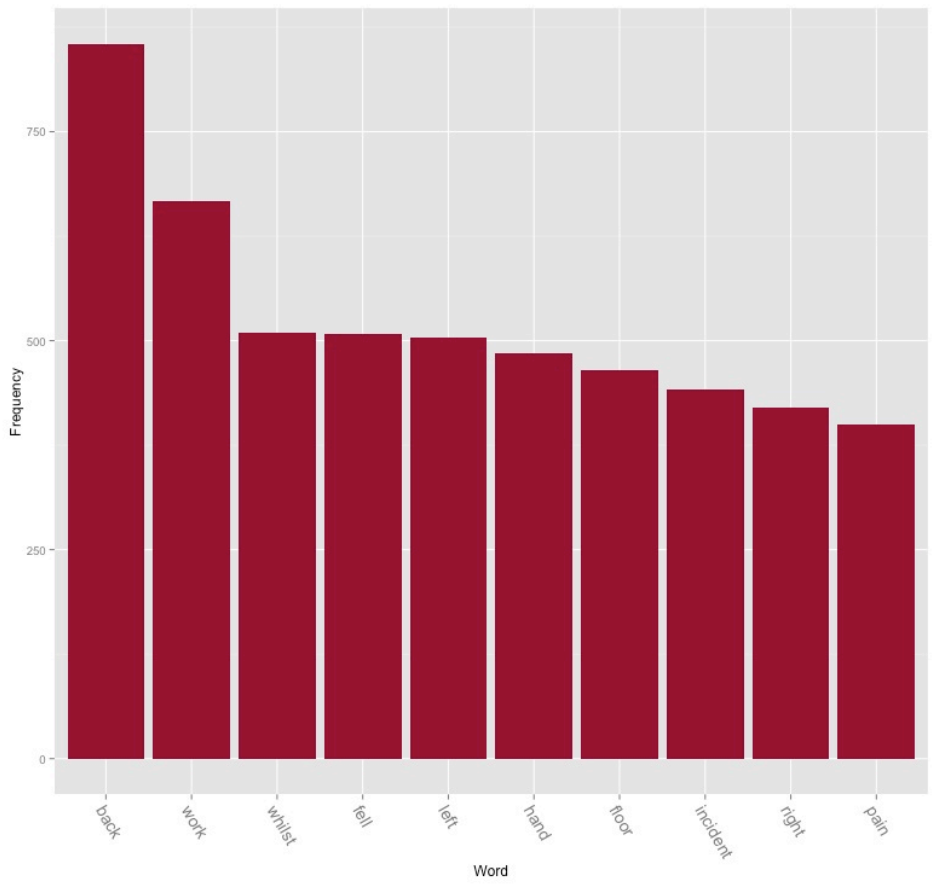


Figure 7.6: Word Cloud of terms appearing at least 200 times- North West

Finally, the data for the North East were extracted, consisting of 1,776 reported accidents. Figures 7.7 and 7.8 show the bar chart and word cloud for the North East RIDDOR data. Again, similar results were found for the North East region as for England and Wales, London and the North West with the terms 'back', 'work', 'hand', 'left', 'floor', 'hospital', 'fell', 'area', 'right' and 'incident' being the top 10 words emerging as the most frequent in the data. Again, visually from the word cloud, the words 'back' and 'work' appear the most pronounced, showing the highest frequency within the text.

Figure 7.7: Top 10 words with highest frequencies- North East

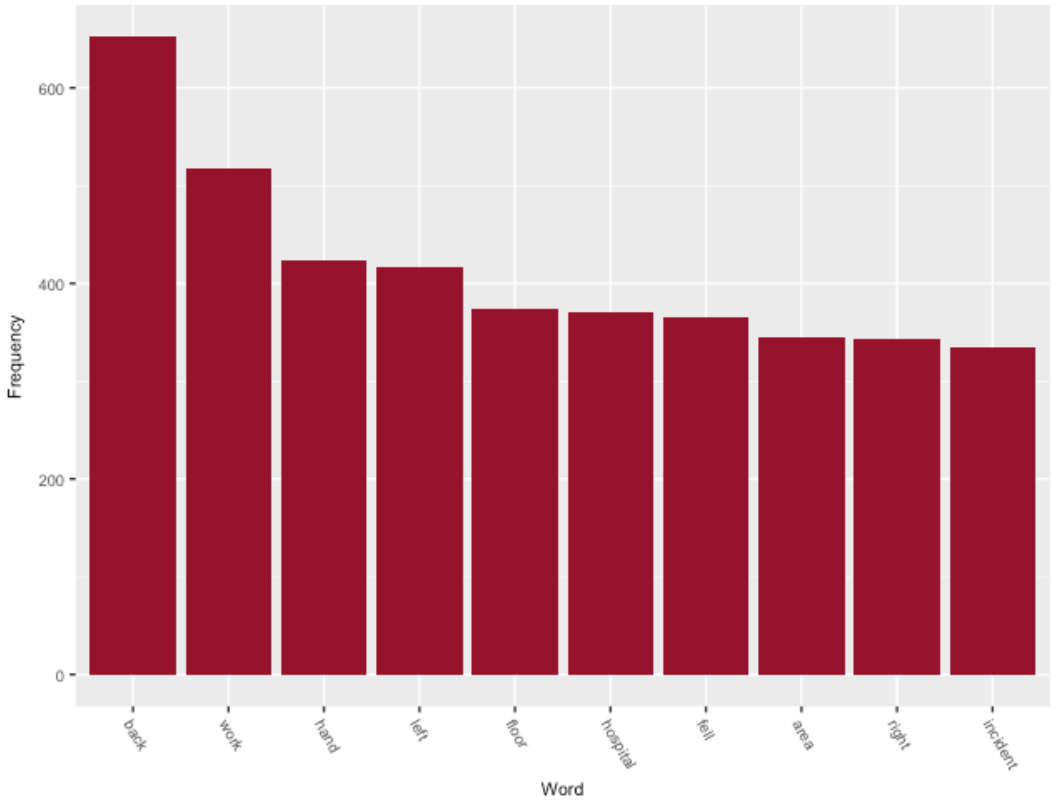


Figure 7.8: Word Cloud of terms appearing at least 200 times- North East

7.4 Topic Modelling: Latent Dirichlet Allocation (LDA)

7.4.1 England and Wales

Having briefly explored the free-text fields of the RIDDOR dataset by identifying the most frequently occurring words within the text for England and Wales and the three case study regions, the next stage was to look at topic modelling to gain more of an understanding of the types of core themes which might exist within the data. LDA was carried out on the RIDDOR dataset, first for England and Wales as a whole area. Several topic model number sizes were tested: 50, 100, 200 and 300. The tables below show the top 10 topics for 50, 100 and 300 topic model numbers that were tested. As can be seen from the tables, no obvious patterns emerge from the output for these topic numbers.

England and Wales: 50 topics

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Work	Work	Hand	Cause	Pain	Work	Work	Work	Cause	Pain
Cause	Hand	Cause	Right	Hand	Right	Back	Cause	Move	Felt
Back	Pain	Report	Hand	Went	Cause	Left	Pain	Staff	Use
Fell	Walk	Left	Left	Walk	Hand	Hand	Hospital	Report	Work
Right	Lift	Went	Hospital	Cause	Fell	Hospital	Staff	Fell	Lift

England and Wales: 100 topics

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Work	Back	Work	Work	Back	Fell	Fell	Work	Work	Cause
Left	Cause	Back	Right	Use	Work	Work	Lift	Hand	Use
Back	Report	Cause	Left	Fell	Cause	Lift	Left	Pain	Report
Hand	Went	Hand	Use	Slip	Floor	Slip	Move	Left	Move
Cause	Staff	Use	Hand	Walk	Incident	Whilst	Carry	Slip	Walk

England and Wales: 300 topics

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Work	Work	Left	Back	Use	Work	Fell	Right	Work	Work
Cause	Left	Pain	Pain	Left	Back	Left	Work	Fell	Weighting
Staff	Went	Right	Hand	Right	Fell	Slip	Pain	Back	Fell
Hospital	Incident	Colleague	Work	Work	Function	Stop	Shoulder	Slip	Back
Back	Accident	Work	Function	Step	Slip	Hit	Injury	Hospital	Staff

Out of all the topic models tested, 200 topics produced the most sensible output of themes. Table 7.1 shows clear themes beginning to emerge from the data. Topic 1 concentrates on words that relate to trips and falls, with key words emerging including: 'trip', 'floor', 'pain' and 'back'. Topic 2 indicates accidents involving cleaning, with terms such as 'vacuum' and 'cleaner' emerging from the text. Other themes which emerge from the data include: slipping on wet floors (Topic 4), accidents involving machinery or equipment (Topic 5), and accidents involving hands whilst wearing gloves (Topic 10).

Table 7.1: Top 10 topics out of 200 topics for England and Wales

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Trip	Vacuum	Pain	Wet	Equipment	Trip	Head	Ankle	Work	Gloves
Pain	Cleaner	Ambulance	Slip	Machine	Desk	Hit	Twist	Back	Hand
Floor	Injure	Hospital	Hospital	Area	Office	Fell	Fall	Hurt	Equipment
Work	Back	Head	Floor	Piece	Cable	Banged	Pain	Hospital	Glove
Back	Hospital	Work	Injure	Vehicle	Floor	Ladder	Work	Patient	Safety

7.4.2 Case Study Regions

Topic modelling was also performed on the RIDDOR dataset that was split into the three case study areas. Below are the tables of results for the top 5 terms for the top 10 topics that appear within the data for London, North West of England and North East of England. Several topic sizes were tested for each case study dataset and 50 topics was chosen as the optimum number, based on the core emerging themes within the results.

Table 7.2: Top 10 topics out of 50 topics for London

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Door	Duties	First	Box	Foot	Pull	Ladder	Work	Plate	Training
Open	Due	Aid	Three	Step	Force	Bruising	Day	Side	Handling
Onto	True	Aider	Days	Site	Muscle	Back	Days	Into	Will
Said	Light	Work	Alleged	Onto	Fire	Work	Following	Unable	Manual
Came	Work	Administered	Shelf	Toe	Shoulder	Sustained	Shift	Note	Risk

Table 7.2 shows the top 10 topics for the London data. Several of the topics contain terms which make reference to indoor or office type facilities. For example, the first term of Topic 1 is 'door', the first and fifth terms of Topic 4 are 'box' and 'shelf'. Other key themes emerging from the results are pulled muscles and suspected injuries to: shoulders, toes and feet. While for London the key themes emerging from the data involved indoor and non-manual work, for the North West region key themes appear to relate to physical objects and manual work (Table 7.3). Terms such as 'container' (Topic 2), 'trailer' (Topic 3), 'cable' (Topic 4), and 'pallet' and 'truck' (Topic 7) all suggest accidents involving physical objects. Accidents involving parts of the body that have emerged from the topic modelling include 'hand' and 'ankle', and problems involving slipping on wet floors, also emerges as a theme (Topic 9).

Table 7.3: Top 10 topics out of 50 topics for North West England

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Work	Container	Trailer	Cable	Staff	Top	Pallet	Risk	Floor	Door
Day	Hospital	Function	Area	Member	Left	Ankle	Assessment	Work	Open
Days	Left	Dog	Loose	Another	Gate	True	Fell	Slipped	Back
Following	Colleagues	Customer	Kitchen	Function	Board	Foot	Back	Fell	Doors
Advised	Back	Garden	Into	Management	Hand	Truck	Lost	Wet	Opened

The results for the North East (Table 7.4) indicate that similarly to the North West, accidents involving physical objects in manual types of jobs tend to be the most common emerging theme for workplace accidents. Terms such as 'machine' (Topic 3), 'vehicle' (Topic 4) and 'pallet' (Topic 5) all indicate accidents involving physical objects. A key theme appearing in Topic 1 is slipping on wet surfaces, with terms emerging including: 'slipped', 'back' and 'wet'. Other emerging themes for the North East are broken fingers using machinery (Topic 3) and pulled muscles (Topic 9).

Table 7.4: Top 10 topics out of 50 topics for North East

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Hospital	Back	Machine	First	Hand	Work	Accident	Manual	Pulled	Work
Slipped	Work	Injured	Door	Pallet	Area	Back	Shoulder	Muscle	Investigation
Work	Fell	Finger	Vehicle	Back	Ankle	Hurt	Back	Day	Day
Back	Investigation	Broke	Site	Training	Leg	Shift	Lift	Injury	Returned
Wet	Day	Accident	Hit	Ground	Lift	Injury	Truck	Following	Around

7.4.3 Seasonal Differences

As with the previous chapter (GWR), the textual data have been split into months to understand more about the sorts of accidents occurring in particular seasons. The data that are currently held runs from April to mid-September, which is a relatively short timeframe. For this reason, topic modelling was carried out for two different months of data: April and August, as these months are the furthest apart in the year, given that the data for September do not include a full uninterrupted month. For this case, August will represent summer months and April for Spring due to the weather generally being less pleasant in April than August. A range of different numbers were specified for topic models, with 50 topics for England and Wales, 40 topics for London, 35 topics for the North West, and 30 topics for the North East region giving the most obvious patterns from the data.

There is a mix of themes emerging from the textual data for England and Wales in April (Table 7.5). Topic 1 suggests accidents involving trips, whilst Topic 2 suggests accidents involving machinery and vehicles. The word 'young' also appears in Topic 2, suggesting potentially younger workers having accidents with machines, or assembly lines. Topic 3 includes terms such as 'patient', 'care', 'nurse' and 'pushed', which indicate accidents within health care. Topic 6 appears to reflect forms of physical abuse with terms such as 'hit', 'colleague', 'force', 'hitting' and 'bruise'. Other themes which emerge include: accidents involving cleaning (Topic 5: 'clean', 'slipped', 'wet', 'floor' and 'mop') and accidents within kitchens (Topic 8: 'hot', 'kitchen', 'pan').

Table 7.5: Top 10 topics out of 50 topics for England and Wales, April

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Trip	Machine	Patient	Work	Clean	Hit	Around	Hot	Fell	Work
Floor	Young	Care	Back	Slipped	Colleague	Work	Kitchen	Duties	Report
Level	Vehicle	Pushed	Hurt	Wet	Force	Site	Pan	Light	Incident
Access	Roller	Assist	Fell	Floor	Hitting	Trip	Unit	Report	Hospital
Ground	Assembly	Nurse	Floor	Mop	Bruise	Ground	Drop	Normal	Safety

Table 7.6 shows the results for England and Wales in August. Here it is noticeable that accidents involving cleaning seem to be a more common occurrence than in April, with it now appearing in Topic 1 ('clean', 'wet', 'slip', 'back'). Other topics which emerge include lifting boxes (Topic 7), accidents involving storage/handling and vehicles (Topic 6), banging head at work (Topic 8) and injuries to hands (Topic 2 and Topic 9).

Table 7.6: Top 10 topics out of 50 topics for England and Wales, August

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Clean	Hand	Work	Area	Slip	Driver	Box	Bang	Work	Carry
Wet	Work	Fell	Site	Back	Storage	Lift	Head	Hand	Back
Cleaning	Slip	Hospital	Fell	Work	Vehicle	Back	Work	Use	Hurt
Slip	Move	Left	Floor	Onto	Gate	Pain	Pain	Report	Work
Back	Side	Back	Hurt	Injure	Back	Work	Hospital	Lift	Return

There is a range of core themes emerging from the topics for the London data for April (Table 7.7). Topic 1 suggests common injuries to the foot and possibly back, due to the words 'back', 'foot' and 'hospital' appearing within the top 5 terms. Topic 2 also suggests accidents involving hands, which could potentially indicate a tool being used or possible strains. Topic 3 suggests that workers are likely to slip over with words 'slipped', 'fell' and 'floor' being within the top 5 terms for that topic.

Table 7.7: Top 10 topics out of 40 topics for London, April

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Back	Work	Incident	Work	Incident	Back	Back	Incident	Right	Staff
Foot	Back	Hospital	Pain	Will	Hand	Hospital	Pain	Hospital	Will
Hospital	Fell	Slipped	Fell	Pain	Will	Pain	Fell	Manager	Hospital
Incident	Hand	Fell	Hospital	Left	Incident	Right	Will	Fell	Right
Staff	Incident	Floor	Manager	Time	Pain	Incident	Foot	Slipped	Went

Table 7.8 shows the results for London in August. The term 'back' comes up often, however as previously mentioned it is difficult to determine the context of some words and whether 'back' is indicating injuries to workers' backs. The term 'door' appears in Topics 1 and 2 suggesting accidents involving doors inside of buildings or offices. Topic 9 suggests workers falling over to be a common theme with terms 'fell', 'pain' and 'hand'. The results are quite repetitive after the first few topics with several terms appearing in more than one topic for example the terms 'back' and 'pain' appear in six and seven of the topics, respectively.

Table 7.8: Top 10 topics out of 40 topics for London, August

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Staff	Incident	Back	Pain	Pain	Staff	Incident	Back	Pain	Back
Door	Will	Work	Incident	Back	Fell	Floor	Hand	Fell	Right
Pain	Back	Right	Left	Incident	Hand	Left	Slipped	Will	Pain
Back	Work	Floor	Floor	Right	Pain	Work	Pain	Hand	Fell
Left	Door	Left	Slipped	Slipped	Work	Hospital	Staff	Area	Work

Table 7.9 below shows the results for the North West in April. Topic 1 suggests that accidents involving workers slipping on something is common with the terms 'slipped', 'floor' and 'back' appearing in the top five words for that topic. Topics 2 and 3 indicate accidents involving workers' hands tends to be a common theme, with the word 'machine' also appearing in Topic 3, indicating accidents involving machinery or tools are likely to be common. Again, there seems to be repetitive terms emerging with 'back', 'pain', 'work' and 'hospital' appearing several times in the terms.

Table 7.9: Top 10 topics out of 35 topics for North West, April

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Work	Hand	Back	Back	Back	Work	Fell	Back	Left	Back
Back	Back	Pain	Pain	Work	Area	Left	Left	Back	Incident
Floor	Causing	Floor	Fell	Fell	Right	Work	Fell	Work	Work
Incident	Area	Machine	Work	Causing	Back	Back	Pain	Working	Left
Slipped	Time	Hand	Floor	Left	Hospital	Hospital	Went	Floor	Hospital

Interestingly, in Topic 1 for the North West in August, the term 'pallet' appears with 'work', 'fell', 'working' and 'accident' (Table 7.10), suggesting that many workers are having accidents involving pallets and are therefore workers in manual types of jobs. Topic 2 suggests workers are slipping on something, possibly injuring their 'foot' or 'hand'. The term 'machine' comes up in Topic 5, again suggesting possibly another physical type of role where workers are having accidents with machinery.

Table 7.10: Top 10 topics out of 35 topics for North West, August

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Work	Slipped	Back	Back	Back	Hand	Left	Back	Slipped	Hand
Fell	Floor	Left	Slipped	Left	Floor	Incident	Hand	Area	Left
Pallet	Foot	Work	Causing	Machine	Pallet	Hand	Whilst	Time	Whilst
Working	Hand	Fell	Working	Hospital	Back	Fell	Left	Back	Right
Accident	Pain	Hospital	Left	Right	Whilst	Patient	Area	Hospital	Taken

Table 7.11 shows the top 10 topics for the North East in April. Topic 1 suggests workers generally are likely to have an injury from slipping on wet surfaces, with the terms 'slip', 'fell', 'wet' and 'area' appearing. Slips in general in the North East in April appear to be a common theme, emerging in Topics 1, 3, 6 and 8. Again the results are quite repetitive with the terms not providing a lot of meaning or differences between topics, with words such as 'report', 'hospital' and 'work' appearing often in the results. Similarly to April, there isn't a large difference between topics for the results for the North East in August (Table 7.12). The term 'vehicle' appears in Topics 1 and 3, suggesting accidents involving moving and driving of vehicles. Injuries involving hands also appear to be a theme in several topics.

Table 7.11: Top 10 topics out of 30 topics for North East, April

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Slip	Work	Hand	Onto	Use	Causing	Fell	Left	Work	Walk
Fell	Hand	Slip	Walk	Hospital	Report	Went	Onto	Use	Will
Wet	Staff	Door	Hospital	Went	Fell	Right	Work	Report	Hospital
Area	Causing	Right	Fell	Report	Hospital	Back	Slip	Walk	Work
Work	Went	Use	Left	Work	Slip	Hospital	Step	Right	Left

Table 7.12: Top 10 topics out of 30 topics for North East, August

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Back	Hand	Vehicle	Right	Work	Right	Fell	Work	Work	Work
Lift	Work	Work	Work	Operations	Work	Went	Hand	Right	Back
Vehicle	Back	Time	Hand	Back	Causing	Right	Patient	Back	Hand
Result	Carried	Causing	Return	Walk	Carried	Back	Hospital	Left	Pain
Move	Put	Staff	Patient	Hand	Left	Hospital	Back	Causing	Move

In general, an issue could be that the datasets are too small when split down into regions and months. There are only approximately 400-700 incidents examined for each of the cases above, as opposed to sometimes several thousand cases which were examined in the previous section before the data were split into months. It may be insightful to examine several terms alone which indicate weather differences or levels of daylight within the case study areas and England and Wales as a whole, to see if there are any patterns which emerge which could indicate seasonal impacts on accidents. The bar charts below show the frequency of words within the RIDDOR text for a selection of contrasting terms which include: 'light' and 'dark', 'hot' and 'cold', 'day' and 'night', and 'dry' and 'wet'.

Figure 7.9 shows the results of the frequency of words of these selected terms for accidents occurring in England and Wales in April and August. Most of the terms appear more frequently in August compared to April. The term 'cold' is found more often in descriptions of accidents occurring in April, potentially indicating weather impacts during this month. The terms 'wet' and 'dark', however, appear more often in August, during the summer months. 'Light' and 'hot' occur more frequently in August, as well as dry, which all indicate summer physical conditions due to weather or daylight.

Figure 7.9: Frequency of hot/cold terms for England and Wales

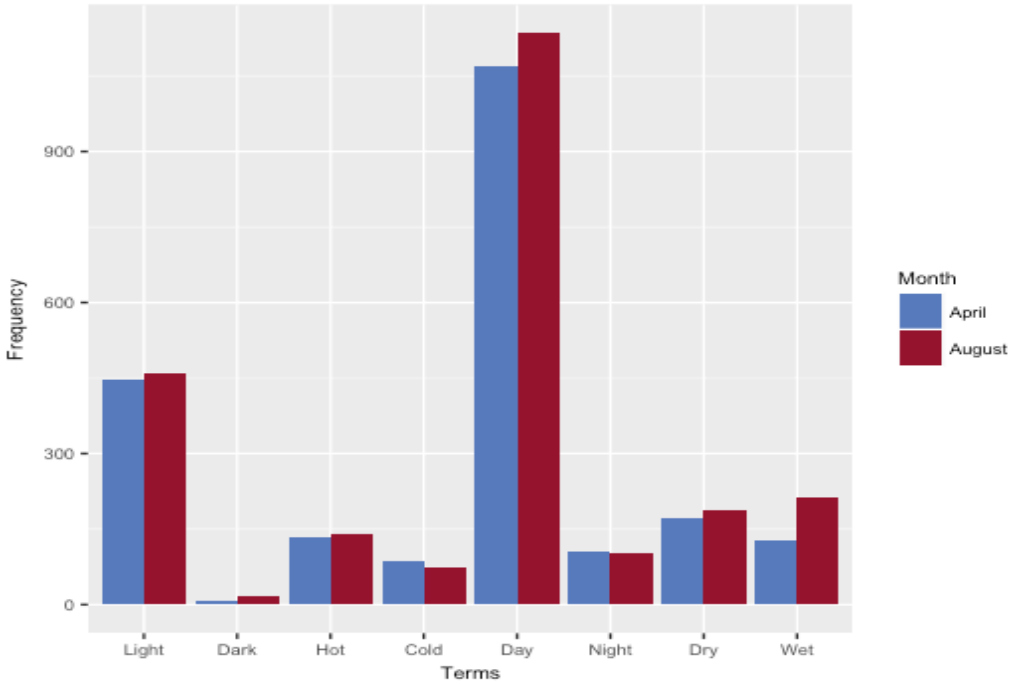


Figure 7.10 shows a bar chart for London for the same terms, identifying their frequencies in reported accidents in April and August. The word 'day' appears most often in April, however similarly, so does the word 'night'. The term 'cold' appears most often in descriptions of reported accidents in April, whilst 'hot' and 'light' appear more often in reports in August. These results again suggest there appears to be some seasonal differences between the accident results in April and August.

Figure 7.10: Frequency of hot/cold terms for London

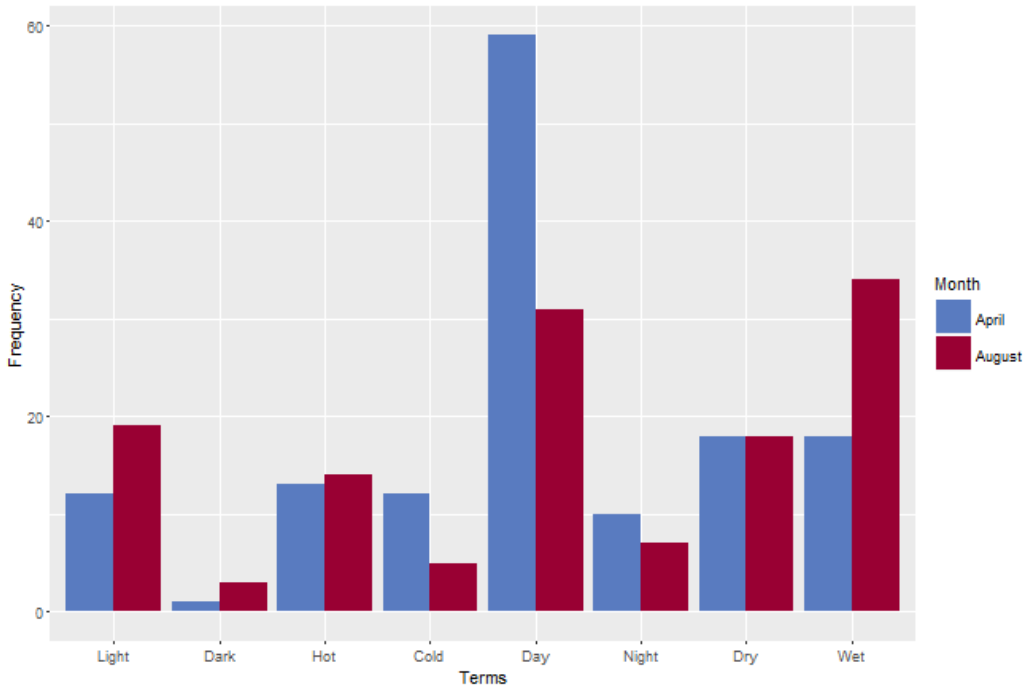
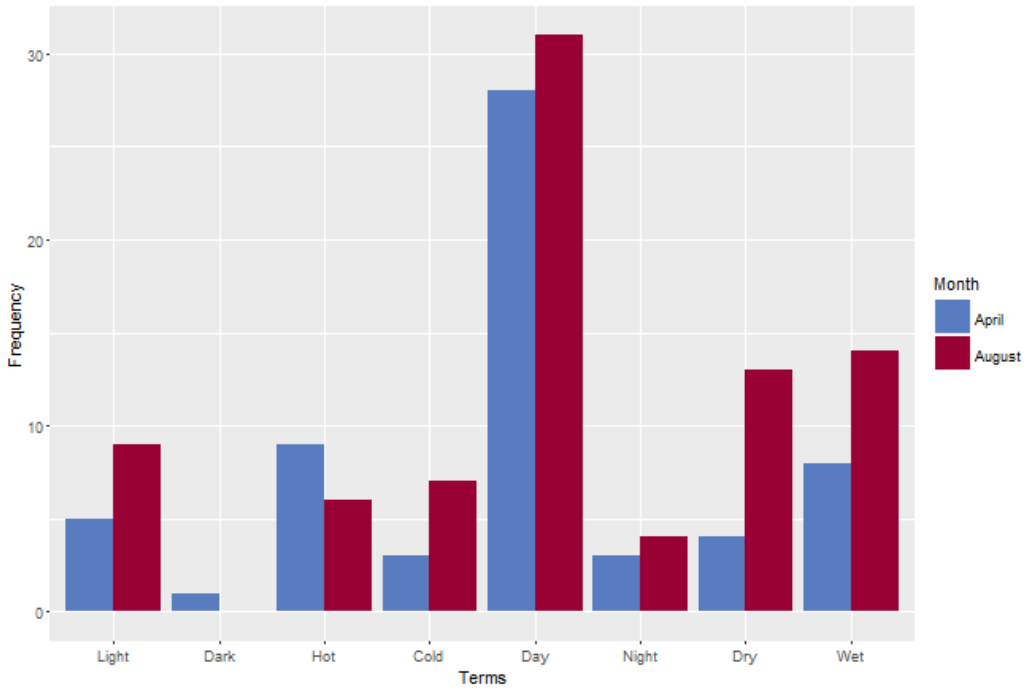


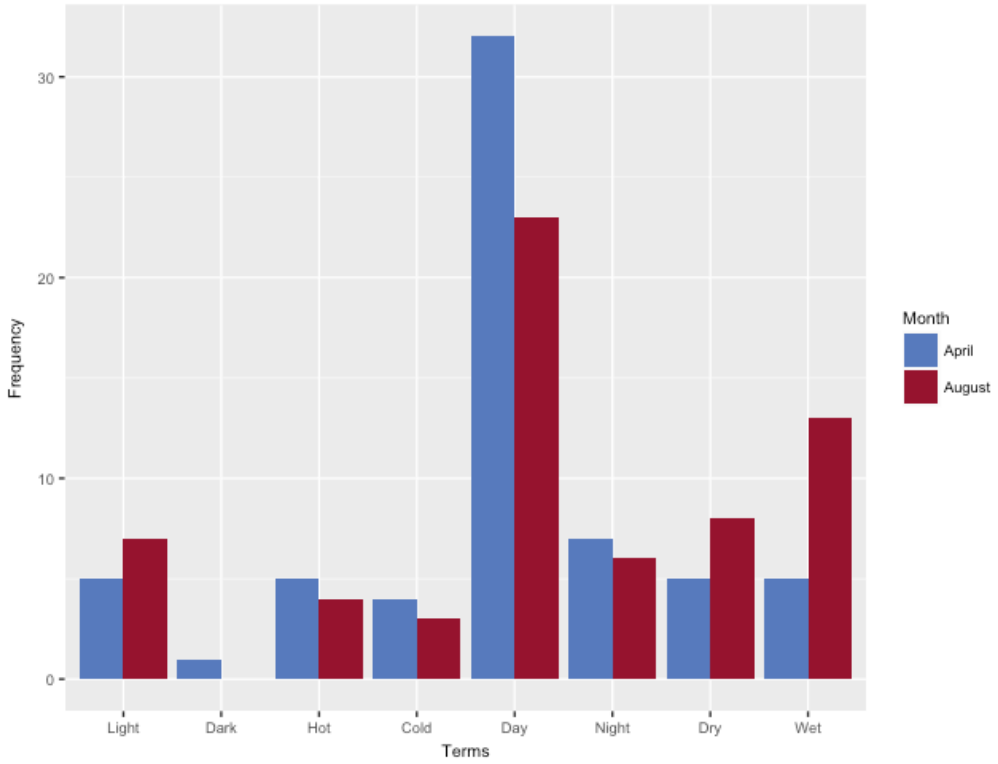
Figure 7.11 shows a bar chart of the frequency of selected terms for accidents reported in the North West. Most of the terms appear more often in August than in April. Terms such as 'wet' and 'dry', and 'day' and 'night' all appear more often in descriptions of accidents occurring in August. This therefore does not reveal any obvious seasonal differences. The term 'dark' appears more in accidents in April than August. Darkness could be a reason for accidents occurring in April in the North West, when there are less hours of daylight in the working day, affecting levels of workers' visibility.

Figure 7.11: Frequency of hot/cold terms for North West



Finally, the results for the North East are shown in Figure 7.12. There is a mix of results shown in the bar chart, with the terms 'dark', 'hot', 'cold', 'day' and 'night' appearing more in April than August. The terms 'light', 'dry' and 'wet' appear more often in descriptions of accidents that occurred in August than April. 'Light' and 'dry' reflect conditions within the summer months, which may reflect accidents occurring outside during hotter days, suggesting some seasonal differences.

Figure 7.12: Frequency of hot/cold terms for North East



7.5 Conclusions

This chapter has explored the free-text fields of the RIDDOR dataset through basic text mining and topic modelling methods. Through text mining, the most common words were extracted from the RIDDOR text, and bar charts and word clouds were created to illustrate the highest frequency of words in the data. The words 'back' and 'work' appear the most often within the free-text fields for accidents occurring in England and Wales as a whole, and for each of the case study areas. The term 'back' could indicate injuries to workers' backs such as pulled muscles and strains caused by accidents in the workplace. A mix of words including 'hand', 'left', 'floor', 'fell', 'incident', 'hospital', 'pain' and 'right' all had the highest frequencies in the data for England and Wales, with no significant differences between the key highest frequency words for the case study areas.

Latent Dirichlet Allocation Analysis, a type of topic modelling, was also carried out and the results were presented for accidents in England and Wales, and again split

into the three case study areas in an attempt to identify any differences spatially in the causes of reported accidents. LDA identified the common themes within the free-text fields of RIDDOR for England and Wales and found that trips and falls tended to be key themes of the first topic, with terms such as 'trip', 'floor', 'pain' and 'back' coming out in the top terms. For accidents occurring in London, words representing indoors and offices such as 'door', 'box' and 'shelf' for example, appeared within the top 10 topics. Injuries such as pulled muscles tended to be a core theme within the data for London also.

For data on accidents occurring within the North West, themes indicating physical objects appeared within the top 10 topics. Terms such as: 'trailer', 'pallet', 'truck' and 'cable' suggest accidents in the North West occurred in manual types of jobs, where workers come into contact with machinery and tools. Within the topics, key themes also emerged suggesting that injuries to workers in the North West tended to be to the hands and ankles. Within the top 10 topics of the LDA output for North East, some of the key terms were 'machine', 'vehicle' and 'pallet'. These terms indicate that accidents are occurring in manual types of jobs in the North East, similarly to the North West of England. Slipping on wet surfaces, broken fingers using machinery and pulled muscles were also found within the top 10 topics.

From the results of the LDA analysis for the case study areas, there are obviously regional differences emerging for the reporting of accidents in the workplace. Specifically, it appears that the core themes in London are generally occurring in indoor types of job roles, with pulled muscles as common injuries. However, in the North of England, results indicate that accidents have occurred due to physical objects which may be associated with manual types of jobs roles, with injuries tending to be fractures and bruising.

Next, to identify any seasonal differences within the data, the data were also split into two months: April and August. Accidents occurring in April were examined as April was considered to be a generally cooler month with more unpleasant weather conditions than warmer, summer months such as August, therefore making April and August the most appropriate months to compare. For England and Wales in

April, the top topics that emerged within the data revolved around the themes of trips, and use of machinery and vehicles. Also emerging was the theme of physical assault, with terms such as 'hit', 'colleague' and force'. In August, the key themes which emerged included accidents related to cleaning and wet surfaces, lifting boxes and handling/storage.

The data were split into case study areas by April and August to identify if there were any seasonal differences that vary spatially from one region to another. Unfortunately, possibly due to the size of the datasets at this level, results were inconclusive and therefore difficult to compare. The frequency of some key terms which might represent contrasting weather conditions and levels of daylight were identified and compared by regions to assess if there were any obvious differences by region for seasons. For London, the word 'cold' appears more in incidents that occurred in April than August, and the words 'hot' and 'light' appear more in incidents that occurred in August than April. This suggests that extreme temperatures may be an influencing factor to accidents and injuries in London.

For the North West, the word 'dark' appears more often in April than August, suggesting that more accidents occurred during the dark hours of the night than the day time, which may indicate an issue with shift work patterns, or lack of daylight in the working day. For the North East, 'light' and 'dry' appear more in descriptions of accidents that's occurred in August than April, whilst 'dark', 'hot', 'cold', 'day' and 'night' all appeared more in April than August.

Overall, there appears to be some seasonal differences with workplace accidents and injuries, which have emerged from the free-text RIDDOR data. These differences appear to reflect the types of weather conditions associated with specific seasons through the year, influencing the cause of accidents and the resulting injuries. Core themes of accidents seem to be different for different regions, suggesting that causes of accidents and injuries do vary spatially.

Chapter 8

Discussion

8.1 Introduction

The primary purpose of this research was to explore the risk factors associated with workplace accidents and to determine whether these risk factors varied geographically and by time. From reviewing the literature surrounding occupational health and safety, it was clear that a debate surrounding the relationships between socio-economic variables associated with workers and workplace accident risk existed and that some factors were more heavily researched than others. Clear gaps were found surrounding the topic, particularly in terms of understanding the risks associated with occupation type of the worker and also the effect of physical factors such as weather conditions or levels of daylight in the working day on accident and injury risk. Most studies generally also lacked any context to the UK.

Chapters 4, 5, 6 and 7 provided the results from a range of methods performed on the RIDDOR dataset which aimed to address the research questions associated with this study. This chapter provides a critical review and summary of the key findings of these chapters and illustrates how they contribute to the wider field of research surrounding occupational health and safety. The chapter also examines the validity of the results, and reviews the strengths and weaknesses of the methods used in analysing the RIDDOR data.

The structure of the chapter is broken down into three sections: results (including policy recommendations), data and methods review, and future work. The first section provides an overall summary of the main results of this study. This is set out

first by socio-economic characteristics associated with workers having a workplace accident, followed by spatio-temporal factors. Following this, policy recommendations are made regarding workplace accident risks. A summary is provided of the types of workers, within particular occupations and geographic areas to be targeted by the HSE with tailored health and safety policy. A review of the data and methods is detailed next, providing an overview of the strengths and weaknesses of the data analysed for this study, as well as the techniques used to analyse the data. Finally, a summary of the future work that could be carried out next on the RIDDOR data, delving further into identifying risk factors associated with workplace accidents.

8.2 Results

The socio-economic risk factors associated with workplace accidents relating to the characteristics of the individual, the job, and the working environment, as defined by the HSE and literature review are outlined below. These factors include the age, gender and socio-economic status of the worker, their occupation and industry type, as well as physical factors such as the weather and daylight hours. The data surrounding these characteristics are treated as aggregated area data. The relationship between these area characteristics and workplace accident risk are discussed based on the relevant findings from Chapters 4, 5, 6 and 7.

8.2.1 Socio-economic characteristics associated with workplace accidents

Age

Age as a socio-economic variable relating to workplace accident risk has been heavily researched. The literature surrounding this topic however is mixed, in that some studies find that the age of a worker has a positive relationship with workplace accidents (Nenonen, 2011; Xiang et al., 2000), whilst others find that a negative relationship exists (Bull et al., 2001; Chau et al., 2014; Root, 1981). The results from

the global regression analysis in Chapter 5 identified that a positive relationship exists between the number of younger workers (age 16 to 25) there are in an area and the number of reported workplace accidents that occurred. The model also identified, however, that a stronger positive relationship exists with older workers (age 56 to 65) and workplace accidents. This suggests that the oldest workers in the workforce are at a higher risk of being involved in a workplace accident or from having a work-related injury. These findings reflect the results from studies carried out by Nenonen (2011) and Xiang et al. (2000) in which older workers were found to have a higher workplace accident risk compared to younger workers. Physical health (Beehr et al., 2000; Christensen, 1955), as well as mental cognitive performance (Park et al., 2012) are found within the relevant literature to be the main factors relating to why this may be the case.

Additionally, the exploratory analyses (Chapter 4) suggested that the severity of injuries increases with age overall. In terms of accidents resulting in injuries whereby the worker is absent from work for greater than 3 days, workers aged 16 to 25 had a higher rate than workers aged 56 to 65. In comparison, rates of major injuries and fatalities were found to be highest amongst the oldest age group studied. Looking into fatalities in greater detail, it was found that rates increase with age, excluding the youngest age group (16 to 25), which was found to have the third highest rate after workers aged 56 to 65 and 46 to 55. These results similarly reflect the work carried out by Peek-Asa et al. (1999) and Salminen (2004) in which a positive relationship between age and severity of injuries was found to exist. Reasons for this are identified by Brorsson (1989) who defined younger workers as having generally greater impact resistance than older workers, and explained that an accident that potentially may result in a fatality to an older worker, might only injure a younger worker.

Gender

Gender, as a socio-economic characteristic of a worker, is believed to be an influencing factor in terms of having a workplace accident or injury (Berecki-Gisolf et al., 2015; Salminen, 1994). The exploratory analyses (Chapter 4) indicate that men have a rate of 2.3% of having a workplace accident, whilst women have a rate of

0.83%. The global regression analysis for England and Wales in Chapter 5, however, found that a negative relationship overall exists between the number of male workers in an area and workplace accident rates. This is most likely to be due to the different proportions of men and women working within different occupational and industrial groups and the associated links between particular jobs and accident risks. The risks by occupational groups, and similarly, socio-economic status groups, were also explored in the global regression analysis, and key results are summarised later in this chapter.

In terms of accident severity, out of all the recorded accidents in the RIDDOR dataset, the highest numbers of fatalities were found for men. These findings reflect similarly to evidence presented by Knestaut (1996), whereby accident severity was found to be higher for male workers than female workers. Literature surrounding gender differences in workplace accident risk suggests that the main reason explaining why men are at a higher risk of having a higher severity injury from a workplace accident is because of the types of work hazards in which men and women come into contact with (DeLeire and Levy, 2001; Weeden, 1998). Particular problem areas were highlighted in industries including: agriculture (Quisumbing et al., 2014) and construction (French and Strachan, 2015) where gender segregation was largely prevalent. From exploring the RIDDOR accident data, it was found that although both men and women have high accident rates in the transportation and storage industry, the majority of reported accidents involving men are in industries including: water supply, sewerage and waste management, and mining and quarrying. For women, on the other hand, accident rates were identified as being highest amongst: public administration and defence, and human health and social work. This suggests that a gender gap appears to exist in terms of gender and occupation type, and the risk associated with the type of industry worked in.

Socio-economic class

Much of the literature surrounding health inequalities is based upon differences in socio-economic classes. Marmot et al. (1997) for example, identifies that health differences can be observed across the entire socio-economic spectrum, with the rate of mortality rising with decreasing socio-economic status. Life expectancy is also

found to be strongly correlated with social class. Rogot et al. (1992) reported that men aged 25 in the labour force, live on average 12 more years longer than those not in the labour force. Also, in a study by Marmot et al. (1978), it was found that men in lower employment grades are shorter, heavier for their height, have higher blood pressure, smoke more and report less leisure-time physical activity than men in higher grades. Workers employed in low-skilled jobs are also exposed to harsher and riskier work environments (Toch et al., 2014). Gillen et al. (2007) found that musculoskeletal disease is more prevalent in low-skilled jobs involving physical work. Other hazards such as exposure to chemicals is also more likely in low-skilled jobs (Alamgir and Yu, 2008).

Much of the work within the existing literature surrounding socio-economic class and occupational health and safety unfortunately lacks context to the UK working population. The global regression analysis which modelled workplace accident rates considered the National Statistics Socio-Economic Classification (NS-SEC). The final global model found that a strong positive relationship exists between the number of workers classed as NS-SEC 7 Routine occupations and the number of workplace accidents and injuries that occur within an area. NS-SEC 7 is the lowest NS-SEC group, excluding NS-SEC 8 (unemployed). This implies that socio-economic class has an effect on workplace accident rate, and that those within the lowest class has a strong positive link with workplace accidents within an area.

Occupation

The literature surrounding occupational health and safety suggests that health quality differences exist between occupational groups. In particular, workers in highly skilled occupations are found to have good levels of overall health, in comparison to low skilled workers having a lower level quality of health (Davidson et al., 2006; Townsend et al., 1988). Although research has been carried out surrounding occupational groups and health quality differences, a clear gap exists in the literature surrounding the differences between occupational type and occupational health and safety, particularly workplace accident risk.

The results from the exploratory analysis of the RIDDOR data reveal that the workers in occupations which are reported as having the most accidents are: Process plant and machine operatives (SOC 8), Skilled workers (SOC 5), and Elementary occupations (SOC 9). The lowest reported accidents involve workers in Professional occupations (SOC 2), Managers, directors and senior officials (SOC 1) and Administrative and secretarial occupations (SOC 4). Similarly, the global regression analysis (Chapter 5), revealed a strong positive relationship exists between Process plant and machine operatives (SOC 8) and Elementary occupations (SOC 9) with workplace accident rates. Alternatively, a strong negative relationship was revealed between Managers, directors and senior officials (SOC 1) and workplace accident rate. The results from this research therefore show that some distinct differences between occupational groups and workplace accident risk exists. Specifically, evidence is presented to show that low-skilled jobs, or jobs that require little education or training to perform the tasks associated with those jobs have an increased risk of a workplace accident, compared to jobs that require some sort of formal training or higher education.

It is evident that different occupations carry different work hazards. Previous work suggests that low skilled jobs are found to be more dangerous, involving physical work such as moving and lifting objects, operating machinery and the use of tools (Ljungberg et al., 1989). Low skilled jobs are also considered to be carried out in the harshest physical environments (Alamgir and Yu, 2008) and workers often face repetitive tasks (Grobler, 2013). The overall main causes of accidents found from the reported RIDDOR incidents include: handling, loading and storing, and slips or trips. Injuries mostly reported include: strains, fractures and bruises. These results suggest that manual work is associated with a higher risk of workplace accidents, which emphasises the existing literature surrounding the prevalence of differences in health and safety quality between occupational groups.

Industry

The literature surrounding workplace safety indicates that the type of industry in which a worker is employed in impacts upon their level of risk of having an accident or injury in the workplace. Examples of industries where workers are found to be

exposed to greater risks in the workplace are manufacturing and construction (Chi and Wu, 1997; Sorock et al., 1993). The industries with the most reported accidents in the RIDDOR data include: Water supply, sewerage, waste management and remediation activities (SIC E), Transportation (SIC H) and Manufacturing (SIC C). There is a gap in the literature, however, on comparisons between industries, particularly, the types of hazards causing workplace accidents and how they vary between industrial sectors.

The analyses suggest that out of all of the reported incidents to RIDDOR, the highest causes of accidents involve: handling, slips and trips (Chapter 4). The most common types of injuries include: strains, fractures and bruises, and the body parts mostly injured include: backs, fingers and upper limbs. Although the causes of the majority of accidents involved slips, trips and handling, some differences were found amongst other reported causes. The third most common cause of accident in SIC C Manufacturing and SIC F Construction was found to be being hit by a moving or falling object. This accounted for approximately 9,000 accidents in manufacturing and 17,000 in construction. Physical assault was reportedly the third most common cause of accident for industries: SIC Q Human health and social work activities (accounting for around 16,000 accidents) and SIC O Public administration and defence; compulsory social security (accounting for around 8,000 accidents). Additionally, approximately 1,300 incidents were reported as being involved in an injury caused by an animal in SIC R Arts, entertainment and recreation (the third most common cause of accident in this industrial sector).

In terms of industries with low accident rates, studies reported other health problems being prevalent. These include for example: repetitive strain injury (Tunwattanapong et al., 2016) and obesity (Levine and Miller, 2007). The industries with the lowest accident rates reported under RIDDOR include: SIC J Information and communication, SIC K Financial and insurance activities, and SIC M Professional, scientific and technical activities. Unfortunately, the RIDDOR data obtained for this research does not include long term illness or chronic injuries, rather individual events involving accidents and injuries. It is therefore not possible to compare the

types of other health problems sustained by industries with low accident rates, as most of the recorded incidents are mainly injuries as a result of physical accidents.

With respect to the global regression modelling, it was difficult to incorporate industrial sectors to understand the relationships between specific industries and workplace accidents. The initial models which included industrial sectors provided results which were not significant. It is believed that the potential overlap of occupations by industrial sector may have affected the results of these models. An example of this would be: skilled workers, and managers, directors and senior officials all working in the construction sector. In this instance, it is difficult to compare the relationship between industries and accident rate, as the tasks associated with these occupational groups can vary significantly.

8.2.2 Physical factors associated with workplace accidents

The working environment can be impacted by environmental factors such as weather conditions and levels of daylight. Often is the case that overall accident risk is increased with poor weather and low levels of daylight, however little is known about workplace accident risk. This section examines the existing literature surrounding working environments and their impacts upon workplace accidents, and discusses how the results from this research compare to the existing literature.

Weather

In terms of environmental factors influencing workplace accident risk, not a large amount of research has been carried out previously. Within the literature surrounding seasonal impacts upon health and safety, studies suggest that winter conditions have an adverse impact upon peoples' ability to carry out daily duties safely. Icy conditions and rain impact upon pavements and roads, making them wet and slippery and therefore resulting in slips and falls (Gao and Abeysekera, 2004). It is not surprising therefore that around 1 in 8 accidents take place on rainy days (Edwards, 1996). Road safety is also compromised. Road accident risk reportedly increases by 2-3 times for wet days compared to dry days (Usman et al., 2012). Alternatively, summer

months have been found to effect the level of health and safety of individuals. Evidence suggests that hot temperatures put strain on the human body to carry out physical and mental tasks (Bridger, 2003). Morabito et al. (2006) also discuss the effects of heat stress and the increased effect it has on workplace accidents specifically.

The calendar heat map (Chapter 4) found that a higher number of accidents occurred during the first week of January and the first two weeks of February in 2009. The Met Office reported that the weather in the UK during these two months in particular was mostly unsettled, with the first half of January being very cold with rainfall across most areas (Met Office, 2013a). It was also stated by the Met Office that February 2009 was reportedly the month that experienced the most widespread snowfall since February 1991 (Met Office, 2013b). Similarly, a higher number of accidents occurred during the first two weeks of January in 2010, as well as early December 2010. The Met Office reported that January 2010 was the coldest January over the UK since 1987, with the beginning of the month faced with widespread snowfalls and some sharp frosts (Met Office, 2013c). In December 2010, it was also exceptionally cold across the UK, being the coldest December in over 100 years. Snowfalls occurred in almost all areas, especially in the first week of the month (Met Office, 2013d). These findings suggest a link exists between weather conditions and workplace accident rates.

Daylight

Many workers generally rely on good levels of daylight to perform tasks related to their jobs. It is often the case, however, that work is carried out through the night. Night time construction, in particular has been reported to be five times more hazardous than day time construction (Arditi et al., 2007). More road fatalities are reported to occur during the night due to poor visibility (Owens, 1993). The months with the highest accident rates are October and January, which experience some of the lowest levels of daylight overall across the year in the UK (National Geographic Society, 2011).

The hours of daylight plots (Chapter 4) show that there are less hours of daylight in the northern parts of Great Britain, and more in the southern parts, during winter months. This appears to link with the mosaic plot (Chapter 4) which shows that more accidents occur in Northern England and Scotland during winter months, whilst less occur in the South of England. In comparison, less hours of daylight are found in the South of England in the summer months, whilst more are found in Northern England and Scotland. This also appears to link with more accidents being reported as occurring in the South West of England, London and Wales during the summer months.

Mustard et al. (2013) found evidence to suggest that accident rate (number of accidents/number of workers working at that time) is highest between 5pm and 5am. Not only is it believed that poor visibility affects accident risk during these times, but fatigue, sleepiness and stress of working during evenings and nights are believed to impact upon workplace health and safety (Dembe et al., 2006; Folkard and Tucker, 2003). Daylight Saving Time (forwarding the clocks by an hour in spring and putting them back by an hour in autumn) is thought to have an effect on our bodies by making them tired and less responsive or alert to danger (Barnes and Wagner, 2009; Varughese and Allen, 2001). The number of accidents reported on a Monday following the clocks being put forward (and in essence an hour of sleep is lost), was not found to be greater, than the number of accidents occurring on a standard 'Monday'.

In terms of days of the week, however, the literature surrounding workplace health and safety indicates that most accidents occur on Mondays than any other day of the week (Butler et al., 2014; Packer and Shaheen, 1993) and this has also been attributed to workers being more tired and less alert on the first day back at work following a weekend. Mondays were found to have the highest number of accidents occurring (as reported under RIDDOR), compared to any other day of the week. The number of accidents then descends through the week to the weekend when the least accidents are reported on Sundays.

8.2.3 Geographic variations in risk factors associated with workplace accidents

It was reported by the HSE that between 2012-2015, rates of workplace accidents and injuries varied by geographic region (HSE, 2015g). The RIDDOR data obtained for this investigation similarly showed that workplace accident and injury rates varied spatially (Chapter 4). Highest rates were found overall in the East Midlands, followed by Yorkshire and the Humber, and the lowest accident rates were found in the South of England (South West, South East and London). By local authority district area, the higher rates of accidents were found in: North Warwickshire (West Midlands), and Corby and Bassetlaw (East Midlands); in comparison, the lowest accident rates were found in the City of London, Isles of Scilly and Tower Hamlets (London). Additionally, overall, the results from the global model suggest that a negative relationship exists with population density and the number of workplace accidents which occur within an area. This suggests that workers within more suburban and rural areas, where the workplace population density is smaller, compared to urban areas, are at a higher risk of a workplace accident.

Work carried out by Davies and Elias (2000) found that characteristics of individuals working within areas which had a high accident rate, varied to those of workers in areas with low accident rates. It was shown that for London, an employee injury rate of 46% below the average across all regions existed. This was attributed to there being a high proportion of young workers employed within London, which was believed to be a low-accident risk group. For Northern England, Davies and Elias (2000) found that average employee injury rates were 38% above the average rate for all regions. This was attributed to the types of industrial and occupational compositions found within the North of England. Specifically, a high proportion of workers were employed within the construction industry, and a relatively high proportion of workers were employed to carry out work in craft related occupations. Work carried out by Cameron et al. (2008) observed clear differences in accident rates between Scotland and the rest of Great Britain. Jobs involving the use of scaffold and bricklaying activities were found to be associated with high accident rates in Scotland.

The exploratory analysis (Chapter 4) revealed clear differences between areas by workers' characteristics, similarly to findings from Cameron et al. (2008) and Davies and Elias (2000). The local authority district with the highest rate of accidents (North Warwickshire) experiences the highest rates in: SIC B Mining and quarrying, SIC H Transportation and SIC N Administrative and support activities. The second district with the highest rates (Corby, Northamptonshire) has the most accidents occurring within: SIC H Transportation, SIC N Administrative and support activities and SIC C Manufacturing. In comparison, the City of London, where the lowest rates of workplace accidents occurred, workers were found to mostly have accidents in: SIC E Water supply, sewerage and remediation activities, SIC F Construction and SIC I Accommodation and food service activities.

Differences between characteristics were also found from the global regression models fit to data across accidents occurring in London, North West England and North East England. The model for London showed a stronger positive relationship with the number of NS-SEC 7 workers, as well as younger workers (age 16 to 25), with workplace accidents compared to the other regions. In comparison, the North East seen a strong positive relationship with SOC 1 Managers, directors and senior officials and workplace accident rates whereas the other regions seen a strong negative relationship. The relationship between SOC 9 Elementary occupations and workplace accidents was also found to be strongest in the North East, compared to London and the North West. It is evident that the socio-economic coefficient estimates from the global models varied between regions. These findings suggest that different socio-economic compositions of areas result in different workplace accidents rates, supporting the work carried out by Cameron et al. (2008) and Davies and Elias (2000). For some coefficients, one or several case study areas were found to have a stronger relationship with workplace accidents, than for England and Wales, as a whole. In other cases, the variables for England and Wales were found to have a stronger relationship with workplace accidents. As a result, evidence was presented that showed it may be unreliable to consider only one global model to represent the workplace population of every part of the country and so a local regression analysis was carried out.

Chapter 6 presented the results for the GWR analysis of several key socio-economic groups that were modelled. These included: SOCs 1, 8 and 9 for London, SOCs 1 and 9 and NS-SEC 7 for the North West, and SOCs 1, 8 and 9 for the North East. Table 8.1 below shows the key areas where a positive relationship was found between each of the coefficients and the number of workplace accidents. In general, the coefficient estimates were found to be largest for the North East, whilst the estimates were similar in size for London and the North West.

Table 8.1: Areas with the largest GWR coefficient estimates out of each of the three regions

	London	**North West**	**North East**
SOC 1 Managers, directors and senior officials	East, North & South West	Merseyside, Lancaster & parts of Manchester	County Durham & Hartlepool
SOC 8 Process, plant and machine operatives	Parts of South London		West and South (including County Durham & Richmondshire)
SOC 9 Elementary Occupations	Central London & North London (including Camden, Westminster, Barnet & Enfield)	South Cheshire & West Cumbria	East and South (including County Durham & parts of Newcastle)
NS-SEC 7 Routine Occupations		Lancashire & South Manchester	

Focusing on SOC1 Managers, directors and senior officials first, although the global model overall found that a negative relationship existed between this coefficient and the number of workplace accidents, positive GWR coefficient estimates were found in small areas in each of the three regions. In London, a positive relationship was found to exist in parts of the East and North and South West of the region, whilst in North West England, positive relationships were found to exist in areas around Merseyside, Lancaster and parts of Manchester. In all of these areas, the percentages

of SOC 1 workers was small. The strongest positive relationships in the North East region were found in County Durham and Hartlepool, where a slightly higher proportion of SOC 1 workers were located. The coefficient estimates for SOC 1 workers in these areas, were the largest out of all three of the regions within the study.

For SOC 9 Elementary occupations, London was found to consist of mostly positive GWR coefficient values. The strongest positive relationships between the number of SOC 9 workers and workplace accidents in London exist in parts of the East and West of the region, where the number of SOC 9 workers is large. Positive GWR coefficients were also found where the number of SOC 9 workers was small, including parts of Central London and North London (Camden, Westminster, Barnet and Enfield). The strongest relationship between the number of SOC 9 workers in an area and the number of accidents in North West England was found in South Cheshire and West Cumbria (where a large number of SOC 9 workers are located). In the North East region, positive GWR Coefficients were found in the East and South (in particular: Scarborough, County Durham and parts of Newcastle). The coefficient estimates for these areas were the largest compared to those in the North West and London.

In London, positive relationships were found between the number of SOC 8 Process, plant and machine operatives in an area and workplace accidents in areas where a small percentage of SOC 8 workers are found (South of the region). Negative relationships were found in the West of London (Hounslow). For the North East, the strongest positive relationships were found in the West and South of the region (County Durham and Richmondshire) and parts of Carlisle. In the North West, positive GWR coefficients were found for NS-SEC 7 Routine occupations in Lancashire and South Manchester, and negative GWR coefficients were found in Central Manchester and parts of Merseyside and Cheshire.

In terms of the text mining analysis carried out in Chapter 7, the most common words appearing in accident descriptions for all three regions were: 'back' and 'work'. Differences between the three regions, became clearer with the topic modelling

analysis. For England and Wales, the most common topics surrounded themes involving: trips and falls, cleaning, slipping on wet surfaces, and machinery/equipment. For London, topics surrounded indoor themes, with terms emerging such as: door, box and shelf. For the North West region, topics surrounded themes of physical objects such as: trailer, cable, pallet and truck, with slipping on wet floors also a clear theme. Finally, for the North East, the theme of physical objects was also prevalent, with terms emerging such as: machine, vehicle and pallet. Another clear theme was slipping on wet floors and broken fingers with machinery.

The local regression analysis and topic modelling analysis has identified that risk factors associated with workplace accidents vary spatially. There is a large gap in the literature surrounding geographic variations in occupational health and safety and so this analysis has been one of the first to examine and report upon differences between socio-economic groups across geographic areas, and shown that variations in accident risk are prevalent within small areas.

8.2.4 Spatio-temporal variations in risk factors associated with workplace accidents

A gap in the literature exists surrounding the spatio-temporal variations in risk factors associated with workplace accidents. Whilst it is clear from the relevant literature, that environmental factors may play an important part in understanding accident risk (Bridger, 2003; Findikyan and Sells, 1965), the significance on workplace accident risk, and more specifically, geographic variations in workplace accident risk, is largely under researched. Although physical factors have been examined on accident risk, and geographic differences in workplace accidents have been examined independently (see previous section), spatio-temporal variations were explored as part of this research, to help fill a gap within the field of occupational health and safety.

The mosaic plot (Chapter 4), illustrated how the accidents reported under RIDDOR varied geographically and by month. As summarised in Section 8.2.2, for the North of England, more accidents were found to occur in December and January, whilst fewer accidents occurred in June and July. In comparison, for London and the South, more accidents were found to occur in June and July and less were found to occur in December and January. The Met Office (2012) explains that in general, areas in the south of the UK tend to be drier, warmer and sunnier than those in the north. These conditions usually also occur more in the spring and summer months than autumn and winter. Similarly, the daylight hours' maps showed how the number of hours of daylight varied across Great Britain, with more hours in the South in winter, and more in the North in summer.

The GWR analysis (Chapter 6), together with the topic modelling analysis (Chapter 7) revealed some interesting spatio-temporal variations of workplace accident risks. Tables 8.2 and 8.3 show the key areas where a positive relationship was found between each of the coefficients and the number of workplace accidents during summer and winter out of each of the three regions. In general, the spatial structure is similar between summer and winter in London for the SOC 1 and SOC 9 estimates. The strongest differences between the summer and winter coefficient estimates are found in the North East.

Table 8.2: Areas with the largest GWR coefficient estimates out of each of the three regions during summer

	London	**North West**	**North East**
SOC 1 Managers, directors and senior officials	South & West (including Hounslow, Ealing & Richmond upon Thames	Merseyside & South Lancashire	South of region
SOC 9 Elementary Occupations	Central & North East	North West Cumbria & West Lancashire	South Northumbria & parts of Darlington

Table 8.3: Areas with the largest GWR coefficient estimates out of each of the three regions during winter

	London	**North West**	**North East**
SOC 1 Managers, directors and senior officials	Central & North West	Central Manchester	West of region (including Carlisle & West Cumbria)
SOC 9 Elementary Occupations	North West (including Barnet, Harrow & Brent)	Merseyside & South Manchester	West Northumbria

The key results highlight that for areas in Merseyside and South Lancashire during the summer, a stronger positive relationship exists between the number of SOC 1 Managers, directors and senior officials and the number of accidents. In the winter, larger coefficients are found in Central Manchester for SOC 1 workers. For SOC 9 workers, the GWR coefficients were the largest in North West Cumbria and West Lancashire during the summer, but largest in the South of the region (including Merseyside and South of Manchester) during winter. In the North East, the strongest positive relationship between the number of SOC 9 Elementary workers and the number of workplace accidents exists in the West of Northumbria in winter, whilst larger GWR coefficients are found in lower Northumberland and parts of Darlington in summer. In winter, a stronger relationship between SOC 1 workers and workplace accident rate is found in the West of the North East region (Carlisle and parts of West Cumbria) compared to the summer. Alternatively, in the summer, higher coefficient estimates are found in the South of the region (in the district of Eden).

The topic modelling revealed differences between accidents reported in April, compared to August. Trips and slips were more prevalent in April for all of England and Wales, London and the North West. The term 'cold' was also more common in descriptions of accidents occurring in April, compared to August for England and Wales and London in April, and the term 'dark' was more common for the North West in April. The terms 'light', 'hot' and 'dry' came out most often in August for England and Wales, with the themes of topics surrounding: cleaning, lifting boxes

and storage/handling.

In general, most of the literature surrounding the physical impacts upon the risk of accidents lacks context specifically to the workplace and no large research study to date examines the spatio-temporal variations of accident risks across England and Wales. This research has been one of the first studies to identify differences between workplace accident rates by season and time, identifying geographic variations in workplace accident risks.

8.3 Policy Recommendations

The HSE's mission is 'the prevention of death, injury and ill health to those at work and those affected by work activities' (HSE, 2015b). Although health and safety in Great Britain's workplaces have improved significantly since the introduction of the Health and Safety at Work etc. Act 1974, it is understood that the HSE's contribution to improving standards of health and safety must continue to ensure the well-being of workers into the future. The HSE recognises that the world we operate in today is markedly different to the one in 1974, when the Health and Safety at Work etc. Act was introduced, and therefore must adapt its policies and regulations to the new challenges that it is faced with. The key areas that the HSE wishes to improve upon are set out within the Science Plan and Business Plan 2015/16:

A. Leading others to improve health and safety in the workplace

B. Providing an effective regulatory framework

C. Securing compliance with the law

D. Reducing the likelihood of low-frequency, high-impact catastrophic events

(HSE, 2015c)

The aim of this research was to help continue the success of the HSE's role in regulating health and safety. In doing this, the overall goal was to provide policy recommendations upon the areas in which the HSE wishes to improve upon, as

listed above. There are several key findings from this study. The first is that a set of socio-economic variables was found to impact upon health and safety in the workplace, which had not been studied in great detail previously. Evidence was next provided to show that these risk factors vary geographically. Again, previous research surrounding this topic was minimal. Finally, the third overall key finding was that these risk factors also varied between months, in that more accidents occurred during certain seasons over others, for particular socio-economic groups in certain areas. In helping to lead others to improve health and safety in the workplace, provide an effective regulatory framework, secure compliance with the law and reduce the likelihood of catastrophic events, the HSE needs to help target the right policies to the right workplaces, with a focus on particular times of the year.

The results of this research found that the more older workers there are in an area, the more likely a workplace accident is to occur. It is therefore clear that the HSE must provide advice to employers and older workers in ensuring extra care is taken during training and when carrying out job-related tasks. The severity of injuries, in particular, was also found to be higher for older workers than younger workers, with more older workers reported to have suffered major injuries and fatalities. When targeting tailored health and safety policies, the HSE must consider the severity of injuries of this high-risk group, which may benefit from increased knowledge surrounding health and safety and prevention of high-risk accidents. In terms of the two age groups analysed, geographic differences were found in the case studies looked at. In London, the number of accidents was found to increase with the number of younger workers, whilst a negative relationship was found between younger workers and accident rates in the North West and North East of England. Although older workers overall have a positive relationship with accident rate for each of the regions, policy recommendations should also be tailored around younger workers in London, to help reduce workplace accidents specifically in that region.

Nationally, more men than women were reported as having a work related accident. Additionally, the rate of fatality as a result of a work-related accident was found to be highest amongst male workers than female workers. The industries in which the majority of accidents occur for just men include: water supply, sewerage and waste

management, and mining and quarrying. For women, on the other hand, accident rates were highest amongst: public administration and defence, and human health and social work. This suggests that the HSE should consider gender alongside industry worked in, in determining how to tailor health and safety policy.

A strong positive relationship was found to exist between the number of workers classed as NS-SEC 7 Routine occupations and the number of workplace accidents and injuries that occur within an area. NS-SEC 7 is the lowest NS-SEC group, excluding NS-SEC 8 (unemployed), indicating that workers within the lowest class have a strongly positive link with workplace accidents. The strongest relationship was found in London, whilst for all areas, a positive relationship was found to exist. Policy implications of this suggest that the HSE needs to provide further assistance in these areas, with particular focus on the London region, in the form of education and increased awareness on work-related health and safety, specifically in areas with high numbers of NS-SEC 7 workers.

The industries with the most reported accidents in the RIDDOR data include: Water supply, sewerage, waste management and remediation activities (SIC E), Transportation (SIC H) and Manufacturing (SIC C). There is a gap in the literature, however, on comparisons between industries, particularly, the types of hazards causing workplace accidents and how they vary between industrial sectors. Although the causes of the majority of accidents involved slips, trips and handling, some differences were found amongst other reported causes. Being hit by a moving or falling object caused the third highest proportion of injuries in manufacturing and construction, whilst physical assault was the third most common cause of accident for industries including human health and social work activities and public administration and defence; compulsory social security. The HSE should tailor its policies to specific industrial sectors, to help reduce work-related accidents that are most common in each sector.

Evidence is presented to show that low-skilled jobs, or jobs that require little education or training to perform the tasks associated with those jobs have an increased risk of a workplace accident, compared to skilled jobs, which require some

sort of formal training or higher education. The most high risk occupational groups were found to be SOC 8 Process plant and machine operatives, SOC 5 Skilled workers and SOC 9 elementary occupations. The relationship between SOC 8 and 9 with workplace accident rates were found to vary geographically, with the strongest link between SOC 8 and accident rate being found in the North West compared to the other two regions, whilst for SOC 9 the strongest relationship was found in the North East region. The number of SOC 1 workers in an area was found to be negatively related to the number of workplace accidents likely to occur, for all areas, other than the North East. The HSE should focus attention on these specific occupations where accident rates are highest in certain areas across England and Wales.

The GWR analysis looked into the small-scale areas found within the three case study regions, to discover whether there were any variations in the relationships between these socio-economic variables and workplace accident rates. The key results were presented, which were for occupational groups SOCs 1, 8 and 9 for London, SOCs 1 and 9 and NS-SEC 7 for North West, and SOCs 1, 8 and 9 for North East. The results indicated that the following key socio-economic groups within parts of the three study regions were found to have an increased risk of having workplace accidents. These included:

- SOC 1 Managers, directors and senior officials in East, North and South West areas of London, as well as Merseyside, Lancaster and parts of Manchester, and County Durham and Hartlepool,
- SOC 8 Process, plant and machine operatives in parts of South London, and West and South of the North East region,
- SOC 9 Elementary occupations in Central and North London, South Cheshire and West Cumbria, and County Durham and parts of Newcastle, and
- NS-SEC 7 workers in Lancashire and South Manchester.

It would therefore be beneficial for tailored health and safety guidance to be targeted locally to these types of workers within these particular areas. Finally, the findings suggest that the HSE should aim to focus attention on weather conditions and hours of daylight. This is particularly prevalent for Northern England in winter months, and Southern England in summer months, which both experience higher rates of accidents when weather conditions are poor, or there are less daylight hours in the working days. Key results were summarised for particular occupational groups within parts of the three case study regions, during summer and winter. These results indicate that it would be beneficial for tailored health and safety guidance to be targeted to the following types of workers during the summer:

- SOC 1 Managers, directors and senior officials in Merseyside and South Lancashire, and South of the North East region, as well as Hounslow, Ealing and Richmond upon Thames (in South and West London), and
- SOC 9 Elementary occupations in North West Cumbria and West Lancashire, South Northumbria, parts of Darlington and Central and North East London.

The following areas should be targeted with health and safety guidance during the winter:

- SOC 1 Managers, directors and senior officials in Central Manchester, and Carlisle and West Cumbria, and
- SOC 9 Elementary occupations in Merseyside, South Manchester, West Northumbria and North West London (including Barnet, Harrow and Brent).

These findings suggest spatial variations exist by season, and that the HSE needs to consider physical and area-level characteristics in determining tailored health and safety guidelines to the right workplaces. Potentially with targeting these particular high-risk groups of workers with tailored guidelines to improve health and safety conditions in the workplace, accidents may be preventable and the overall number of accidents could be significantly reduced. Identifying key risk factors of workers, as well as where workers are geographically, should provide the HSE with a significant

level of information to be able to draw up new policies and guidance to reach out to particular workplaces which may be most affected by poorer health and safety standards. This overall contributes significantly to the aims of the Business and Science Plans that the HSE has in place to help towards reducing the number of occurrences of accidents and injuries in the workplace.

8.4 Review of Data and Methods

The data obtained for this study consisted of a snapshot of the RIDDOR database obtained from the HSE which contained recorded workplace accident data consisting of information with regards to the worker, place of employment and the type of injury sustained as a result of the accident. The dataset consisted of more than 822,000 records spanning across 6-years' worth of workplace accidents, making it a relatively accurate representation of the population of workers having workplace accidents. There are several issues surrounding the data, however, that should be addressed. Although it is a legal requirement to report workplace accidents that occur across Great Britain, there may be issues surrounding workplaces that do not report accidents. Reporting rates may be biased towards those that have a stricter and more rigorous health and safety workplace policy. In particular, those working from home, or those working 0-hour contracts may not be accurately represented within the sample. Also, within the literature, health quality emerged as a key topic surrounding those working in the harshest environments (Alamgir and Yu, 2008; Davidson et al., 2006) as well anxiety and stress being a factor for those working in professional roles (Tunwattanapong et al., 2016). The RIDDOR data obtained for this study does not contain records of long-term illnesses as a result of work, which could add a significant amount of information to the study regarding occupational health and safety.

For the Census data, workplace populations by selected socio-economic characteristics were chosen to be modelled for the study. These were obtained by Workplace Zone geographies, which are a new geography created to provide

accurate boundaries of defined workplace populations. The selected data was from the most recent 2011 Census. Although this is now approximately 6 years old, the data accurately represents the workplace population at the time the accidents were recorded under RIDDOR (2006-2011). One of the main limitations of the Census data, however, was that workplace populations were unavailable by Workplace Zones for Scotland, which meant that the study had to be restricted to cover solely England and Wales. This was detrimental to the study as the RIDDOR data covered accidents occurring across Great Britain and it would have been hugely beneficial to the HSE to provide policy recommendations across the entirety of Great Britain considering a range of geographic areas consisting of different socio-economic compositions.

The aim of the exploratory analysis (Chapter 4) was to gain an insight into the details of the reported accidents and injuries stored within the RIDDOR dataset. The reasoning behind this was to discover basic information and trends that might have helped to indicate the types of risk factors associated with having workplace accidents. This chapter provided a concise preliminary analysis to the research that highlighted key socio-economic factors to explore further with creating a global regression model. Chapter 5 focused on gaining a 'global' perspective on accidents in the workplace based on socio-economic area characteristics across England and Wales. This was carried out via global regression modelling whereby alternative forms of Generalised Linear Models (Poisson, quasi-Poisson and negative binomial regression models) were fitted to a set of socio-economic variables based on the characteristics of the data. The final global model was constructed through testing a range of variables inferred from the literature surrounding occupational health and safety and through a data-driven based variable selection process.

The variables within the final model included: the number of workers who were male, aged 16 to 25 and 56 to 65, classed within SOC 1 Managers, directors and senior officials, SOC 8 Process plant and machine operatives and SOC 9 Elementary occupations, as well as NS-SEC 7 Routine occupations, and the total population density. Although regression modelling is widely used in studies modelling accident data (Hilton and Whiteford, 2010; Park et al., 2012), there is no large study to date that utilises regression modelling in analysing a large dataset specifically of accidents

occurring within the workplace, to understand the key socio-economic risk-factors relating to workplace health and safety. An important contribution of this research is therefore that a model has been developed to identify risk factors associated with having a workplace accident.

The main issue surrounding the global model was that it represented the key socio-economic risk-factors associated with workplace accidents across the whole of England and Wales, and failed to pick up any geographic differences that may have been present within the data. In particular, the exploratory analysis (Chapter 4) identified particular areas where workplace accident rates were highest. It was clear overall from the mosaic plot and mapping the accident rates by local authority district workplace populations, that there were geographic variations in workplace accidents. Additionally, the residuals map of the global model shown in Chapter 6 revealed that the model did not accurately predict accidents across the whole of England and Wales. A local regression analysis was therefore carried out on the RIDDOR data to explore in further detail the risk factors of these workplace accidents and how they varied locally.

Geographically Weighted Regression analysis (Chapter 6) was performed using the standalone GWR4 Windows-based application on data spanning across three regions: London, North West England and North East England. The reason for splitting into case study regions was because running the GWR analysis on all WZs across England and Wales failed due to the size of the data, and similarly failed attempts were made using R packages **spgwr** and **GWmodel**. The three case study areas were chosen based on their spread across the country as well as varying socio-economic compositions. This helped to provide an accurate representation of how well GWR performed across different areas. The GWR pseudo R^2 and residuals maps of these regional models showed that they predicted accident rates relatively accurately across each of the three regions.

The GWR analysis successfully provided a more detailed analysis of workplace accident risk because it revealed the relationships between socio-economic area characteristics on accident rates at a smaller level geography (WZs). This allowed

specific socio-economic groups located within particular smaller geographic areas to be highlighted. The result of this is that high-risk groups within the workplace population can be targeted with tailored health and safety policies by the HSE to help reduce workplace accident rates. Alternative spatial modelling techniques could have been utilised as part of this study, in particular multilevel modelling or Bayesian spatial analyses. These methods were explored in greater detail in Chapter 3, however it was decided that GWR would be utilised for this study as it built upon the previous standard global regression model and was a relatively simple method to implement.

The free-text RIDDOR data obtained as part of this study was first explored via text mining techniques, to gain an understanding of what words appeared most often within workplace accident descriptions. Topic modelling was carried out next, specifically Latent Dirichlet Allocation, on descriptions of accidents occurring across England and Wales and then individually by case study region. A set of topics were found for each area, with their associated terms that most accurately described the accidents reported to the HSE. These topics were used to explore differences between geographic areas and between seasons, to elaborate on the findings from the GWR analysis and to discover any new seasonal variation in the reported accident data.

The main issue with the LDA analysis, however, was that the textual dataset obtained for the research was small in size, in that it only covered 6 months of recorded accidents. This made it difficult to compare findings across seasons as results were compared for accidents which took place in either April or August as these were the months that were the furthest apart in the dataset. It was also difficult to gain any detailed insight from the topics which were generated from the seasonal data, split by regions, as the datasets became very small. The LDA analysis, overall however, provided an alternative view of the RIDDOR data, and allowed the free-text dataset to be explored in detail. No large study to date has analysed textual descriptions of workplace accidents and in general, mining techniques utilised within previous studies are often not performed alongside quantitative analyses. The possibilities for knowledge discovery using a range of mixed methods has been demonstrated in this

research study by utilising topic modelling techniques alongside regression and GWR techniques.

8.5 Future Work

There are limitations associated with the research methods and the data used for this research analysis, as outlined in the previous section. It is therefore useful to address some of these in further detail and discuss possible extensions to the study that should be considered if the research is to be built on in the future. In terms of data, if workplace populations become available in the future for Scotland by WZs, then Great Britain should be modelled, rather than England and Wales. This would be beneficial to the HSE to ensure policies stretched across the whole of Great Britain (where workplace accidents are recorded), rather than restricted to just England and Wales.

A limitation of the GWR analysis was that the software was unable to cope with very large datasets, and that smaller case study areas had to be considered instead. The problem with this is that a full picture was not painted of England and Wales, instead only small-areas within three regions across the country were investigated. In the future, should software permit, it would be worthwhile to carry out the GWR analysis on the data for the whole of Great Britain (if data becomes available for Scotland), rather than split into case study areas. This could reveal new features in areas that have not been considered in this study.

In terms of the textual analysis, the main limitation was the size of the data. The textual RIDDOR dataset obtained by the HSE contained only reported accidents covering six months. This meant that carrying out analyses on the effects of seasonal differences was difficult, in that April and August were chosen as contrasting months, rather than comparing months from summer and winter. It would have been beneficial to have the textual dataset covering the same timeframe as the first RIDDOR dataset obtained, which covered six years' worth of accident data. This would have more accurately reflected the quantitative work carried out in chapters 5 and 6.

Finally, to extend upon the research into occupational health and safety, it would be interesting to analyse data from one of the alternative RIDDOR data sources, for example cases of diseases reported as a result of work. This might offer an alternative view on workplace occupational health and safety, whereby diseases and illnesses may be associated with different socio-economic characteristics, and be more common in particular seasons and geographic areas, compared to workplace accidents. Potentially this could further expand on tailoring health and safety policy to specific areas by the HSE, which in effect could also help to improve standards of occupational health and safety.

8.6 Summary

This chapter has provided a summary of the results from Chapters 4, 5, 6 and 7 in relation to the existing literature surrounding occupational health and safety. It is clear that there are specific socio-economic groups that are more at risk of having a workplace accident or injury over others, and that risk factors vary geographically and seasonally. Particular socio-economic groups within parts of the case-study regions were identified and policy recommendations were outlined for the HSE to consider in tailoring occupational health and safety guidance.

A review of the data and methods was provided next, detailing the strengths and limitations of each. The chapter concluded with details of possible approaches for future work, taking into consideration the key limitations of the data and methods utilised for this study. The following and final chapter recaps the main findings and refers back to the initial aims and objectives of the study, detailing how the analyses have helped answer some important questions around occupational health and safety.

Chapter 9

Conclusions

9.1 Introduction

This chapter is concerned with summarising the research by highlighting the main results and how they fit in to the wider literature surrounding occupational health and safety. The aims and objectives from Chapter 1 are revisited and evidence is presented to show how each of these have been met throughout this research and a summary of the overall original findings are made.

9.2 Research Summary

The analyses undertaken comprised three parts. First, workplace accident rates in England and Wales were studied, based on a set of socio-economic area characteristics, which identified the types of workers most likely to have a workplace accident. Secondly, the relationships between these socio-economic characteristics and workplace accident rates were explored geographically to identify spatial variations in workplace accident risks. Temporal factors were also considered such as the impact of weather conditions and levels of daylight brought by different seasons through the year. Thirdly, text mining the accident reports themselves provided a different perspective on the types of accidents occurring, giving a further source of information to reflect on the key results from the quantitative analyses. Each component is novel and original and together comprise a highly innovative and revealing programme of research.

In Chapter 1 the aims for this research project were set out:

1. To understand which socio-economic risk factors relate to workplace accidents and injuries,

2. To investigate whether the relationship between these socio-economic characteristics and workplace accidents and injuries vary geographically; and

3. To explore the impact that environmental conditions have upon workplace accident and injuries; investigating how these affect the relationships between the socio-economic risk factors on accident rates geographically.

In order to achieve these aims, eight research objectives were formulated. These are outlined below with summaries of how they were achieved, and the research findings from each relevant chapter.

Objective 1. Review the existing body of literature surrounding the socio-economic and physical determinants of workplace accidents, paying particular attention to geographic variations.

Chapter 2 provided a review of the literature surrounding risk factors associated with workplace accidents. The three aspects explored which the HSE considers to be important in helping to improve health and safety standards included characteristics relating to: the worker, the type of job, and working conditions. The socio-economic variables related to a worker which were found to have an impact upon workplace accident risk included: age, gender, socio-economic class and occupation and industry type. The physical factors outlined included weather conditions brought by seasonal variations and levels of daylight in the working day. Lastly the chapter provided a history and background to health and safety policy and legislation.

Clear gaps were found in the literature surrounding workplace health and safety. There is a lack of research focusing on the UK as a whole and a mix of evidence surrounding the effects of different socio-economic characteristics on accident rates is revealed in published work. The comparisons across occupational and industrial groups are considerably under researched. In addition, a clear gap existed on the effect of seasonal impacts on workplace accidents across different occupational and socio-economic groups. Lastly, geographic variations in workplace accidents and injuries were also shown to be largely under researched, with no large study to date having attempted to capture the effect of seasonal changes on geographic differences in workplace accident risks. Overall, this chapter provided context to the research study and identified issues needing to be addressed and investigated to gain a greater understanding about occupational health and safety.

Objective 2. Review the suitability of methods for analysing the RIDDOR data to achieve the aims of this research.

Chapter 3 provided an overview of the data and methodological approaches used in the research study. The details surrounding the RIDDOR data and Census data were outlined first. The strengths and limitations of the data were discussed, as well as the overall structure and list of variables contained within the datasets. The methods used for the study were discussed next, giving an overview as to the available approaches that could be undertaken to analyse the RIDDOR data.

It was highlighted in the literature review (Chapter 2) that a set of socio-economic variables may impact workplace accident risk. To investigate this further, regression analysis, specifically Generalised Linear Modelling (GLM), a flexible statistical framework for investigating the relationship between a set of variables, was explored. Examples surrounding the application of GLMs were evaluated and methods of performing regression analysis in R were outlined in detail. To explore the geographic determinants of workplace accidents, a form of local regression modelling (Geographically Weighted Regression(GWR)) was examined. GWR is capable of exploring spatially-varying relationships within data. The global regression

analysis produced a set of outputs highlighting the similarities between socio-economic variables with accident rates, however GWR highlighted the spatial variations between these variables. Chapter 3 explored this method in detail, outlining the application of GWR using the GWR4 application and the overall benefits of using this approach to model local variations in accident rates.

Finally, topic modelling as a method for investigating the free-text fields of the second RIDDOR dataset obtained for the study was explored. Topic models are outlined as probabilistic models for uncovering the underlying structure of a group of unorganised texts. More simply, a 'topic' consists of a cluster of words that occur frequently together in a set of documents. For investigating the themes within the free-text RIDDOR data, Latent Dirichlet Allocation (LDA) was introduced in Chapter 4 as a type of topic modelling technique. LDA has previously been used in a variety of research fields and is growing in popularity. The chapter makes reference to the application of LDA in R, outlining the R packages used for carrying out the analysis and the capabilities of LDA as a textual analysis method.

Objective 3. Explore the RIDDOR data for key themes to gain a background to reported workplace accidents and injuries

Chapter 4 explored the background statistics of the RIDDOR dataset. Basic trends were summarised in terms of the characteristics of the workers having accidents: their age, gender, and the types of occupations and industries worked in. The chapter gave an overview of the types of accidents that commonly occurred, including causes of accidents, types of accidents and the severity of the injuries obtained. A brief overview of the location of the accidents was provided. Specifically, the rates of accidents by workplace populations by Local Authority District area were highlighted and a mosaic plot illustrated the accidents grouped by region and by month. These revealed key features in terms of geographic variations in workplace accidents and how these rates vary by time of the year. The chapter illustrated the emerging key themes found within the RIDDOR data, which provided an introduction to the following analyses chapters (Chapters 5, 6 and 7).

Objective 4. Model the RIDDOR data to gain an insight into the relationships between the key socio-economic determinants of workplace accidents nationally.

Chapter 5 focused on gaining a global perspective on accidents in the workplace based on a set of socio-economic area characteristics. This was carried out via global regression modelling. Alternative forms of GLMs (Poisson, quasi-Poisson and negative binomial) most suited to the characteristics of the data were fitted. Socio-economic variables were selected on the basis of those found to be potential determinants of workplace accidents, as identified in the literature review (Chapter 2).

The final model included a range of variables: age, gender, socio-economic class, population density and three different occupational groups. The coefficient estimates revealed the relationships between the socio-economic variables and workplace accident rates. The number of younger and older workers, the number of workers within SOCs 8 and 9, as well as the number of workers classed as NS-SEC 7 within an area were found to have positive relationships with the number of workplace accidents. The number of male workers, population density, and the number of SOC 1 workers were found to have a negative relationship with the number of workplace accidents within an area. The model therefore provides a comprehensive overview of the likely socio-economic area characteristics that have an effect on workplace accidents.

Objective 5. Carry out a local regression analysis of the RIDDOR data to explore the geographic variation in accident risks.

While chapter 5 explored the relationship between socio-economic variables and workplace accident rates, at a global national scale, the next stage of the analysis involved identifying whether these relationships varied geographically. Chapter 6 involved performing a GWR analysis on the RIDDOR data and census data in an attempt to provide a more nuanced model of the determinants of workplace

accidents across England and Wales. GWR was performed on three case study areas: London, North West England and North East England, to identify the spatial variation in the socio-economic coefficients within each of these regions.

The key GWR local coefficient estimates were mapped for each of the regions, which included SOCs 1, 8 and 9 for London and the North East, and SOCs 1 and 9 and NS-SEC 7 for the North West. The key findings were highlighted which identified particular high-risk groups of workers in certain areas of the regions. For example, the number of SOC 8 Process, plant and machine operatives was found to have a positive relationship with the number of workplace accidents occurring in parts of County Durham and Richmondshire in the North East, where percentages of SOC 8 workers were small. In Merseyside, Lancaster and parts of Manchester, the number of SOC 1 Managers, directors and senior officials had a positive relationship with the number of workplace accidents occurring.

The final part of the chapter discussed the GWR results for the models which were fitted to the case study regional data, split into the summer and winter seasons. The aim of this was to identify any differences between socio-economic characteristics and workplace accidents based on particular times of the year and geographically. Key results were again highlighted showing variations between seasons. The chapter illustrated that spatial variations exist between workplace accident rates and that the relationships between the determinants vary geographically and by time of the year.

Objective 6. Explore the free-text RIDDOR data, obtained in addition to the large RIDDOR dataset, to identify any key trends and to elaborate the findings from the previous global and local models.

Chapter 7 explored the free-text RIDDOR data through basic text mining techniques and topic modelling methods. The aim of this was to gain an insight into the causes and determinants of workplace accidents, based on an alternative source of data which contained descriptions of all the reported incidents. Through text mining, words with the highest frequencies were extracted from the RIDDOR text and bar

charts and word clouds were created to illustrate the terms which appeared the most often in the text.

Topic modelling was explored next by carrying out a Latent Dirichlet Allocation (LDA) analysis. LDA was performed on the RIDDOR text to identify the core themes within the data. These core themes were extracted for England and Wales, London, North West England and North East England and comparisons were made between them. Additionally, the textual data was further split into seasons to identify whether any core seasonal themes emerged. Although the results were inconclusive, further analyses were undertaken which compared the frequency of key terms that may have indicated issues brought by different weather conditions or daylight levels. Overall, there appeared to be some seasonal differences in workplace accidents between regions. The LDA analysis explored individual experiences of workplace accidents and the key findings from this chapter reflected the results from the previous chapters which looked at quantitative methods of analysing rates of accidents.

Objective 7. Identify the original findings of the research, where the results fit into the wider field of literature on occupational health and safety and make policy recommendations for the HSE.

The results of the study were outlined in Chapters 4, 5, 6 and 7. The findings were then further evaluated in Chapter 8, the Discussion. The Discussion provided a critical review and summary of the previous chapters and illustrated how the key results contributed to the wider field of research surrounding occupational health and safety. The chapter also examined the validity of the results and provided policy recommendations for the HSE. The key findings from this study were that a set of socio-economic variables were found to impact upon health and safety in the workplace, and evidence was shown that these factors vary geographically and between seasons. More specifically, these findings include:

- Older workers are more likely to have a work related accident, whilst younger workers are less likely. This is the case across England and Wales, and for London and the North West specifically. For younger workers, a positive relationship exists in London with workplace accident rate, however is negative in both Northern regions.

- Injury severity increases with age, with more major injuries and fatalities found for older workers than younger workers.

- Men overall reportedly have more workplace accidents compared to women.

- Causes of workplace accidents vary by industry, with examples such as high rates of physical assault in human health and social work, and workers being hit by a moving or falling object in manufacturing and construction.

- Workers in low-skilled jobs (SOC 8 Process Plant and Machine Operatives and SOC 9 Elementary Occupations) have an increased risk of workplace accidents, compared to skilled jobs, or jobs associated with formal training and education (SOC 1 Managers, directors and senior officials). These relationships were identified as varying geographically. Within the selected case study regions, high risk areas in particular include:

  - SOC 1 workers around Merseyside, Lancaster and parts of Manchester (where small percentages of SOC 1 workers are located),
  - SOC 8 workers around parts of South London, and County Durham and Richmondshire in North East England, and
  - SOC 9 workers in Central and North London (including the Camden, Westminster, Barnet and Enfield districts), South Cheshire and West Cumbria, and parts of the East and South of the North East region.

- Socio-economic class is a key determinant of workplace accidents, with positive relationships found between NS-SEC 7 Routine occupations and

workplace accidents in all regions. Particular high-risk areas identified in the North West of England include: Lancashire and South Manchester.

- Accidents occur more frequently in Northern England in winter, and South England in summer, when weather conditions are generally poorer and there are less daylight hours in the working days. High risk areas in particular include: SOC 9 workers in North West Cumbria and West Lancashire in summer, SOC 9 workers in West of Northumbria in winter and SOC 1 workers in Central Manchester in winter.

Objective 8. Review the success of the research project, making recommendations for improvements and possible future research

The Discussion chapter provided a review of the data and methods used as part of this research. The main limitations of the study were identified, which included the lack of availability of workplace population Census data for Scotland by Workplace Zone. This meant that the global regression analyses could only model accidents occurring within England and Wales. Additionally, the textual data obtained as part of this research only covered accidents occurring within a 6-month period, making comparisons between different seasons difficult. The chapter outlined potential steps in taking this research further in the future. These included: obtaining a larger sample of the textual data to explore seasonal trends in greater detail, running the GWR analysis on the whole of Great Britain, should software permit, as well as expanding on occupational health and safety research in general by obtaining data on recorded illnesses and diseases as a result of work. This could add a further dimension to the study in helping the HSE improve health and safety standards.

9.3 Final Remarks

Occupational health and safety have largely been under researched and this is mostly due to there being a lack of access to workplace accident data. In addition, generally, research surrounding occupational safety is mixed in that contrasting evidence has been found on the associations between socio-economic characteristics of individuals and their impacts on workplace accident rates. Previously, also, little work has been carried out surrounding how determinants of workplace accidents vary geographically, and whether physical conditions, such as the weather, impact negatively upon workplace safety.

This thesis is an attempt to provide a greater understanding of the determinants of workplace accidents, delving deeper into where and when they are likely to occur. This is achieved by exploring data generated under RIDDOR, a large database of workplace accident records, to identify the key risk factors associated with workplace accidents. A mixed-methods approach to analysing the RIDDOR records was utilised, using both quantitative techniques to explore the workplace accidents records and qualitative techniques to text mine the textual descriptions of workplace accidents. This study also utilises Workplace Zone boundary data, a new geography generated by the ONS to explore workplace population trends. No other large study to date has utilised WZs in the occupational health and safety field to examine the socio-economic characteristics of workers associated with workplace accidents. This study is therefore novel and original and comprises a highly innovative and revealing programme of research.

The overall aim was to equip the HSE with the knowledge of the key high-risk groups within the workplace population, and this is achieved through setting out clear policy recommendations, which are outlined in Chapter 8. It is hoped that the results will allow a closer focus on these workers so as to further reduce workplace accidents in the future.

Bibliography

Akaike, H. (1973) Information theory as an extension of the maximum likelihood principle, in: Petrov, B. and Csaki, F. (Eds) *2nd Symposium of Information Theory*, Akademiai Kiado: Budapest, pp.267-81

Akerstedt, T. (2003) Shift work and disturbed sleep/wakefulness, *Occupational Medicine,* 53(2), pp.89-94

Alamgir, H. and Yu, S. (2008) Epidemiology of occupational injury among cleaners in the healthcare sector, *Occupational Medicine*, 58(6), pp.393-399

Amin, M.D, Zareie, A. and Amador-Jimenez, L. (2014) Climate change modelling and the weather-related road accidents in Canada, *Transportation Research*, 32(1), pp.171-183

Amoros, E., Martin, J. L. and Laumon, B. (2003) Comparison of road crashes incidence and severity between some French counties, *Accident Analysis and Prevention*, 35(2), pp.537-547

Anderson, D., Burnham, K. and Thompson, W. (2000) Null Hypothesis Testing: Problems, Prevalence, and an Alternative, *Journal of Wildlife Management*, 64(4), pp.912-923

Arcaya, M., Brewster, M., Zigler, C.M. and Subramanian, S.V. (2012) Area variations in health: A spatial multilevel modelling approach, *Health Place*, 18, pp.824–831

Arditi, D., Lee, D. E. and Polat, G. (2007) Fatal accidents in nighttime vs. daytime highway construction work zones, *Journal of Safety Research*, 38(4), pp.399-405

Ashbury, F. D. (1995) Occupational repetitive strain injuries and gender in Ontario, 1986 to 1991, *Journal of Occupational and Environmental Medicine*, 37(4), pp.479-485

Baker, C. (2015) *Accident and Emergency Statistics Briefing Paper*. House of Commons Library, SN06964

Barling, J., Loughlin, C. and Kelloway, E. K. (2002) Development and test of a model linking safety-specific transformational leadership and occupational safety, *Journal of Applied Psychology*, 87, pp.488-496

Barnes, C. M. and Wagner, D. T. (2009) Changing to Daylight Saving Time cuts into sleep and increases workplace injuries, *Journal of Applied Psychology*, 94(5), pp.1305-1317

Beehr, T. A., Glazer, S., Nielson, N. L. and Farmer, S. J. (2000) Work and nonwork predictors of employees' retirement ages, *Journal of Vocational Behaviour*, 57(2), pp.206-225.

Berecki-Gisolf, J., Smith, P. M., Collie, A. and McClure, R. J. (2015) Gender differences in occupational injury incidence, *American Journal of Industrial Medicine*, 58(3), pp.299–307

Bilimoria, P., (2009) "Sick leave" is mainly taken on Mondays and in January, *Occupational Health*, 61(12), pp.4-8

Bivand, R. and Yu, D. (2017) *spgwr: Geographically Weighted Regression*, R package

Björnstig, U., Björnstig, J. and Dahlgren, A. (1997) Slipping on ice and snow- elderly women and young men are typical victims, *Accident; Analysis and Prevention*, 29(2), pp.211-215

Blei, D. M. (2012) Probabilistic topic models, *Communications of the ACM*, 55(4), pp.77–84

Blei, D. M. and Lafferty, J. D. (2009) Topic Models, in: Srivastava, A. and Sahami, M. (Eds), *Text Mining: Classification, Clustering, and Applications*, Chapman & Hall/CRC Data Mining and Knowledge Discovery Series

Blei, D. M., Ng, A. Y. and Jordan, M. I. (2003) Latent Dirichlet Allocation, *Journal of Machine Learning Research*, 3(1), pp.993-1022

Bonnet, M. H. and Arand, D. L. (1995) We are Chronically Sleep Deprived, *American Sleep Disorders Association and Sleep Research Society*, 18(10), pp.908-911

Bortkiewicz, L. von (1898) *The Law of Small Numbers*. Leipzig: G. Teubner

Bridger, R. S. (2003) *Introduction to ergonomics*, 2nd edition. London: Taylor & Francis

Brodsky, H. and Hakkert, A. S. (1988) Risk of a road accident in rainy weather, *Accident Analysis & Prevention*, 20(2), pp.161-176

Brorsson, B. (1989) Age and Injury Severity, *Scand. Journal Public Health,* 17, pp.287–290.

Brown, L.A. and Jones, J.P. (1985) Spatial variation in migration processes and development: a Costa Rican example of conventional modelling augmented by the expansion method, *Demography*, 22, pp.327–352

Brunsdon, C., Fotheringham, S. and Charlton, M. (1998) Geographically Weighted Regression, *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47, pp.431-443

Brunsdon, C. and Singleton, A.D. (2015) Geocomputation : a practical primer. London : Sage, 2015.

Bull, N., Riise, T. and Moen, B.E. (2001) Occupational injuries to fisheries workers in Norway reported to insurance companies from 1991 to 1996, *Occupational Medicine*, 51(5), pp.299-304

Burnham, K. P. and Anderson, D. R. (2002) *Model selection and multimodel inference: a practical information-theoretic approach*, New York: Springer

Butler, R. J., Kleinman, N. and Gardner, H. H. (2014) I Don't like Mondays: Explaining Monday work injury claims, *Industrial & Labour Relations Review*, 67, pp.762-783

Cahill, M. and Mulligan, G. (2007) Using Geographically Weighted Regression to explore local crime patterns, *Social Science Computer Review*, 25, pp.174-193

Cameron, I., Hare, B. and Davies, R. (2008) Fatal and major construction accidents: A comparison between Scotland and the rest of Great Britain, *Safety Science, Construction Safety*, 46, pp.692-708

Case, A. and Paxson, C. (2005) Sex Differences in Morbidity and Mortality, *Demography*, 42, pp.189-214

Casetti, E. (1997) The Expansion Method, Mathematical Modelling, and Spatial Econometrics, *International Regional Science Review*, 20, pp.9–33

Casetti, E. (1972) Generating Models by the Expansion Method: Applications to Geographical Research, *Geographical Analysis*, 4, pp.81–91

Cellier, J.-M., Eyrolle, H. and Bertrand, A. (1995) Effects of age and level of work experience on occurrence of accidents, *Perceptual and Motor Skills*, 80, pp.931-940

Chang, J. (2015) *lda: Collapsed Gibbs Sampling Methods for Topic Models*, R Package

Charlton, M., Fotheringham and A.S., Brunsdon, C. (2016) GWR4 for Windows

Chau, N., Dehaene, D., Benamghar, L., Bourgkard, E., Mur, J. M., Touron, C. and Wild, P. (2014) Roles of age, length of service and job in work-related injury: a prospective study of 63,620 person-years in female workers, *American Journal of Industrial Medicine*, 57, pp.172-183

Cherry, N., Parker, G., McNamee, R., Wall, S., Chen, Y., Robinson, J. (2005) Falls and fractures in women at work, *Occupational Medicine*, 55, pp.292–297

Chi, C.F., Chang, T.-C. and Hung, K.-H. (2004) Significant industry–source of injury–accident type for occupational fatalities in Taiwan, *International Journal of Industrial Ergonomics*, 34, pp.77–91

Chi, C. F., Chang, T. C. and Ting, H. I. (2005) Accident patterns and prevention measures for fatal occupational falls in the construction industry, *Applied Ergonomics*, 36, pp.391-400

Chi, C. F. and Wu, M. L. (1997) Fatal occupational injuries in Taiwan- Relationship between fatality rate and age, *Safety Science*, 27, pp.1-17

Christensen, E. H., (1955) Physical working capacity of old workers and physiological background for work tests and work evaluations, *Bulletin World Health Organization*, 13, pp.587-593

Clark, R. E. (1961) The limiting hand skin temperature for unaffected manual performance in the cold, *Journal of Applied Psychology*, 45, pp.193-194

Connors, J.P., Lei, S. and Kelly, M. (2012) Citizen Science in the Age of Neogeography: Utilizing Volunteered Geographic Information for Environmental Monitoring, *Annals Association American Geography*, 102, pp.1267–1289

Consul, P. C (1989) *Generalized Poisson Distributions: Properties and Applications*, Volume 99 of Statistics: Textbooks and Monographs, New York: Marcel Dekker Inc.

Courville, J., Vézina, N. and Messing, K. (1991) Comparison of the work activity of two mechanics: A woman and a man, *International Journal of Industrial Ergonomics*, 7, pp.163–174

Creed, P. A., Macintyre, S. R. (2001) The relative effects of deprivation of the latent and manifest benefits of employment on the well-being of unemployed people, *Journal of Occupational Health Psychology*, 6(4), pp.324-331

Cubbin, C. and Smith, G. S. (2002) Socioeconomic inequalities in injury: critical issues in design and analysis, *Annual Review of Public Health,* 23(1), pp.349-375

Davidson, R., Kitzinger, J. and Hunt, K. (2006) The wealthy get healthy, the poor get poorly? Lay perceptions of health inequalities, *Social Science and Medicine*, 62, pp.2171-2182

Davies, R. and Elias, P. (2000) *An Analysis of Temporal and National Variations in Reported Workplace Injury Rates*, Institute for Employment Research

DCLG (2010) *Fire and Rescue Service Operational Statistics Bulletin for England 2009-10*, DCLG

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K. and Harshman, R. (1990) Indexing by latent semantic analysis, *Journal of the American Society of Information Science*, 41, pp.391-407

Defence Statistics (2014) *UK Armed Forces Annual Personnel Report*, National Statistics

DeLeire, T. and Levy, H. (2001) *Gender, Occupation Choice and the Risk of Death at Work (Working Paper No. 8574)*, National Bureau of Economic Research

Dembe, A. E., Erickson, J. B., Delbos, R. G. and Banks, S. M. (2006) Nonstandard shift schedules and the risk of job-related injuries, *Scandinavian Journal of Work, Environment & Health*, 32(3), pp.232-240

Derksen, S. and Keselman, H.J. (1992) Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables, *British Journal Mathematical Statistical Psychology*, 45, pp.265–282

Diez-Roux, A. V. (1998) Bringing context back into epidemiology: Variables and fallacies in multilevel analysis, *American Journal of Public Health*, 88, pp.216-222

Dijk, D. J., Duffy, J. F. and Czeisler, C. A. (1992) Circadian and sleep/wake dependent aspects of subjective alertness and cognitive performance, *Journal of Sleep Research*, 1(2), pp.112-117

Doran, T. (2004) Is there a north-south divide in social class inequalities in health in Great Britain? Cross sectional study using data from the 2001 census, *BMJ*, 328, pp.1043-1045

Dowswell, T. and Towner, E. (2002) Social deprivation and the prevention of unintentional injury in childhood: a systematic review. *Health Education Research* 17(2), pp. 221-237

Dredze, M., Paul, M.J., Bergsma, S. and Tran, H. (2013) Carmen: A Twitter Geolocation System with Applications to Public Health. *AAAI Workshop Expand. Boundaries Health Inform. Using AI HIAI*, pp.20–24.

Drever, F. (1997) *Health Inequalities: Decennial Supplement*, Stationery Office Books

Edwards, J. B. (1996) Weather-related road accidents in England and Wales: a spatial analysis, *Journal of Transport and Geography*, 4(3), pp.201-212

ESRI, 2013. ESRI (Environmental Systems Resource Institute) (2013). *ArcGIS Resource Center*, Version 10.2. ESRI, Redlands, California. URL: http://www.arcgis.com/.

Estabrooks, P. A., Lee, R. E. and Gyurcsik, N. C. (2003) Resources for Physical Activity Participation: Does Availability and Accessibility Differ by Neighborhood Socioeconomic Status? *Annals of Behavioral Medicine*, 25(2), pp.100-109

Eurostat (2012) *ESAW*, URL: http://ec.europa.eu/eurostat/web/health/health-safety-work (last accessed 8th June 2016)

Feinerer, I. and Hornik, K. (2015) Text Mining Infrastructure in R, *Journal of Statistical Software*, URL https://www.jstatsoft.org/article/view/v025i05 (last accessed 18th August 2016).

Fellows, I. (2014) *wordcloud: Word Clouds*, R package.

Feyer, A. M. and Williamson, A. (1998) *Occupational injury, risk, prevention, and intervention*, London: Taylor & Francis

Findikyan, N. and Sells, S. B. (1965) *Cold stress: parameters, effects, mitigation*, PN

Fingerhut, M., Driscoll, T., Imel Nelson, D., Concha-Barrientos, M., Punnett, L., Pruss-Ustin, A., Steenland, K., Leigh, J., Corvalan, C. and (2005) Contribution of occupational risk factors to the global burden of disease, *SJWEH Supplements*, 1, pp.58-61

Fisher, D. L., Pollatsek, A. P. and Pradhan, A. (2006) Can novice drivers be trained to scan for information that will reduce their likelihood of a crash? *Injury Prevention*, 12, pp.25-29

Fletcher-Lartey, S. M. and Caprarelli, G. (2016) Application of GIS technology in public health: successes and challenges, *Parasitology*, 143, pp.401-415

Folkard, S. (1997) Black times: Temporal determinants of transport safety, *Accident Analysis and Prevention*, 29, pp.417-430

Folkard, S., Tucker, P. (2003) Shift work, safety and productivity, *Occupational Medicine*, 53(2), pp.95-101

Fotheringham, A.S, Brunsdon, C. and Charlton, M. (2000) *Quantitative geography : perspectives on spatial data analysis*, SAGE

Fotheringham, A. S., Brunsdon, C. and Charlton, M. (2002) *Geographically weighted regression: the analysis of spatially varying relationships*, Chichester: Wiley

Fotheringham, A. S., Charlton, M. E. and Brunsdon, C. (2001) Spatial Variations in School Performance: A Local Analysis Using Geographically Weighted Regression, *Geographical and Environmental Modelling*, 5(1), pp.43-66

Fotheringham, A. S and Pitts, T. C (1995) Directional variation in distance decay, *Environment and Planning*, 27(5), pp.715-729

French, E. and Strachan, G. (2015) Women at work! Evaluating equal employment policies and outcomes in construction, *Equality, Diversity and Inclusion*, 34(3), pp.227-243

Friendly, M. (2000) *Visualizing Categorical Data*, SAS Institute

Frone, M.R. (1998) Predictors of work injuries among employed adolescents, *Journal of Applied Psychology*, 83(4), pp.565-576

Frost, P., Kolstad, H. and Bonde, J. (2009) Shift work and the risk of ischemic heart disease – a systematic review of the epidemiologic evidence, *Scandinavian Journal of Work, Environment & Health,* 35(3), pp.163-179

Gagnon, M. and Lortie, M. (1987) A biomechanical approach to low-back problems in nursing aides, *Trends in Ergonomics/Human Factors IV,* Holland: Elsevier Science Publishers, pp.795-802

Gao, C. and Abeysekera, J. (2004) A systems perspective of slip and fall accidents on icy and snowy surfaces, *Ergonomics*, 47, pp.573-598

Gelfand, A. E, Kim, H-J, Sirmans C.F. and Banerjee, S. (2003) *Spatial Modeling with Spatially Varying Coefficient Processes,* Journal of the American Statistical Association, 98(462), pp.387-396

Gilbert, A. and Chakraborty, J. (2011) Using geographically weighted regression for environmental justice analysis: Cumulative cancer risks from air toxics in Florida, *Social Science Research*, 40, pp.273-286

Gillen, M., Yen, I., Trupin, L., Swig, L., Rugulies, R., Mullen, K., Font, A., Burian, D., Ryan, G., Janowitz, I., Quinlan, P., Frank, J. and Blanc, P. (2007) The association of socioeconomic status and psychosocial and physical workplace factors with musculoskeletal injury in hospital workers, *American Journal of Industrial Medicine*, 50(4), pp.245-260

Girardin, F., Calabrese, F., Fiore, F.., Ratti, C. and Blat, J. (2008) Digital footprinting: Uncovering tourists with user-generated content, *Pervasive Comput*ing, IEEE 7, pp.36–43

Gollini, I., Lu, B., Charlton, M., Brunsdon, C. and Harris, P. (2015) *GWmodel*: an R Package for Exploring Spatial Heterogenity using Geographically Weighted Models

Goodchild, M. (2007) Citizens as sensors: the world of volunteered geography, *GeoJournal*, 69, pp.211–221

Graham, H. (1985) Women, Health and the Family, *Feminist Review*, 20, pp.107-109

Grazier, S. and Sloane, P. J. (2008) Accident risk, gender, family status and occupational choice in the UK, *Labour Economics*, 15(5), pp.938-957

Grobler, C. (2013) Self-reported work-related musculoskeletal injuries and isometric handgrip strength, *Occupational Medicine*, 63(3), pp.210-216

Grun, B. and Hornik, K. (2011) *topicmodels: An R Package for Fitting Topic Models*, R package

Hale, A. R. and Glendon, A. I. (1987) Individual behaviour in the control of danger, *Industrial safety series: 2*, Amsterdam: Elsevier

Harris, C. R., Jenkins, M. and Glaser, D. (2006) Gender Differences in Risk Assessment: Why Do Women Take Fewer Risks Than Men? *Judgment and Decision Making*, 1(1), pp.48-63

Hasselberg, M. and Laflamme, L. (2009) How do car crashes happen among young drivers aged 18-20 years? Typical circumstances in relation to license status, alcohol impairment and injury consequences, *Accident Analysis Prevention,* 41(4), pp.734-738

Health and Safety at Work etc. Act 1974, c37, URL: http://www.legislation.gov.uk/ukpga/1974/37/contents (last accessed: 6th June 2015)

Heinrich, H. W. (1931) *Industrial Prevention: A Scientific Approach*, McGraw-Hill

Herd, D. R., Agent, K. R. and Baker, R. (1980) Traffic accidents: Day versus night. In Traffic Accident Analysis and Application of System Safety, *Transportation Research Record*, 753, pp.25-30

Hertel, A. W. and Mermelstein, R. J. (2016) Smoker Identity Development Among Adolescents Who Smoke, *Psychology of Addictive Behaviors,* 30(4), pp. 475-483

Hilton, M. and Whiteford, H. (2010) Associations between psychological distress, workplace accidents, workplace failures and workplace successes, *International Archives of Occupational & Environmental Health*, 83(8), pp.923-933

Hoffman, T. (1999) Probabilistic latent semantic indexing. Proc. 22nd Annu. Int. SIGIR Conf.

Holland, N. and Hinze, J. (2000) Daylight Savings Time Changes and Construction Accidents, *Journal of Construction Engineering and Management*, 126(5), pp.404-406

Hollenstein, L., Purves, R. (2010) Exploring place through user-generated content: Using Flickr tags to describe city cores, *Journal Spatial Information Science*, pp.21–48

HSE (2016) *The health and safety toolbox: How to control risks at work*, URL: http://www.hse.gov.uk/toolbox/index.htm (last accessed 6th July 2016)

HSE (2015a) *Health and safety statistics*, URL: http://www.hse.gov.uk/statistics/ index.htm (last accessed 24th April 2016)

HSE (2015b) *Business Plan 2015-16*, URL: http://www.hse.gov.uk/aboutus/strategiesandplans/ businessplans/plan1516.pdf (last accessed 9th August 2016)

HSE (2015c) *Summary Science Plan 2015-2018*, URL: http://www.hse.gov.uk/ research/content/ science-plan-2015-18.pdf (last accessed 2nd May 2015)

HSE (2015d) *Human factors/ergonomics- Introduction to human factors*, URL: http://www.hse.gov.uk/humanfactors/introduction.htm (last accessed 10th May 2016)

HSE (2015e) *Health and Safety Statistics: Annual Report for Great Britain 2014/15,* URL: http://www.hse.gov.uk/statistics/overall/hssh1415.pdf (last accessed 30th April 2016)

HSE (2015f) *Work-related Musculoskeletal Disorder (WRMSDs) Statistics*, URL: http://www.hse.gov.uk/statistics/causdis/musculoskeletal/msd.pdf (last accessed 19th May 2016)

HSE (2015g) *Health and safety in Great Britain 2015 Work-related ill health and injury statistics*, URL: http://www.hse.gov.uk/statistics/regions/country-and-region.pdf (last accessed 7th June 2016)

HSE (2015h) *Historical Picture- Trends in work-related injuries and ill health in Great Britain since the introduction of the Health and Safety at Work act 1974*, URL: http://www.hse.gov.uk/STATISTICS/history/historical-picture.pdf (last accessed 24th April 2016)

HSE (2012a) *HSE: Timeline*, URL http://www.hse.gov.uk/aboutus/timeline/ (last accessed 3rd July 2013)

HSE (2012b) *A guide to the Reporting of Injuries, Diseases and Dangerous Occurrences Regulations 1995*, (4th Ed), HSE

Hsieh, C. C. and Pugh, M. D. (1993) Poverty, Income Inequality, and Violent Crime: A Meta-Analysis of Recent Aggregate Data Studies, *Criminal Justice Review*, 18(2), pp.182-202

Hurvich, C.M. and Tsai, C.-L. (1990) The Impact of Model Selection on Inference in Linear Regression, *American Statistician*, 44, pp.214–217

Huyghebaert, B., Lorencowicz, E., Uziak, J., Pawlak, H. and Nowakowicz-Debek, B. (2015) Farm Machinery and Processes Management in Sustainable Agriculture, *Agriculture and Agricultural Science Procedia*, 7, pp.209-214

Idler, E. L. and Benyamini, Y. (1997) Self-Rated Health and Mortality: A Review of Twenty-Seven Community Studies*, Journal of Health and Social Behavior*, 38(1), pp.21-37

Istre, G. R., McCoy, M. A., Osborn, L., Barnard, J. J. and Bolton, A. (2001) Deaths and Injuries from House Fires, *New England Journal of Medicine*, 344(25), pp.1911-1916

Jackson, L. L. (2001) Non-fatal occupational injuries and illnesses treated in hospital emergency departments in the United States, *Injury Prevention: Journal Of The International Society For Child And Adolescent Injury Prevention*, 7 Suppl 1, pp.i21-i26

Joe, H. and Zhu, R. (2005) Generalized Poisson distribution: the property of mixture of Poisson and comparison with negative binomial distribution, *Biometrical Journal*, 47(2), pp.219-229

Johnson, C.J., Seip, D.R. and Boyce, M.S. (2004) A quantitative approach to conservation planning: using resource selection functions to map the distribution of mountain caribou at multiple spatial scales, *Journal Applied Ecology*, 41, pp.238–251

Jones, J.P., Casetti, E. (1992) *Applications of the Expansion Method*. Routledge.

Kamdar, B. B., Tergas, A. I., Mateen, F. J., Bhayani, N. H. and Oh, J. (2013) Night-shift work and risk of breast cancer: a systematic review and meta-analysis. *Breast Cancer Res. Treat*, 138(1), pp.291-301

Kennedy, P. (2003) *A Guide to Econometrics*, MIT Press.

Kjellstrom, T., Holmer, I. and Lemke, B. (2009) Workplace heat stress, health and productivity- an increasing challenge for low and middle-income countries during climate change, *Global Health Action*, 2, pp. 1-6

Knestaut, A. (1996) Fewer women than men die of work- related injuries, data show, *Compensation and Working Conditions*, 6, pp. 45-47

Konecni, V., Ebbeson, E.B. and Konecni, D.K. (1976) Decision processes and risk taking in traffic: Driver response to the onset of yellow light, *Journal of Applied Psychology*, 61(3), pp.359-367

Kong, P. W., Suyama, J. and Hostler, D. (2013) A review of risk factors of accidental slips, trips, and falls among firefighters, *Safety Science*, 60, pp.203-209

Krieger, N., Chen, J., Waterman, P., Soobader, M., Subramanian, S. and Carson, R. (2002) Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: Does the choice of area-based measure and geographic level matter? *American Journal of Epidemiology*, 156(5), pp.471-482

Kunst, A. E., Groenhof, F. and Mackenbach, J. P. (1998) Occupational Class and Cause Specific Mortality in Middle Aged Men in 11 European Countries: Comparison of Population Based Studies, *British Medical Journal*, 1636(7145), pp.1636-1641

Laberge, M., Calvet, B., Fredette, M., Tabet, N., Tondoux, A., Bayard, D. and Breslin, C. (2016) Unexpected events: Learning opportunities or injury risks for apprentices in low-skilled jobs? *Safety Science*, 86, pp.1-9

Lambert, D. (1992) Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing. *Technometrics*, 34(1), pp.1-14

Leckie, G. and Goldstein, H. (2015) A multilevel modelling approach to measuring changing patterns of ethnic composition and segregation among London secondary schools, 2001–2010. *J. R. Statistics in Society,* 178, pp.405–424

Leclerc, A., Lert, F. and Fabien, C. (1990) Differential mortality: some comparisons between England and Wales, Finland and France, based on inequality measures, *International Journal of Epidemiology*, 19(4), pp.1001-1010

Legree, P.J., Heffner, T., Psotka, J., Martin, D. and Medsker, G. (2003) Traffic Crash Involvement: Experiential Driving Knowledge and Stressful Contextual Antecedents, *Journal of Applied Psychology*, 88, pp.15-28

Leventhal, T. and Brooks-Gunn, J. (2000) The Neighborhoods They Live in: The Effects of Neighborhood Residence on Child and Adolescent Outcomes, *Psychological Bulletin*, 126(2), pp.309-337

Levine, J. A. and Miller, J. M. (2007) The energy expenditure of using a "walk-and-work" desk for office workers with obesity, *British Journal of Sports Med*, 41(9), pp.558-561

Ljungberg, A. S., Kilbom, Å. and Hagg, G. M. (1989) Occupational lifting by nursing aides and warehouse workers, *Ergonomics* 32(1), pp.59-78

Lloyd, S., Lawton, J., Caraher, M., Singh, G., Horsley, K. and Mussa, F. (2011) A tale of two localities: Healthy eating on a restricted income, *Health Education Journal*, 70(1), pp.48-56

Lord, D., Washington, S. P. and Ivan, J. N. (2005) Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory, *Accident Analysis & Prevention*, 37(1), pp.35-46

Lu, B., Charlton, M. and Fotheringham, A. S. (2011) Geographically Weighted Regression Using a Non-Euclidean Distance Metric with a Study on London House Price Data, *Procedia Environmental Sciences*, 7, pp.92-97

Lu, B., Harris, P., Charlton, M., Brunsdon, C., Nakaya, T. and Gollini, I. (2017) *GWmodel*: Geographically-Weighted Models

Ma, L., Yan, X., Qiao, W., Ma, L., Yan, X. and Qiao, W. (2014) A Quasi-Poisson Approach on Modeling Accident Hazard Index for Urban Road Segments, *Discrete Dynamics in Nature and Society*, 2014(3), pp.1-8

Macquet, P., Degueldre, C., Delfiore, G., Aerts, J., Peters, J. M., Luxen, A. and Franck, G. (1997) Functional Neuroanatomy of Human Slow Wave, *Journal of Neuroscience*, 17(8), pp.2807-2812

Marmot, M. G., Rose, G., Shipley, M. and Hamilton, P. J. (1978) Employment grade and coronary heart disease in British civil servants, *Journal of Epidemiology and Community Health*, 32(4), pp.244-249

Marmot, M., Ryff, C. D., Bumpass, L. L., Shipley, M. and Marks, N. F. (1997) Social inequalities in health: Next questions and converging evidence, *Social Science & Medicine*, 44(6), pp.901-910

Mayes, J. and Wheeler, D. (2013) Regional weather and climates of the British Isles- Part 1: Introduction, *Weather*, 68(1), pp.3-8

McCaig, L. F, Burt, C. W. and Stussman, B. J. (1998) A comparison of work-related injury visits and other injury visits to emergency departments in the United States, 1995-1996, *Journal of Occupational and Environmental Medicine*, 40(10), pp.870-875

Met Office (2015) *The Heatwave of 2003*, URL: http://www.metoffice.gov.uk/ learning/learn-about-the-weather/weather-phenomena/case-studies/heatwave (last accessed 17th May 2016)

Met Office (2013a) *January 2009*,URL: http://www.metoffice.gov.uk/climate/uk/summaries/2009 /january (last accessed 2nd April 2017)

Met Office (2013b) *February 2009*, URL: http://www.metoffice.gov.uk/climate/uk/summaries /2009/february (last accessed 2nd April 2017)

Met Office, (2013c) *January 2010*,URL: http://www.metoffice.gov.uk/climate/uk/summaries /2010/january (last accessed 2nd April 2017).

Met Office, (2013d) *December 2010*, URL: http://www.metoffice.gov.uk/climate/uk/summaries /2010/december (last accessed 2nd April 2017)

Met Office (2012) *UK regional climates*, URL: http://www.metoffice.gov.uk/ climate/uk/regional-climates (last accessed 5th August 2016)

Michael, K. and Ben-Zur, H. (2007) Risk-taking among adolescents: Associations with social and affective factors, *Journal of Adolescence*, 30(1), pp.17-31

Ministry of Defence (2016) *Climatic illness and injury in the Armed Forces: Force protection and initial medical treatment*, JSP 539 v2.5

Möller, H., Haigh, F., Harwood, C., Kinsella, T. and Pope, D. (2013) Rising unemployment and increasing spatial health inequalities in England: further extension of the North-South divide, *Journal of public health*, 35(2), pp.313-321

Morabito, M., Cecchi, L., Crisci, A., Modesti, P. A. and Orlandini, S. (2006) Relationship between work-related accidents and hot weather conditions in Tuscany (central Italy), *Industrial health*, 44(3), pp.458-464

Morassaei, S., Breslin, F. C., Ibrahim, S. A., Smith, P. M., Mustard, C. A., Amick, B. C., Shankardass, K. and Petch, J. (2013) Geographic variation in work injuries: a multilevel analysis of individual-level data and area-level factors within Canada, *Annals of Epidemiology,* 23(5), pp.260-266

Morassaei, S., Smith, P. M. (2010) Switching to Daylight Saving Time and work injuries in Ontario, Canada: 1993–2007, *Occupational and Environmental Medicine*, 67(12), pp.878-880

Muller, A. (2002) Education, income inequality, and mortality: a multiple regression analysis, *BMJ*, 324(7328), pp.23-25

Mustard, C. A., Chambers, A., McLeod, C., Bielecky, A. and Smith, P. M. (2013) Work injury risk by time of day in two population-based data sources, *Occupational and Environmental Medicine*, 70(1), pp.49-56

Nakaya, T., Charlton, M., Brunsdon, C., Lewis, P., Yao, J. and Fotheringham, A.S. (2016) GWR4.09 User Manual

National Geographic Society (2011) *Season,* URL: http://nationalgeographic.org /encyclopedia/ season/ (last accessed 5th August 2016)

Neff, R. A., Curriero, F. C. and Burke, T. A. (2008) Just in the wrong place…?: Geographic tools for occupational injury/illness surveillance, *American Journal of Industrial Medicine*, 51(9), pp.680-690

Nelder, J. A. and Wedderburn, R. W. M., (1972) Generalized Linear Models, *Journal of the Royal Statistical Society. Series A*, 135(3), pp.370-384

Nenonen, N. (2011) Occupational accidents in the Finnish local government sector: Utilisation of national statistics, *International Journal of Injury Control and Safety Promotion*, 18(4), pp.321-329

Neter, J., Wasserman, W. and Kutner, M.H. (1989) Applied linear regression models. Irwin.

Newman, D., Chemudugunta, C., Smyth, P. and Steyvers, M. (2006) Statistical Entity-Topic Models, *KDD '06 Proceedings of the 12th ACM SIGKDD International conference on Knowledge discovery and data mining*, pp.680–686

ONS (2015*) Statistical bulletin: National Life Tables, United Kingdom: 2012-2014*, URL: http://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies /bulletins/nationallifetablesunitedkingdom/2015-09-23 (last accessed 14th July 2016)

ONS (2011) *National Population Projections, 2010-Based Statistical Bulletin*, URL http://www.ons.gov.uk/ons/rel/npp/national-population-projections/2010-based-projections/stb-2010-based-npp-principal-and-key-variants.html (last accessed 13th June 2016)

ONS (2010) *SOC2010 volume 3: the National Statistics Socio-economic classification (NS-SEC rebased on SOC2010)* URL: http://www.ons.gov.uk/ons/guide-method/classifications/current-standard-classifications/soc2010/soc2010-volume-3-ns-sec--rebased-on-soc2010--user-manual/index.html (last accessed 30th May 2016)

Owens, D. A. (1993) The role of reduced visibility in night-time road fatalities, *Transportation Research Institute*, Report No: UMTRI-93-33

Packer, G. J. and Shaheen, M. A. (1993) Patterns of hand fractures and dislocations in a district general hospital, *The Journal of Hand Surgery: British & European Volume*, 18(4), pp.511-514

Park, S., Yang, D., Shin, J., Park, S., Oh, H., Choi, E., Park, Y., Park, B., Park, M. and Yang, Y. (2012) Workplace Accident Prevention and Improvement of Work Ability in an Aging Society, *Journal of Physical Therapy Science*, 24(1), pp.143-148

Peek-Asa, C., Erickson, R. and Kraus, J.F. (1999) Traumatic occupational fatalities in the retail industry, United States 1992–1996, *American Journal of Industrial Medicine*, 35(2), pp.186-191

Perger, C., Fritz, S., See, L., Schill, C., van der Velde, M., McCallum, I. and Obersteiner, M. (2012) A campaign to collect volunteered geographic information and land cover and human impact, in: Jekel, T., Car, A., Strobl, J., Griesebner, G. (Eds.), *Presented at the GI_Forum 2012: Geovizualisation, Society and Learning*

Perrenoud, P. (1999) Gestion de l'imprévu, analyse de l'action et construction de competences, *Éducation Permanente*, 140(3), pp.123-144

Pickett, K. E., Kelly, S., Brunner, E., Lobstein, T. and Wilkinson, R. G. (2005) Wider Income Gaps, Wider Waistbands? An Ecological Study of Obesity and Income Inequality, *Journal of Epidemiology and Community Health*, 59(8), pp.670-674

Pickett, K. E. and Pearl, M. (2001) Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review, *Journal of Epidemiology and Community Health*, 55(2), pp.111–122

Polus, A. and Cohen, M. (2012) A new, non-canonical Poisson regression model for the prediction of crashes on low-volume rural roads, *IATSS Research*, 35:2 pp. 98-103

Pouliakas, K. and Theodossiou, I. (2010) An Inquiry into the Theory, Causes and Consequences of Monitoring Indicators of Health and Safety at Work (IZA Discussion Paper No. 4734), *Institute for the Study of Labor (IZA)*

Prior, J. and Kendon, M. (2011) The UK winter of 2009/2010 compared with severe winters of the last 100 years, *Weather* 66(1), pp.4-10

Probst, T. M. (2002) Layoffs and tradeoffs: Production, quality, and safety demands under the threat of job loss, *Journal of Occupational Health Psychology*, 7(3), pp.211-220

Quisumbing, A. R., Meinzen-Dick, R., Raney, T.L., Croppenstedt, A., Behrman, J.A. and Peterman, A. (2014) *Gender in agriculture : closing the knowledge gap*, Dordrecht: Springer

R Core Team (2015) *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria

Ramsey, J.D., Burford, C.L., Beshir, M.Y. and Jensen, R.C. (1983) Effects of workplace thermal conditions on safe work behaviour, *Journal of Safety Research*, 14, pp. 105–114

Reading, R., Langford, I. H., Haynes, R. and Lovett, A. (1999) Accidents to preschool children: comparing family and neighbourhood risk factors, *Social Science & Medicine*, 48(3), pp.321-330

Ready, A. E., Boreskie, S. L., Law, S. A. and Russell, R. (1993) Fitness and Lifestyle Parameters Fail to Predict Back Injuries in Nurses, *Canadian Journal of Applied Physiology*, 18(1), pp.80-90

Robinson, A. M. and Smallman, C. (2006) The contemporary British workplace: a safer and healthier place? *Work, Employment & Society,* 20(1), pp.87-107

Rogot, E., Sorlie, P. D. and Johnson, N. J. (1992) Life Expectancy by Employment Status, Income, and Education in the National Longitudinal Mortality Study, *Public Health Reports*, 107(4), pp.457-461

Root, N. (1981) Injuries at work are fewer among older employees, *Monthly Labor Review*, 104(3), pp.30

Rose, N. L. (1990) Profitability and Product Quality: Economic Determinants of Airline Safety Performance, *Journal of Political Economy*, 98(5), pp.944-964

Ruser, J. W. (1998) Denominator choice in the calculation of workplace fatality rates, *American Journal of Industrial Medicine*, 33(2), pp.151-156

Salminen, S. (2004) Have young workers more injuries than older ones? An international literature review, *Journal of Safety Research*, 35(5), pp.513-521

Salminen, S.T. (1994) Epidemiological Analysis of Serious Occupational Accidents in Southern Finland, *Scandinavian Journal of Public Health*, 22(3), pp.225-227

Saper, C. B., Scammell, T. E. and Lu, J. (2005) Hypothalamic regulation of sleep and circadian rhythms, *Nature*, 437(7063), pp.1257-1263

Schuring, M., Robroek, S. J. W., Lingsma, H. F. and Burdorf, A. (2015) Educational differences in trajectories of self-rated health before, during, and after entering or leaving paid employment in the European workforce, *Scandinavian Journal of Work, Environment & Health,* 41(5), pp.441-450

Shapiro, R., Siegel, A. W., Scovill, L. C. and Hays, J. (1998) Risk-taking patterns of female adolescents: what they do and why, *Journal of Adolescence*, 21(2), pp.143-159

Sherman, G. D., Lee, J. J., Cuddy, A. J. C., Renshon, J., Oveis, C., Gross, J. J. and Lerner, J. S. (2012) Leadership is associated with lower levels of stress, *Proceedings of the National Academy of Sciences,* 109(44), pp.17903-17907

Smith, M. E., McEvoy, L. K. and Gevins, A. (2002) The Impact of Moderate Sleep Loss on Neurophysiologic Signals during Working-Memory Task Performance, *Sleep*, 25(7), pp.784-794

Smith, P. M. and Mustard, C. A. (2004) Examining the associations between physical work demands and work injury rates between men and women in Ontario, 1990–2000, *Occupational and Environmental Medicine*, 61(9), pp.750-756

Sorock, G. S., Smith, E. O. and Goldoft, M. (1993) Fatal occupational injuries in the New Jersey construction industry, 1983 to 1989, *Journal of Occupational Medicine*, 35(9), pp.916-921

Theodossiou, I. (1998) The effects of low-pay and unemployment on psychological well-being: A logistic regression approach, *Journal of Health Economics*, 17(1), pp.85-104

Toch, M., Bambra, C., Lunau, T., Wel, K. A. van der, Witvliet, M. I., Dragano, N. and Eikemo, T.A. (2014) All Part of the Job? The Contribution of the Psychosocial and Physical Work Environment to Health Inequalities in Europe and the European Health Divide, *International Journal of Health Services*, 44(2), pp.285-305

Townsend, P., Phillimore, P. and Beattie, A. (1988) *Health and deprivation: inequality and the North*, Croom Helm, London; New York

Tunwattanapong, P., Kongkasuwan, R. and Kuptniratsaikul, V. (2016) The effectiveness of a neck and shoulder stretching exercise program among office workers with neck pain: a randomized controlled trial, *Clinical Rehabilitation*, 30(1), pp.64-72

Urberg, K. A., Degirmencioglu, S. M. and Pilgrim, C. (1997) Close friend and group influence on adolescent cigarette smoking and alcohol use, *Developmental Psychology*, 33(5), pp.834-844

Usman, T., Fu, L. and Miranda-Moreno, L. F. (2012) A disaggregate model for quantifying the safety effects of winter road maintenance activities at an operational level, *Accident; Analysis And Prevention*, 48, pp.368-378

Vaalavuo, M. (2016) Deterioration in health: What is the role of unemployment and poverty? Scand. J. Public Health, 44(4), pp.347-353

Vahtera, J., Kivimäki, M. and Pentti, J. (2001) The role of extended weekends in sickness absenteeism, *Occupational and Environmental Medicine*, 58(12), pp.818-822

Valentin, V., Mannering, F., Abraham, D. and Dunston, P. (2010) Evaluation of the Visibility of Workers' Safety Garments during Night-time Highway-Maintenance Operations, *Journal of Transportation Engineering*, 136(6), pp.584-591

Van Exel, N. J. A., de Graaf, G. and Brouwer, W. B. F. (2006) "Everyone dies, so you might as well have fun!" Attitudes of Dutch youths about their health lifestyle, *Social Science & Medicine*, 63(10), pp.2628-2639

Varughese, J. and Allen, R. P. (2001) Fatal accidents following changes in daylight savings time: the American experience, *Sleep Medicine*, 2(1), pp.31-36
Venables, W. N. and Ripley, B. D. (2002) *Modern Applied Statistics with S: Fourth Edition*, Springer: New York

Verbrugge, L. M. (1989) The Twain Meet: Empirical Explanations of Sex Differences in Health and Mortality, *Journal of Health and Social Behavior*, 30(3), pp.282-304

Ver Hoef, J. M. and Boveng, P. L. (2007) Quasi-poisson vs. negative binomial regression: how should we model overdispersed count data? *Ecology*, 88(11), pp.2766-2772

Waldron, I., McCloskey, C. and Earle, I. (2005) Trends in gender differences in accidents mortality: Relationships to changing gender roles and other societal trends, *Demographic Research*, 4(17), pp.415-454

Wedderburn, R. W. M. (1974) Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method, *Biometrika* 61(3), pp.439-447

Weeden, K. A. (1998) Revisiting occupational sex segregation in the United States, 1910–1990: Results from a log-linear approach, *Demography*, 35(4), pp. 475-487

Wei, X. and Croft, W. B. (2006) LDA-Based Document Models for Ad-hoc Retrieval, *Proceedings of the 29th Annual International ACM SIGIR Conference on Research & Development on Information Retrieval (SIGIR 2006)*, pp. 178-185

Wells, T. (1999) Changes in Occupational Sex Segregation during the 1980s and 1990s, *Social Science Quarterly*, 80(2), pp.370-380

Wilkins, K. and Mackenzie, S.G. (2007) Work injuries, *Health Report Statistics Canada*, 18(3), pp.25-42

Wilkinson, P., Pattenden, S., Armstrong, B., Fletcher, A., Kovats, R. S., Mangtani, P. and McMichael, A. J. (2004) Vulnerability to winter mortality in elderly people in Britain: population based study, *BMJ*, 329(7467), pp.647

World Health Organization (2012) *Falls. Fact sheet N°344*, URL: http://www.wmo.int/pages/mediacentre/press_releases/gcs_2011_en.html (last accessed 4th June 2016)

Xiang, H., Wang, Z., Stallones, L., Keefe, T. J., Huang, X. and Fu, X. (2000) Agricultural Work-Related Injuries Among Farmers in Hubei, People's Republic of China, *American Journal of Public Health*, 90(8), pp.1269-1276

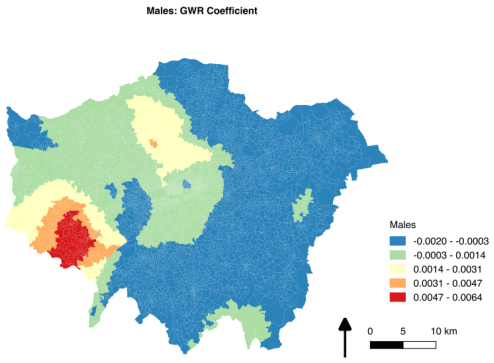
Yerushalmi, E. and Polingher, C. (2006) Guiding Students to Learn from Mistakes, *Physics Education*, 41(6), pp.532-538

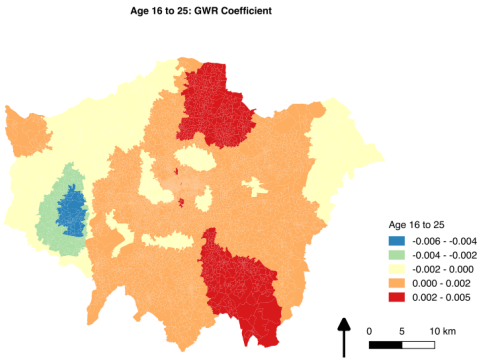
Zweig, J. M., Lindberg, L. D. and McGinley, K. A. (2001) Adolescent Health Risk Profiles: The Co-Occurrence of Health Risks among Females and Males, *Journal of Youth and Adolescence*, 30(6), pp.707-728

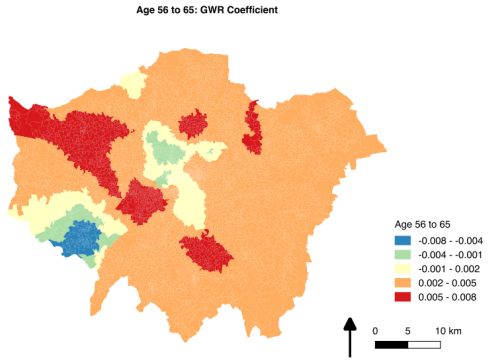
Appendix

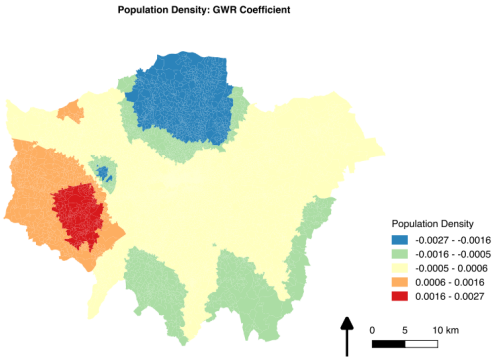
Additional GWR Maps

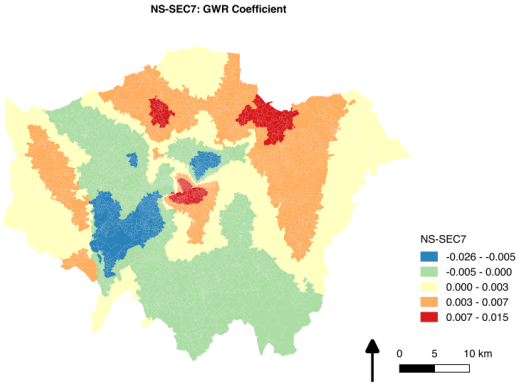
1. London



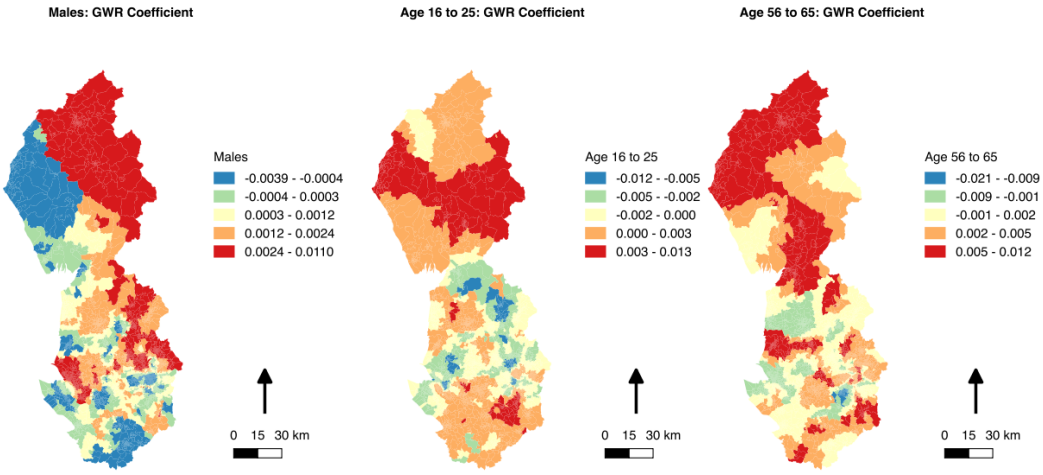


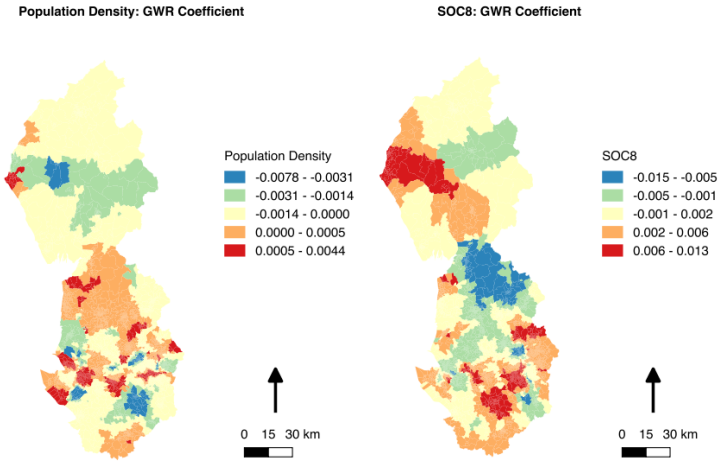






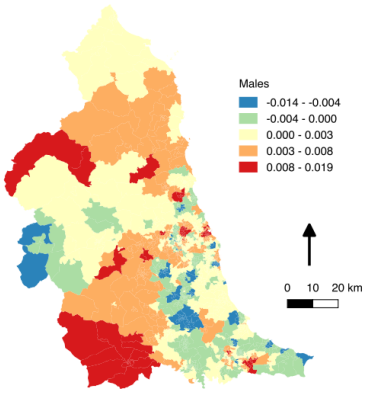
2. North West


Males: GWR Coefficient
Males
-0.0039 - -0.0004
-0.0004 - 0.0003
0.0003 - 0.0012
0.0012 - 0.0024
0.0024 - 0.0110
0 15 30 km
Age 16 to 25: GWR Coefficient
Age 16 to 25
-0.012 - -0.005
-0.005 - -0.002
-0.002 - 0.000
0.000 - 0.003
0.003 - 0.013
0 15 30 km
Age 56 to 65: GWR Coefficient
Age 56 to 65
-0.021 - -0.009
-0.009 - -0.001
-0.001 - 0.002
0.002 - 0.005
0.005 - 0.012
0 15 30 km

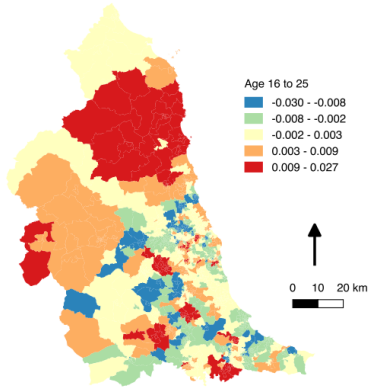

Population Density: GWR Coefficient
Population Density
-0.0078 - -0.0031
-0.0031 - -0.0014
-0.0014 - 0.0000
0.0000 - 0.0005
0.0005 - 0.0044
0 15 30 km
SOC8: GWR Coefficient
SOC8
-0.015 - -0.005
-0.005 - -0.001
-0.001 - 0.002
0.002 - 0.006
0.006 - 0.013
0 15 30 km

3. North East

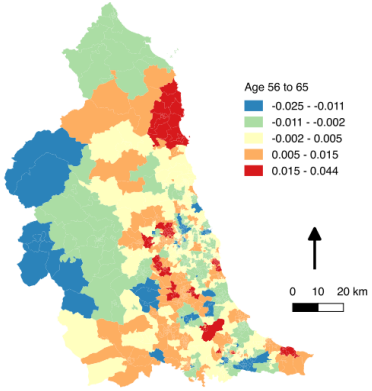
Males: GWR Coefficient



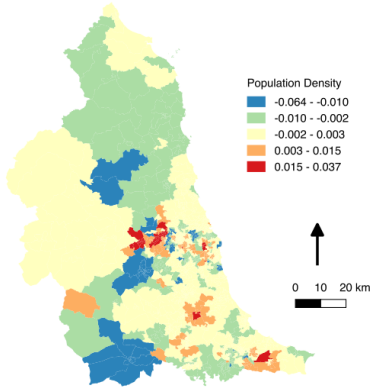
Age 16 to 25: GWR Coefficient



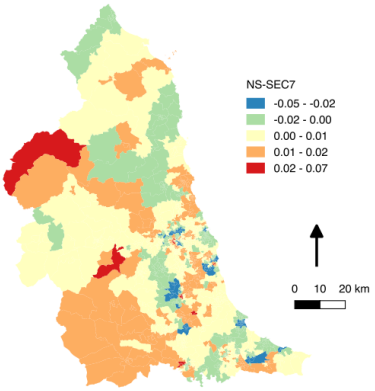
Age 56 to 65: GWR Coefficient



Population Density: GWR Coefficient

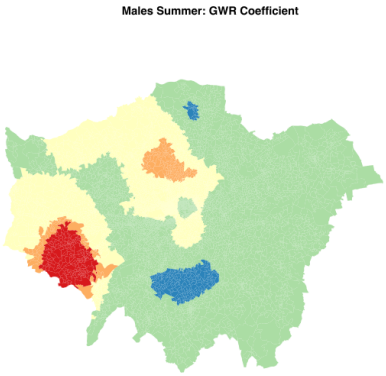


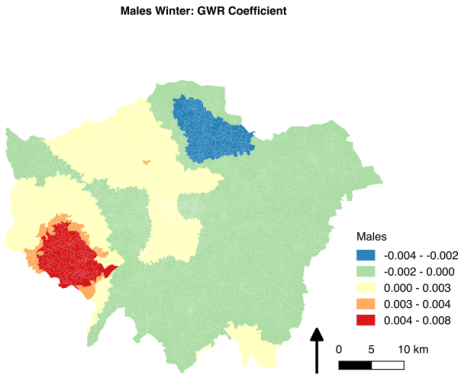
NS-SEC7: GWR Coefficient

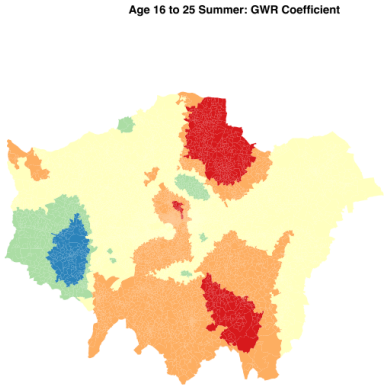


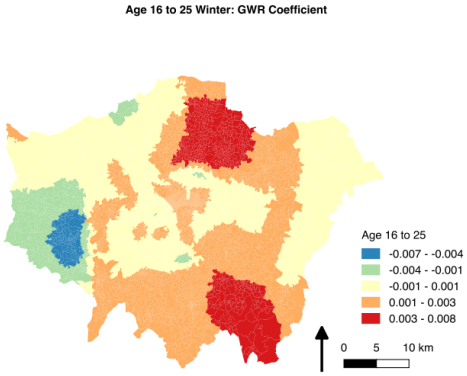
4. Seasonal

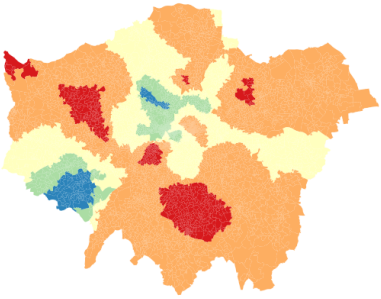
4.1 London

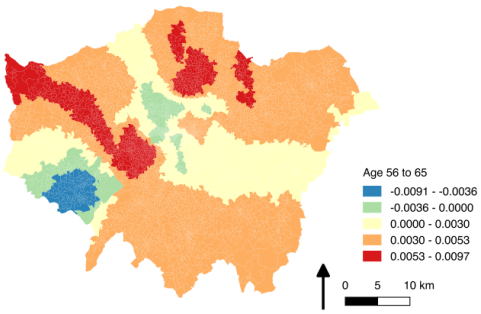




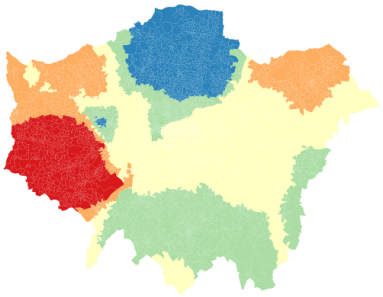




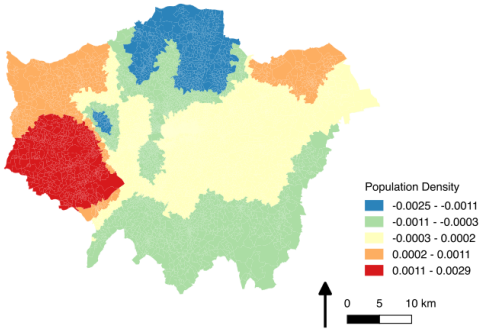




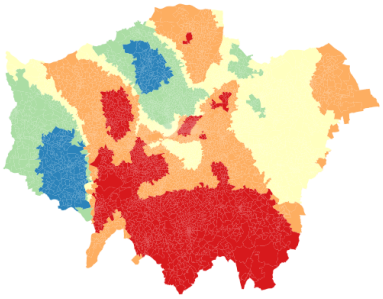
Population Density Summer: GWR Coefficient



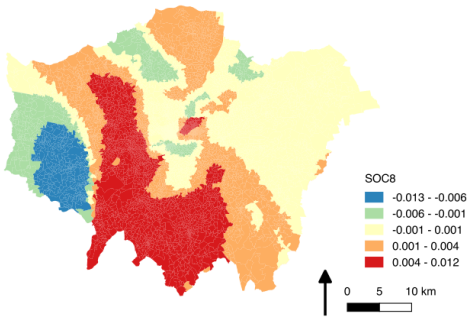
Population Density Winter: GWR Coefficient



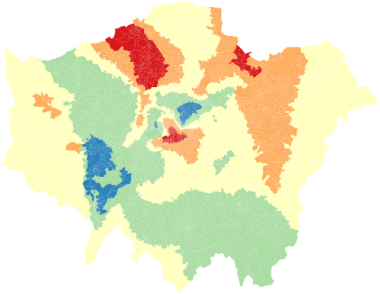
SOC8 Summer: GWR Coefficient



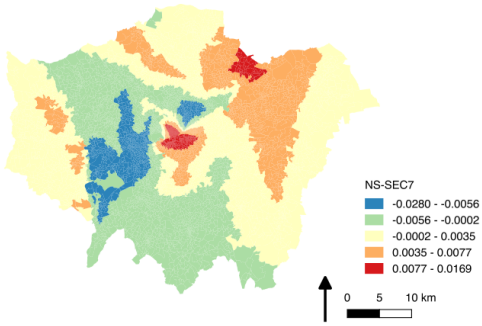
SOC8 Winter: GWR Coefficient



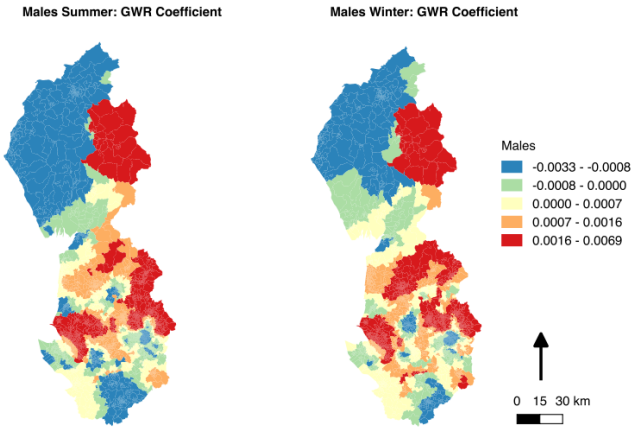
NS-SEC7 Summer: GWR Coefficient

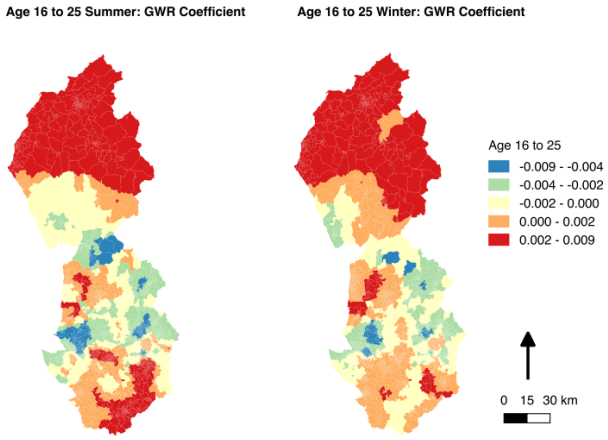


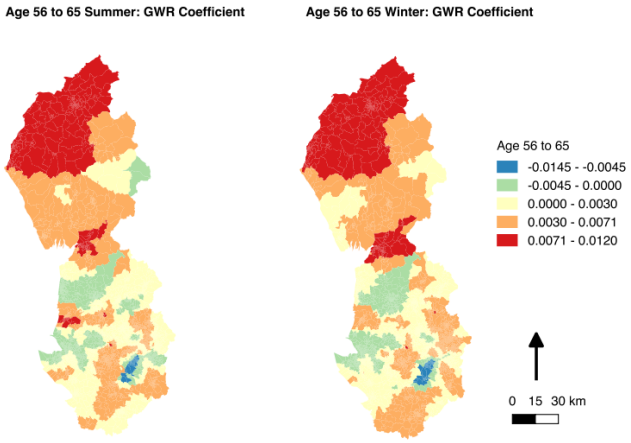
NS-SEC7 Winter: GWR Coefficient

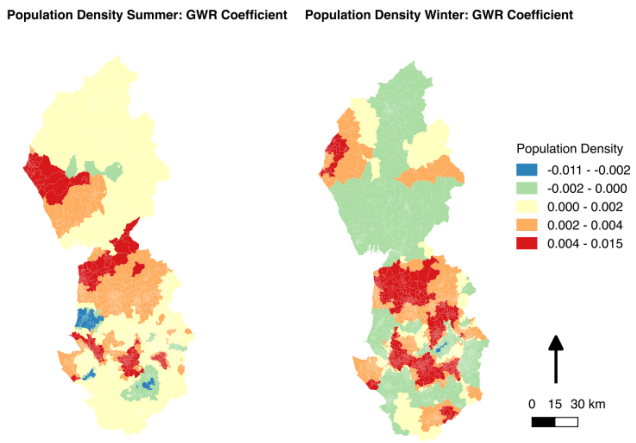


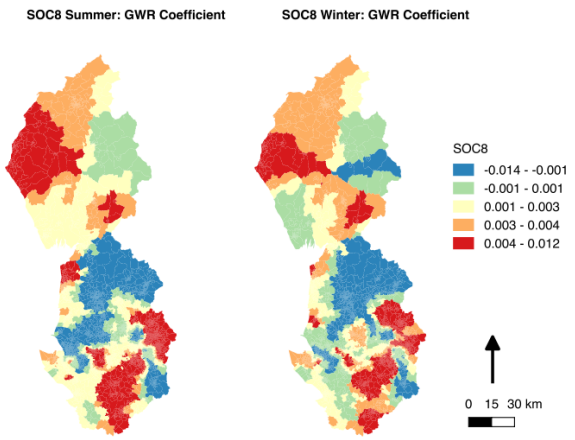
4.2 North West

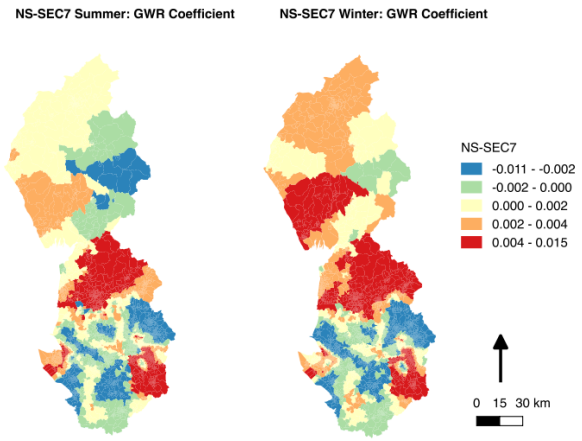




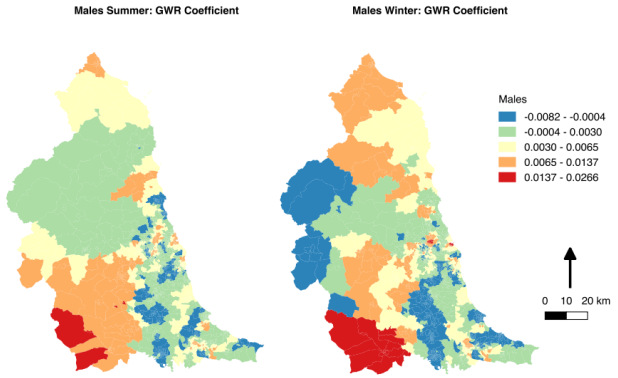


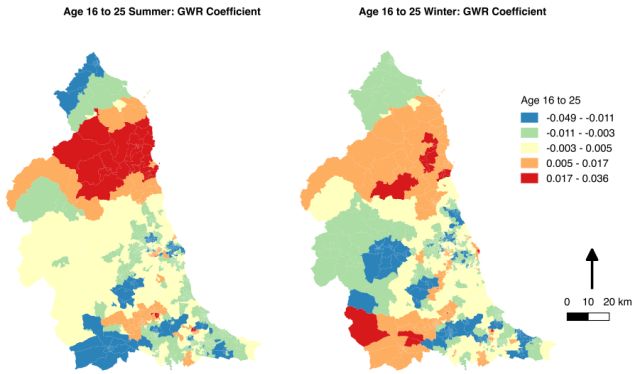

Population Density Summer: GWR Coefficient
Population Density Winter: GWR Coefficient
Population Density
-0.011 - -0.002
-0.002 - 0.000
0.000 - 0.002
0.002 - 0.004
0.004 - 0.015
0 15 30 km

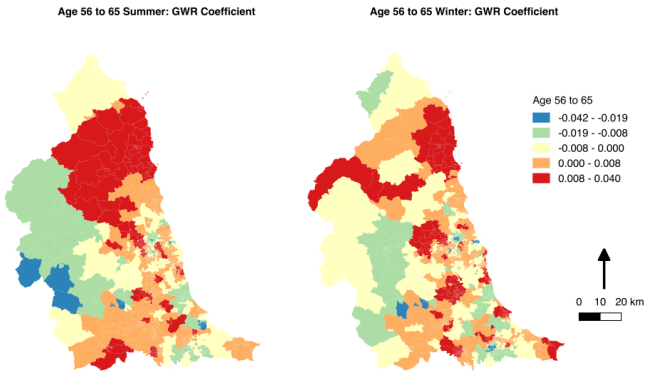

SOC8 Summer: GWR Coefficient
SOC8 Winter: GWR Coefficient
SOC8
-0.014 - -0.001
-0.001 - 0.001
0.001 - 0.003
0.003 - 0.004
0.004 - 0.012
0 15 30 km

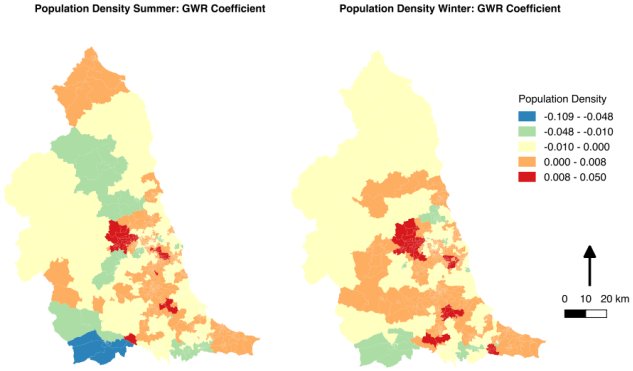

NS-SEC7 Summer: GWR Coefficient
NS-SEC7 Winter: GWR Coefficient
NS-SEC7
-0.011 - -0.002
-0.002 - 0.000
0.000 - 0.002
0.002 - 0.004
0.004 - 0.015
0 15 30 km

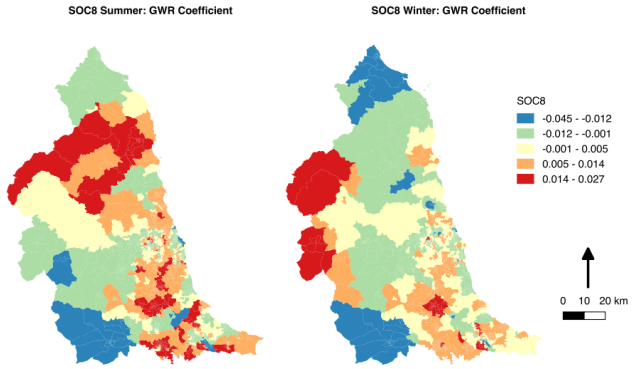
4.3 North East

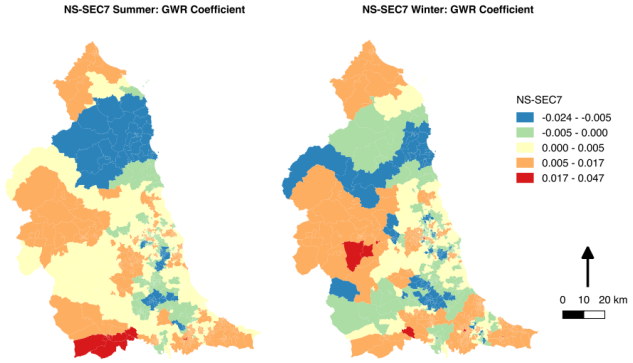







Population Density Summer: GWR Coefficient
Population Density Winter: GWR Coefficient
Population Density
-0.109 - -0.048
-0.048 - -0.010
-0.010 - 0.000
0.000 - 0.008
0.008 - 0.050
0 10 20 km


SOC8 Summer: GWR Coefficient
SOC8 Winter: GWR Coefficient
SOC8
-0.045 - -0.012
-0.012 - -0.001
-0.001 - 0.005
0.005 - 0.014
0.014 - 0.027
0 10 20 km


NS-SEC7 Summer: GWR Coefficient
NS-SEC7 Winter: GWR Coefficient
NS-SEC7
-0.024 - -0.005
-0.005 - 0.000
0.000 - 0.005
0.005 - 0.017
0.017 - 0.047
0 10 20 km