A PRICE NOT WORTH PAYING:
USING CAUSAL EFFECT MODELLING TO EXAMINE THE
RELATIONSHIP BETWEEN WORKLESSNESS AND
MORTALITY FOR MALE INDIVIDUALS IN SCOTLAND

Thomas Laurie Clemens

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews

2012

Full metadata for this item is available in
Research@StAndrews:FullText
at:
http://research-repository.st-andrews.ac.uk/

Please use this identifier to cite or link to this item:
http://hdl.handle.net/10023/4036

This item is protected by original copyright
A PRICE NOT WORTH PAYING: USING CAUSAL EFFECT MODELLING TO EXAMINE THE RELATIONSHIP BETWEEN WORKLESSNESS AND MORTALITY FOR MALE INDIVIDUALS IN SCOTLAND

THOMAS LAURIE CLEMENS

A THESIS SUBMITTED FOR THE DEGREE OF PhD

AT THE

UNIVERSITY OF ST ANDREWS

2012
1. Candidate’s declarations: I, Thomas Clemens, hereby certify that this thesis, which is approximately 63,000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree. I was admitted as a research student in September 2006 and as a candidate for the degree of Ph.D in September 2007 the higher study for which this is a record was carried out in the University of St Andrews between 2007 and 2012.

Date signature of candidate

2. Supervisor’s declaration: I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Ph.D in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date signature of supervisor

3. Permission for electronic publication: In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below. The following is an agreed request by candidate and supervisor regarding the electronic publication of this thesis:

Access to all of printed copy but embargo of all of electronic publication of thesis for a period of three years on the grounds publication would preclude future publication.

Date

signature of candidate

signature of supervisor
ACKNOWLEDGEMENTS

I would like to thank firstly my supervisors, particularly Prof Paul Boyle, who started me on this process and Dr. Frank Popham, both of whom must take much credit for ensuring I made it to the end. Their combined help and guidance proved invaluable on many occasions. I would also like to thank other supervisors who have helped me along the way including Prof. Robin Flowerdew and Dr. Chris Dibben. I would also like to thank the examiners, Prof. Clare Bambra and Dr. Donald Houston, for their extremely valuable feedback, suggestions and revisions all of which have greatly improved the final piece of work.

Formally, I would like to acknowledge the help and support provided by staff of the Longitudinal Studies Centre – Scotland (LSCS), particularly Dr. Petke Feijten, Dr. Lee Williamson, Joan Nolan, Claire Boag and Dr. Fiona Cox. The LSCS is supported by the Economic and Social Research Council (ESRC)/Joint Information Systems Committee, the Scottish Funding Council, the Chief Scientist's Office and the Scottish Government. The author alone is responsible for the interpretation of the data. Census output is Crown copyright and is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

I must also mention members of the ESRC funded transitions and choices at older age’s project group whom I had the privilege to work with briefly at the beginning of my PhD. In particular, I would like to thank Shayla Leib and Kevin Lynch at the Office for National Statistics for providing the data from the ONS Longitudinal Study which was used in Chapter Four.

A great deal of thanks and appreciation go to my Mum and Dad for giving me the opportunities that have allowed me to do this work. As well, many friends and colleagues have helped me at various points and I am very grateful to everyone. My two flatmates, Fergus Neville and David Small, deserve special mention for their encouragement and camaraderie at different stages of the thesis.

My final thanks go to Hannah, for her constant support and amazing capacity for patience during the latter stages of completing the thesis. It has been an enormous help and I am very grateful. I look forward to returning the favour sometime soon.
‘Rising unemployment and the recession have been the price that we have had to pay to get inflation down. That price is well worth paying’

Norman Lamont, as Chancellor of the Exchequer, to the House of Commons

16th May 1991
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONTENTS</td>
<td>I</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>III</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>IV</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>VII</td>
</tr>
<tr>
<td>CHAPTER 1: INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 BACKGROUND</td>
<td>1</td>
</tr>
<tr>
<td>1.2 DEFINITIONS OF WORKLESSNESS USED IN THE THESIS</td>
<td>2</td>
</tr>
<tr>
<td>1.3 STRUCTURE AND FORMAT OF THE THESIS</td>
<td>3</td>
</tr>
<tr>
<td>CHAPTER 2: WORKLESSNESS AND MORTALITY: A REVIEW OF THE LITERATURE</td>
<td>7</td>
</tr>
<tr>
<td>2.1 INTRODUCTION</td>
<td>7</td>
</tr>
<tr>
<td>2.2 THE UK LABOUR MARKET AND WELFARE STATE</td>
<td>8</td>
</tr>
<tr>
<td>2.2.1 The Political economy and the welfare state</td>
<td>9</td>
</tr>
<tr>
<td>2.2.2 The UK economy, labour market and welfare state</td>
<td>12</td>
</tr>
<tr>
<td>2.2.3 Health related worklessness</td>
<td>16</td>
</tr>
<tr>
<td>2.2.4 Summary</td>
<td>24</td>
</tr>
<tr>
<td>2.3 THE ASSOCIATION OF WORKLESSNESS WITH MORTALITY</td>
<td>26</td>
</tr>
<tr>
<td>2.3.1 Ecological Evidence</td>
<td>26</td>
</tr>
<tr>
<td>2.3.2 Individual mortality and worklessness</td>
<td>29</td>
</tr>
<tr>
<td>2.3.3 Summary</td>
<td>34</td>
</tr>
<tr>
<td>2.4 THE CAUSAL EXPLANATION</td>
<td>35</td>
</tr>
<tr>
<td>2.4.1 Financial Disadvantage</td>
<td>35</td>
</tr>
<tr>
<td>2.4.2 The 'latent functions' of work</td>
<td>39</td>
</tr>
<tr>
<td>2.4.3 Physical health and stress responses</td>
<td>40</td>
</tr>
<tr>
<td>2.4.4 Health Behaviours</td>
<td>42</td>
</tr>
<tr>
<td>2.4.5 Summary</td>
<td>43</td>
</tr>
<tr>
<td>2.5 NON-CAUSAL FACTORS</td>
<td>44</td>
</tr>
<tr>
<td>2.5.1 Direct health selection</td>
<td>44</td>
</tr>
<tr>
<td>2.5.2 Indirect Selection</td>
<td>46</td>
</tr>
<tr>
<td>2.5.3 Testing for selection</td>
<td>50</td>
</tr>
<tr>
<td>2.5.4 Summary</td>
<td>54</td>
</tr>
<tr>
<td>2.6 CONCLUSIONS</td>
<td>55</td>
</tr>
<tr>
<td>2.6.1 Summary of Chapter two</td>
<td>55</td>
</tr>
<tr>
<td>2.6.2 Empirical applications of the thesis</td>
<td>57</td>
</tr>
<tr>
<td>CHAPTER 3: A METHODOLOGICAL FRAMEWORK</td>
<td>59</td>
</tr>
<tr>
<td>3.1 INTRODUCTION</td>
<td>59</td>
</tr>
<tr>
<td>3.2 COUNTERFACTUAL CAUSALITY</td>
<td>60</td>
</tr>
<tr>
<td>3.2.1 The fundamental problem of causal inference</td>
<td>60</td>
</tr>
<tr>
<td>3.2.2 Potential outcomes</td>
<td>61</td>
</tr>
<tr>
<td>3.2.3 Exchangeability and randomisation</td>
<td>62</td>
</tr>
<tr>
<td>3.3 BIAS</td>
<td>63</td>
</tr>
<tr>
<td>3.3.1 Common causes and common effects</td>
<td>64</td>
</tr>
<tr>
<td>3.3.2 Bias in the worklessness and mortality relationship</td>
<td>66</td>
</tr>
<tr>
<td>3.3.3 Over-adjustment</td>
<td>68</td>
</tr>
<tr>
<td>3.3.4 Structural confounding</td>
<td>70</td>
</tr>
<tr>
<td>3.4 PROPENSITY SCORE MATCHING AND THE RUBIN CAUSAL MODEL</td>
<td>72</td>
</tr>
<tr>
<td>3.4.1 The Rubin Causal Model (RCM)</td>
<td>73</td>
</tr>
<tr>
<td>3.4.2 Matching and the propensity score</td>
<td>74</td>
</tr>
</tbody>
</table>
CHAPTER 4: UNEMPLOYMENT, MORTALITY AND THE PROBLEM OF HEALTH-RELATED CONFOUNDING................................................. 89

4.1 INTRODUCTION ......................................................................................................................................................... 89
4.2 METHODS ................................................................................................................................................................. 91
    4.2.1 Data and outcome variables ............................................................................................................................... 91
    4.2.2 Analysis .............................................................................................................................................................. 92
4.3 RESULTS ..................................................................................................................................................................... 94
4.4 DISCUSSION .............................................................................................................................................................. 103

CHAPTER 5: THE CAUSAL EFFECT OF UNEMPLOYMENT ON ALL-CAUSE MORTALITY ....................................................... 105

5.1 BACKGROUND ............................................................................................................................................................ 105
5.2 METHODS ................................................................................................................................................................. 107
    5.2.1 Data and sample .................................................................................................................................................. 107
    5.2.2 Study Design ...................................................................................................................................................... 107
5.3 RESULTS ..................................................................................................................................................................... 110
5.4 DISCUSSION .............................................................................................................................................................. 115

CHAPTER 6: THE CAUSAL EFFECT OF ECONOMIC INACTIVITY DUE TO SICKNESS FOR MORTALITY ................................. 119

6.1 INTRODUCTION ............................................................................................................................................................ 119
6.2 METHOD ...................................................................................................................................................................... 120
    6.2.1 Data and sample .................................................................................................................................................. 120
    6.2.2 Propensity Score matching .................................................................................................................................. 121
    6.2.3 Study Design ...................................................................................................................................................... 121
    6.2.4 Mortality Analysis ................................................................................................................................................ 122
6.3 RESULTS ..................................................................................................................................................................... 123
6.4 DISCUSSION .............................................................................................................................................................. 126
    6.4.1 Principal findings ................................................................................................................................................ 126
    6.4.2 Strengths and limitations .................................................................................................................................... 127
    6.4.3 Interpretation ...................................................................................................................................................... 130
6.5 CONCLUSIONS ........................................................................................................................................................... 132

CHAPTER 7: DISCUSSION AND CONCLUSIONS ................................................................................................................................. 133

7.1 INTRODUCTION ............................................................................................................................................................ 133
7.2 KEY CONTRIBUTIONS TO THE WORKLESSNESS AND MORTALITY DEBATE ................................................................. 134
    7.2.1 Selection versus causation .................................................................................................................................. 134
    7.2.2 Mortality and health related labour market inactivity ......................................................................................... 140
    7.2.3 Summary .............................................................................................................................................................. 145
7.3 LIMITATIONS OF THE STUDIES ..................................................................................................................................... 146
    7.3.1 Residual Confounding ......................................................................................................................................... 146
    7.3.2 Census and Mortality Data .................................................................................................................................. 147
7.4 WIDER IMPLICATIONS ..................................................................................................................................................... 148
    7.4.1 Data and methodological innovation .................................................................................................................. 148
    7.4.2 Implications for theory and policy ....................................................................................................................... 150
    7.4.3 Summary .............................................................................................................................................................. 159
7.5 CONCLUSIONS ............................................................................................................................................................. 161
    7.5.1 Summary of main findings and key contributions ............................................................................................. 161
    7.5.2 Future research ................................................................................................................................................... 162
    7.5.3 Concluding remarks ............................................................................................................................................. 164

REFERENCES ............................................................................................................................................................................ 167
List of Tables

TABLE 3.1 THE FOUR ‘CAUSAL TYPES’ IN A HYPOTHETICAL EXPOSURE/O U T C O M E STUDY OF UNEMPLOYMENT AND MORTALITY ................................................................. 61
TABLE 3.2 HYPOTHETICAL TABLE TO EVALUATE BALANCE IN TERMS OF DEPRIVATION BETWEEN UNEMPLOYED AND EMPLOYED GROUPS IN AN UNMATCHED AND MATCHED SAMPLE USING T-TESTS .......................................................................................................................... 77
TABLE 3.3: LABOUR MARKET STATUS BY PRESENCE OF HEALTH PROBLEM FOR WORKING AGE MEN. RAW CELL COUNTS WITH CELL PERCENTAGES IN BRACKETS’ . ................................................................................................................................. 83
TABLE 3.4: LABOUR MARKET STATUS BY PRESENCE OF HEALTH PROBLEM FOR WORKING AGE WOMEN. RAW CELL COUNTS WITH CELL PERCENTAGES IN BRACKETS’ . ................................................................................................................................. 86
TABLE 4.1 LENGTH OF WEAR-OFF PERIOD, MORTALITY PERIODS AND MORTALITY FREQUENCIES FOR EACH WEAR-OFF PERIOD BY SCOTLAND AND ENGLAND & WALES ......................................................................................... 92
TABLE 4.2 FREQUENCIES AND PERCENTAGES FOR ECONOMIC POSITION, SEX, SOCIAL CLASS, MARITAL STATUS, CARSTAIRS DEPRIVATION, ETHNICITY AND LIMITING LONG-TERM ILLNESS BY SCOTLAND AND ENGLAND & WALES ......................................................................................................................... 93
TABLE 5.1 DISTRIBUTION OF DEATHS DURING FOLLOW-UP PERIOD (2001-2006) BY UNEMPLOYED AND IN WORK.................................................................................. 110
TABLE 5.2 CHARACTERISTICS OF FULL SAMPLE IN TERMS OF SOCIO-ECONOMIC COVARIATES PREDICTING UNEMPLOYMENT IN 2001 AND STATISTICAL TESTS OF DIFFERENCE BETWEEN 2001 UNEMPLOYED AND IN WORK GROUPS .......................................................................................................................... 111
TABLE 5.3 CHARACTERISTICS OF FULL SAMPLE IN TERMS OF HEALTH COVARIATES PREDICTING UNEMPLOYMENT IN 2001 AND STATISTICAL TESTS OF DIFFERENCE BETWEEN 2001 UNEMPLOYED AND IN WORK GROUPS .......................................................................................................................... 112
TABLE 5.4 CHARACTERISTICS OF CASE-MATCHED SAMPLE IN TERMS OF HEALTH COVARIATES PREDICTING UNEMPLOYMENT IN 2001 AND STATISTICAL TESTS OF DIFFERENCE BETWEEN 2001 UNEMPLOYED AND IN WORK GROUPS .......................................................................................................................... 114
TABLE 5.5 CHARACTERISTICS OF CASE MATCHED SAMPLE IN TERMS OF COVARIATES PREDICTING UNEMPLOYMENT IN 2001 AND STATISTICAL TESTS OF DIFFERENCE BETWEEN 2001 UNEMPLOYED AND IN WORK GROUPS .......................................................................................................................... 115
TABLE 5.6 MORTALITY RISKS OF UNEMPLOYMENT RELATIVE TO EMPLOYMENT DURING FOLLOW-UP PERIOD 2001-2006 .......................................................................................................................................................... 116
TABLE 6.1 DISTRIBUTION OF DEATHS BY IN WORK AND INACTIVE SICK .................................................................................................................................................................. 123
TABLE 6.2 SOCIO-ECONOMIC COVARIATE COMPARISONS (MEANS AND PROPORTIONS) BETWEEN TREATMENT AND IN WORK GROUPS ACROSS THE FULL SAMPLE .................................................................................................................................. 124
TABLE 6.3 HEALTH COVARIATE COMPARISONS (MEANS AND PROPORTIONS) BETWEEN TREATMENT AND IN WORK GROUPS ACROSS THE FULL SAMPLE .................................................................................................................................. 126
TABLE 6.4 SOCIO-ECONOMIC COVARIATE COMPARISONS (MEANS AND PROPORTIONS) BETWEEN TREATMENT AND IN WORK GROUPS ACROSS THE PROPENSITY SCORE CASE MATCHED (NEAREST NEIGHBOUR) RESTRICTED SAMPLE .................................................................................................................................. 129
TABLE 6.5 HEALTH COVARIATE COMPARISONS (MEANS AND PROPORTIONS) BETWEEN TREATMENT AND IN WORK GROUPS ACROSS THE PROPENSITY SCORE CASE MATCHED (NEAREST NEIGHBOUR) RESTRICTED SAMPLE .................................................................................................................................. 130
TABLE 6.6 COX HAZARD RATIO ESTIMATES OF TREATMENT (COMPARED TO IN WORK) AND CONFIDENCE INTERVALS FOR PROPENSITY SCORE MATCHED SAMPLES AND SENSITIVITY ANALYSIS (WEAR-OFF SELECTION) FOR FOLLOW-UP PERIOD 2001-2006 .................................................................................................................................. 131
TABLE 7.1 COMPARISON OF EFFECT ESTIMATES FROM CHAPTERS FOUR TO SIX ILLUSTRATING THE POTENTIAL OVER-ADJUSTMENT .................................................................................................................................. 139
List of Figures

Figure 1.1 Categorisation of labour market status according to the ILO definition of economic activity and inactivity ................................................................. 2
Figure 1.2 Definitions of labour market status and acronyms (bracketed) used in the thesis ........................................................................................................... 3
Figure 2.1: UK unemployment rate (%) 1900-1999 .......................................................................................................................... 12
Figure 2.2 Numbers employed in sectors of the UK economy (000’s), males and females. .......................................................... 17
Figure 2.3 Proportion of men employed by occupational social class groups 1973-1993 .............................................................. 19
Figure 2.4 Numbers in employment by socio-economic group and presence of limiting long-term illness 1973-1993 ................................................................. 20
Figure 2.5 Distribution of sickness claimants in England and Wales and Scotland in August 1996 as a percentage of total working age male population ..................... 22
Figure 2.6 UK annual average of the gross unemployment benefit replacement rates for period 1961 to 2007 ........................................................................ 36
Figure 3.1 Causal diagram representing a causal effect of exposure E on outcome D ...... 64
Figure 3.2 Causal diagram representing a confounded association between exposure E and outcome D confounded by L ......................................................... 65
Figure 3.3 Causal diagrams representing a confounded association between exposure E and outcome D confounded by the conditioning variables C and M .................. 65
Figure 3.4 Healthy worker bias expressed as a common effect causal structure ............ 66
Figure 3.5 Explanations for the association between unemployment and health .......... 66
Figure 3.6 Over adjustment for an intermediate variable M on the causal pathway between unemployment E and mortality D ......................................................... 69
Figure 3.7 Over-adjustment for the proxy variable M of the intermediate variable U on the causal pathway between unemployment E and mortality D ................. 70
Figure 3.8 Histograms comparing the predicted probability of unemployment between actually employed and unemployed individuals for a hypothetical study. The graphs illustrate an extreme example of the off-support problem ........................................ 72
Figure 4.1 Mortality odds ratios for levels of economic position within a given wear-off period with further adjustment for age, age² and sex in Scotland ...................... 95
Figure 4.2 Mortality odds ratios for levels of economic position within a given wear-off period with further adjustment for age, age² and sex in England & Wales .......... 96
Figure 4.3 Mortality odds ratios for levels of economic position within a given wear-off period with further adjustment for age, age², sex, social class, marital status, Carstairs deprivation and ethnicity in Scotland ......................................................... 97
Figure 4.4 Mortality odds ratios for levels of economic position within a given wear-off period with further adjustment for age, age², sex, social class, marital status, Carstairs deprivation and ethnicity in England & Wales ........................................... 98
Figure 4.5 Mortality odds ratios for levels of economic position within a given wear-off period with further adjustment for age, age² and sex and additionally limiting long-term illness in Scotland ......................................................... 99
Figure 4.6 Mortality odds ratios for levels of economic position within a given wear-off period with further adjustment for age, age² and sex and additionally limiting long-term illness in England and Wales ............................................. 100
Figure 4.7 Mortality odds ratios for levels of economic position within a given wear-off period with further adjustment for age, age², sex, social class, marital status, Carstairs deprivation and ethnicity and additionally limiting long-term illness in Scotland ................................................................. 101
Figure 4.8 Mortality odds ratios for levels of economic position within a given wear-off period with further adjustment for age, age², sex, social class, marital status,
CARTAIRS DEPRIVATION AND ETHNICITY AND ADDITIONALLY LIMITING LONG-TERM ILLNESS IN ENGLAND AND WALES........................................................................................................................................ 102

FIGURE 5.1 SAMPLE SELECTION CRITERIA (CORRESPONDING SAMPLE SIZE IN BRACKETS) AND OUTLINE OF COVARIATES USED TO ESTIMATE THE PROPENSITY SCORE FOR UNEMPLOYMENT IN 2001........ 108

FIGURE 5.2 DISTRIBUTION OF THE PROPENSITY SCORE FOR UNEMPLOYMENT IN 2001 BY IN WORK AND UNEMPLOYED GROUPS IN 2001 .................................................................................................. 113

FIGURE 6.1 SAMPLE SELECTION CRITERIA, STUDY DESIGN AND DETAILS OF THE COVARIATES USED .. 122

FIGURE 6.2 DISTRIBUTION OF THE PROPENSITY SCORE BY TREATMENT AND IN WORK GROUPS IN 2001 ........................................................................................................................................ 127
The research conducted in this thesis examines the relationship between forms of worklessness (both active unemployment and inactivity due to sickness and disability) and mortality for working age men. Previous research has shown that being out of work is associated with a greater risk of mortality relative to being in work. However, there remains debate as to whether this association is the result of a causal pathway leading from worklessness to mortality or whether it reflects the ‘selection’ of individuals who are already at greater risk of mortality from pre-existing poor health or other characteristics. In the UK, many studies rely on the use of ‘wear-off’ periods in which mortality events occurring within five years after the observation of employment status are ignored to allow the confounding effects of selection to diminish. Generally these studies concluded in support of a causal relationship. In contrast, more recent studies making use of innovative methodological designs such as natural experiments and linked register and health datasets have found less evidence for this explanation with many emphasising the role of confounding and selection. The thesis aims to firstly, examine the effectiveness of wear-off periods and secondly, to develop an alternative counterfactual approach to examine the relationship between worklessness (both active unemployment and health related inactivity) and mortality. These questions are addressed in three stand-alone papers.

In the first paper, data from the Scottish Longitudinal Study and the England & Wales Longitudinal Study was used in logistic regression models which estimated the odds of death in a given time period after the 1991 Census for those aged 35–64 in 1991. The odds ratios for the different economic positions (in work, unemployed, retired, permanently sick and other inactive) were compared, as well as the changes in risk associated with cumulatively increasing the length of wear-off prior to follow-up. No evidence was found of health related selection for the unemployed in 1991 suggesting that the use of the five year wear-off period in many studies of mortality and unemployment may be an ineffective and unnecessary technique for mitigating the effects of health-related selection.

The second paper examined men aged between 35 and 54 who were in work in 1991. Subsequent employment status in 2001 was observed (in work or unemployed) and the relative all-cause mortality risk of unemployment between 2001 and 2007 was estimated. To account for potential selection into unemployment of those in poor health, a counterfactual propensity score matching framework was used to construct unbiased and comparable
Abstract

samples of in work and unemployed individuals. Matching was based on a wide range of explanatory variables including health status prior to year of unemployment (hospital admissions and self-reported limiting long term illness) as well as measures of socio-economic position. The findings showed that unemployment was associated with a doubling (hazard ratio 2.1 95% CI 1.30 - 3.38) of the subsequent risk of mortality from all causes relative to employment. This scale of effect was consistent across different samples and was robust controlling for prior health and socio-demographic characteristics. These findings were interpreted as evidence that the often observed association between unemployment and mortality may contain a causal component.

The second paper implemented a similar analytical design to address the lack of evidence for the independent mortality effect of inactivity due to sickness. The results showed that the mortality risk of economic inactivity due to sickness relative to active employment was significant (HR. 3.18, 95% CI 2.53-3.98) and suggest that economic inactivity due to sickness poses a mortality risk that is independent of prior health. The findings could be interpreted in two ways; either economic inactivity due to sickness is worse for health than actively seeking work or previous studies of unemployment and mortality have underestimated the true effect of being out of work generally.

Across the three studies, the main contribution of the thesis is to reassert the importance of worklessness as a determinant of individual mortality. In doing so the studies also found little evidence of systematic confounding by either health or other characteristics. The thesis concludes with a comprehensive discussion of the wider implications of the findings in relation to both general methodological issues in observational epidemiology and possible policy interventions that could be implemented to tackle work-related inequalities in male mortality.
Chapter 1

INTRODUCTION

1.1 Background

Individuals who are out of work commonly suffer poorer health than those in stable employment (Dooley, Fielding & Levi 1996; Bartley, Ferrie & Montgomery 2006). However, there remains great debate about the nature of this relationship and within this debate, two distinct camps have emerged. For some, the psychosocial stresses of unemployment, coupled with the decline in income and material circumstances and the associated physiological stress responses, represent plausible pathways between worklessness and health & mortality. Within this view, worklessness is therefore believed to exert a causal effect on health and mortality (Bartley, Ferrie & Montgomery 2006). However, the alternative explanation, commonly referred to as selection, describes the process whereby ill or socio-economically deprived members of the population are more likely to become unemployed compared to both healthier and wealthier individuals (Bartley and Ferrie 2001). As a result, it is difficult to isolate the direct health effects of labour market disadvantage. The question of which of these explanations is most important has been and remains the focus of a great deal of international research (Bartley 1988; Martikainen and Valkonen 1996; Bartley, Ferrie & Montgomery 2006) with the most recent evidence emphasising the selection explanation (Lundin et al. 2010). The research conducted in the thesis contributes to this debate by using recent methodological advances in observational epidemiology to test for a causal relationship between worklessness and mortality amongst male individuals in Scotland. This contribution is especially prescient, coming at a time when concerns over rates of unemployment and worklessness have resurfaced following the global financial crisis of 2008 which has had lasting effects on national economies and labour markets.
1.2 Definitions of worklessness used in the thesis

An important theme throughout the thesis is the distinction between different forms of worklessness and how these may differ in their effects for mortality. It is therefore important, from the outset, to ensure that a consistent definition of worklessness is identified that can be applied throughout. It was decided to adopt the almost universally applied definition from the International Labour Organisation (ILO) which is the definition adopted by the Office for National Statistics (ONS) and the one upon which labour market statistics from the UK labour force survey are based (ONS 2010). This definition divides the working age (aged 16 to 64 for men) labour force into two broad categories; the economically active and the economically inactive. This is illustrated in Figure 1.1.

Figure 1.1 Categorisation of labour market status according to the ILO definition of economic activity and inactivity.

Source: ONS report 2010

Within these categories, the unemployed are defined as individuals who, during a reference week, are without a job, who want a job, have actively sought work in the last four weeks and are available to start work in the next two weeks. This definition also includes individuals who are out of work but have found a job and are waiting to start in the next two weeks. Throughout the thesis, this group of individuals will therefore be referred to as the active unemployed or unemployed. The criterion for being in employment is the completion of at least one hour of paid work in the reference week. Thus, individuals who are out of work but do not meet these criteria are considered economically inactive.
Chapter 1: Introduction

The economically inactive are defined as individuals who have not been looking for work in the last four weeks or who are not able to start work in the next two weeks. This definition results in a much larger and more heterogeneous group than the active unemployed and includes individuals looking after family or home, students, retirees and other inactive groups. It also contains individuals who are not seeking work due to illness or sickness (ONS 2010). This latter group will be referred to as ‘inactive due to sickness’ with the remaining economically inactive referred to as ‘other inactive’. To distinguish between individuals in employment and those either unemployed or inactive due to sickness the term ‘workless’ and ‘worklessness’ shall be used for the latter. Finally, because the empirical work in the thesis is based on the England and Wales and Scottish censuses, it is necessary to introduce the term ‘permanently sick’ which is the category in the employment status question of the census which, in combination with information from the limiting long term illness question also in the census, most closely resembles the inactive due to sickness definition. These definitions are summarised in Figure 1.2.

Figure 1.2 Definitions of labour market status and acronyms (bracketed) used in the thesis

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active unemployed (AU)</td>
<td>Individuals reporting themselves as unemployed in the census and assumed to be actively seeking work. Also includes those waiting to start employment in the next two weeks.</td>
</tr>
<tr>
<td>Inactive due to sickness (IDS)</td>
<td>Individuals who are not seeking work due to illness or sickness and recorded as permanently sick in the census.</td>
</tr>
<tr>
<td>In work</td>
<td>Individuals who are currently in employment and report themselves as in employment in the census.</td>
</tr>
<tr>
<td>Workless</td>
<td>Individuals either inactive due to sickness or active unemployed but not in work.</td>
</tr>
</tbody>
</table>

1.3 Structure and format of the thesis

The research conducted within this thesis is structured around three research questions which emerged from a review of the prevailing methodological and empirical literature. These questions are as follows:

1. In studies of employment status and mortality, to what extent is health selection of the workless mitigated by the use of wear-off periods?
2. Is there evidence of a causal link between AU and mortality in Scotland using individual level data?
3. Is there evidence of a causal link between IDS and mortality in Scotland using individual level data?
These questions are addressed by way of three original manuscripts that have either been published or are in the process of being submitted for publication in various academic journals. Each of the manuscripts is presented as a separate chapter (chapters four to six). The chapters can be read in isolation; however, they are all linked by the broad aim of testing for a causal relationship between worklessness and mortality. Collectively, each chapter develops a number of observations that can only be examined and discussed when read in the context of the thesis as a whole. Chapters’ four to six are, therefore, bookended by a literature review (chapter two), a methodological framework (chapter three) and a discussion and conclusions section (chapter seven). These chapters allow for the thesis to be read as a coherent whole and therefore set the context within which the empirical chapters can sit.

The main aim of the research questions is to test for evidence that the relationship between worklessness and mortality is causal. In chapter two, literature that is relevant to this aim is examined and reviewed and is structured around the following topics. Firstly, and in order to inform a wider discussion in chapter seven about the possible implications of the thesis for theory and policy, the chapter begins with a brief outline of political economy explanations of health inequalities that places work, worklessness and the welfare state at its heart. This theoretical perspective is used to examine the historical context of the UK labour market and how it has shaped contemporary patterns of labour market participation and welfare state provision. In the second section, empirical literature on the relationship between worklessness and mortality at both the ecological and individual levels is reviewed. Thirdly, the question of whether the relationship is a causal one is addressed through an examination of possible mechanisms by which worklessness may cause mortality. This section is an important component in determining the plausibility of a potential causal effect. The fourth section presents an outline of evidence illustrating how this causal explanation is complicated by confounding effects, particularly in light of recently published evidence. At the conclusion of this chapter, the main findings are summarised and the three key empirical contributions of the thesis are outlined.

Chapter three outlines a methodological framework for the thesis, and, similarly to chapter two, aims to present a more detailed discussion and review than is possible in each of the empirical chapters. Details of the specific methods used in the analyses can be found in each of the separate journal papers and thus did not warrant inclusion in this section. However, the main theme of this chapter is the methodological issues and difficulties associated with the identification of causal effects from observational data generally and the worklessness and mortality relationship specifically. The chapter begins by explaining the notion of causality from within a counterfactual framework and introduces the concept of the counterfactual and how these may be analysed using a potential outcomes approach. It then discusses the
specific problems of bias that result from the problem of confounding in observational studies generally and illustrates with examples from the worklessness and health context specifically. Finally, the chapter concludes with a discussion of the Rubin Causal Model which is a broad framework of steps and analytical techniques for the examination of causal effects and one that is used in chapters five and six. Given the methodological focus of the thesis, it was decided to outline the precise research questions of the thesis at the conclusion of this chapter. This allowed for the fact that the three research questions grew out of both the empirical literature in chapter two and the technical methodological literature in chapter three and therefore helps create a narrative to the thesis. Two key limitations of the data that is used in the thesis are also outlined in this section.

In the first of the results chapters, the issue of health related selection of various labour market groups is addressed. It examines in detail a particular approach to mitigating health selection that has been used in previous studies of worklessness and mortality. In the second and third results chapters I examine the question of causality in the relationship between different forms of worklessness and mortality. The results from the first study are used to justify an alternative analytical approach based on the framework identified in chapter three. The final chapter of the thesis brings together and summarises the various findings from chapter’s four to seven and relates these to the original aims of the thesis. In addition, it identifies a number of common ideas that run through the findings and provides a discussion of these themes and their implications for the wider literature. These build on concepts that are developed in chapter three including over-adjustment in statistical models and issues of comparability and exchangeability as it applies to comparisons of different forms of worklessness. The final chapter also places these findings in the context of wider debates in social policy and welfare reform and reflects on their wider implications for theory, policy, and possibilities for future research. A discussion of the limitations of the studies and some final summary remarks conclude the thesis.
Chapter 2

WORKLESSNESS AND MORTALITY: A REVIEW OF THE LITERATURE

2.1 Introduction

The main aims of this thesis are to develop and explore recent methodological advances in observational epidemiology in order to test for a causal relationship between labour market disadvantage and mortality. An underlying theme that will develop throughout the thesis generally, and chapter three particularly, is the analytical difficulties associated with achieving this. However, it is argued that although many of these difficulties can be overcome through the use of advanced analytical methods, an important component of these methods, and the underlying theoretical framework that underpins them, is the need for careful and detailed understanding of the causal mechanism that links the variables of interest. The process of estimating a causal effect is impossible without this understanding (Rubin 2008). This represents the main aim of this chapter.

As well as meeting this objective, the chapter also contains material that is relevant for chapter seven, where the findings are interpreted and the wider implications in terms of theory, policy and methods outlined. As such, the chapter has two broad aims which structure the chapter. In section two, and in order to lay the foundations for the discussion in chapter seven, a political economy perspective is used to examine changes in the UK economy during the post-war period and the associated labour market and social policy responses. It provides the context for a discussion of the increase in health related worklessness in the UK which represents an important component of the research contained in the thesis. This section also notes the contemporary economic context following the global financial crisis of 2008 which has seen increasing interest in the issue of unemployment and
worklessness. Much of the discussion in the final chapter will return to these ideas in more
detail in order to outline the broader contribution of the findings in the thesis.

In section three onwards, prevailing empirical literature that is relevant for the main aim of
thesis is examined. Thus, it begins with a broad overview of the scope and extent of
empirical research that has documented the statistical association between worklessness and
health and includes both ecological studies and those conducted at the micro or individual
scale. Here, the material is organised into groups depending on the country of study in order
to determine the state of the evidence base in the UK in comparison to other countries. This
section will also begin to touch on the difficulties of interpreting this association for the
reasons that are expanded upon in the subsequent sections which focus on the notions of
causality and selection in sections four and five respectively. These latter sections, and the
evidence that is examined within them, are the crucial parts of the causal effects jigsaw.
They unpack and make clear both the hypothesised mechanism of the causal effect and the
difficulties in isolating that effect, all of which are necessary moving into the methodological
discussion in chapter three.

The literature covers a number of disciplines including public health, sociology, economics
and geography, and contains both quantitative and qualitative empirical evidence. The
approach to identifying appropriate empirical studies was to search for relevant terms on
various academic databases and search engines followed by hand searches of bibliographies
and reference lists. This particular area benefits from a number of systematic reviews, meta-
analyses and general review papers and these were used to help identify key material. They
also precluded carrying out an original meta-analysis of the literature. Each section contains
a brief summary which will flag the importance of the material for the main aims of the
thesis. To conclude this chapter, and to take forward the remainder of the thesis, the key
areas in which the thesis will contribute are drawn together; these are then developed more
fully alongside the research questions at the conclusion of chapter three.

2.2 The UK labour market and welfare state

Politics and the organisation of the welfare state are important determinants of health and
health inequalities (Bambra, Fox & Scott-Samuel 2005). For example, if some of the health
effects of worklessness originate in changes to income and material wealth, then the level of
wage replacement provided by state welfare support becomes an important factor. It is
therefore important to outline firstly, a broad based theoretical framework of how these
wider political and economic factors might influence health more generally and secondly, in
light of this, an account of how historical processes in the UK have shaped and influenced
contemporary labour market patterns. These historical processes have been heavily
influenced by two key periods in the history of capitalism since the Second World War. Each can be seen as the result of two crises of capitalism, both of which fundamentally restructured the capitalist system. The crises and subsequent reforms of the welfare state markedly changed the state’s relationship towards the economy and labour market.

This section will examine the broader policies that were implemented during the period after the Second World War, and the potential implications of these changes in terms of the ways in which they reflected wider changes in the societal and political attitude towards unemployment and worklessness. The methodological focus of the thesis precludes a thorough examination of these issues but, nonetheless, this section aims to illustrate and frame the remainder of the chapter by linking the effect of worklessness for health and mortality to these wider societal and political changes. In doing so, it also sets up the discussion in chapter seven in which the findings from the thesis are linked to wider contemporary debates in welfare and labour market policy.

2.2.1 The Political economy and the welfare state

Since the publication of the Black report in 1980, which highlighted significant disparities in health outcomes across a range of social and demographic groups, the issue of health inequalities has become a focus for health research. The persistent observation of elevated rates of poor health amongst workless individuals is one such inequality. The Black report (1982) offered four explanations for these inequalities: cultural-behavioural, materialist, psychosocial and life course (Bartley 2004). These four explanations are developed in more detail in later sections of the thesis in a discussion of the mechanism by which worklessness is theorised to effect health (section 2.4). More recently however, increasingly political explanations have begun to be advanced in the literature which instead focus on structural and institutional influences on health that are separate from more individual-focused explanations (Navarro and Shi 2001; Bambra, Fox & Scott-Samuel 2005; Beckfield and Krieger 2009; Bambra 2011b; Muntaner et al. 2012; Navarro et al., 2006). Though health may be influenced by both access to certain material resources (materialist) and a lack of control and self-determination (psychosocial) that can accumulate and act at various points in the life course, it is in fact upstream factors such as the political system, the economy and the configuration of the labour market and work that are the real root of social inequalities in health (Bambra 2011b).

In Bambra’s (2011a and 2011b) model of the political economy of health inequalities, the social determinants of health and the unequal exposure to these arise as an outcome of socio-economic class, which itself drives health inequalities through the manifestation of poorer health at the individual scale amongst those of lower socio-economic position. Positioning of
the social determinants of health in this way allows for both the overarching and upstream effect of politics and the political system (Bambra, Fox & Scott-Samuel 2005). These act to condition and mediate exposure to the social determinants of health through exerting control over the differentiation of social classes via the mechanisms of state (de)regulation of the economy, the organisation of work and the labour market and the roles of the welfare state. As a result, ‘political power relations and the economic system set the landscape and the structural parameters within which the social determinants of health operate’ (Bambra 2011a pg. 748). Politics and the political system operates as both a direct macro influence on structures and interventions but also as the overarching structure which conditions each of the intervention points along the pathway to public health outcomes. Thus, public health and patterns of social inequalities in health become inextricably intertwined within both the economic and political spheres (Bambra 2011a). From this perspective, work, worklessness and the socio-economic structures within them become crucial components of the underlying root cause of health inequalities and, as a result, the welfare state becomes one of the key political and economic interventions in the study of public health (Bambra, Fox & Scott-Samuel 2005).

The precise configuration of welfare state systems varies considerably between countries and is the subject of a branch of theoretical work known as welfare state regime theory (Esping-Anderson 1990). Welfare state regime theory suggests that certain countries can be grouped in terms of differential or divergent configurations of national welfare states to the crises of capitalism that occurred during the 20th century (Esping-Andersen 1990). In this work, the key discriminating factor for these groupings is the notion of de-commodified labour, which refers to the extent to which “individuals and families can maintain a normal and socially acceptable standard of living regardless of their market performance” (Eikemo and Bambra 2008 pg. 4). Conversely, the degree of labour commodification refers to the “extent to which workers and their families are reliant upon the market sale of their labour” (Eikemo and Bambra 2008 pg. 4).

Early work on this topic has suggested that welfare regimes with more de-commodified labour markets are associated with healthier populations and a lesser degree of health inequality (Black 1982). However, more recent work which has focused specifically on the latter of these two public health concerns has found less evidence for this. For example, it has been shown that the greatest income related health inequalities were not always to be found in the liberal regime countries with many Scandinavian countries also showing significant health inequalities (Eikemo et al. 2008a; Eikemo et al. 2008b; Mackenbach et al. 2008). Similarly, a recent development in the literature and one of significance for this thesis is the role of the welfare state in mediating health inequalities between the workless and the
in work. For example, Bambra and Eikemo (2008) studied the relationship between unemployment and two self-reported health measures (limiting long term illness and general health) and compared the findings between five different welfare state regimes throughout Europe. Though no distinct pattern was apparent in the incidence of limiting long term illness, the findings were interesting in the sense that the welfare regimes with the greatest health inequalities in terms of self-rated general health included the Scandinavian countries, as well as those in the Anglo Saxon (UK and Ireland) and Bismarckian (e.g. Austria, Germany, Netherlands) regimes. In contrast, countries in the Southern Europe regime (e.g. Greece, Italy, and Portugal) and Eastern Europe regime showed the fewest differences in health between the unemployed and in work (Bambra and Eikemo 2008). Whether this is evidence that the welfare state has little effect on the health outcomes of the unemployed is difficult to determine; there is likely to be strong country-specific cultural influences that may affect self-rated health measures and mask effects that may occur with more objective measures such as mortality. Such work has yet to be undertaken in the literature.

In terms of absolute health outcomes, evidence for a relationship between welfare state regime types and health appears to be stronger. For example, Coburn (2004), using the welfare state typology developed by Navarro and Shi (2001), utilises data on both infant and general mortality to conclude that, in general, the neo-liberal policy agendas that have been pursued by countries in liberal type welfare regimes have resulted in worse performances in terms of these absolute indicators. Similarly, Bambra (2006) and Chung & Muntaner (2007) also find evidence that rates of infant mortality vary by welfare state regime type with similar conclusions to those of Coburn (2004). Elsewhere, another study suggested that, of the cross-national variation in self-rated health, around half could be attributed to welfare state regime type with those in the Anglo-Saxon and Scandinavian regions showing significantly reduced levels of poor self-reported health (Eikemo, et al. 2008a).

This literature represents an alternative approach to the traditional explanations of health inequalities outlined by Bartley (2004). Rather than focusing on individual and collective behaviours, it places the determinants of health squarely within the realm of the political and the economic spheres. In other words, it moves beyond explanations that have traditionally been focused more on the unit of the individual as a site of intervention and instead argues more for changes at a broader structural level. Importantly, the control and regulation of the structures of work and worklessness is crucial in political economy explanations. It therefore provides a useful framework with which to understand and explain the relationship between worklessness and mortality and an important lens through which to interpret the findings of the thesis. The next section builds on this broad outline of the political economy approach
and presents a historical account of economic and political change in the post war period in the UK and the concomitant effects for the labour market and welfare state policy.

2.2.2 The UK economy, labour market and welfare state

Having outlined the broad concept of political economy explanations for health and health inequalities and the role of the welfare state within this, the discussion will now move on to the historical context of the UK labour market, viewed through the lens of the political economy. Figure 2.1 illustrates the UK unemployment rate from 1900 up until 1999. It illustrates, in the post-war era, two distinct periods of the economic history of the UK; the full employment period from 1945 until the early 1970s when unemployment rates remained low and stable, followed by the period from the mid to late-1970s onwards when rates began to rise following marked shifts in the organisation of economic production. In the following section these periods are examined in order to understand the context of current labour market patterns in the UK, particularly in terms of the growth in the numbers of individuals who are out of work due to sickness or disability, which will be discussed in more detail in the subsequent section. In particular, the discussion in this section will examine changes to the provision and entitlement of welfare and state support for the workless during these periods. A central and overarching tension in the debate around the provision of state-led welfare benefits payments is the degree to which entitlement should be based on citizenship or whether they should be based on a system of individual responsibility. This section will chart the development of certain welfare state policies in relation to this key debate.

Figure 2.1: UK Unemployment rate (%) 1900-1999.

![UK Unemployment Rate Graph](image)
2.2.2.1 Full employment: 1945 to the early 1970s

Figure 2.1 illustrates the trend in unemployment rates in the UK since 1900 until 1999 and shows that, after the Second World War, unemployment rates in the UK remained stable at around 1-2% (Broadberry 1991). This low rate of unemployment occurred alongside the post-war adoption of Keynesian economic policies that put ‘full employment’ as a key objective of social policy. In a white paper produced in 1944 on Employment Policy it was stated that ‘the government accept as one of their primary aims and responsibilities the maintenance of a high and stable level of employment after the war’ (cited in Broadberry 1991 pg. 232). The commitment to full employment during this period formed part of what came to be known as the post-war settlement and was a key component of the boom period following the Second World War. During this time, negotiation with unions was successful in reducing wage demands amongst workers, which, in turn, allowed control over inflation. Because low inflation could be maintained by these wage negotiations, stable low unemployment rates or ‘full employment’ was achieved which remained stable without the normally concomitant effects of inflationary pressure. The post war settlement therefore reflected governments’ and unions’ desire for full employment at the cost of lower wages and perhaps reduced productivity (Broadberry 1994).

During this period, access and entitlement to welfare state benefits were made on the basis of citizenship. This notion of welfare provision has its roots in the work of Marshall (1996) who developed the concept of citizenship based around three key elements; the civil, the political and the social, of which the social is of the greatest concern for this thesis. For Marshall, the continued operation of capitalist modes of production necessitated the presence of a welfare system based on these principles to ensure a degree of social equality. This was deemed important for the maintenance and reproduction of the capitalist system because it provided a check on its worst excesses and cushioned its tendency for inequity (Dwyer 2002). Bambra (2011b) notes that maintaining the purchasing power of a surplus population in a Fordist system through the provision of relatively generous benefit payments was beneficial rather than in opposition to the demands of capitalism in its Fordist form. This is because it was necessary for the maintenance of political legitimacy and for placating labour groups as well as stimulating domestic demand for the mass-produced goods and services that characterise domestic Fordist production (Bambra 2011a).

This perspective indicates the essence of a Fordist mode of production whereby increasingly efficient methods of mass production (which are able to produce lower cost goods in greater quantities) service increased consumer demand which itself is kept high through the provision of generous, extensive and unconditional provision of social security (Bambra 2011b). These policies include demand side and redistributive economic policies that
ensured that workless members of the labour market were able to benefit from increased productivity and a restriction on collective wage bargaining which allowed continued accumulation of capital through sustained control of wages (Bambra 2011b). As a result, the workless were able to access basic social welfare payments on a passive and unconditional basis throughout much of the post war period and up until the early 1990s.

Inflationary pressure during the late 1960s brought the period of low unemployment and low inflation to an end. Increased union activity in terms of negotiating for rises in workers’ wages was deemed to have increased inflationary pressure and contributed to poor productivity figures in the UK. This thinking shifted economic policy from striving for full employment to a focus on controlling inflationary pressures. Here, it is worth noting the general influence of the concept of the Non-Inflationary Rate of Unemployment (NAIRU) in neo-classical economics which describes the relationship between unemployment and inflation (Friedman 1968). Specifically, the NAIRU describes the threshold rate of unemployment below which inflation is no longer increasing (Layard and Nickell 1985). The crucial component to this, and a clear departure from the Keynesian ideal of full employment, is the notion that unemployment is a necessary precursor for lower and thus controlled inflation. By no means an uncontested theory within the broader macro-economic literature (e.g. Mitchell 2000; Lindsay 2003; Stockhammer 2004), it nonetheless represented the dominant approach to neo-classical macro-economic policy in the period after the post war consensus.

2.2.2.2 High-unemployment: Late-1970s onwards
The large-scale decline of Fordist modes of production following the capitalist crisis in the 1970s brought about the end of the full employment era, and, as illustrated in Figure 2.1, marked the beginning of a period in which unemployment rose dramatically during the period between the late 1970s and mid-1980s. Although unemployment began to decline again in the period before the recession of the early 1990s, it did not do so to levels which were comparable to the 2-3% level which characterised the post war-period. Instead, this period was characterised by unemployment rates that were far more variable and volatile, and that never dipped below 5% until the early 2000s. Thus, in terms of unemployment, the capitalist crisis of the 1970s brought about significant changes to the ways in which unemployment was viewed politically and socially, and reflected fundamental changes to the ways in which capitalism operated. In particular, it ushered in an era of increasingly neo-liberal economic policies.

Broadly, neo-liberalism and the new post-Fordist modes of production sought to implement systematic deregulation of national economies, the opening up of domestic markets and changes to production systems that included a new emphasis on innovation and economic
Chapter 2: Worklessness and mortality: A review of the literature

growth in the goods and services sectors (Jessop 1994). Correspondingly, the labour market became characterised by increasingly short-term and temporary contracts, greater wage flexibility and moves to encourage the young and workless to take lower wage jobs (Dicken 2003). In the UK, this context was an important precursor to Margaret Thatcher arriving in power and helped to usher in the particular economic policies of the 1980s which, from a neo-classical economics perspective, saw a recovery in productivity and declining inflation at the expense of high rates of unemployment (Broadberry 1991).

A crucial component of the shift from Fordism to post-Fordism was a change in the nature and origin of economic demand for goods and services. As mentioned earlier, a characteristic of the Fordist system was the need to maintain productivity and growth through the maintenance of domestic demand for the new mass production of goods. As demand from international markets and exports grew, the need for domestic demand was reduced. Thus, post-Fordism came to be characterised by new classes of workers, which replaced the mass numbers of affluent workers in the previous era. These new classes were divided into groups of affluent skilled workers and the more precariously and flexibly employed semi- and unskilled workers, together with a surplus of workers who represented the structural workless (Jessop 1994).

This reorganisation of capitalism coincided with (or co-produced) a neo-classical economic discourse which emphasised the role of inflation for economic policy (Tomlinson 1994). Then Chancellor of the Exchequer, Geoffrey Howe, suggested in 1979 that ‘nothing in the long-run could contribute more to the disintegration of society and the destruction of any sense of national unity than continuing inflation’ (in Tomlinson 1994 pg. 263). This fear of inflation, combined with the shift towards a more competitive and productive economy and a flexible deregulated labour market, resulted in a period of tight labour market conditions in the UK (Figure 2.1). As Bartley (1996) argues, although these monetarist policies did indeed produce a “leaner and fitter workforce” (pg. 448), the social costs for those left behind were severe in a number of ways including, perhaps most importantly, for the health and wellbeing of individuals affected by tighter labour market conditions.

These costs to health and wellbeing occurred alongside marked changes to the level of welfare state support for the workless which fell in line with developing discourses of neoliberalism. As the welfare state was no longer required to guarantee domestic demand for the mass produced goods of the Fordist era, the need to guarantee the incomes of the surplus workers was also reduced (Bambra 2011b). Thus, by the time of the Thatcher government in the 1980s, a shift from a citizen-based entitlement to a system which emphasised individual responsibilities had occurred in welfare state discourses (Dwyer 2004). In doing so the welfare state system was subordinated by the priorities of capital accumulation and domestic
competiveness which emphasised the activation of individuals as workers with responsibilities to contribute to this economic system (Dwyer 1998). Thus, rather than focusing on demand side solutions, supply side active labour market policies began to become popular with emphasis on individual self-sufficiency through training, education, job-search and work experience schemes (Houston and Lindsay 2010). Importantly, receipt of any state welfare benefits became conditional on participation in these schemes, and thus, entitlement based on discourses of individual responsibilities became the norm (Heron and Dwyer 1999). These trends in social welfare policy have continued into the 21st century and, as the current period of economic recession continues, are likely to become increasingly important in terms of mediating the relationship between worklessness and health (Bambra 2010).

2.2.3 Health related worklessness

The third research question that is addressed in the thesis examines the relationship between economic inactivity due to sickness and subsequent mortality. In the previous section we have seen the historical context within which unemployment began to rise after the post-war period and have briefly reviewed the 1980s as the period during which unemployment rose to levels unprecedented in the post-war period. However, in order for this question to be addressed adequately, a deeper understanding of the complex relationship between de-industrialisation, demand in local labour markets and patterns of ill health amongst the working population is needed. An important aspect of this is the extent to which the social costs associated with the economic conditions of the 1980s were distributed equally across different groups in society, in particular, to what extent the costs of flexible labour market policies fell disproportionally on the unskilled and less educated workers and those in the manufacturing sector and amongst those with poorer health. In the following section these ideas are investigated together with an examination of the current wave of welfare state reform which has as its main focus individuals who are IDS.

2.2.3.1 Hidden unemployment and sickness

A striking feature of the 1980s was the decline in the numbers employed in the manufacturing and extractive industrial sectors of the economy. Figure 2.2 illustrates this decline and shows that the majority of the unemployment that occurred in the 1980s was concentrated in the manufacturing sector. For example, between 1980 and 1990 total numbers employed in the manufacturing sector and extractive industries, declined by around a fifth compared to around a 20% increase in the service sector. The majority of jobs in the manufacturing sector required unskilled manual labour and as a result, the demand for unskilled labour in the UK during the 1980s collapsed (Nickell and Bell 1995).
Alongside this, the numbers of people who are registered IDS has been rising steadily since 1981 amongst both male and female members of the population (Beatty et al. 2009). In 1981, the number of male long-term sickness claimants was 1.6 million compared to 2.5 million by the end of the 1990s (Beatty, Fothergill & MacMillan 2000). This would appear, at first glance, to suggest that levels of poor health have been rising amongst the general population. This conclusion can be challenged given absolute decreases in mortality (Leyland et al. 2007) and, as a result, an alternative explanation has been developed that focuses on demand side deficiencies in local labour markets.

Many studies have charted the rise in incapacity benefit claimants (Berthoud 1998; Gregg and Wadsworth 1998; Armstrong 1999; Sutherland 1999; Tomaney, Pike & Cornford 1999). Some have highlighted the discrepancy between rising inactivity and unemployment in specific cities in the UK (Turok and Edge 1999), or have attempted to build an alternative definition of unemployment for certain areas in the UK (Green and Owen 1998). Much of this work has suggested that, rather than a deterioration of population health, a more likely
explanation might be that increasing numbers of the unemployed are being ‘hidden’ amongst the group of sickness claimants. This explanation has gained further empirical grounding (Beatty and Fothergill 1996; Beatty and Fothergill 1997; Beatty et al. 1997; Beatty and Fothergill 1999a; Beatty and Fothergill 1999b; Beatty and Fothergill 2000; Beatty and Fothergill 2003a; Beatty and Fothergill 2003b) and has been theorised by Beatty, Fothergill and Macmillan (2000).

The key component to the idea of ‘hidden unemployment’ is the idea that at any given time ill health or disability is spread amongst all three groups of in work, AU and IDS. Although prevalence is highest amongst the AU and IDS, the group of in work will also contain individuals who have managed to hold down a job despite illness or disability. Some of these individuals will be no less healthy than those who are recorded as IDS but are able to remain in employment because their illness or disability does not impact on their ability to carry out their work. Beatty et al. (2000) define the severity of this illness as ‘sufficient for them to qualify for sickness-related benefits should they wish to do so’ (pg. 9). Put simply, the measure of sickness, as it is measured in terms of the assessment for sickness benefit, is likely to apply to individuals residing in all of the employment groups rather than exclusively the IDS group. Sick members of the AU group therefore constitute a group of ‘hidden sick’.

Bartley and Owen (1996) used data from the period from 1973 to 1993 to show that poorer health was strongly associated with increased job insecurity and risk of job loss particularly for those in manual occupations (Figure 2.3). This suggests that during periods of mass redundancies, although both sick and healthy individuals will lose their jobs, the ‘hidden sick’ and those in lower social classes will suffer disproportionately. The healthy redundant will be recorded as AU together with a limited number of sick individuals who continue to look for alternative employment and therefore reside in the group of AU. Conversely, those individuals who suffer a ‘significant illness’ or one that qualifies them for sickness benefit will flow into the group of IDS during periods of tight labour market conditions. Beatty et al. (2000) refer to this initial process as the redundancy effect, which can be summarised as an increase in not just AU but IDS as a result of movement of ‘hidden sick’ on to incapacity and sickness benefits.
A second stage of hidden unemployment, referred to as the benefit shift, occurs as a result of increased competition for jobs. Beatty et al. (2000) describe the sorting of AU individuals into more or less advantaged positions with respect to job prospects. This competition is a queue in the sense that the most advantaged and therefore healthy members of the labour force will join the job queue nearer the front and the least healthy nearer the back. It has been established for example that potential employers are often reluctant to take on individuals with a mental illness on the basis of concerns around trust, supervision and inability to deal with the public (Biggs et al. 2010).

Therefore, in a slack labour market the pool of ‘sick’ AU are likely to remain AU for a considerable period of time cultivating feelings of disillusionment associated with prolonged separation from employment. Eventually these individuals will flow into the group of IDS. The two effects together constitute both ‘hidden unemployment’ and ‘hidden sickness’ which can be imagined as an increase in recorded sick over and above that which might be expected in a fully employed economy. As Beatty et al. (2000) note, it is the notion that levels of sickness are determined by the benefits system and job loss as well as health.
Finally, it is necessary to highlight a number of additionally important elements of this account of AU and IDS. The first of these are the implications for definitions of unemployment for the purposes of national labour market statistics. For example, enumerating the unemployed on the basis of claimant counts is likely to underestimate the true level of AU. Furthermore, this underestimation is likely to affect definitions of AU from the International Labour Organisation (ILO) which is often regarded as the preferred measure of AU. A key component of the ILO definition of AU is the need to be actively searching for work which, in the UK context, would remove this group of hidden unemployed from the ILO AU figures. A second and highly important consideration is the notion that the model does not suggest a change in the overall level of sickness amongst the population beyond that which may be attributed to the deleterious health effects of labour
market separation. Rather, previously invisible sickness that was hidden amongst the group of in work may appear as these individuals become eligible for sickness benefits. These elements combined would appear to suggest that the official measures of unemployment are of less relevance as a measure of labour market slack in areas of greater labour market disadvantage (MacKay 1999).

The geographic distribution of hidden unemployment reflects very specifically the various trends in economic change during the 1970s, 1980s and early 1990s that were discussed earlier. The maps below illustrate this geography, particularly the high concentration in areas such as Scotland, Wales and Northern England. Unsurprisingly, it is these areas that have suffered the most in terms of the decline in manufacturing and heavy industry. For example, despite great decline in the coal mining industries in the decade 1981 to 1991, the levels of unemployment in these areas remained constant or returned to a lower level (Beatty and Fothergill 1996). According to the theory of hidden unemployment the majority of these former workers bypassed AU and diverted into IDS (Beatty and Fothergill 1996). Conversely, as the maps highlight, in areas such as London which are considered on average to be the most prosperous, the prevalence of sickness is around 5% of the male working age population, compared to around 20% in areas that have experienced consistent and large scale job loss such as South Wales, North East England, Merseyside and Clydeside (Beatty and Fothergill 1997).

Beatty et al. (2000) explain this geographic pattern by considering two hypothetical areas A and B. Area A is characterised by a predominance of heavy industry and a slack labour market because of high rates of pre-existing unemployment. The nature of work in the industries in this area is likely to result in high rates of cancers and respiratory illness which will increase the level of both recorded and hidden sickness and therefore premature death. In area B there is a tighter labour market resulting in lower pre-existing unemployment with much less heavy industry and therefore lower rates of both hidden and recorded sickness. The allocation of individuals onto unemployment or sickness benefits will clearly not be uniform across these two areas. In area A, a decline in the manufacturing industry would release a large number of hidden sick who are unable to compete for jobs in an already slack labour market. The likelihood is that the majority of these will flow onto sickness benefits but the overall rate of recorded unemployment will remain fairly constant. In the example of Area B, a similar level of job loss will release individuals into a less competitive job market and with fewer numbers of hidden sick amongst them it is unlikely that the numbers of IDS will rise, whereas the rates of recorded AU will almost certainly go up.
Despite these ideas of hidden unemployment there has been a lack of engagement with the possible independent health effects of IDS (Bartley, Ferrie & Montgomery 2006). The few studies that have examined this question are mostly to be found in Scandinavian countries (Quaade et al. 2002; Åhs and Westerling 2006; Wallman et al. 2006; Karlsson et al. 2007). In the UK for example, the census measure of permanent sickness in the census has been little used to study the effect of economic inactivity due to sickness for mortality outcomes, apart from at older age groups (Akinwale et al. 2010). This therefore represents a key area to which this thesis can contribute. However, the interpretation of the findings from such a study need to be carefully considered in light of both the current welfare reform agenda and wider discourses of dependency, paid work and the reserve army of labour hypothesis. To conclude this first section, the current welfare state reform agenda is examined in more detail to provide a context to a study of the health effects of IDS.
2.2.3.2 The current UK welfare reform agenda

In section 2.2.2 it was noted that, alongside the crisis of capitalism that occurred during the 1970s, a number of profound changes occurred to the general system, provision and entitlement of welfare state support for the workless. These broad discourses of rights versus responsibilities had largely ignored the group of IDS who had for the most part retained unconditional entitlement to welfare benefits. However, more recently and in light of the dramatic rise in numbers of incapacity benefit claimants, the latest rounds of welfare reform have switched their focus to reform of the system of welfare payments to the IDS (Lindsay and Houston 2011).

The general shift towards more active labour market policies and this more recent shift in attention towards the IDS has occurred against the backdrop of the growing strength of a number of important discourses. Chief amongst these is the notion that the unconditional rights-based entitlement to welfare that was described by Marshall (1992) is no longer appropriate and likely to create passive welfare dependency (Dwyer 2004). In conjunction with this, there is the longstanding discourse that emphasises the benefits of paid work in terms of health and wellbeing (Cole 2008), as well as accounts of worklessness from neo-classical economics that echo the Marxist concept of the reserve army of labour. Collectively these discourses provide a powerful contemporary legitimacy for policies that seek to ‘activate’ individuals on incapacity benefit back into the workforce where previously such policies would have been considered unjust (Dwyer 2004).

Political economy accounts of the welfare state indicate its role as a disciplining mechanism that seeks to reorganise the labour market in ways that make it conducive to the very specific economic conditions of the post-Fordist capitalist system. From this perspective it is perhaps not surprising that the UK has seen a gradual ‘creeping conditionality’ in terms of welfare entitlement (Dwyer 2004). Just as a passive type of welfare entitlement gave way to more active-responsibility based provision for the AU, the current transition appears to be culminating in the extension of workfare type policies into the domain of the IDS (Bambra 2011a). The latest round of welfare reform resulted in complete overhaul of the incapacity benefit payment, which the labour government replaced with the employment and support allowance (ESA), which itself will be replaced in time by a coalition government policy of universal credit (Lindsay and Houston 2011). The ESA requires that all existing claimants undergo a medical examination with all but the most seriously ill (the majority) being moved onto a lower rate benefit payment. Receipt of this lower rate is on the basis of participation in employability initiatives; failure to do so results in transfer to the lower rate job seekers allowance (JSA). The remaining individuals deemed too sick to work to receive the unconditional ESA plus a support premium (Bambra 2011a; Lindsay and Houston 2011).
Broadly, the implications of these reforms are that the majority of individuals previously receiving incapacity benefit will experience a real decrease in benefit payments coupled with increasingly stringent and conditional entitlement rights. Within the group of IDS, this new system is therefore characterised by newly defined groups of deserving/undeserving sick and able/unable to work which will become crucial in determining access to unconditional welfare (Bambra 2011a). On the basis of what has been outlined in this section, the findings in this thesis will need to be interpreted carefully in light of the potential political and policy applications that they may speak to. In light of the context of welfare state reform for the IDS, interpreting any findings of the health effects of this form of labour market disadvantage are, therefore, complex. The discussion section in chapter 7 will return to these complex debates to provide a discussion of the findings in this thesis in relation to this policy agenda.

2.2.4 Summary

This section has summarised the key debates around economic change and capitalist reform and the associated reforms to the provision of welfare state support for the workless. The main purpose of this section has been to provide both a theoretical framework through which an analysis of the relationship between worklessness and health can be interpreted, and also to provide a context for a later discussion in chapter seven in which the policy implications of the findings in this thesis are discussed.

To begin, this chapter outlined a broad conceptual model that sought to explain the social determinants of health from a political economy perspective, highlighting the importance of the welfare state in determining and producing social inequalities in health. This perspective allowed for different theories of the role of the welfare state in maintaining, facilitating and reproducing the capitalist system of production and labour regulation. Given the importance of the welfare state in political economy accounts of health inequalities, literature examining the potential for differences in population health outcomes across welfare state regimes was then reviewed. Using the ideas from this conceptual model, attention then turned to the ways in which this welfare state capitalism and the capitalist system in general experienced crisis during the twentieth and twenty first centuries and the implications of this in the UK setting. In particular, it highlighted the rise in worklessness during the 1980s that accompanied more neo-liberal and neo-classical economic policies. It then highlighted the role that deindustrialisation played in the collapse in demand of unskilled labour during the 1980s, which resulted in a period of greater health related marginalisation in the labour market. This manifested in greater numbers of IDS individuals claiming incapacity benefit which, according to the theory of hidden unemployment, reflected spatially patterned demand side
problems in certain local labour markets. Furthermore, welfare state provision came to be ensconced within discourses of individual responsibility which contrasted with more passive forms of provision which had gone before. These discourses have combined to push reforms in the direction of attaching increasing conditionality to the entitlement to welfare benefit payments for the IDS with potentially important implications for the independent health risks of this form of labour market disadvantage.

In conclusion to this section, there is a clear need to review these ideas in light of the current crisis that afflicts much of the industrialised western world. In section 2.2.2 the capitalist crises of the 20th century were outlined in terms of their effects and ramifications for welfare state provision and the wider labour market and economy. Though there is debate as to what constitutes a capitalist crisis and how they might be distinguishable from the boom and bust cycles of modern economies, the events of 2008 would appear to bear the hallmarks of another such crisis (Gamble 2009). As unemployment rates continue to rise at the time of writing (BBC 2012), it is likely that this new crisis represents a period in which the health effects of worklessness will again rise to prominence in discussions of health inequalities and policy interventions more generally. In the UK, as will be highlighted in subsequent sections of this chapter, much of what is known about unemployment and health originates from studies conducted during previous periods of recession. However, there is good reason, certainly on the basis of what has been outlined in this section, to suspect that the health effects of this most recent crisis may well be significantly different to those that have gone before. This section has highlighted that, compared to previous recessions in the UK, the unemployed in the current era have less of a welfare safety net, are increasingly stigmatised and will perhaps have less social support (Bambra 2010). This, in essence, provides the overarching motivation for the research contained within this thesis which has, as its main aim, to test for a causal relationship between worklessness and mortality during a period of increasing economic uncertainty and recession. Much of the discussion in the final chapter of this thesis will focus on the issues outlined in this section, particularly in terms of the wider implications of the findings in the context of the current economic crisis and direction of welfare state reform and policy. However, for the remainder of this chapter the focus moves to an examination of the relationship between worklessness and mortality in order to inform the main aim of the thesis, which is to test for a causal effect between worklessness and mortality.
2.3 The association of worklessness with mortality

The study of the relationship of AU with mortality in the UK has a long history stretching back around 40 years. During this period, as discussed in the previous section, the post-war economic boom, which saw unemployment rates of less than 2%, began to unravel in the 1960s and 1970s and culminated in large rises in unemployment rates during the Thatcher years (Broadberry 1994). In this time, two distinct strands of research have emerged. In the first, the effects of recession measured at the national scale as a trend in economic activity are correlated with trends in mortality rates. Within this literature there exists significant unresolved debate. In the second, micro-level individual data has been used to measure the association between individual labour market status and mortality relative to individuals in work. In this section, evidence from both will be reviewed. The main purpose of this section is to review both sets of evidence in order to allow for a more detailed examination in sections 2.4 and 2.5 of the two competing explanations for the relationship between labour market position and health and mortality: causality and selection.

2.3.1 Ecological Evidence

For some, continued economic growth during the twentieth century has contributed to the on-going secular decline in mortality rates throughout the developed Western world (Brenner 2005). However, a key debate that has developed within this area is the relationship between economic recession and short term mortality effects. This work is potentially informative for this thesis as in many cases unemployment is often used as the proxy measure for business cycle activity as it better captures the negative effects of economic change than do, for example, measures of economic output such as gross domestic product (Granados 2003). In the following section, these debates are illustrated to highlight firstly, the importance of individual level data to tease out these various effects and, secondly, to highlight between country differences in the effects. Thus, this section reviews the relationship between aggregate or ecological measures of health and mortality across and within countries. In particular, it will focus on how national economic cycles, or alternate periods of economic expansion and recession, co-vary with trends in aggregate mortality rates (Granados 2003).

2.3.1.1 Mortality and economic fluctuations

Harvey Brenner is credited with the earliest attempts to link short term trends in the economy to population-level health (Brenner 1976; Brenner 1979; Brenner 1983; Brenner 1984; Granados 2003). These studies produced the “Brenner hypothesis” which suggests that during periods of economic recession or contraction mortality rates often rise (Brenner 2005; McAvinney 1984; Junankar 1991; Bunn 1979). This pattern has been termed counter-cyclical mortality and has attracted considerable critical scrutiny from a number of
researchers despite the fact that it appears intuitive given the widely held notion that economic growth and development is beneficial for health (McKeown 1976a; McKeown 1976b).

The roots of this debate lay firstly with some critiques of the methods applied by Brenner (Wagstaff 1985; Søgaard 1992) and secondly with those that highlighted contradictory empirical evidence that suggested that mortality rates appeared to rise rather than decline during economic periods of expansion. This pattern of mortality, often termed pro-cyclical mortality, is not a recent empirical discovery and can be traced back to the 1920s (Ogburn and Thomas 1922; Thomas 1925), though the counterintuitive nature of the findings meant that it was given little credence (Granados 2005a; Bezruchka 2009). However, the finding began to be replicated in later empirical work and consequently gained further credibility (Eyer 1977a; Higgs 1979; Adams 1981). Furthermore, during the 2000s, the idea of counter-cyclical mortality grew with evidence from a number of sources highlighting the apparent short-term decline in mortality rates during periods of recession (Ruhm 2000; Brenner 2005; Gerdtham and Johannesson 2005; Granados 2005a; Gerdtham and Ruhm 2006; Granados 2008; Tapia and Diez 2009).

Various attempts at theorising the evidence of pro-cyclical mortality have been developed. Perhaps the most important is referred to as the “Eyer stress model” (Granados 2003 pg. 931) which suggests that surges in the economy often bring about changes in both the ability and desire to practice healthier behaviours such as exercise and social interactions as well as to increase consumption of alcohol and tobacco as a response to increased amounts of working stress (Eyer and Sterling 1977; Eyer 1977a; Eyer 1977b; Sterling and Eyer 1981). More recent evidence has also substantiated this model, with Ruhm and Black (2002) highlighting a decrease in levels of drinking during economic contraction, Ruhm (2005) finding an increase in the level of physical activity and a decrease in excess weight and levels of smoking, and Svensson and Hagquist (2010) finding that adolescent drinking rates decline when the unemployment rate increases. In analyses on both US (2005a) and Japanese (2008) data, Granados has found that mortality from cardiovascular and renal diseases, cancer, traffic injuries, flu, pneumonia and liver cirrhosis was particularly likely to increase during expansions.

However, not all causes of death exhibit a counter-cyclical pattern. For example, it was found that rates of suicides often increase during periods of economic contraction (Granados 2005a). This procyclical pattern of suicide mortality was further replicated in a study comparing national rates of unemployment with rates of mortality in a number of countries in Europe (Stuckler et al. 2009). This study found that unemployment rate increases, in the main, were associated with corresponding increases in the rates of suicide with the exception
Chapter 2: Worklessness and mortality: A review of the literature

of Sweden and Finland. These countries were especially noteworthy in that they managed to ‘decouple’ the unemployment rate trend from the corresponding suicide rate increases. This was interpreted to reflect greater spending on active labour market policies in these countries compared to other countries (Stuckler et al. 2009). These findings together, it is argued, reinforce the ‘stress model’. During cycles of economic expansion, the increased pace of life and work pressures associated with increased productivity increase levels of stress and fatigue. These pressures are manifested in the greater prevalence of chronic diseases which arise from the decreasing levels of health promoting behaviours that occur with a hectic working lifestyle. Conversely, for those that remain in secure employment during recession, the sudden shock of job loss or the persistent anxiety over the possibility of job loss that might accompany economic contraction is not present, which perhaps explains the decline in rates of suicide (Granados 2003).

Another way to investigate these types of questions is to examine the process of rapid privatisation during the early 1990s in some of the communist former Soviet Union countries which provides a natural experiment of the effect of rapid economic and welfare state reform for population health. In Russia in particular, a ‘post-communist mortality crisis’ has become apparent, in which emergence from communism has coincided with rapid decreases in life-expectancy and equally rapid increases in mortality rates (Shkolnikov and Meslé 1996; Leon and Shkolnikov 1998; Shkolnikov and Vladimir 1998; Walberg et al. 1998; Leon 2004; Stuckler, King & Mckee 2009). For example Stuckler, King & Mckee (2009) note that life-expectancy declined by five years between 1991 and 1994 compared to declines of one year in countries undergoing similar transition such as Croatia and Poland. The authors argue that differences in the degree and pace of reform may be responsible for the particularly severe health trends seen in Russia compared to other countries. Russia’s transition to a market based economy was achieved by means of what has become known as ‘shock therapy’, which consisted of the rapid implementation of three main components; price and trade liberalisation, inflation suppression policies and mass privatisation (Fischer and Gelb 1991). Stuckler et al. (2009) show that the rapid adoption of these policies and the concurrent rise in unemployment rates in many of the former Soviet Union countries was associated with short-term increases in male working age mortality. Therefore, in contrast to the pro-cyclical pattern of mortality, this work would appear to suggest that unemployment associated with economic change at the ecological scale appears to hold damaging rather than protective effects for mortality. Thus, for a number of commentators, the suggestion that economic development can actually be harmful for population health is counterintuitive, paradoxical and apparently contradictory to a wealth of literature (Brenner 2005; Catalano and Bellows 2005).
2.3.1.2 Ecological versus individual scales of analyses

As periods of recession or economic contraction are often concomitant with rises in the rate of unemployment, a pro-cyclical pattern of mortality could be interpreted as evidence that the effect of unemployment for mortality may be overstated. However, in reality, it is difficult to justify this interpretation from associations between ecological measures because they only reveal insights about an association between employment and health at an aggregate level rather than the association at an individual scale. They also do not permit analyses of the relative effects of various forms of labour market disadvantage. Thus, critiques regarding the substantive impact of aggregate studies such as these have been widespread (Kasl 1979; Kasl 1982; Goff 1990; Catalano and Bellows 2005; Neumayer 2005). Whilst the stress model of economic expansion that is used to account for pro-cyclical mortality is plausible for those in work, the effect on those who are unemployed is less so. The reduction in levels of stress and the decline in working hours that might accompany economic downturn are all factors that could be imagined to improve an individual’s wellbeing and health, thus giving rise to an improvement in population mortality trends. However, none of these factors would accrue positive health effects amongst the unemployed. This is logical, given that they are not in work and therefore do not benefit from the potentially positive effects that accompany decreased economic activity which seem to mostly focus on improved working conditions.

The observation that overall mortality rates decline during economic contraction is likely to be an observation of the proportionally greater number of individuals who are either in employment or who are economically inactive relative to the absolute numbers of unemployed. Clearly, even in countries with relatively high rates of unemployment, the unemployed will still number significantly less than the rest of the population. As a result, according to the stress model, health improvements resulting from economic contraction amongst the employed will produce a net improvement in health across the whole population despite decreasing health amongst the unemployed. As a result, it tells us nothing of the individual mortality risk of unemployment or worklessness; for that, individual level data and an examination of relative excess risks is required.

2.3.2 Individual mortality and worklessness

Analysing the relationship between mortality and worklessness at the individual scale is more difficult, mostly owing to the difficulties in obtaining appropriate data for what is a relatively infrequent outcome. Thus, in order to test the hypothesis, that, for example, the unemployed may have higher death rates to any degree of statistical rigour requires a large initial sample size which may not be available in many social survey data sources. Secondly,
measurement of death events requires a substantial period of follow-up; this again is often not possible in many traditional social surveys. As a result, death information that is linked to adequate socio-economic contextual information, including employment status, is often more difficult than for other health outcomes. Unsurprisingly, owing to these challenges, the literature examining the effect of worklessness for mortality is slightly less comprehensive than that examining other health outcomes.

The following section reviews this existing evidence. It does so in order to understand the extent of empirical evidence on the topic in order to begin to determine where the thesis can potentially contribute. As such, the purpose of this section is to identify the extent and strength of the reported associations, in order to inform the later sections of the discussion which focus on the causality and selection explanations respectively. Given suggestions in the literature covered in section 2.2.1 that the worklessness experience may be mediated by the various types of welfare state regimes (Lahelma and Arber 1994; Bambra and Eikemo 2008; Stuckler et al. 2009) and to compare evidence from the UK with other countries, the material is organised according to country of origin. Finally, this section will borrow heavily from a recently conducted meta-analysis of the worklessness and mortality literature (Roelfs et al. 2011) which was a useful guide for the comprehensive and systematic identification of relevant studies.

2.3.2.1 Sweden

Sweden is well represented in worklessness and mortality research. Much of this research is made possible by the personal registration number that is issued to all Swedish citizens which has allowed the linkage of various different data sources to a greater degree than has been possible elsewhere (Bartley and Ferrie 2010). A good example of this, and one that is key in terms of this thesis, is a recent study conducted by Lundin et al. (2010) in which information for employment status, socio-economic position, health, health behaviours and other markers of disadvantage across the life course were linked to subsequent mortality information. Findings from this study suggest that amongst individuals currently in work, those who experienced periods of worklessness in the recent past were at a significantly increased risk of all-cause mortality in subsequent follow-up periods. However, with adjustment for various confounding factors this excess mostly disappeared though a significant 30% excess still remained. When examining cause specific mortality, initial excess deaths from suicide and other violent deaths were found to be insignificant when other factors were accounted for. The author’s interpretation of these findings was to conclude that ‘a substantial part of the increased relative risk of mortality associated with worklessness may be attributable to confounding by individual risk factors’ (Lundin et al.
Elsewhere, other findings from alternative data sources seem remarkably consistent with the findings from Lundin et al. (2010). For example, an early analysis of the Swedish Survey of Living Conditions reported a mortality risk of AU relative to a population of employed individuals of 37% which translated into a figure 61% for men and 14% for women (Stefansson 1991). Furthermore, evidence from the Swedish Survey of Living Conditions showed that the unemployed were at a 50% excess risk of mortality particularly for deaths from external causes (Gerdtham and Johannesson 2003; Åhs and Westerling 2006). Another source of evidence in Sweden is the Swedish twin registry which contains information for all twins born in Sweden in the period 1926-1958. Analyses of this data suggest that worklessness again maintained a 50% excess risk of mortality for men (Nylen, Voss & Floderus 2001). Interestingly, for women the effect was higher at around a 100% excess risk (Nylen, Voss & Floderus 2001). A later study of the same data showed similar results that were consistent with longer periods of follow-up (Voss et al. 2004). The evidence from Sweden would, therefore, suggest that the risk of mortality associated with worklessness is roughly between 30% and 50%, a figure which is consistent across a variety of studies and data sources. However, for suicides, the figure has been reported to be higher (Johansson and Sundquist 1997).

### 2.3.2.2 Finland

Similarly to Sweden, Finland also benefits from the availability of data which is extremely useful for the study of worklessness and mortality, the most important of which appears to be based around routinely collected census data. The distinct advantage of many of these studies is the full population samples that they can draw on, which number in the millions of individuals in most cases. The earliest of these studies which was based on census data from the 1980 census and a follow-up period from 1981-1985 found the relative risk of mortality from worklessness to be around 93% (Martikainen 1990).

However, more recent analyses suggest this figure may be exaggerated. For example, an analysis of the 1990 census, with the addition of information from the labour force data files, suggested that the magnitude of the worklessness mortality risk depended upon whether the economy was in recession. For example, they noted that the relative risk varied from over 100% in 1990 when overall worklessness was low to 35% in 1992 during the period of recession (Martikainen and Valkonen 1996). This finding has also been replicated more recently (Martikainen, Maki & Jäntti 2007) although other recent studies put the relative risk at a much greater figure, around 150% or greater in some cases (Pensola and Martikainen 2003; Pensola and Martikainen 2004; Kivimäki et al. 2003c). The evidence from Finland
would appear to suggest that a proportion of the excess mortality in the unemployed can be accounted for by other characteristics. This conclusion, which is supported by the Lundin et al. (2010) analysis represents potential confounding of the worklessness mortality relationship which is a theme developed in section 2.5.

2.3.2.3 United States

Compared to the quantity of evidence available from the Scandinavian countries, evidence from the US examining the relationship of worklessness with mortality is relatively sparse. Evidence that has been collected does, however, suggest that the risks may be higher in the US than has been reported in Scandinavia though this is not consistent across all of the studies examined. The National Longitudinal Mortality Study in particular provided early evidence and benefits from substantial sample sizes. Evidence from this study suggested that the magnitude of the relative risk of mortality associated with worklessness was between 1.6 and 2.2 (Sorlie and Rogot 1990). Later analyses showed consistency with this finding with an identical risk magnitude but also highlighted age variation with the effect decreasing in older age groups (Sorlie, Backlund & Keller 1995). Other data showed significantly higher mortality risks associated with worklessness with, for example, analyses of the US Panel Study of Income Dynamics reporting around a 200% excess risk (Lavis 1998).

A very recent study which found no evidence of a worklessness effect for mortality would appear to highlight inconsistencies in this relationship. This study followed up mortality for individuals made voluntarily as well as those made involuntarily redundant following the collapse of the Pan-Am airline corporation. The findings suggested that, when compared to the general American population, the mortality risk for these individuals was actually lower (Steenland and Pinkerton 2008). However, findings from studies such as these which rely on factory or company downsizing or collapse, are open to critique, a point that will be returned to in section 2.5.3 (Morris and Cook 1991).

2.3.2.4 The UK

Early studies conducted in the UK were somewhat ground-breaking in nature owing to the innovative use of the nationally representative England and Wales longitudinal census study provided by the Office of Population Censuses and Surveys (OPCS). The studies were innovative in the sense that they were amongst the first to utilise the power of linkages between mortality and vital events records and the national census. These linked studies allowed comparison of subsequent mortality risk between those in and out of employment at each census and were thus amongst the first to identify such relationships in samples that could be generalised to the population. Two studies, published in the 1980s, were particularly important. The first reported results from the 1971 census and found that the
unemployed had a mortality rate that was 36% higher than would be expected for all men in the study (Moser, Fox & Jones 1984). Furthermore, only 20-30% of this excess could be explained by the social circumstances of the unemployed (Moser, Fox & Jones 1984). In a later study of the 1981 census these findings were repeated (Moser et al. 1987) and then confirmed with increasing duration of follow-up (Bethune 1996a; Bethune 1996b).

More recent findings from the England and Wales longitudinal study takes a cause specific mortality approach (Lewis and Sloggett 1998) or focus on older age groups (Akinwale et al. 2010). In the first, the unemployed were 2.6 times more likely to commit suicide, though this association was unable to account for various confounding factors including most crucially previous episodes of poor mental health (Lewis and Sloggett 1998). Finally, analysis of the British Household Panel Survey, which although smaller in terms of sample size, benefitted from a rich collection of contextual variables and highlighted the mortality risk associated with increasing years of worklessness (Gardner and Oswald 2004).

2.3.2.5 Evidence from elsewhere

The remainder of the evidence comes from a range of different countries including Australia (Morrell et al. 1999), Italy (Costa and Segnan 1987), Switzerland (Gognalons-Nicolet et al. 1999), Denmark (Iversen et al. 1987) and New Zealand (Blakely, Collings & Atkinson 2003). The evidence from many of these studies is largely similar to the findings reported in the previous countries although a couple warrant further discussion.

The study in Italy is notable in that it was conducted around the same time as the original UK census studies though the magnitude of the effect in Italy was considerably higher with a doubling of the risk (Costa and Segnan 1987). The study from Denmark was conducted on the full population of unemployed (around 22,000) and employed (around two million) and reported adjusted mortality excesses in the unemployed of around 40-50% with deaths from suicides and accidents especially high. These effects were adjusted for a range of individual characteristics including geographic area. Adjustment for geographic area suggested that the effect was reduced in areas of higher rates of worklessness and was more pronounced in areas of lower rates (Iversen et al. 1987). This finding will be discussed in more detail in section 2.5.3.

More recent findings from Denmark also confirmed the relationship between worklessness and suicide though the effect was modest (3%) compared to other studies (Mortensen et al. 2000). A further interesting finding from this study related to the effect of adjustment for previous mental health (from hospital admissions data) which increased the risk of suicide by 45% and reduced the excess risk of suicide amongst the unemployed by over a half. The strength of this association lead the authors to suggest that the relationship between
worklessness and mortality reported in many other studies (e.g. Lewis and Sloggett 1998) ‘may be confounded and overestimated owing to the lack of adjustment for the association with mental disorders’ (Mortensen et al. 2000 pg. 1).

2.3.3 Summary

This section has examined the extent and scope of empirical evidence that details the statistical association between worklessness and mortality and has looked at both ecological and individual level evidence. The main findings to draw from the ecological evidence were that there remains debate about the effect of recession for population health which is often measured in terms of rates of worklessness and that the evidence and types of study conducted make it almost impossible to discern the individual level mortality effects of worklessness. Evidence from individual level studies, grouped according to the country in which they were carried out, was then examined. The main findings from this body of evidence were that (1), in the main studies report a strong and persistent association between worklessness and mortality and suicides in particular. (2) There was significant variation in terms of the origin of the studies though the majority were conducted in Sweden and Finland. (3) There appeared to be significant variation in the magnitude of the effect between countries and studies which, (4), probably reflected significant differences in methods and available control variables. (5) One source of this variation lay with the observation that effect sizes tended to differ in periods and geographical locations of higher and lower rates of worklessness. (6) A final important finding was the importance of serious mental health episodes prior to becoming worklessness.

Although only briefly mentioned in this section, the concepts of confounding and selection represent a significant challenge to the notion that the observed association between worklessness and mortality is a causal one. Thus, alternative explanations focus on the possibility that the unemployed may contain higher numbers of ill people as well people in lower or more deprived socio-economic positions. Although labelled a false antithesis in some quarters, the extent to which worklessness ‘causes’ death is a matter of continuing debate between those that argue for selection and those that maintain that unemployment holds independent and deleterious effects for health and mortality (Bartley 1988).

The observation of significant inconsistencies in the magnitude of the observed effect in many of the individual studies reflects, in the most part, the extent to which the studies under review have to a lesser or greater degree attempted to mitigate the effects of confounding, in order to arrive at a causal interpretation. An examination of these approaches will follow in section 2.5 but for now it is suffice to say many of the studies that benefitted from these more innovative techniques and comprehensive data were to be found in countries such as
Chapter 2: Worklessness and mortality: A review of the literature

Sweden, Finland and Denmark. This suggests a clear need for such studies to be replicated in other countries, particularly the UK and US where studies like these have been less prevalent. This represents the key area to which this thesis will contribute, to test for the presence of a causal relationship between worklessness and mortality in the UK context where such evidence is restricted to studies carried out in the 1980s. This research is especially prescient, given very recent evidence from Lundin et al. (2010) that emphasises the non-causal components of the worklessness and mortality relationship. In the next section, the mechanism through which worklessness is theorised to cause mortality is outlined.

2.4 The causal explanation

As I have already suggested, a key component of any attempt to establish causal relationships is ‘plausibility’ with respect to theoretical and empirical prior knowledge (Hill 1965). The purpose of the following section is to examine this prior knowledge and to assert the plausibility of a potential causal link between worklessness and mortality. This is a crucial part of the analytical process that is outlined in chapter three and is one stage of the three-stage analytical approach that forms the underlying methodology of the thesis. This section is structured as follows: firstly it will discuss the role of financial disadvantage associated with worklessness and include a review of the various ways in which income and material wealth has been linked to a variety of health outcomes. It will then discuss the ‘latent functions’ of work which are conceptualised as the non-financial benefits for working individuals which can have an immediate shock effect when lost due to redundancy. The stress related aspect of this shock is also discussed in relation to possible physiological responses in the body including immune system responses. Finally, the roles of health behaviours amongst the unemployed will be discussed. Much of the broader theories and empirical material linking worklessness to health and mortality maps very closely to the often discussed explanations for health inequalities more generally which were first outlined in the Black report (Black 1982). These ideas have been mentioned briefly in section 2.2.1 but are expanded upon here in relation to the specific relationship between worklessness and mortality.

2.4.1 Financial Disadvantage

To a large extent, the potential for unemployment-related financial disadvantage depends on the levels of state unemployment benefits relative to previous income. As discussed in section 2.2, the period following the capitalist crisis of the early 1970s saw a consistent decrease in the average provision of state benefits for the UK unemployed (Figure 2.6).
The effect of the trend illustrated in Figure 2.6 was a real decline in wage rates of the poorest and, therefore, an increase in inequality across the income distribution. For example, between 1979 and 1991 despite rises in average household income of between 36%, the corresponding figure for the poorest 10% of households actually fell by 14%. Furthermore, by 1991 the share of total income amongst the lowest 10% of households on the income distribution stood at 2.9%, a decline of 1.2% from the corresponding figure of 4.1% in 1981. On the other hand the richest households had corresponding figures of 26.2% in 1991 and 21.3% in 1981 (Bartley, Ferrie & Montgomery 2006). The changes in the labour market during the 1980s and corresponding changes to the welfare system would appear, therefore, to have dramatically worsened the financial situation of the poorest individuals and especially those out of work. Unsurprisingly, the effect of unemployment has been shown to dramatically reduce the living standards of families which continue to decline with
increasing duration of unemployment (Bradshaw, Cooke & Godfrey 1983). The following sections investigate the relationship of income and material wealth to a range of health outcomes.

2.4.1.1 Mental Health

The relationship between financial difficulty and psychological ill health has been demonstrated by a number of studies in a range of countries (Elder Jr and Liker 1982; Elder Jr, Liker & Jaworski 1984; Kessler, Turner & House 1989; Tahlin 1989; Mullis 1992; Lundberg and Fritzell 1994; Lynch, Kaplan & Shema 1997; Thiede and Traub 1997; Benzeval and Judge 2001). In one study from the US which used a longitudinal approach to allow the measurement of financial position over multiple time points, it was found that periods of economic hardship had a dose-response type effect on psychological and social functioning (Lynch, Kaplan & Shema 1997). Evidence from Sweden also appears to illustrate a clear role of income in determining the instances of psychosocial illness although it is also noted that relative income changes are also very important (Lundberg and Fritzell 1994) with similar findings reported in Germany (Thiede and Traub 1997). In terms of understanding a causal link between unemployment and health, this could be important due to the likelihood of relative income to be the important causal pathway. A further interesting finding suggests that long term levels of wealth rather than short term fluctuations are the driving influence for health. In other words, short term fluctuations or shorter periods of income reduction are much less damaging than the longer term patterns of lower income (Benzeval and Judge 2001).

In the UK the British Household Panel Survey (BHPS), which is a longitudinal panel study, has produced a great deal of research on the role of income and financial hardship for individual health. It includes extensive information on health outcomes including the general health questionnaire GHQ which, together with longitudinal information on socio-economic characteristics, makes it a very powerful resource. Some of this work has focused on the reverse effect of increased income by looking at incidences of lottery wins. This is potentially a very useful approach as it enables the researcher to examine the effect of a sudden injection of wealth that is likely to be independent of other factors. Findings from this work showed an improvement in mental wellbeing as measured by the GHQ which occurred two years after the lottery win although initially, and perhaps unsurprisingly, individuals experienced a negative effect during the period immediately after the lottery win (Gardner and Oswald 2007).

Although this evidence suggests that wealth is an important determinant of mental and psychological health and emotional functioning, it is perhaps not as easy to assume that all unemployed will therefore suffer these health outcomes as a result of their poorer relative
financial situation. As we noted earlier, the extent to which unemployment leads to financial disadvantage depends on the type of job that was lost and the level of state benefit that is available to compensate this cost. The unemployment specific knock-on effect of financial problems for the health of the unemployed has been studied in the UK particularly with regard to psychological health as measured by the GHQ. Jackson and Warr (1984) for example, find a link between unemployment related changes in household income and an individual’s GHQ score. Other evidence links instances of financial borrowing with increased risk of depression (Heady and Smyth 1989). In some cases, financial hardship was shown to account for almost all of the association between unemployment and psychological illness (Rodgers 1991). A body of evidence from countries other than the UK including the Netherlands and the US also appear to confirm these findings (Kessler, Turner & House 1987; Klein-Hesselink and Spruit 1992; Leeflang, Klein-Hesselink & Spruit 1992).

2.4.1.2 Mortality and Life Expectancy

Mortality has also been linked with financial disadvantage and economic inequality (Zick and Smith 1991; Menchik 1993; Wolfson et al. 1993; Smith and Zick 1994; Duncan 1996; Pritchett and Summers 1996; McDonough et al. 1997; Rossum 2000). In the US, wealth, permanent income and transitory income were all associated with a greater risk of mortality in older men. Furthermore, a raised mortality risk was also found as the number of spells of poverty increased (Menchik 1993). In a longitudinal study based on routinely collected data which allowed for information on career earnings, a significant mortality gradient was found across the whole distribution of incomes (Wolfson et al. 1993). To smooth out potential short-term fluctuations Duncan (1996) took average family income over five years which showed a strong association with mortality in the US. In addition to this, losses and instability of income over long periods (relative to stable income) were also found to be highly predictive of mortality (Duncan 1996). Similar US findings with regard to income loss were found by McDonough et al. (1997) with an additional finding pointing to the importance of income instability for mortality in middle ranking income bands but not as much in lower bands. Slightly different findings were shown in a longitudinal study where, after adjustment for health, a decline in income was strongly predictive of mortality whereas initial income level was not (Kaplan and Haan 1989). Contrasting with much of this evidence, a Canadian study suggested that baseline income level was the most significant predictor of mortality while income decline or change was found to be insignificant (Hirdes and Forbes 1989). In the UK, the association between individual income and mortality is less well documented due to the lack of a question on income in the national census which is an often used source of individual mortality studies.
2.4.1.3 Self-Assessed and Physical Health

Level of income has also been strongly related to self-rated health (Ettner 1996; Ecob and Davey Smith 1999; Frijters, Haisken-DeNew & Shields 2005; Jones and Wildman 2008) and physical health (Kessler, House & Turner 1987a; White 1991). Evidence from the UK has linked physically measured health indicators such as height, waist–hip ratio and respiratory function as well as self-assessed measures of ‘malaise symptoms’ and limiting long term illness to income. The authors of this study also demonstrated that increases in log income produced a corresponding health effect of a similar positive magnitude though the effect plateaued at the upper end of the income distribution (Ecob and Davey Smith 1999). In Sweden, changes to income both relative and absolute has been linked to physical health although the relationship is dependent upon the level of initial income (Lundberg and Fritzell 1994).

Similarly to the earlier study of lottery wins and health, a study by Frijters et al. (2005) used the reunification of Germany as a natural experiment of the effect of the sudden and dramatic increase in income that was experienced by many individuals in East Germany, on self-assessed health. The findings from this work suggested a modest health improvement (Frijters, Haisken-DeNew & Shields 2005). Again similarly to the findings from studies of psychological outcomes, longer term levels of income appear more important in terms of self-rated health than do short term fluctuations (Contoyannis, Jones & Rice 2004).

2.4.2 The ‘latent functions’ of work

There is little doubt that job loss is a stressful and abrupt event with real implications for health. However, it is unlikely that the resultant levels of stress associated with unemployment are entirely associated with economic hardship, particularly as unemployment is often associated with feelings of loneliness, social isolation and alienation from family, friends and work colleagues (Fineeman 1979; Jahoda 1979; Baum, Fleming & Reddy 1986; Fryer and Ullah 1987; Warr and Bryan 1987; Ezzy 1993; Bartley, Ferrie & Montgomery 2006). These effects have been termed by Jahoda (1979) as the ‘latent functions’ of work which describe those effects of the employment experience which provide more far reaching benefits than those derived solely from financial security. These functions include daily time structures, shared experiences with colleagues, a purpose with goals and aims over and above personal interest and continued and constant activity as well as social and political status. For example, the work environment provides a daily social network which develops through the day-to-day interaction with other colleagues (Jahoda 1979; Jahoda, Lazarsfeld & Zeisel 2002). The daily time structure and routine is also important in terms of ‘traction’, which is the notion that work provides a daily reason to get
up and go on through the day (Bartley, Ferrie & Montgomery 2006). Thus, as Rodriguez (2001) notes from an analysis of different types of welfare state entitlement, the availability of adequate material resources does not entirely remove the negative health effects of unemployment which was found to be worse for means tested benefits compared to non-means tested entitlement.

### 2.4.2.1 Mental Health

The latent factors associated with work, together with the increased opportunities for mental and physical activity that are present during employment, work to improve and instil individual self-esteem and the feeling of respect from others (Warr and Bryan 1987). Much work on this area focuses on the extent to which unemployment can impact on emotional functioning commonly measured by the General Health Questionnaire (GHQ) (Banks and Jackson 1982; Claussen, Bjorndal & Hjort 1993; Hannan, Ó Riain & Whelan 1997; Andres 2004; Thomas, Benzeval & Stansfield 2005), measures of life satisfaction (Winkelmann and Winkelmann 1998), indicators of levels of depression (Burgard, Brand & House 2007) and other clinical measures of common mental disorders (Ford et al. 2010).

Evidence from the UK has found links between unemployment and more severe episodes of poor mental health that require medical consultation or hospitalisation (Montgomery et al. 1999). However other evidence from Denmark suggests very little difference in rates of hospitalisation between those recently displaced from their job and those remaining in work (Browning, Dano & Heinesen 2006). Some work suggests that these effects may be mediated by the context of the job held prior to unemployment including aspects such as the salary level and the level of psychosocial dependence on the job for social status and identity (Nordenmark and Strandh 1999). Furthermore, it is also suggested that the negative effect for mental health can be reversed with re-employment (Kessler, Turner & House 1989) although this may depend fairly significantly on the potential for this re-employment to resolve an individual’s economic problems as well as the extent to which the new job provides perceived longer-term stability and security (Strandh 2000). Furthermore, the possibility of re-employment, measured by individual notions of ‘re-employability’, has also been shown to reduce the negative mental health effects of unemployment (Green 2011).

### 2.4.3 Physical health and stress responses

Given that work provides such a wide range of personal and social functions including financial security and other latent functions, it is perhaps expected that we might find evidence of a physiological response that can be linked directly to the unemployment experience, independently of other factors. However, evidence of a direct relationship of this nature is less prevalent than other health outcomes (Korpi 2001). A number of studies using
Chapter 2: Worklessness and mortality: A review of the literature

data from the Health and Retirement Survey in the US have found evidence of effects of unemployment for physical functioning (Gallo et al. 2000) and heart attacks and strokes (Gallo et al. 2006; Gallo et al. 2004). Furthermore evidence from a wider age range suggests that general visits to hospital or the doctors was 208% and 57% higher respectively amongst the unemployed (Beale and Nethercott 1985; Beale and Nethercott 1987).

Perhaps more interestingly, recent evidence has begun to examine specific unemployment related physiological responses in the body (Kemeny and Schedlowski 2007). For example a recent study has linked the stress of unemployment with negative effects for the immune system, in particular, in terms of the increased probability of low grade tissue inflammation (Hintikka et al. 2009). The authors of this study noted that links between stress and the immune system are important due to the increased risk of a number of other diseases that result from, or are caused by, a weakened immune system (Hintikka et al. 2009). A similar conclusion was reached in a study of women, where some aspects of the immune system appeared to be affected by job loss 9 months earlier (Arnetz et al. 1987). The unemployed have also been shown to differ in cortisol secretion levels, which are markers of a physiological stress reaction, on a daily basis, (Ockenfels et al. 1995) and to have higher blood pressure during spells of unemployment (Janlert 1992). Other studies exploring this link between stress and physiological responses have found similar results (Sorrells and Sapolsky 2007; Capuron et al. 2008; Janicki-Deverts et al. 2008; Picciotto et al. 2008; Leonard and Myint 2009; Miller, Maletic & Raison 2009).

On the whole it appears that the stress associated with unemployment, in particular the effects associated with the decline in stocks of mental wellbeing and psychological health is linked, in a number of ways, to a physiological response. This physiological response is often on the pathway to a number of other illnesses due to its effect on the immune system. It is difficult to isolate exactly what order along the causal pathway much of these effects operate and it is likely that in some cases stress may be a causal factor for unemployment rather than solely an outcome of it. For example, one study of construction workers conducted in Finland showed exactly this pattern; work stress was a contributory factor preceding unemployment but the unemployment experience significantly exacerbated this (Leino-Arjas et al. 1999). The evidence that directly links physiological responses in the immune and other systems with the types of stresses that are apparent in the lives of unemployed individuals does, however, lend weight to the notion that unemployment has an independent effect for health.
Chapter 2: Worklessness and mortality: A review of the literature

2.4.4 Health Behaviours

In this section, I examine evidence that specifically links the unemployment event with subsequent changes in patterns of health behaviour and worsening of self-rated health measures. Increases in smoking, alcohol consumption, drug taking and diet often form part of a ‘coping strategy’ that is noted throughout the literature as one of the main routes through which unemployment may cause mortality (Hammarström and Janlert 1994; Bartley, Ferrie & Montgomery 2006; Roelfs et al. 2011). These changes in health behaviours may therefore form an important part of the causal chain between unemployment and mortality. I will discuss each of the main areas in turn; smoking, alcohol consumption and other behaviours.

2.4.4.1 Smoking

Smoking as an outcome of unemployment has been demonstrated in a number of studies (Hammarström and Janlert 1994; Montgomery et al. 1998; Falba et al. 2005). In Britain, unemployment in terms of past accumulation of more than three years and recent unemployment in the previous year was associated with a 100% and 200% increase in the risk of smoking after adjustment for a variety of factors (Montgomery et al. 1998). In Sweden the finding was replicated amongst a sample of school leavers where twenty weeks or more of unemployment was associated with odds ratios of 2.44 and 3.45 for men and women respectively when compared to employed or people with only short term unemployment previously (Hammarström and Janlert 1994). This gender effect was also evident in a study from the US (Weden, Astone & Bishai 2006). Amongst older workers, involuntary job loss was associated with a greater risk of smoking amongst previous non-smokers than those who remained in work and was also related to increased consumption amongst current smokers who were made redundant (Falba et al. 2005). Elsewhere, however, the effect was only identified amongst younger individuals (Reine, Novo & Hammarström 2004).

2.4.4.2 Alcohol consumption

A number of studies have suggested that a common response to unemployment is to increase levels of alcohol consumption (Janlert and Hammarström 1992; Viinamäki, Niskanen & Koskela 1997; Montgomery et al. 1998; Claussen 1999; Mossakowski 2008; Virtanen et al. 2008; Bolton and Rodriguez 2009). Claussen (1999) examined evidence relating to alcohol use as well as medical examinations relating to alcohol related illnesses in Norway and found significant excesses amongst the unemployed which were reduced with re-employment during the follow-up period. Virtanen et al. (2008) examined employment trajectories in Finland and reported that those individuals on a downward employment status trajectory or those in chronic unemployment states increased their alcohol consumption in
the period of follow-up. Similarly Janlert and Hammarstrom (1992) adjusted for previous alcohol consumption and noted a remaining effect of unemployment for increased alcohol amongst young individuals in Sweden.

In other studies the effect was less strong and was restricted to the involuntarily unemployed (Ettner 1997). Elsewhere, the magnitude of the effect appears to be moderated by education (Broman et al. 1995) or levels of psychic stress (Viinamäki, Niskanen & Koskela 1997) providing evidence that, similar to smoking, alcohol consumption is indeed part of a coping mechanism for stress.

### 2.4.4.3 Other behaviours

Other behaviours that have been linked to unemployment include substance use (Peck and Plant 1986; Hammer 1992; Crofts et al. 1996; Fergusson, Horwood & Lysnkey 1997; Merline et al. 2004), poor diet and eating behaviour (Laitinen, Ek & Sovio 2002), parasuicide (Platt and Kreitman 1984; Platt and Kreitman 1985), criminal behaviours and pregnancy (Fergusson, Horwood & Woodward 2001) and reduced physical activity (Popham and Mitchell 2007). In one analysis of the British Regional Heart Study, the effect of smoking and alcohol consumption was found to be less significant than that for unemployment associated weight gain (Morris, Cook & Sharper 1992). In a study by Peck and Plant (1986), similar negligible findings were observed for smoking and drinking but in addition their remained a moderately significant excess in illegal drug abuse amongst the unemployed. Linked to this, increases in illegal drug-taking, particularly cannabis, was also reported to be higher amongst the unemployed (Hammer 1992).

### 2.4.5 Summary

In this section, evidence which supports the plausibility of a causal link between worklessness and mortality has been outlined. This evidence includes links between health and financial disadvantage, the latent functions of work and mental health, stress and physiological stress responses and changes in health behaviours. The main findings suggest (1); the financial penalty of unemployment is a strong mediating factor and has potential pathways to a number of health outcomes. (2); The loss of the ‘latent functions’ of work associated with unemployment has links to mental health and self-esteem and can also lead to (3); stress responses which have been shown to manifest themselves in physiological responses in the immune system as elevated blood pressure. Finally, (4); the unemployment experience has been shown to elicit a coping response that can lead to changes in various health damaging behaviours.

These factors, collectively, provide a plausible causal pathway between worklessness and mortality but, in order for that causal pathway to be estimated, the role of various ‘selection’
factors that were touched on briefly in section 2.3, need to be examined. This is addressed in
the following section.

2.5 Non-causal factors

Having outlined a plausible pathway through which unemployment could cause mortality, it
is now necessary to focus in more detail on the notion of selection, in particular the direct
and indirect forms of selection that are regularly highlighted in the literature. The first
sections will examine a variety of health conditions and outcomes that are related to
unemployment. In the second section, evidence for the more indirect forms of selection will
be discussed including socio-economic confounding, the accumulation effects of these
factors along the life-course and, finally, the role of personality traits in determining
unemployment. In the final section, I will examine some of the data and methodological
approaches that test this selection effect and will re-examine some of the evidence identified
in section 2.3.2 in relation to this. Similarly to the previous section, the material discussed is
of great importance for the development of the methodological approach in chapter three.
Much of the methodological theory used in the thesis relies on imagining the data as if it had
been generated from a randomised experiment. In order to do this, it is crucial to understand
not just the plausibility of a causal explanation but also the mechanism by which individuals
become unemployed and the factors or characteristics that influence this.

2.5.1 Direct health selection

Direct health selection describes the pattern whereby individuals in poorer health are more
likely to become unemployed than those in better health (Bartley, Ferrie & Montgomery
2006). As a result, the observed excess of poor health amongst the unemployed could itself
have caused unemployment. The association between unemployment and mortality then
simply reflects the ‘selection’ of ill individuals into unemployment rather than the effect of
that unemployment event for health. One of the key findings from section 2.3 for example,
was that poor mental health is often associated with an increased risk of unemployment. The
potential for direct health related selection therefore presents a significant challenge to the
causal explanation outlined in the previous section. The following section briefly
summarises the empirical evidence for health as a determinant of unemployment focusing on
self-rated health, health behaviours and physical & mental health.

2.5.1.1 Physical, mental health and self-rated health

In Finland, poorer health indicated by self-rated measures has been shown to increase the
likelihood of unemployment amongst a cohort of men who were in fixed term employment
at baseline (Virtanen et al. 2005c). Similar evidence was uncovered in a longitudinal study
Chapter 2: Worklessness and mortality: A review of the literature

from Finland using a five-point general health question. Although the health status for those who experienced unemployment was generally worse than those that did not experience unemployment, the difference was apparent from baseline and therefore not a consequence of unemployment itself (Böckerman and Ilmakunnas 2009). A further study from Canada has suggested that ill health related selection of the unemployed or workless was related to the duration of the spell of worklessness (Stewart 2001). This provides evidence of direct health selection because at any given time the stock of unemployed is likely to contain greater numbers of health-related workless because these individuals are more likely to remain in unemployment than individuals who were made workless for reasons other than health (Stewart 2001).

Elsewhere evidence from the Netherlands links many forms of poor health including chronic conditions as well as self-rated health indicators to an early exit from the workforce (van de Mheen et al. 1999). Mental and psychiatric illness was also found to precede unemployment in Norway (Mastekaasa 1996). In the US, the problem of reverse causality in the relationship between increased alcohol consumption after unemployment was highlighted (Ettner 1997).

2.5.1.2 Health Behaviours

Although the role of health behaviours has been shown to worsen in section 2.4.4 following unemployment, it is also likely that individuals exhibiting risky health behaviours may be more likely to lose their jobs in the first place. Empirical evidence persists of the role of health behaviours preceding unemployment, particularly with regard to levels of alcohol consumption and cigarette smoking (Ryan, Zwerling & Jones 1996; Leino-Arjas et al. 1999; Lundin et al. 2010). As well as finding links between self-rated health and unemployment in France, Jusot et al. (2008) suggested that, firstly, men who smoked at baseline were more likely to be unemployed 4 years later and, secondly, that women experienced the same effect for obesity. In a survey of US postal workers, those who were current smokers at the time of hiring were at greater risk of job loss than non-smokers (Ryan, Zwerling & Jones 1996) and in Sweden, smoking and alcohol use in the past were all associated with future unemployment around 20 years later (Lundin et al. 2010). Drug taking behaviours have also been shown to precede unemployment. In a New Zealand birth cohort, cannabis use strongly predicted future unemployment (Fergusson and Boden 2008) a finding replicated in Norway (Hammer 1997).

Alternative evidence suggests that behaviours improve with unemployment owing to the financial restraints associated with job loss (Temple et al. 1991; Roelfs et al. 2011). Studies of other behaviours include alcohol consumption (Hammer 1992; Leino-Arjas et al. 1999), cigarette smoking (Fagan et al. 2007) and levels of physical activity (Matoba, Ishitate &
Chapter 2: Worklessness and mortality: A review of the literature

Noguchi 2003; Jurj et al. 2007) all found improvements following unemployment (Rehm and Gmel 1999; Rodriguez and Chandra 2006; Goel 2008; Virtanen et al 2008).

2.5.2 Indirect Selection

It has been shown that health itself is often a precursor to worklessness; however, health interacts with a variety of other socio-economic factors to bring about an exit from the labour market. For example, as highlighted earlier, Bartley (1996) found that amongst individuals with a limiting long term illness, those in manual occupations were more likely to lose their job than those in non-manual occupations. Furthermore, Arrow (1996) found that, in an analysis of a German cohort of workers, it was not necessarily the effect of health, per se, that preceded unemployment, but the interaction between health and other factors such as gender and the nationality of the worker. As McDonough and Amick (2001) note in a panel study of US workers, ‘evidence that the hazard of labour market exit in the context of perceived ill-health depended on gender, race and education, but in ways that were not constant across each of these social positions’.

Of course these processes, in conjunction with the ‘healthy worker effect’ (in which conversely, healthy members of the population are selected into employment, (Arrighi and Hertz-Picciotto 1994; Li and Sung 1999)) bring great difficulty when trying to ascertain the independent effect of worklessness for mortality. The notion that health interacts with other factors leads us to a discussion of the more indirect factors that may determine job loss. Broadly speaking, indirect selection refers to those forms of confounding that result not from differences in the prevalence of disease but ‘some other characteristics that make an individual more prone to both unemployment and early mortality’ (Bartley and Ferrie 2010).

In attempting to estimate a causal effect, the analysis needs to address the potentially confounding effects of indirect selection which are detailed in the following section.

2.5.2.1 Socioeconomic confounders

The most common of the indirect selection effects are the group of socio-economic and demographic characteristics that are often used in studies of health inequalities; social class, income, marital status, education, housing tenure and ethnicity for example. Typically, in many studies of the unemployment and health relationship, these characteristics (and others) are accounted for because, in almost all cases, the unemployed are relatively disadvantaged in many of these characteristics compared to those in work (Lewis and Sloggett 1998; Morrell et al. 1999; Blakely, Collings & Atkinson 2003; Gerdtham and Johannesson 2003; Akinwale et al. 2010; Lundin et al. 2010; Schmitz 2010; Popham, 2010). The recognition of this, of course, is not new, and the factors identified here are often a source of confounding in many epidemiology studies. However, there remains some debate as to the extent to which
these sorts of factors constitute mediating effects or are on the causal pathway between unemployment and mortality (Mäki and Martikainen 2010). In this respect, the timing of events is crucial and this theme will develop throughout much of the remaining discussion and will be picked up again in chapter three.

2.5.2.2 Life-course accumulation of social disadvantage

In addition to contemporary risk factors it has been suggested that an important source of selection may arise from the accumulation of disadvantage along an individual’s life course. In other words, determinants of adult disease or poor health should not be considered solely from cross-sectional information but also information which documents instances of social disadvantage that occurs from birth up until adulthood (Kuh and Ben-Shlomo 2004). Furthermore, it has been suggested that this effect may operate in a cumulative fashion in which social chains of risk act to condition adult health outcomes (Wadsworth 1991; Holland et al. 2000). Such an approach has been applied to various health outcomes including mortality and life-expectancy (Kuh and Smith 1993; Smith et al. 1997; Cambois, Robine & Hayward 2001), physical health (Hallqvist et al. 2004) and self-rated health (Power, Matthews & Manor 1996; Power, Manor & Matthews 1999). Moreover, the possibility of in-utero influences on later life has also been noted (Barker 1992; Barker 1994). The potential for these life-course effects to influence movement into unemployment is therefore an important aspect of the unemployment and health relationship. It suggests that the group of unemployed may contain a larger proportion of individuals who are disadvantaged not simply in terms of contemporary circumstances, but also in terms of cumulative instances of disadvantage beginning perhaps as early as the gestation period and onto adult life.

A significant body of work has investigated these ideas. One study has found that the group of unemployed contained a significantly greater proportion of individuals who had experienced socio-economic disadvantage (proxied by slow growth and behavioural maladjustment) during childhood (Montgomery et al. 1996). The effect of unemployment during the early years of life has also been found to have lasting effects on the health and socio-economic chances for individuals (Wadsworth, Montgomery & Bartley 1999). This finding reflects evidence from the economics literature that shows that past unemployment is strongly predictive of current unemployment (Heckman and Borjas 1980; Heckman 1991). Moreover, a study of the England and Wales longitudinal study also suggested that unemployment in 1971 was independently associated with illness 20 years later suggesting that the health effects of unemployment can manifest themselves much later in life (Bartley and Plewis 2002). The accumulation of periods of unemployment has also been shown to determine limiting long term illness twenty years later, independently of contemporary circumstances (Bartley and Plewis 2002). The upshot of much of this work, therefore, is to
consider the enduring effects of various measures of social position at various points in the life course. Selection of individuals into unemployment is likely to reflect an accumulation of these factors in conjunction with contemporaneous circumstances.

2.5.2.3 Personality traits

A recent addition to the indirect selection hypothesis concerns the role of personality traits in determining social position and therefore as a possible explanation for health inequalities. The link between certain personality traits and various health outcomes including hypertension, coronary heart disease and mortality has been well studied in the psychological and medical literature (Barefoot, Dahlstrom & Williams 1983; Shekelle et al. 1983; Barefoot et al. 1995; Everson et al. 1997). Furthermore, these traits also appear to be associated with various measures of socio-economic status (Carroll et al. 1997; Christensen et al. 2004). As a result the potential for these sorts of measures to help explain health inequalities in mortality is being studied within the health inequalities literature (MacKenbach 2010a). Evidence from the US suggests that personality traits, measured in terms of the big five factors of neuroticism, extraversion, openness to experience, agreeableness and conscientiousness accounted for around 20% of the mortality risk that was associated with lower socio-economic status (Chapman et al. 2009). Similar evidence from France, although using slightly different measures of personality, found the corresponding reductions to be around 34% of the association for education and around 28% for income (Nabi et al. 2008).

In terms of the potential selection effect into unemployment for certain personality traits, less evidence is apparent. Although often suggested as a candidate confounder in many studies, it has rarely been used or examined in the worklessness and health literature apart from in passing (Bartley, Ferrie & Montgomery 2006). Two exceptions, both originating in Finland have uncovered strong evidence of, firstly, a moderating effect of the hostility trait in terms of the relationship between unemployment and subsequent health (Kivimäki et al. 2003a) and, secondly, that this trait is also associated with a greater likelihood of ending up in temporary and insecure work as opposed to more secure and lengthy employment (Virtanen et al. 2005a).

The possibility for personality traits to partly explain health inequalities has been controversial due to the potential for such an explanation to undermine the more structural explanations which place at their heart notions of poverty, environmental factors and (a lack of) social justice. However, as noted by Mackenbach (2010a), the development of these traits is not necessarily independent of structural inequalities in health and is likely to develop at various critical points along the life course. In other words, socio-economic position may be associated with the development of certain personality traits which can thus be considered a structural inequality in its own right. The degree to which the studies conducted in this thesis
are able to account for these effects is, however, limited by the lack of information in the census.

2.5.2.4 Place effects for health

A further source of potential confounding, though one that is rarely considered in much of the unemployment and health or mortality literature, may arise from the notion that the area within which people live might exhibit an independent effect for health. These area, or contextual effects, are imagined to operate over and above effects that occur at the individual scale (population composition) (Graham et al. 2004). Thus far, the indirect selection effects that have been examined are based on individual characteristics, but if an individual’s area of residence exerts independent effects for health, this might make them more likely to experience unemployment as well as, of course, raising their risk of mortality.

Teasing out the relative importance of contextual and compositional effects remains problematic. A common method for attempting to disentangle place effects on health involves controlling for individual characteristics so as to analyse any remaining variation as potentially a contextual effect. This has produced mixed results. Sloggett and Joshi (1994), for example, question the link between ward deprivation and premature mortality after controlling for individual characteristics relating to traditional deprivation indicators. They suggest that while it is correct to suppose a relationship between ward level deprivation and premature mortality, this does not warrant conclusions suggesting place effects on mortality. Furthermore, ‘the evidence does not confirm any social miasma whereby the shorter life expectancy of disadvantaged people is further reduced if they live in close proximity to other disadvantaged people’ (Sloggett and Joshi 1994 pg. 4).

Elsewhere, other evidence seems to suggest otherwise and in a lengthy discussion of a number of empirical studies, Curtis and Jones (1998) conclude that ‘contextual effects operating at the level of places seem to have some power to explain health inequalities, independently of the strong effects of individual attributes’ (pg. 666). For example, Wiggins et al. (1998) carried out a study focusing on limiting long term illness and concluded that area variations in morbidity were not entirely explained after controlling for individual effects (see also Gould and Jones 1996). Furthermore, Duncan et al. (1999) demonstrate how smoking behaviour can be influenced by area neighbourhood effects, in particular the levels of deprivation in that area. In a further observation, the authors also note that the propensity to smoke appeared uniform through differing groups of people suggesting that, in addition to a ‘social miasma’ effect associated with other people smoking in the area, there may be further effects related to ‘unpleasant, undesirable, unsafe environments in which there are fewer opportunities for making healthy choices’ (Duncan et al. 1999 pg. 503). This link between deprivation and health may well be important in terms of studying the effect of
worklessness for health and mortality as it has also been found that living in more deprived
neighbourhoods increases the chances of unemployment independently of individual
circumstances (van Ham and Manley 2010).

As this evidence suggests, there remains debate about the existence of independent area level
effects for health and much of this disagreement likely reflects theoretical or methodological
differences in the concept of an area effect. For example, Macintyre et al. (2002) suggest that
these differences might reflect ‘differing conceptualisations and operationalisations of area
effects, and in particular to differences in whether certain features of individuals or local
areas (such as baseline health status, smoking, or cardiovascular risk factors) are seen as
‘confounding’ or ‘intervening’ variables’ (pg. 11). Related to this, there is evidence that
the geographical aggregation of the area unit may also influence the magnitude of an area or
neighbourhood effect (Flowerdew, Manley & Sabel 2008). More technically, the common
approach to identification of such effects is often to use multi-level regression modelling
which relies on comparisons between poorer individuals in more deprived areas with
similarly poorer individuals in less deprived areas in order to infer a causal or independent
relationship. Such a scenario is difficult to envisage on the ground and it can, therefore, be
difficult to conclude that there is a causal link between area of residence and health. This
issue will be explored in more detail in chapter three alongside a discussion of the limitations
of regression modelling for inferring causal relationships more generally. In terms of the
implications for this thesis, the literature does suggest that area based deprivation may be an
important source of indirect selection and the empirical analyses will reflect this with chapter
three providing a greater discussion of how this is implemented.

2.5.3 Testing for selection

It is clear that selection is an important component of the association between worklessness
and mortality. This has led to a number of studies that have aimed to test for these selection
effects. In order to do so, many of these studies have employed a range of innovative
approaches and techniques, both in terms of analytical design and the linking and analysis of
different sources of data. The key feature of this work is that it uses techniques that might be
considered quasi-experimental as they often utilise naturally occurring patterns, trends or
events within the labour market. The following section discusses the evidence from these
studies. In particular, it will examine, firstly, the potential for labour market dynamics to
moderate the strength of the unemployment effect and, secondly, the evidence from natural
experiments such as factory closures or company downsizing or collapse. Thirdly, I will look
at those studies that have taken account of the timing of various events relative to
unemployment and, finally, I will examine the analytical concept of the ‘wearing off’ of
Chapter 2: Worklessness and mortality: A review of the literature

selection. All of these studies directly address the difficulties of estimating a causal effect in the worklessness and mortality relationship in various ways. Thus, all of these studies are important in terms of the main aims of the thesis which are to develop and use an innovative methodological framework to contribute to this literature.

2.5.3.1 Labour market dynamics

A number of studies have examined whether the association of unemployment with mortality is weaker during periods of higher unemployment. The hypothesis underlying this theory is that when the overall unemployment rate is high, the group of unemployed will contain a greater proportion of individuals who were made redundant for reasons other than poor health. During periods of lower overall unemployment, the proportion of sick and unhealthy individuals would be much higher compared to the group of employed individuals because healthier individuals would be more likely to gain employment during a period of slack labour market conditions. Therefore, if health selection were the driving mechanism behind the relationship between unemployment and mortality, one would expect this relationship to weaken during periods of higher unemployment (Martikainen and Valkonen 1996; Bartley, Ferrie & Montgomery 2006). Martikainen and Valkonen (1996) calculated mortality ratios for the unemployed in Finland during 1990 when overall unemployment rates were low and compared them to corresponding ratios for those that experienced unemployment for the first time in 1992 when overall unemployment rates were high. They found that the ratio of mortality amongst the unemployed reduced significantly during these two periods, suggesting that studies carried out during periods of lower unemployment may be subject to confounding due to selection. Around the same time a similar finding has also been shown to exist in the UK (Bethune 1996a).

This initial evidence has given rise to a number of further studies taking the same approach including another study in Finland examining suicide deaths (Mäki and Martikainen 2010) and one from Sweden examining both total and cause specific mortality (Åhs and Westerling 2006). Interestingly, these later studies seem to find less support for the selection explanation. In the Finnish study, although the effect of unstable employment for suicides was slightly less during periods of higher national unemployment rates, this effect was not seen amongst the long-term unemployed (Mäki and Martikainen 2010).

2.5.3.2 Natural Experiments

Another approach which makes use of so-called ‘natural experiments’, is studying what happens to workers who are made redundant when entire companies cease trading completely. This scenario is a useful one for studying the effect of mortality on health as the researcher can be sure that the mechanism by which an individual loses their job is likely to
be free of the selection characteristics outlined in sections 2.5.1 and 2.5.2 across all members of the workforce (Morris and Cook 1991). Similarly to the work above, this results in a diluting of the selective ‘unhealthy-ness’ of the unemployed allowing for more robust estimates of the mortality risk attached to unemployment.

Recently, one such study was carried out by Steenland and Pinkerton (2008) following the bankruptcy of Pan American World Airways in 1991. They compared the mortality rates of individuals who left work voluntarily, involuntarily and due to illness with the US population generally. They found little evidence of a heightened mortality risk for both the voluntarily and involuntarily unemployed, although they did find strong evidence of a significant healthy worker effect (Steenland and Pinkerton 2008). A similar study carried out in Finland using population registration data also found strong evidence of direct health selection with the risk of mortality declining significantly in the context of workplace downsizing (Martikainen, Maki & Jäntti 2007). Evidence from New Zealand also supports these findings (Keefe et al. 2002). Other evidence from Finland using the example of large scale downsizing within a single town showed that health was most likely to deteriorate amongst those who had been made unemployed following the downsizing process compared with, for example, those who were re-employed (Kivimaki et al. 2003a).

Studies such as these are not without significant shortcomings. Firstly, in many cases the samples tend to be fairly homogenous and selective, often excluding the most vulnerable members of society. For example in many instances samples tend to, by definition, include only the healthiest individuals compared to the general population due to the healthy worker effect. They also tend to capture two distinct effects that may both affect mortality; the shorter term shock associated with job loss as well as the potential longer term effects of unemployment. Furthermore, the potential for an individual to feel self-recrimination or blame is more likely amongst individuals who suffer redundancy in circumstances other than factory closure or company collapse (Morris and Cook 1991). All of these shortcomings combined would suggest that studies based on factory closures perhaps have a tendency to underestimate the ‘true’ effect of unemployment on health and mortality and may not be generalizable as a result.

2.5.3.3 Timing of health events and analytical approaches

As has been illustrated throughout this section, an important consideration when estimating the causal effect of worklessness for health is accounting for the timing of important socio-economic ‘events’ relative to the worklessness event. This is achieved through the use of longitudinal data. In the Nordic and Northern European countries, where access and availability of data is considerably more comprehensive than elsewhere, studies are well equipped to do this. Perhaps the best example comes from the Lundin et al. (2010) study.
The crucial benefits of this study were that it was able to link information on social background, psychological factors, health and substance use to a range of other data sources which provided additional information such as childhood social circumstances, psychiatric diagnoses from hospital discharge data, adult socioeconomic position and sickness absence, labour market participation and incidences of unemployment. All of these were linked to mortality outcome data. As Lundin et al. (2010) report, these factors explained most but not all of the elevated mortality amongst the unemployed.

Allied to this, an increasing number of studies are beginning to make use of recent analytical advancements emanating predominantly from the econometrics discipline. These new techniques have produced some very worthy contributions to the unemployment and health and mortality debate. Examples include difference in differences approaches in which changes to employment status in time are used to create a treatment and control group contrast (Böckerman and Ilmakunnas 2009) and, more commonly, propensity score matching (Eliason and Storrie 2009; Browning, Dano & Heinesen 2006; Kim et al. 2008; Böckerman and Ilmakunnas 2009). These techniques have at their core a more ‘quasi-experimental’ type approach to data analysis which is often better equipped to deal with the processes of selection and confounding. However, the extent to which these approaches have been implemented in the UK is limited and the majority of work within the UK has focused on a technique known as the ‘wearing off’ of selection in mortality studies.

2.5.3.4 ‘Wearing off’ selection

In the UK, as we have seen, attempts to identify the effect of unemployment on mortality have been dominated by the Office for National Statistics longitudinal study because of the ability to link socio-demographic information from censuses to subsequent mortality events (Moser, Fox & Jones 1984; Moser et al. 1986; Moser, Goldblatt and Fox 1987; Moser et al. 1987; Bethune 1996b). However, apart from the self-reported census measures of limiting long-term illness and general health, controlling for both prior health and the timing of any health problems is difficult in UK census data. One possible solution was proposed in which it was suggested that comparing mortality trends at differing intervals after baseline would reveal the extent of health selection. The theory proceeds as follows; if the raised mortality in the unemployed were due to higher prevalence of acute and potentially life-threatening illnesses then mortality amongst the unemployed would be expected to be higher immediately during the initial period of follow-up and would then proceed to decline with increasing duration of follow-up (Fox, Goldblatt & Adelstein 1982). This follows because those members of the unemployed group who became unemployed due to an acute illness which prevented them from working would eventually die (or recover) with this process likely taking place in the earlier periods of follow-up. Thus, if selection were playing a
significant role then we would expect the mortality rates to converge at a point later in the follow-up period (Moser, Fox & Jones 1984; Moser et al. 1986; Moser et al. 1987).

This reasoning led to the development of a technique known as the wearing off of selection in which mortality events are ignored for the first given number of years of follow-up in order that less biased estimates may be obtained (Moser, Fox & Jones 1984; Moser et al. 1986; Moser et al. 1987). An obvious disadvantage to this approach is that it results in five years of wasted mortality data which does not contribute to the analysis. Mortality is a relatively infrequent event and requires a sufficiently large baseline data set in order to capture the potential relationship with employment status. Removing the first five years of mortality further increases the amount of both baseline data and mortality data that is required. Thus, for a ten year follow-up mortality study, fifteen years of mortality data would be required. Furthermore, it is not immediately evident that this technique is effective at reducing health selection for the AU who, in contrast to the permanently sick in the census, show little evidence of a health selection shape during mortality follow-up (Bartley 1994). However, these ideas are yet to be explicitly tested alongside a measure of health such as the limiting long term illness measure that was introduced in the 1991 census.

As a result, although much of the early work on the longitudinal study suggested a strong link between worklessness and mortality, there remains debate as to whether these studies were able to fully account for health selection using the wear-off period approach. The more recent evidence highlighted above, which was able to adjust for these factors in various ways, casts further doubt on the findings of the earlier and original UK analyses by emphasising selection and confounding explanations. Thus, there appears a need to re-examine the issue of worklessness and mortality more generally in the UK which takes direct account of health selection and which also utilises more sophisticated analytical methods than has been implemented in previous studies. Furthermore, there also appears to be a need to test the effectiveness of the selection wear-off approach specifically, in light of this new literature, particularly given its continued usage in more recent studies in the UK (Akinwale et al. 2010).

2.5.4 Summary

This section has outlined how unemployment and job loss is the result of a number of factors which may also influence subsequent health and mortality. It has thus illustrated how the association between worklessness and mortality may be the product of these various non-causal factors rather than the causal effect of unemployment per se. In particular, the effect of a range of pre-existing health outcomes was found to be associated with unemployment including self-rated health, health behaviours and physical and mental health. As a result,
distinguishing between these pre-existing effects and the effects of unemployment is very
difficult. Furthermore, the concept of indirect selection in terms of socio-economic
confounding factors, life course effects and personality indicators was also investigated.
Much of this evidence has described how many of the poorer socio-economic characteristics,
disadvantaged life histories or negative personality traits are to be found disproportionately
amongst the unemployed. As these characteristics can all be plausibly linked to poor health
and mortality these indirect factors also represent a confounding mechanism of the
unemployment and mortality association.

In order to overcome these confounding issues a number of different approaches have been
employed including natural experiment type designs and the use of comprehensive
administrative type data sources. Many of these studies, most of which were conducted
outside of the UK, argue that much of the relationship between labour market disadvantage
and mortality could be attributed to selection and confounding. However, similar evidence
from the UK relies heavily on the concept of the ‘wearing off’ of selection, where deaths are
ignored for a period after baseline in order for selection effects to diminish. The
effectiveness of this approach and the application of alternative techniques for uncovering
the causal effect of worklessness for mortality thus represent key areas to which this thesis
will contribute.

2.6 Conclusions

2.6.1 Summary of Chapter two

This chapter has been broken into two broad sections. In the first, a general political
economy framework by which to understand the social determinants and social inequalities
in health was outlined. This perspective emphasised the importance of work, non-work and
the role of the welfare state in shaping patterns of health and health inequalities. This
theoretical framework then provided a lens through which a critical historical account of
post-war patterns of worklessness and welfare state reform was developed. One of the key
features of this period were the fundamental changes to the welfare state that acted as a
precursor to more recent labour market patterns such as the growth in numbers of IDS and
the continued retrenchment of the welfare state. The implications of a political economy
perspective for both the results in this thesis and wider debates in the fields of public health
and social epidemiology are reflected on in chapter seven, particularly in terms of the policy
implications of the findings in this thesis. Furthermore, taken together, these broad changes
to the welfare system and labour market are likely to mean that the experience of
unemployment and worklessness may well be very different than it has been in the past. As
the UK and much of the Western world faces a prolonged period of economic uncertainty and recession, greater clarity and continued research around the potential causality of the health effects of worklessness remains crucial and provides strong motivation for the work contained in the thesis.

Section 2.2.3 then devoted significant attention to the health related workless who, for a number of reasons, represent an important group in the context of this thesis. A significant number of individuals in this group can be considered hidden unemployed which may have important implications for capturing the full effect of unemployment for mortality and health generally when restricting to the AU. For example, many of the studies reviewed in this chapter focus attention on the AU and ignore the IDS members of the workless group. This is problematic as these individuals are likely to be relatively vulnerable compared to the larger population and thus are an important part of any study that aims to examine the health effects of worklessness. This is as yet, an under-researched area with very few studies that examine the relationship between this form of labour market disadvantage and mortality.

The second broad section, which contained the remainder of the chapter, was devoted to unpacking and examining the extent of the analytical problems associated with isolating a causal component of the worklessness and mortality association. Though there is a strong association between worklessness and mortality which is theorised to contain a causal component through the factors highlighted in section 2.4, identifying these causal pathways is often a complex exercise owing to the direct and indirect selection of individuals into worklessness on the basis of characteristics that are also linked to mortality. A number of approaches have been implemented to adjust for this selection problem including natural experiment type designs which utilise naturally occurring events such as recession or company bankruptcy and data linkage approaches that utilise information from a wide variety of administrative and register based datasets. Much of this evidence, particularly more recent studies, tends to suggest that selection is the dominant explanation for the association between mortality and worklessness that was outlined in section 2.3. However, there are reasons to suspect that these findings may not be sufficient to reject the hypothesis of causal relationship between worklessness and mortality. For example, studies of factory closures and company downsizing or collapse may not be representative of a wider population and may not be completely free of residual selection effects. Furthermore, it was noticeable that much of the evidence using alternative methods was conducted in countries which benefit from improved access to administrative and register based datasets such as Finland and Sweden. It is possible that evidence from these countries may not be representative in terms of the severity of the effect of unemployment in other countries.
owing to the relatively generous and comprehensive provision of state welfare in these countries.

Contemporary comparative evidence that explicitly tests these ideas in countries such as the UK and the US, where such welfare state provision is generally less comprehensive, are less prevalent. Those studies that have been conducted in the UK commonly use a technique known as a wear-off period, in which deaths occurring within five years of commencement of follow-up are ignored to allow selected individuals to either die or recover. Such an approach results in a significant quantity of wasted mortality data and as a result requires a dataset with a significant amount of follow-up. Furthermore, despite the ubiquitous nature of this approach, explicit verification of its effectiveness in reducing the effects of selection has yet to be published.

2.6.2 Empirical applications of the thesis

To conclude this chapter, the specific applications of the studies in this thesis will be outlined together with a rationale for their study on the basis of the literature reviewed in the chapter. These three areas will be expanded and elaborated upon at the conclusion of chapter three following an extensive review of the methodological issues that are relevant to the topic. The specific research questions to emerge will also be outlined at the conclusion of chapter three.

2.6.2.1 Wear-off periods and health selection

In light of the findings in this chapter regarding the common use of wear-off periods in many UK based studies of worklessness and health, this study proposes to test the effectiveness of this approach to explicitly examine its effectiveness in reducing the effects of health related selection. The study argues that such an approach may be unnecessary and not as effective as other approaches outlined in chapter three. It therefore holds potentially important implications for future mortality follow up studies by determining the benefits of the wear-off approach relative to the significant costs that are associated with its implementation such as the requirement of significant amounts of mortality follow-up data and the loss of five years of data during the wear-off period.

2.6.2.2 The selection and causation debate

The extent to which the association between worklessness and mortality is reflective of a causal relationship remains in question. Recent evidence from both natural experiment type studies and those with access to register based data sources, mostly from the Scandinavian countries, has cast doubt on the thesis that worklessness exerts a significant causal influence on mortality. Earlier studies in the UK consistently found raised mortality risks amongst
workless individuals but these studies were limited in their ability to adjust for the potential effects of selection, particularly health related selection, and typically relied on the use of wear-off periods. However, contemporary evidence from the UK has been lacking. This study contributes to the ongoing debate and reasserts the thesis that active unemployment is causally related to mortality through the use of linked administrative data and an innovative analytical design. The methods underpinning this analytical design are fully discussed and developed in chapter three.

2.6.2.3 *Hidden unemployment and economic inactivity due to sickness*

The final study in the thesis draws on the finding that, amongst the worklessness and mortality literature, there has been relatively little attention paid to the effects of economic inactivity due to disability or sickness. This is despite a trend in the UK since the 1980s which has seen increasing numbers of individuals claiming incapacity benefit at a rate which does not match population changes in morbidity. As highlighted in this chapter, many of these individuals constitute a group of hidden unemployed in the sense that they are capable of work but are unable to find suitable employment in local labour markets that have been strongly affected by deindustrialization. Furthermore, many are suffering significant illness and poor health relating to their previous employment in manual labour type industrial occupations which qualifies them for incapacity benefit and related sickness payments. Studies examining the effect of worklessness for health that neglect to consider individuals in this group therefore risk missing perhaps the more vulnerable members of the workless population. However, the problems of selection that have been outlined in this chapter are likely to be greater in this group than for the AU. This study therefore extends the analysis from the previous study to individuals who are IDS to test for an independent and causal relationship to mortality.
Chapter 3

A METHODOLOGICAL FRAMEWORK

3.1 Introduction

The underlying theme of much of chapter two was the debate that exists around the relationship of worklessness for health and mortality. Though section 2.3 illustrated that the statistical association between them is pervasive, a causal interpretation is often precluded by problems variously described as confounding and selection or bias more generally. The following section aims to examine these general concepts in more detail and develops an argument to justify the need for a more nuanced methodology in studies of worklessness and mortality that is informed by current methodological developments in the field of observational epidemiology.

The chapter will be structured as follows; firstly, it will introduce the counterfactual framework of causal inference which, although not uncontroversial, remains a compelling but perhaps underused approach to estimating causal effects. It will then discuss the concept of randomisation and the fundamental problem of causal inference in observational data which will then lead on to the issues of selection and confounding and the apparent confusion in the usage of these two terms in the literature discussed in chapter two. This will lead onto a discussion highlighting both structural confounding and the possibility of over-adjustment in many studies of mortality and worklessness whereby causal effects of interest may be obscured by intermediary variables. This line of reasoning will set the context for a brief outline of the statistical approach known as propensity score matching which is used extensively in the thesis but is perhaps less widely covered within the general literature. The overarching aim of this chapter therefore is not to outline the precise methods used (these will be covered in the subsequent empirical chapters) but more to discuss the broader analytical framework that is used to address the challenges of causal inference with observational data and to tally these ideas with the study proposals set out in section 2.6.2 in order to formulate the precise research questions of the thesis.
3.2 Counterfactual causality

3.2.1 The fundamental problem of causal inference

David Hume wrote that ‘we may define a cause to be an object followed by another...where, if the first object had not been, the second never had existed’ (Hume 1748 pg. 115). The key implication of this statement is that the first object is a necessary condition for the second implying that the absence of the first would result in the absence of the second. In other words, we can only observe a causal effect if the outcomes of both exposure and non-exposure can be observed on the same unit at the same time. In the context of the causal effect of worklessness for mortality, this equates to the observation of the same individual in both exposure states (workless and in work) at the same time in order to observe the effect on the outcome (mortality). Of course, such a condition could only be observed in a parallel universe resulting in a paradox that is often referred to as the fundamental problem of causal inference (Holland 1986). Out of this fundamental problem has arisen one of the firmest notions of epidemiological research, that association does not equal causation (Barnard 1982). This important property of statistical analysis has resulted in a great deal of cross-disciplinary, technical and theoretical literature.

Early approaches attempted to formalise the notion of causality heuristically. For example, Austin Bradford-Hill (1965), presented nine criteria that could be used to assess the presence of a causal effect; strength of association, consistency, specificity, temporality, biological gradient, plausibility, coherence, experiment and analogy (Hill 1965). The major critique of such an approach is that they attempt to reduce the complexity of any causal system to a degree that undermines the importance of theoretical understanding and a grasp of the consequences of the possible actions that might be taken as a result of the findings (Höfler 2005). Causal systems are necessarily complex to the extent that in many cases heuristic criteria aimed at establishing them simply do not exist. Furthermore, “causal inference in epidemiology is better viewed as an exercise in measurement of an effect rather than as a criterion-guided process for deciding whether an effect is present or not” (Rothman and Greenland 2005 pg. 1).

Although never originally intended as a checklist by Bradford-Hill, these nine criteria have since been increasingly misused in original empirical research for this purpose, perhaps because the job of establishing a causal effect becomes considerably easier should such criteria exist. Despite this critique, the underlying ideas offered by the Bradford-Hill criteria represent key concepts in modern accounts of causality and causal inference in epidemiology, an area which remains highly relevant and contentious. Much of this contemporary discussion is beyond the scope of this thesis but it is important to note that a
consensus is still far from being reached regarding a universal working definition of causal effects or a generalised methodology by which to study them. However, the presence of ‘checklists’ such as the Bradford-Hill criteria through which they can be identified has been widely rejected (Dawid 2000; Greenland 2000; Robins and Greenland 2000; Poole 2001; Karhausen 2001a; Karhausen 2001b; Olsen 2003; Rothman and Greenland 2005).

One general approach, widely attributed to the seminal work of Neyman in 1923 (later translated in 1990) and subsequently developed by a number of scholars in both statistics and philosophy over the last 20 to 30 years (Copas 1973; Lewis 1973; Rubin 1974; Rubin 1978; Rubin 1990; Lewis 2001; Maldonado and Greenland 2002), uses the concept of alternative events or counterfactuals. This framework approaches the problem of causal inference by re-conceptualising the fundamental problem of the unobservable event as a quasi-observable event through the concepts of substitution and exchangeability (Maldonado and Greenland 2002). This underlying principle remains controversial (Dawid 2002; Elwert and Winship 2002; Kaufman and Kaufman 2002; Shafer 2002) but, despite these debates, what consensus has been reached would appear to support a counterfactual based approach (Holland 1986; Greenland 2000; Parascandola and Weed 2001; Poole 2001; Maldonado and Greenland 2002; Flanders 2006; VanderWeele and Hernán, Hernández-Díaz & Robins 2004). Thus, this thesis adopts the counterfactual framework for the methodological and analytical approaches to causal inference. This is explored in the following sections.

### 3.2.2 Potential outcomes

To begin to understand a causal effect in terms of the counterfactual, we need to imagine the full set of potential outcomes that could result from a given study question. Greenland and Robins (1986) introduce the four possible causal types for a single unit in a classic study of treatment exposure and outcome response. These are summarised as (i); no effect doomed, (ii); exposure causative, (iii); exposure preventative and (iv); no effect immune. To understand these, Morabia (2010) applies the four types to the example of a single unit (an elderly patient) in a study of the efficacy of a flu vaccine and its effect on the prevention of the flu outcome (Morabia 2010). Table 3.1 outlines this example but modified to fit the unemployment and mortality relationship. It is important to remember that these four types originally imagined by Greenland and Robins (1986) represent all possible interpretations of a study on one individual.

| Table 3.1 The four ‘causal types’ in a hypothetical exposure/outcome study of unemployment and mortality |
If we imagine that the subject is ‘administered’ with unemployment and is then observed to die, we might assume that the unemployment experience is causative of mortality and conclude for $C$. However, in the absence of the ‘potential outcome’, or the outcome of the same individual at the same time but unexposed, we are unable to rule out $D$, that the individual might have died irrespective of their employment status. Similarly, if the subject were to survive having been unemployed, we would be equally unable to deduce $P$ that unemployment was protective of mortality because the subject may have been ‘immune’ to death irrespective of their unemployment, summarised by $I$. The potential outcome, therefore, is the unobservable outcome of the alternative exposure on the same individual at the same time. As Greenland and Robins note (1986), this problem is known as non-identifiability; ‘different possibilities for the effect predict identical data distributions, and so we cannot identify the effect from the data’ (pg. 414).

### 3.2.3 Exchangeability and randomisation

Simply introducing an in work control subject so that it is at least theoretically possible to observe two outcomes for each causal type (i.e. both subjects die, exposed dies, unexposed dies or neither dies) holds the same problem as before in terms of controlling for the possibilities of $D$ and $I$. The problem arises that we cannot be sure that the control is similar enough to the individual to be imagined as identical, i.e. the same person at the same time. This introduces the concept of exchangeability. Exchangeability can be imagined as the extent to which the unobserved potential outcome can be substituted for the outcome on a different individual in a different treatment. In other words, it is the extent to which two different individuals are similar in terms of the mechanism by which they do or do not receive the treatment (worklessness). Although observation of the true counterfactual event is impossible, we can substitute the potential outcome for an exchangeable individual which removes the possibility of outcomes $D$ and $I$ in Table 3.1 because the two individuals are similar in terms of underlying frailty/robustness. The most common approach to the identification of causal effects through exchangeability in this way is through the randomised experiment or randomised control trial RCT (Pearl 2000).

<table>
<thead>
<tr>
<th>Types</th>
<th>N</th>
<th>Description</th>
<th>Outcome</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>In Work</td>
<td>Unemployed</td>
</tr>
<tr>
<td>Doomed</td>
<td>D</td>
<td>Frail independently of employment status</td>
<td>Died</td>
<td>Died</td>
</tr>
<tr>
<td>Preventive</td>
<td>p</td>
<td>Unemployment protective against mortality</td>
<td>Died</td>
<td>Survived</td>
</tr>
<tr>
<td>Causative</td>
<td>C</td>
<td>Unemployment causes mortality</td>
<td>Survived</td>
<td>Died</td>
</tr>
<tr>
<td>Immune</td>
<td>I</td>
<td>Insensitive to unemployment</td>
<td>Survived</td>
<td>Survived</td>
</tr>
</tbody>
</table>
To satisfy the notion of exchangeability in an RCT, the researcher implements a key tenet of the experimental approach. This is the ability to randomise the allocation of treatment status to substitute for the unobservable potential outcome. The randomisation element removes the possibility that treatment status may be allocated for reasons that may be related to the outcome. For example, it removes the possibility that the allocation of treatment status is related to a third variable that is also related to the outcome. Trialling a new drug in a sample of individuals, this would translate into the random allocation of subjects to either drug or placebo, thus ensuring that in all respects, other than treatment assignment, both groups are identical.

As with many studies assessing the effect of social phenomena on various outcomes, experimental randomisation is simply not possible. Particular barriers include the practicality of such a large scale randomised experiment and the length of the period of follow-up that would be needed as well as, perhaps most crucially, the ethics of randomising periods of worklessness to individuals. As a result, studies of social phenomena often rely on the use of observational data. The key difference between randomised experiments and observational data is, as the name suggests, the difference between data that is observed without experimental manipulation and data in which the researcher is able to modify treatment allocation. As a result, individuals are observed as being workless through the collection of survey data, rather than being randomly assigned as workless in an experimental setting. The result is that the allocation of treatment (worklessness) is non-random and may be the product of both observed and perhaps unobserved factors. These factors, which include the direct and indirect selection characteristics outlined in chapter two, introduce bias and present challenges for the analysis of causal effects in observational data.

### 3.3 Bias

Bias represents a significant challenge for the identification of causal effects in both randomized experiments and observational studies. However, in observational studies the challenges are considerably more acute. In this section I draw on a paper by Hernán, Hernández-Díaz & Robins (2004) who develop a structural classification of bias into three categories; reverse causation, common effects and common causes. Though the first one is important in studies in which reverse causation is a possibility I focus on the latter two categories as the first is not a realistic possibility when using mortality as an outcome. Adopting this concept of bias allows for greater clarity in the terminology particularly in the workless and health literature where terms and concepts appear to be used interchangeably. The distinction between the two forms of bias is therefore used to critique current misrepresentations of these concepts that were apparent in the literature reviewed in chapter two.
and feeds into a discussion of the related concepts of over-adjustment and structural confounding both of which hold important implications for the analytical framework of the studies in this thesis.

### 3.3.1 Common causes and common effects

In the literal sense of the word, the term confounding connotes a mixing process whereby the effects of certain exposures of interest are mixed or pooled with the effects of other variables. This presents significant difficulties when trying to separate and identify a causal link between the exposure of interest and an outcome (Morabia 2010). Hernán, Hernández-Díaz and Robins (2004) consider the three causal structures which may arise in studies in which an association between an exposure E (unemployment) and an outcome D (mortality) has been observed. These structures are illustrated in the diagrams below in the form of causal diagrams which are often used to represent causal effect structures (Greenland, Pearl & Robins 1999).

The key feature of causal diagrams is that they illustrate direct causal relationships between variables which are represented by an arrow between the relevant variables. Thus, a variable X causes Y if, and only if, it is connected with an arrow leading from X into Y (Glymour 2006). Hence, Figure 3.1 can be imagined as an ideal randomised trial in which the observed association can be interpreted as a direct causal relationship because of the absence of any other potentially causal variables in the diagram (Hernán, Hernández-Díaz & Robins 2004).

**Figure 3.1 Causal diagram representing a causal effect of exposure E on outcome D**

![Figure 3.1](Image)

(Source Hernán, Hernández-Díaz & Robins 2004 pg. 616)

In Figure 3.2, the interpretation of the same association between E and D is complicated by a common cause of both outcome and exposure represented by L, which can be imagined to be health status. This structure is often termed the classical definition of confounding and is summarised according to the associational criteria. This formally states that the two variables exposure E (unemployment) and outcome D (Mortality) in Figure 3.1 are confounded if a vector L (Health for example) affects both E and D (Pearl 2000). Thus in Figure 3.2 the only true causal effect operating on the outcome is that of L → D and the association of E and D is entirely due to the common cause through L (represented by the lack of an arrow between E and D) and is therefore confounded by L (Hernán, Hernández-Díaz & Robins 2004). If the confounding variable or vector L (health) is not observed, or is measured imprecisely, the
result is that \( L \) determines jointly \( E \) and the potential outcomes \( D \). Consequently, it cannot be assumed that we are making unbiased counterfactual comparisons between treated individuals and their corresponding control, thus precluding a causal interpretation between \( E \) and \( D \).

**Figure 3.2 Causal diagram representing a confounded association between exposure \( E \) and outcome \( D \) confounded by \( L \)**

![Causal diagram](image)

(Source Hernán, Hernández-Díaz & Robins 2004 pg. 616)

Diagrams a and b in Figure 3.3 represent the common effects causal structure in which an estimate is conditioned on a common effect of both exposure and outcome with the conditioning variable represented by a square symbol. Typically, this is referred to as selection bias (Hernán, Hernández-Díaz & Robins 2004). A simple description of this spurious causal structure is to suggest that the effect estimate of exposure is an artefact of the conditioning variable.

**Figure 3.3 Causal diagrams representing a confounded association between exposure \( E \) and outcome \( D \) confounded by the conditioning variables \( C \) and \( M \)**

a)

![Causal diagram](image)

b)

(Source Hernán, Hernández-Díaz & Robins 2004 pg. 616)

This occurs when the effect estimate within strata of the conditioning variable is different to the estimate that would have been observed on an unrestricted sample. In other words, in both diagrams in Figure 3.3, knowing something about the conditioning variable provides some information about the exposure \( E \). There are many examples of selection bias including
the selection of poor control cases in a case control study, non-response or missing data bias, volunteer or self-selection bias and survivor bias. Another example that is directly relevant to this thesis is the healthy worker bias, represented in Figure 3.4, which can be seen as an extension of both diagrams in Figure 3.3. It illustrates a study of workers which failed to find an association between exposure to workplace chemical E and mortality D because the study was effectively conditioned on individuals currently at work at the ascertainment of the outcome (indicated by C). Thus, because presence at work is related to underlying but unmeasured health U, even adjustment for health measurements L of the workers does not produce an unbiased estimate. This is because exposure E is dependent on being present at work. As a result, those suffering ill health effects U as a result of exposure to the chemical are less likely to be included in the study because of ill health related absence from work. The example fits the common effects structure because the conditioning variable is a direct effect of the exposure and indirectly related to the outcome.

**Figure 3.4 Healthy worker bias expressed as a common effect causal structure**

![Figure 3.4 Healthy worker bias expressed as a common effect causal structure](Source Hernán, Hernández-Díaz & Robins 2004 pg. 617)

### 3.3.2 Bias in the worklessness and mortality relationship

Figure 3.5 illustrates the three explanations for the association between worklessness and mortality that are commonly encountered in the literature. The causation explanation is straightforward and is the relationship we might infer if the estimate of the effect of unemployment for health were the result of a perfectly designed random experiment. Thus this explanation corresponds to the causal diagram in figure 3.1. Such an example is only useful as a thought experiment as the experimental randomisation of unemployment is both unethical and impractical and imagining the causal structure in this way is therefore hypothetical.

**Figure 3.5 Explanations for the association between unemployment and health**
The terminology of the remaining explanations, indirect and direct selection, is however, slightly misleading. This can be illustrated by returning to the example of the healthy worker effect presented in figure 3.4. Hernán, Hernández-Díaz & Robins (2004) note that ‘the term “healthy worker” bias is also used to describe the bias that occurs when comparing the risk in certain groups of workers with that in a group of subjects from the general population’ (pg. 619). In both cases the reason for the presence of bias is essentially the same; that workers are more likely to be healthy than non-workers and thus unemployed or workless are more likely to be unhealthy. However, in the factory example in Figure 3.4, bias occurs as a result of the inappropriate selection of the study sample (i.e. from workers present at work) rather than the systematically unequal distribution of poor health or other factors across the labour force throughout the general population. In this sense, the factory study lacks generalizable findings because the sample is not representative of the population of workers who were exposed to the chemical.

On the other hand and assuming that the findings have been generated from population representative samples\(^1\), direct and indirect selection represent threats to the internal validity of the study because of direct confounding by health or indirect confounding by factors such as socio-economic status or individual personality characteristics. In other words, the findings are repeatable in other samples but the interpretations are biased because of the unobserved systematic effects of these confounders. Hence, and in accordance with Hernán, Hernández-Díaz & Robins (2004), the difference between direct and indirect selection is

---

\(^1\) In the most part, this assumption reflects the literature where most of the studies are conducted using representative datasets such as the census or other social survey sources. However factory closure studies for example are often critiqued on the basis of the representativeness of the samples as they rely on single factories or plants.
reduced to a problem of confounding and intermediary confounding respectively, rather than one of sample selection. As a result, I argue that the terms direct health confounding and indirect confounding more accurately capture the biasing effects of health and other characteristics in the worklessness and mortality relationship.

### 3.3.3 Over-adjustment

Thus far, I have outlined the issues of direct health confounding and indirect confounding and distinguished them from the related but separate concept of selection bias. The most common approach to mitigating these sources of bias is to apply analytical tools such as regression modelling. Typically, models are constructed around the data which predicts the outcome as a function of both the exposure and a vector of covariates that are specified, *a priori*, to confound the relationship between exposure and outcome. Thus, for each of the confounding variables, a correction is applied that directly adjusts the data for the effect of the covariates. However, problems may arise when the nominated confounding variables are poorly specified such that they may not confound the causal pathway between worklessness and mortality. This problem is known as over-adjustment and is a concept that has received relatively little formal attention in the literature that was reviewed in chapter two (Schisterman, Cole & Platt 2009).

Over adjustment is defined in the dictionary of epidemiology as “statistical adjustment by an excessive number of variables or parameters uninformed by substantive knowledge (e.g. lacking coherence with biologic, clinical, epidemiological or social knowledge). It can obscure a true effect or create an apparent effect where none exists” (Porta 2008). In the context of this thesis, the possibility of obscuring a true effect is of the greater significance. Figure 3.6 illustrates the simplest form of over-adjustment for an intermediate health variable (M) on the causal pathway between unemployment (E) and mortality (D).

To help interpretation of the diagrams, it is pertinent to highlight the time precedence element of causal chains and the consequently acyclic nature of the causal diagrams that represent them. One of the key criteria for causality is the simple notion that cause must precede effect. Thus, when representing cause and effect in causal diagrams there can be no chains or series of arrows that link back to a previous variable as this would indicate a relationship that is cyclical with a system where descendant variables feedback to antecedent variables. Thus, all causal diagrams must be acyclic in the manner set out by Pearl (1995). Figure 3.6 depicts a study of an acyclical causal relationship between unemployment and mortality which has been conditioned on a descendant variable M which can be considered as some measure of health. Schisterman, Cole & Platt (2009) highlight that conditioning on the health variable M introduces a bias operating towards the null hypothesis that
unemployment has no effect on mortality. The reason for this is that the variable M is a descendant of unemployment and therefore a descendant link in the causal chain leading from unemployment to mortality. The degree of bias towards the null hypothesis depends upon the proportion of the causal effect that flows through the direct link between unemployment and mortality relative to that which flows through the intermediate M. If, for example, all of the causal effect from E to D was due to M, the effect estimate of E to D would be zero conditional on M and therefore a biased estimate (Schisterman, Cole & Platt 2009). In Figure 3.7, the same result would occur if conditioning on a proxy variable M of the intermediate variable U. However, in this example the degree of bias would depend again on the proportion of the causal effect moving through the intermediary but also the degree to which the variable M is a true proxy for the intermediate health variable U.

**Figure 3.6 Over adjustment for an intermediate variable M on the causal pathway between unemployment E and mortality D.**

(Source: Schisterman, Cole & Platt 2009)

What Figures 3.6 and 3.7 illustrate is the important distinction between direct and indirect causal effects. In effect, Figure 3.6 has simply moved the confounder L in Figure 3.2 to a position in which it occurs after exposure. It is thus no longer a confounder and becomes an intermediate variable. Whether the effect of worklessness is direct or mediated by a descendant variable, the effect remains causal and, therefore, conditioning on the intermediate will produce a biased estimate. The implications of this are clear and highlight the need for longitudinal data to ensure that the timing of events relative to exposure and outcome can be accounted for.
An example of this from the literature can be found with two studies of the relationship between unemployment and suicide; one in Denmark (Mortensen et al. 2000) and the other in the UK (Lewis and Sloggett 1998). Both reported excess suicide risk amongst the unemployed but, in the former, much of this was accounted for by prior mental health problems. However, in the latter, unemployment was still considered the causal link despite a lack of adjustment of their estimates for prior mental health. In this instance the role of mental health as either a confounder or an intermediary held important consequences for the interpretation of the study. Elsewhere, some studies report cross sectional findings where confounding effects such as health and socio-economic status are measured concurrently with unemployment prior to follow-up making it impossible to distinguish between those effects occurring before and those occurring after unemployment (Nylen, Voss & Floderus 2001; Gerdtham and Johannesson 2003; Gardner and Oswald 2004). Other studies such as that conducted by Lundin et al. (2010) are similarly able to adjust for a large number of confounders but, in addition, designed their study to ensure that many of these occurred as antecedents to the unemployment exposure. For the studies conducted in this thesis, similar attention will therefore be paid to the analytical design and, by way of discussion, the findings compared both to the studies highlighted here and other studies to understand the effect of adjustment for antecedent and cross-sectionally measured variables.

### 3.3.4 Structural confounding

In some cases, confounding can be introduced into a study as the result of inherent and structural differences within the population from which the data is observed. This is known as structural confounding. A good example with which to illustrate this is to draw on recent interest in the potential for neighbourhoods to affect individual health outcomes independent of the composition of the individuals and their characteristics within that area. A vigorous debate has developed between proponents of the multi-level regression approach (Diez Roux...
Chapter 3: A methodological framework

2004; Subramanian 2004; Chaix et al. 2006) and those that argue that such approaches are limited in their ability to generate meaningful results (Oakes 2004; Oakes 2006; Oakes and Johnson 2006; Messer, Oakes & Mason 2010). Much of this debate is centred on the issue of structural confounding. For example, the paper by Chaix et al. (2006) examines the effect for individual mortality of both neighbourhood population density and the socio-economic environment. The paper uses a multi-level regression model in which individual effects are controlled for, allowing examination of the independent effects for health of combinations of these various neighbourhood characteristics. The relevance of this debate for the thesis is that it neatly illustrates the problem of structural confounding in the context of regression analysis.

Oakes (2006), in a commentary of the Chaix et al. (2006) paper, suggests that it is important to try to imagine the ‘idealised experiment the investigators aim to mimic’ (pg. 644). Thus, with two exposure factors (population density (PD) and Socio-economic conditions (SES)) their experiment contains four conditions; (1) high PD and high SES, (2) high PD and low SES, (3) low PD and high SES and (4) low PD and low SES. They have one observed outcome and three unobserved counterfactual outcomes that have to be substituted for exchangeable units. In other words, if they observe the outcome from condition (1) they would need to find areas for each of the remaining conditions that were identical to enable estimation of the causal effects of each of the treatment levels. According to Oakes (2006), the problem arises when, as is evident from the paper, neighbourhoods with both high PD and low SES do not exist within the sample and thus cannot be compared unless relying entirely upon imputed values based on the model estimates.

Perhaps of more critical concern is the role of individual level confounding effects. A crucial component of the study was to adjust for the individual composition of the different areas in order to make un-confounded comparisons between exchangeable individuals. It is useful to consider the concept of the underlying propensity score that describes the probability of residing in a given area (or more generally the probability of treatment group membership in any study). If this probability is known, it is possible to compare the distributions of the propensity score across these poorer and richer areas. If they differ markedly with little overlap it could suggest that the areas are structurally confounded. If, for example, more deprived areas contain on average greater numbers of socio-economically deprived people it is likely that the majority of comparisons between deprived areas A and wealthy areas B would be making comparisons between two structurally different and therefore incomparable units that are made up of greater numbers of poorer and richer individuals respectively. Rather than basing effect estimates on actual comparisons between similar neighbourhoods (similar that is in terms of demographic composition but different in terms of neighbourhood
classification), they rely instead on the model based average effect estimates for the individual confounders. However, these effects are likely to be endogenous as the mechanism by which individuals are assigned to an area and not another (socio-economic status, income etc.) is likely to reflect the same variables that confound the relationship to health. Thus, adjusting for these confounders leaves zero variance with which to apportion to the neighbourhood effect should one exist. In essence, the neighbourhood effect is therefore unidentifiable because the assumption that an individual living in one area is exchangeable with another individual in a different area is violated.

**Figure 3.8 Histograms comparing the predicted probability of unemployment between actually employed and unemployed individuals for a hypothetical study. The graphs illustrate an extreme example of the off-support problem.**

The implications for the study of worklessness and mortality are clear. Determining whether an unemployment or workless effect for health is independent of structural confounding factors depends on the extent to which we are able to identify an overlap in the underlying possibility of unemployment or worklessness between people actually unemployed/workless and those employed. Figure 3.8 shows a hypothetical distribution of the probability of unemployment for employed individuals compared to unemployed individuals and illustrates an extreme (and very unlikely) example of structural confounding. In this instance, the two distributions show no evidence of any overlap and we would be forced to conclude that there are no (or very few) unemployed individuals who were exchangeable with employed individuals. Such a pattern in reality would preclude the potential for a causal contrast to be made between the two groups of interest. The graphs in Figure 3.8 rest on the concept of an underlying propensity score for unemployment or worklessness which in reality is never
observed and thus has to be estimated. This concept of an underlying propensity score forms
the basis for the causal inferences within this thesis and is outlined in the next sections.

3.4 Propensity score matching and the Rubin Causal Model

As we have seen, in order to adequately construct the potential outcome, or counterfactual
event, we need to produce a situation in which the two groups upon which the causal contrast
is being made are exchangeable. In other words, substituting them for each other would not
alter the estimate of modifying effect of the exposure on the outcome. In order to be
exchangeable, the two groups therefore need to be comparable with respect to the
confounding variables. To use an analogy, when interpreting exposure effect estimates are
we comparing for example two apples in our exposure group with two pears in our control
group or a mixture of the apples and pears in both the control and exposure groups? In the
former, though we may adjust for a confounding apple ‘effect’ we are still comparing
fundamentally different entities (apples and pears) which preclude our assumption of
exchangeability and therefore our ability to identify a causal effect. As we are not comparing
like with like, we are reliant on the model estimates and the resulting extrapolated or
imputed values rather than comparisons between empirically observed similar groups. As a
result, our inferences are “off support” of the data used to generate the estimates in a manner
similar to that illustrated in Figure 3.8 (Oakes 2004; Oakes 2006; Oakes and Johnson 2006).
However, despite the importance of exchangeability in causal effect estimates, a general
routine within the regression toolbox for establishing comparability is not explicit and rarely
considered in empirical studies. For example, in the Lundin et al. (2010) study, though they
are able to adjust for a wide range of potentially confounding effects and though the
distribution of these variables between unemployed and in work groups is presented in a
table, explicit tests of balance or histograms such as those in Figure 3.8 are missing. Thus,
the reader is unable to ascertain the degree to which the estimates that are derived from the
regression analyses are reliant on model extrapolation rather than genuine overlap in the
potentially confounding characteristics. In the following section, I will outline a general
model for overcoming these limitations known as the Rubin Causal Model or RCM before
going on to discuss a specific application of the RCM which is an approach known as
propensity score matching.

3.4.1 The Rubin Causal Model (RCM)

The main guiding principle of the RCM is that when attempting to infer causal effects from
observational data, it is important to imagine that the data is the result of an idealised
randomised experiment. From here, the design of the ‘experiment’ is forced to reflect similar
decisions and processes to those in an experimental study, even though the data has already been observed. In this sense the ‘design trumps the actual statistical analysis’ (Rubin 2008). In this outline and the discussion of propensity score matching I will borrow heavily from two papers by Donald Rubin (2008; 2007) as well as a book chapter from Oakes (2006).

As Rubin (2008) suggests, the first stage of the RCM is an exercise in making clear the mechanism about which we are making inferences on the basis of theory and prior knowledge so that a plausible pathway can be identified. This is relatively straightforward and merely suggests that the researcher be clear about the plausibility of the mechanism under study. This stage was achieved from the detailed review carried out in chapter two, particularly the discussion in section 2.4. In the second stage however, the researcher is urged to consider the precise ‘assignment mechanism’ by which treatment is conferred on an individual or unit. This stage is more problematic in the context of this thesis due to the issues of confounding that were outlined throughout chapter two, but particularly in section 2.5. Rubin (2008) suggests that the most useful way to think about this second stage is to imagine that the observed dataset under study is the result of a large and complex randomised experiment for which the various treatment and control assignment rules have been lost. In order for the study to be successful, the researcher is tasked with reconstructing these rules. To do this a number of considerations are suggested. Firstly, it is important to think about the ways in which unemployment is distributed or ‘assigned’ to individuals in our study in terms of, for example, poor health and lower socio-economic position which are all characteristics that can cause unemployment. This assignment mechanism is constructed on the basis of prior knowledge from the literature reviewed in chapter two.

A second step is to remove all outcome data from the dataset. This is a crucial stage in the process of reconstructing the ‘experiment’ as it prevents ad-hoc and outcome mediated decisions in the design stage of the study which may bias the final results. This is logical as in an experiment, outcome data is only observed after completion of the design stage. Finally, it is necessary to ascertain whether this assignment mechanism is such that valid comparisons between treatment and control groups can be made. This decision is made on the basis of the distribution of key covariates across treatment and control groups to determine balance and comparability. If the two groups are significantly imbalanced in terms of these covariates then the comparisons become structurally confounded and the ‘experiment’ design from which the data were generated becomes irreparably broken.

### 3.4.2 Matching and the propensity score

The application of these concepts is achieved through matching generally and, more specifically, by matching on the propensity score. Following the steps of the RCM, the basic
goal of matching is to mimic randomisation in treatment and control studies using observational data where actual randomisation is impossible. The way to achieve this is to resample the data to construct unbiased samples of treatment (unemployed/workless) and control (in work). This re-sampling is carried out on the basis of similarities in a range of *a priori* identified confounding variables (from chapter two) which are used to match individuals. Matching techniques therefore represent a quasi-experimental approach for ascertaining causality in treatment (unemployment) and control (employment) studies that use non-randomised, observational data. As set out in the various rules of the RCM, the aim is to mimic a randomised experiment by identifying a counterfactual group of employed individuals who are otherwise as similar as possible to the unemployed individuals. This satisfies the condition that treatment status is hypothetically exchangeable between treatment and control groups. A necessary assumption for this is to argue selection on observables. The concept of selection on observables simply suggests that the treatment assignment mechanism that we are reconstructing has been observed within the data and is not a function of unobserved variables. If we satisfy this assumption we can then demonstrate that our treatment and control groups differ only in their treatment status (unemployment) and not on the set of observed confounding variables (Rubin 2007; Rubin 2008).

Matching within exact cells of the variable matrix is desirable. This involves constructing groups of workless who have corresponding matches with identical observed characteristics within the in work group. For example if the only confounding effect of interest was sex, it would be necessary to match each unemployed male with a male in employment and each unemployed female with a female in employment. This would be a relatively straightforward procedure with even a small sample. However as the number of confounding variables increases the number of unique combinations of those variables also increases dramatically and the possibility of finding an identical match decreases. In most cases sample sizes relative to the number of covariates is therefore too small to allow for exact matching (Rosenbaum and Rubin 1984). Thus matching is implemented on the basis of a propensity score for treatment. This requires the estimation (via logistic or probit regression) of a probability for treatment assignment which is calculated as a function of the observed variables with the resulting probability corresponding to the propensity score for treatment assignment for each case within the sample.

Oakes summarises two ways in which the propensity score can be utilised; sub-classification and direct matching² (Oakes and Johnson 2006). Sub-classification on the propensity score produces an effect estimate within each quintile of the propensity score which is then

---

² A third approach, known as regression weighting on the propensity score is also used in some cases but can pose problems with off support inferences and thus will not be considered.
weighted and averaged over all of the quintiles. Evidence from the technical literature suggests that the corresponding average effect size is consistently effective in reducing 90% of bias (Rosenbaum and Rubin 1984). The advantage of this approach is that it uses all of the cases within the dataset and does not discard unmatched cases. A disadvantage is that by retaining all cases in the study we may introduce bias by including some cases that are not adequately comparable to a control case. Thus, we run the risk of making inferences off support of the data when the treated and control groups become more imbalanced.

Another approach is to pool the propensity score and match individual control cases with individual treated cases through an approach known as nearest neighbour caliper matching (Rosenbaum and Rubin 1985). This technique implements the following steps. Firstly, the data is randomly ordered. Then, a treatment case is selected and is then matched to its closest neighbour from the reservoir of control case on the basis of the value of the propensity scores. To ensure that matched cases are appropriately similar, the matching algorithm incorporates a caliper of specified width. This caliper imposes an interval around the estimated propensity score for each of the treated or exposed cases. Unexposed or control cases whose propensity score falls outside this interval are disregarded for matching. Thus, the caliper ensures a treated case is matched to its nearest neighbour if, and only if, its nearest neighbour has a propensity score that falls within the specified caliper interval.

The value of this interval or caliper is defined by the researcher but, reflecting the scale of the propensity score as a probability, it must fall between 0.0 and 1.0. Thus, a caliper of width 1.0 would be, in effect, an unrestricted caliper and a caliper 0.0 would only match cases with an identical propensity score. To illustrate, consider an unemployed (treated) individual with a very high propensity score for unemployment of 0.8 whose nearest in work (control) neighbour has a corresponding score of 0.71. With a caliper width of 0.05 the acceptable interval around the treated case is 0.75 to 0.85 which would prevent a match being made to its nearest neighbour. However, with a caliper of width 0.1 the interval becomes 0.7 to 0.9 thus ensuring a match. In other words, setting a wide caliper ensures a greater number of matches but increases the risk of imbalanced treatment and control groups and thus increases the potential for bias. Conversely, a narrow caliper ensures good and unbiased matching but increases the number of unmatched and therefore unused treatment cases. The magnitude of the interval thus represents the boundaries within which we are prepared to accept that a treated and control subject are similar enough as to be considered exchangeable. In other words, the interval is narrow enough to be confident that differences in the outcome between treated and control subjects is due to the exposure but wide enough to allow for cases to be matched (Oakes and Johnson 2006).
The exact configuration of the matching algorithm is an iterative process with each specification evaluated using t-tests on comparisons of the prevalence of the variables used to estimate the propensity score by treatment and control groups pre- and post-matching. Thus, a better specification is one that removes significant differences across variables between the groups of employed and unemployed. Table 3.2 gives a hypothetical example of this procedure by comparing the distribution of an area based deprivation variable that may confound the unemployment mortality relationship.

Table 3.2 Hypothetical table to evaluate balance in terms of deprivation between unemployed and employed groups in an unmatched and matched sample using t-tests

<table>
<thead>
<tr>
<th>Deprivation quintiles</th>
<th>Un-Matched Sample</th>
<th>Matched sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unemployed (%)</td>
<td>Employed (%)</td>
</tr>
<tr>
<td>Least deprived</td>
<td>17</td>
<td>27***</td>
</tr>
<tr>
<td>2nd</td>
<td>23</td>
<td>30***</td>
</tr>
<tr>
<td>3rd</td>
<td>19</td>
<td>22***</td>
</tr>
<tr>
<td>4th</td>
<td>23</td>
<td>13***</td>
</tr>
<tr>
<td>Most Deprived</td>
<td>18</td>
<td>8***</td>
</tr>
</tbody>
</table>

*This variable ranks areas of residence according to the level of area based deprivation. The ranked distribution is then divided into quintiles thus giving information on whether an individual lives in a most through to least deprived area.

$t$-test significance: *** $p<0.001$, ns no significant difference

It shows that across unemployed and employed individuals in the unmatched sample, a much higher proportion of unemployed individuals live in more deprived areas with the t-tests indicating significant differences. However, the matched sample shows, with the removal of both employed and unemployed cases with no match, these significant differences are attenuated and balance is achieved. Once the matching specification is adequately completed unmatchable treated cases and those control cases without a matched treated case are excluded from the new sample. This new sample is then free to be analysed using standard estimation procedures. The precise implementation of the propensity score approach is detailed in the manuscripts where appropriate.

3.5 Summary

The purpose of this chapter has been to outline a framework by which a causal effect of worklessness and mortality can be tested. It has outlined a number of important methodological theories and relatively new and innovative analytical techniques that are relevant to the broader area of labour market status and health. The main areas that were covered include the application of counterfactual and potential outcome concepts of causal effects to the worklessness and health literature, the importance of the concept of over-
adjustment when attempting to control for confounding effects and the potential utility of the propensity score approach to the worklessness and mortality relationship. To conclude the chapter, this section summarises the material in order to draw out the research questions that are addressed in chapter’s four to six.

3.5.1 Selection versus causation

There still remains considerable debate on the subject of causation versus selection explanations for the unemployment and mortality relationship. A plausible theory exists, as outlined in section 2.4, by which unemployment may occupy a causal pathway to mortality, however, significant evidence using a number and combination of rich datasets and quasi-experimental approaches shows mixed results, though few find no effect at all. Those that do not find an effect tend to be from studies using factory closures as a natural experiment. Typically, results from these studies are difficult to generalise and may not represent the wider population. In many cases, evidence from other approaches suggests that a significant portion of the excess mortality in the unemployed is related to selection, with the conclusion that the resultant association is a reflection of confounding factors.

In the UK, much of the evidence attempting to reduce the effects of prior health and other confounding effects has relied on the use of wear-off periods. These result in significant losses of data and do not appear to fully account for all of the confounding effects and do not appear as effective as the more theoretically nuanced methods outlined in this chapter. As a result, after testing the effectiveness of these wear-off periods, the subsequent analyses will look to implement these more advanced methods. Rather than ignoring large amounts of mortality data through a wear-off period approach, these studies will apply counterfactual theory and an innovative analytical design.

3.5.2 Counterfactual causality and potential outcomes

The central tenet of the counterfactual approach is the fundamental problem of causal inference which describes the impossibility of observing the counterfactual or the outcome (death/not death) on the same individual at the same time under different treatment assignments (in work/workless). The counterfactual analytical approach imagines the counterfactual as a potential outcome which can be reconstructed in various ways. In the randomised experiment, this is achieved by substituting an exchangeable control (in work) subject so that the two cases theoretically differ only in their allocation of treatment assignment. Any remaining differences are assumed to be random. In non-experimental observational settings such as the worklessness and mortality relationship, experimental randomisation is impossible and thus inferring causal relationships between variables of
interest is considerably more challenging. A particular aim of the thesis, therefore, is to explicitly examine the degree of exchangeability of both the AU and the IDS in terms of the mechanisms of confounding and selection that were highlighted in chapter two.

3.5.3 Bias and over-adjustment

In chapter three, a distinction was made between bias due to confounding and bias due to selection in terms of the common effects and common causes causal structures outlined by Hernán, Hernández-Díaz & Robins (2004). This was important because of the sometimes confusing terminology in the unemployment and health & mortality literature. As a result, health selection was reduced to a problem of confounding with health representing a common cause of both unemployment and mortality. Furthermore, the notion of indirect selection which, again, is used extensively in the unemployment and health & mortality literature can also be regarded as confounding with the intermediary indirect variable such as personality and health both considered as confounders.

The importance of the timing of confounding effects was also outlined in order to avoid the problem of over-adjustment. For example, when unemployment and health are measured at the same time in, for example, a cross-sectional study, the relationship between the two is difficult to determine because some of the nominated indirect confounding characteristics may themselves be outcomes of the experience of unemployment depending on their occurrence in time relative to the unemployment event. Thus, variables that are descendant to the treatment event of interest (unemployment) should not be considered confounders due to their occurrence after treatment allocation. Similarly, antecedent events to the treatment should be considered as confounders. This reiterates the importance of the timing of these various confounding factors relative to unemployment. For example, factors that have been linked to health such as marital status (Holt-Lunstad, Birmingham & Jones 2008; Liu and Umberson 2008; Molloy et al. 2009) and housing tenure (Filakti and Fox 1995; Ellaway and Macintyre 1998; Macintyre et al. 2003) that occur prior to unemployment (antecedent) could be imagined to cause unemployment whereas those occurring after unemployment (descendant) could easily be considered as outcomes of unemployment. Adjustment for the descendant factors may remove the significant effect of unemployment and, as a result may under-estimate its relationship to health. However, adjustment for the antecedent factors is crucial in order to remove their potential effects on unemployment and health jointly. The studies in this thesis will take account of potential over-adjustment and will compare the prevalence of any factors antecedent to unemployment allowing greater understanding of some of the important characteristics of individuals that become workless.
3.5.4 Reconstructing the ‘experiment’

Key to the counterfactual framework of causality in observational data is the process of reconstructing the imaginary counterfactual based experiment from which the data was generated. This approach, known as the Rubin Causal Model (RCM), is for the most part a thought exercise encouraging the researchers to think through the data as if it were a genuine experiment. This specifically involves imagining the hypothetical rules governing allocation to treatment (worklessness) which can then be reconstructed in terms of nominated confounding variables. During this process it is important that all of the outcome data is stripped from the dataset to avoid the deterministic fitting of prior hypotheses. This approach was extended into the discussion of a matching approach based on the propensity score. This involves producing matched pairs of treated (workless) and control (in work) subjects who are determined to be exchangeable on the basis of an estimated propensity score for treatment (worklessness). The propensity score is estimated on the basis of a number of important confounding variables which are antecedent to the worklessness event.

3.6 Research questions and Data

3.6.1 Research questions

Out of the reviews presented in chapters two and three has emerged three research questions that will be addressed in chapter’s four to six of this thesis. These questions are presented below in the form of short abstracts for each of the papers

3.6.1.1 In studies of employment status and mortality, to what extent is health selection of the unemployed mitigated by the use of wear-off periods?

Testing whether unemployment causes health deterioration is complicated because failing health may increase the probability of unemployment. In some previous studies of unemployment and mortality in the UK, a ‘wear-off’ period, after employment status is observed, is used which ignores the first few years of mortality events. It is assumed that selection effects will wear-off during this period. This study tests the effectiveness of using wear-off periods. Using data from the Scottish Longitudinal Study (SLS) and the England & Wales Longitudinal Study (ONS LS) logistic regression models will be used to model the odds of death in a given time period after the 1991 census for those aged 35 – 64 in 1991. Odds ratios for the different employment statuses will be compared alongside a comparison of the changes in risk associated with cumulatively increasing the length of wear-off prior to follow-up.
3.6.1.2 Is there evidence of a causal link between unemployment and mortality in Scotland using individual level data in a causal effects framework with adjustment for prior health?

Unemployment is associated with mortality. However, whether this relationship is causal is a matter of continuing debate with confounding (factors correlated with unemployment) and direct selection (poor health causing unemployment) offered as alternative explanations. This study will utilise routinely collected population level administrative data from Scotland and a causal research design (propensity score matching) to explore whether unemployment is causally related to mortality. The sample includes men aged between 35 and 54 who were in work in 1991. Subsequent employment status in 2001 will be observed (in work or unemployed) and the all-cause mortality risk of unemployment between 2001 and 2006 estimated. To account for potential selection into unemployment of those in poor health the study will use propensity score matching to construct comparable samples of in work and unemployed individuals. Matching will be based on a wide range of explanatory variables including health status prior to year of unemployment (hospital admissions and self-reported limiting long term illness) as well as measures of socio-economic position.

3.6.1.3 Is there evidence of a causal link between worklessness due to permanent sickness and mortality in Scotland using individual level data with adjustment for prior health?

Labour market disadvantage in terms of active unemployment has been consistently linked to poor health and mortality though whether this association is causal remains a topic of debate. Similar evidence for the independent mortality effect of inactivity due to sickness is limited despite strong evidence that this group contains high numbers of ‘hidden’ unemployed. This study will incorporate a causal effect framework that is very similar to the previous study to examine the mortality risk of non-retired working age males who are prevented from working due to illness and disability. Longitudinal analysis of 19,165 men from the Scottish Longitudinal Study which contains socio-economic information will be linked to mortality and hospital records data. The analysis was carried out again using propensity score matching and cox proportional hazards models.

3.6.2 Data

Data for each of the studies was obtained from the Scottish Longitudinal Study (SLS). The SLS is a census based study containing a 5.3% sample of the Scottish population. These individuals are selected randomly at the 1991 census and linked to their corresponding records in the 2001 census. Additional linkages are made to hospital admissions data in chapters five and six. The precise configuration and variable selection in each of the studies
is covered in more detail in each of the chapters together with detailed outline of the study design and statistical analysis. However, there are two issues that over-arch each of the papers in terms of the choice of data and variables that warrant further discussion prior to the empirical chapters. These are, firstly, general problems around the classification of economic inactivity in the census and, secondly, the difficulties associated with classifying female participation in the labour market.

3.6.2.1 Classification of employment status in the census

The measurement of employment status in the thesis is based around the self-reported economic position question in the census. In both 1991 and 2001, this question differentiates between the in work, the AU, the permanently sick and a number of other categories including retired, looking after family or home and an ‘other inactive’ category. This may introduce problems, particularly for the group of permanently sick, in terms of discrepancies between this census measure and the actual numbers of individuals claiming incapacity benefit, which is the real population of interest. As has been discussed in the preceding sections, this group is likely to be relatively heterogeneous and the lack of additional information in the census could lead to the analysis wrongly excluding certain individuals and missing the real population of interest. For example, some economically inactive individuals with a disability or illness may record themselves as ‘other inactive’ which comprises individuals who do not reside in any of the other remaining inactive categories. Furthermore, there may be other sick and disabled individuals who may have responsibilities such as single parenthood or who may be early retirees but whom may otherwise consider themselves capable and desiring of work. None of these individuals would be captured in the permanently sick group and thus the analysis would not be able distinguish them from others in their respective groups.

The extent of this problem can be examined using information on presence of limiting illness or disability, employment status and desire for work; all of which can be found in data from the labour force survey. Table 3.3 presents this data in a cross-tabulation of presence of ill health across more detailed levels of labour market activity for working age men only. In terms of the potential for wrongly excluding individuals on the basis of using the permanently sick category there are to main groups of interest; individuals who have a health condition but who would like to work amongst both the retired and those looking after family. In both cases however, the prevalence of men in either of these labour market states is very low (< 1% of the entire sample) relative to other forms of labour market disadvantage which suggests that this issue is unlikely to be significant in the census analyses.
Table 3.3: Labour market status by presence of health problem for working age men. Raw cell counts with cell percentages in brackets.

<table>
<thead>
<tr>
<th>Labour market status</th>
<th>Presence of health issue (cell %)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td><strong>In employment</strong></td>
<td></td>
</tr>
<tr>
<td>In employment</td>
<td>4,397</td>
</tr>
<tr>
<td></td>
<td>(15.17)</td>
</tr>
<tr>
<td><strong>ILO unemployed</strong></td>
<td></td>
</tr>
<tr>
<td>ILO unemployed</td>
<td>433</td>
</tr>
<tr>
<td></td>
<td>(1.49)</td>
</tr>
<tr>
<td><strong>Looking after family</strong></td>
<td></td>
</tr>
<tr>
<td>Seeking but unavailable to work</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
</tr>
<tr>
<td>Not Seeking work but would like work</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
</tr>
<tr>
<td>Not Seeking and would not like work</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>(0.28)</td>
</tr>
<tr>
<td>Total</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>(0.4)</td>
</tr>
<tr>
<td><strong>Retired</strong></td>
<td></td>
</tr>
<tr>
<td>Not Seeking but would like work</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
</tr>
<tr>
<td>Not Seeking and would not like work</td>
<td>612</td>
</tr>
<tr>
<td></td>
<td>(2.11)</td>
</tr>
<tr>
<td>Total</td>
<td>640</td>
</tr>
<tr>
<td></td>
<td>(2.17)</td>
</tr>
<tr>
<td><strong>Sick</strong></td>
<td></td>
</tr>
<tr>
<td>Seeking work but unable to work due to temporary sickness or injury</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
</tr>
<tr>
<td>Seeking work but unable to work due to long term sickness or disability</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
</tr>
<tr>
<td>Not Seeking work due to temporary sickness or injury but would like work</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>(0.24)</td>
</tr>
<tr>
<td>Not seeking work due to long term sickness or disability but would like work</td>
<td>536</td>
</tr>
<tr>
<td></td>
<td>(1.85)</td>
</tr>
<tr>
<td>Not seeking work due to temporary sickness or injury and would not like work</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>(0.18)</td>
</tr>
<tr>
<td>Not seeking work due to long term sickness or disability and would not like work</td>
<td>1,394</td>
</tr>
<tr>
<td></td>
<td>(4.81)</td>
</tr>
<tr>
<td>Total</td>
<td>2,074</td>
</tr>
<tr>
<td></td>
<td>(7.15)</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td>7,664</td>
</tr>
<tr>
<td></td>
<td>(31.31)</td>
</tr>
</tbody>
</table>

*Cell percentages reflect contents of that cell as a percentage of all individuals in the table (n=28402)
Source: Labour force survey, authors own analysis. Data from 2006.

As well as possibly wrongly excluding individuals, the use of the permanently sick measure in the census to capture economic inactivity due to sickness might also risk wrongly *including* certain individuals. For example, it could be argued that the analysis of research question three should only focus on those individuals who are seeking or who would like to work but are unable to do so due to long or short-term sickness or disability. In other words, the analysis should remove those individuals who are not seeking nor want to work due to...
their illness or disability. On this basis, table 3.3 suggests that analyses relying on the permanently sick census measure would include a figure of around 5% of the population of working age men who would not be relevant for the analysis. This figure represents around 72% of the entire population of IDS. However, I argue that the negative effects of labour market disadvantage that were outlined in section 2.4 are likely to impact on all individuals in the group of IDS irrespective of their desire for a job. From this perspective it is important to keep these individuals in the analysis as they provide a more complete representation of the experience of IDS and therefore its relationship to poor health and mortality. Furthermore, the configuration of the propensity score analysis means that all individuals in the IDS group who are significantly different to any individual in the group of employed in terms of the observed characteristics will be dropped from the analysis for the reasons of comparability and exchangeability that have been discussed in this chapter. This further reduces the possibility that we will be wrongly including significant numbers of individuals in the analysis. Nonetheless, it would be a useful addition to the analysis to be able to distinguish between different forms of IDS in order to allow for subgroup analysis to compare the different forms or experiences of IDS in terms of their effects for health.

3.6.2.2 Female labour market participation

It is widely recognised that female labour market participation differs compared to men in terms of reduced labour market attachment and greater involvement in household responsibilities such as looking after a family (Weatherall, Joshi & Macran 1994; Rubery, Fagan & Maier 1996; Gonzalo and Saarela 2000; Bivand 2005; Roelfs et al. 2011). Furthermore, some argue that female identities are less tied to work and income (Hakim 1991; Paul & Moser 2009; Strandh et al. 2012). However, female labour market participation is, and has been, changing (Grant, Price & Buckner 2006). For example, there has been a steady increase in the numbers of female participants in the labour market since 1984, rising by 7% by 2007 (Beatty et al. 2009). This suggests that the labour market has begun to draw on the reserve labour pool of previously economically inactive women which was primarily dominated by those looking after families on a full time basis (Beatty et al. 2009). Furthermore, as deindustrialisation took hold, the increase in the number of jobs associated with the service sector has led to more opportunities for work for women to the extent that patterns of female and male labour market patterns continue to converge (Beatty et al. 2009). Alongside these patterns, recent evidence also suggests that within areas of heavy industry decline, female rates of incapacity benefit claimants often match those of men, suggesting significant linkages between male and female labour markets. These patterns have occurred alongside qualitative evidence that, actually, women suffer from the effects of labour market detachment in much the same way as men (Grant, Price & Buckner 2006).
These patterns point to the fact that, increasingly, understanding the effect of worklessness for the health of women represents an important component for understanding the contemporary public health effect of worklessness (Grant, Price & Buckner 2006; Leeflang et al., 1992). Despite previous evidence that the mortality effect of unemployment is worse for men than women (Roelfs et al. 2011), Bambra and Eikemo (2009) note that, for self-rated health, this pattern is less certain. Furthermore, the authors note that the potential for health related confounding amongst women might well be far greater than for men (Bambra and Eikemo 2009).

Nevertheless, examining these issues using traditional census data is often hampered by the nature of the labour market status question in the census, which does not provide detailed information about the voluntary or involuntary nature of family roles and responsibilities. Although the census question contains a separate category for ‘looking after home/family’, it contains no additional information about the extent to which the women carrying out these roles are doing so voluntarily, or whether they are also actively desiring and seeking work. As a result, it is difficult to ascertain the extent to which some women may be hidden from both the AU group and the IDS group through their self-identification as carers and mothers, for example. To shed some light on the extent of this issue, Table 3.4 presents data from the labour force survey in 2006 illustrating labour market status and presence of an illness or health issue for working age women. It shows that, when compared to the table for men (Table 3.4), there are significantly more women than men who carry out responsibilities in the home with just under half (42%) of all non-employed women reporting responsibilities at home and for the family as their main economic activity. Furthermore, 20% of those with home responsibilities would like to be working and over a quarter have a health issue. As a result, use of census categories to examine female labour market status and mortality risks miss-classifying and miss-representing many women’s labour market experiences. Moreover, the precise configuration of the analyses in chapters three and four in particular means that baseline characteristics and covariates are observed at the 1991 census. Given the convergence in men’s and women’s labour market participation patterns since the mid-1980s, this means that the patterns identified in Table 3.4 (from 2006) are likely to be more pronounced at the time when individuals are selected into the studies.
Table 3.4: Labour market status by presence of health problem for working age women. Raw cell counts with cell percentages in brackets.

<table>
<thead>
<tr>
<th>Presence of health issue (%)</th>
<th>Labour market status</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>In employment</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4,332</td>
<td>15,050</td>
</tr>
<tr>
<td></td>
<td>(15.24)</td>
<td>(52.95)</td>
<td>(68.19)</td>
</tr>
<tr>
<td></td>
<td>ILO unemployed</td>
<td>292</td>
<td>764</td>
</tr>
<tr>
<td></td>
<td>(1.03)</td>
<td>(2.69)</td>
<td>(3.72)</td>
</tr>
<tr>
<td>Looking after family</td>
<td>Seeking but unavailable</td>
<td>23</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.26)</td>
<td>(0.34)</td>
</tr>
<tr>
<td></td>
<td>Not Seeking work but would like work</td>
<td>218</td>
<td>527</td>
</tr>
<tr>
<td></td>
<td>(0.77)</td>
<td>(1.85)</td>
<td>(2.62)</td>
</tr>
<tr>
<td></td>
<td>Not Seeking and would not like work</td>
<td>772</td>
<td>2,147</td>
</tr>
<tr>
<td></td>
<td>(2.72)</td>
<td>(7.55)</td>
<td>(10.27)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>1,013</td>
<td>2,749</td>
</tr>
<tr>
<td></td>
<td>(3.57)</td>
<td>(9.66)</td>
<td>(13.23)</td>
</tr>
<tr>
<td>Retired</td>
<td>Not Seeking but would like work</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.02)</td>
</tr>
<tr>
<td></td>
<td>Not Seeking and would not like work</td>
<td>292</td>
<td>299</td>
</tr>
<tr>
<td></td>
<td>(1.03)</td>
<td>(1.05)</td>
<td>(2.08)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>296</td>
<td>301</td>
</tr>
<tr>
<td></td>
<td>(1.04)</td>
<td>(1.06)</td>
<td>(2.10)</td>
</tr>
<tr>
<td>Sick</td>
<td>Seeking work but unable to work due to temporary sickness injury</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.01)</td>
<td>(0.04)</td>
</tr>
<tr>
<td></td>
<td>Seeking work but unable to work due to long term sickness or disability</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.00)</td>
<td>(0.01)</td>
</tr>
<tr>
<td></td>
<td>Not Seeking work due to temporary sickness or injury but would like work</td>
<td>55</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>(0.19)</td>
<td>(0.07)</td>
<td>(0.27)</td>
</tr>
<tr>
<td></td>
<td>Not seeking work due to long term sickness or disability but would like work</td>
<td>381</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>(1.34)</td>
<td>(0.03)</td>
<td>(1.37)</td>
</tr>
<tr>
<td></td>
<td>Not seeking work due to temporary sickness or injury and would not like work</td>
<td>70</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>(0.25)</td>
<td>(0.04)</td>
<td>(0.28)</td>
</tr>
<tr>
<td></td>
<td>Not seeking work due to long term sickness or disability and would not like work</td>
<td>1,286</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>(4.52)</td>
<td>(0.05)</td>
<td>(4.58)</td>
</tr>
<tr>
<td></td>
<td>Totals</td>
<td>1,802</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>(6.34)</td>
<td>(0.20)</td>
<td>(6.55)</td>
</tr>
<tr>
<td>Totals</td>
<td>7,735</td>
<td>18,922</td>
<td>26,657</td>
</tr>
<tr>
<td></td>
<td>(29.02)</td>
<td>(70.98)</td>
<td>(100)</td>
</tr>
</tbody>
</table>

Cell percentages reflect contents of that cell as a percentage of all individuals in the table (n=)
Source: Labour force survey, authors own analysis. Data from 2006.

These data limitations and the primarily methodological focus of the thesis therefore preclude an examination of female labour market participation. As a result the findings are restricted in terms of their scope and are not generalizable to the wider working age female population. As such, the reader should be careful to remember that the observed effects in the empirical chapters should only be interpreted as a mortality effect for the population of
working age men. The exclusion of women from the analysis, however, does not undermine what remains a crucial component of the worklessness and health & mortality relationship. As Bambra (2010) reiterates, in light of the current recession and the growing importance of work and worklessness for the health and wellbeing of women as well as men, there is an ongoing imperative for research to include women in studies of labour market disadvantage and mortality.
Chapter 4

UNEMPLOYMENT, MORTALITY AND THE PROBLEM OF HEALTH-RELATED CONFOUNding: EVIDENCE FROM THE SCOTTISH AND ENGLAND & WALES (ONS) LONGITUDINAL STUDIES

4.1 Introduction

The potential for unemployment to negatively affect an individual’s health status has been the focus of much research (Bartley, Ferrie, & Montgomery, 2006). Associations between a spectrum of health outcomes and unemployment have been empirically borne out in the literature including; mental health (Nordenmark & Strandh, 1999; Thomas, Benzeval, & Stansfeld, 2005), substance use and teenage pregnancy (Fergusson, Horwood, & Woodward, 2001), suicidal behaviours (Blakely, Collings, & Atkinson, 2003) and limiting long-term illness (LLTI) (Bartley, Sacker, & Clarke, 2004). In addition to these outcomes much work has sought to investigate associations between unemployment and mortality (Fox, Goldblatt, & Adelstein, 1982; Gerdtham & Johannesson, 2003; Morrell et al., 1999; Moser, Fox, & Jones, 1984; Moser, Goldblatt, & Fox, 1987; Stefansson, 1991; Voss et al., 2004).

3 Please note that the formatting in this chapter, particularly the reference style, may vary compared to other chapters due to specific formatting requirements of the journal in which this paper has been published. Full citation of this paper is as follows: Clemens, T., P. Boyle, et al. (2009). "Unemployment, mortality and the problem of health-related selection: Evidence from the Scottish and England & Wales (ONS) Longitudinal Studies." HSQ 43(1): 7-13.
Whilst many of these studies report statistically strong associations between unemployment and poor health, establishing this as a causal relationship poses a greater challenge as they rely on observational rather than experimental studies (Oswald, 2007; Rubin, 2008). One problem stems from the possibility that health may be both an outcome of, and a cause of, unemployment; ill members of the population may be selected into unemployment so any observational analysis could exaggerate the direct effect of unemployment on health or mortality (Bartley, 1996; Bartley & Ferrie, 2001). If such health selection takes place, even strong relationships between unemployment and health cannot be regarded as causal. More generally, of course, there may be other (unobserved) factors that select people into unemployment that could potentially bias any claims that unemployment per se can increase the risk of subsequent poor health and death.

A number of techniques have been applied in attempts to overcome such selection issues. Commonly, longitudinal data is used and the various approaches include controlling for baseline health (Martikainen, 1990; Montgomery et al., 1999), studying whether the impact of unemployment differs at times of low and high overall society-level unemployment when selection effects might be expected to vary (Martikainen, Mäki, & Jäntti, 2007; Martikainen & Valkonen, 1996, 1998), and studying the impact of unemployment experimentally when whole workforces are made or threatened with redundancy (Ferrie et al., 2001; Steenland & Pinkerton, 2008).

Another common approach, based on the work of Fox, Goldblatt & Adelstein (1982) is to exclude deaths in the first few years of follow-up after employment status is observed. If the relative mortality risk is lowered when these deaths are excluded this suggests that health selection may have biased upwards the overall mortality risk. As Bartley (1994 pg. 333) comments:

“In a cohort study, any group selected for physical illness should exhibit high mortality in the early years of follow up which returns towards the level of the rest of the cohort later on as those who are very ill die and the rest recover.”

There are reasons to suppose that this approach may not be as appropriate nowadays. In the UK, despite a downward trend in unemployment rates (defined as those out of work who are actively seeking work), there has been a steady increase in those of working age who are permanently sick with some evidence suggesting that many of those who may previously have identified themselves as unemployed have been diverted to sickness related benefits (Beatty & Fothergill, 2005). This may be typical of those with a limiting long-term illness (Bartley & Owen, 1996).
A consequence of this is that by the time of the 1991 census those classifying themselves as unemployed may be less likely to have been selected into unemployment because of poor health than may have been the case in the 1970s or 1980s when studies such as those by Fox, Goldblatt & Adelstein (1982) were conducted. Also, the addition of a question on work related LLTI in the 1991 UK census provides the opportunity to control for health selection in a way that was not possible in previous UK census-based studies.

In this study, we therefore exploit the Scottish Longitudinal Study (SLS) and the England and Wales ONS Longitudinal Study to explore possible health-related selection into unemployment using post-1991 mortality events linked to individuals whose employment status was recorded in the 1991 census. We assess the association of unemployment with mortality using various lengths of wear-off period and controlling for limiting illness, recorded in 1991.

4.2 Methods

4.2.1 Data and outcome variables

The data for Scotland was extracted from the SLS which links census records from 1991 and 2001 for 5.3% of the Scottish population. In addition, corresponding vital events registry data, from which we draw our mortality events, is also linked for this period and beyond up to 2003 (the last year for which mortality events had been recorded at the time this study was undertaken). The data for England and Wales were extracted from the ONS Longitudinal Study. The ONS LS links decennial census information for 1% of the population in England and Wales from 1971, along with their registration data. Information from the 1971, 1981, 1991 and 2001 Censuses is linked with vital events information (births, deaths and cancer registration), from which we draw our mortality events.

These longitudinal studies are dynamic samples; some members are lost to the study through emigration (moving to another country and settling), and death, and new members enter the study through birth and immigration (Blackwell et al., 2003; Boyle et al., 2009). Sample selection is based on birth dates, using twenty dates (day and month) in Scotland and four dates (day and month) in England and Wales and to create samples which are representative of the population.
Table 4.1 Length of wear-off period, mortality periods and mortality frequencies for each wear-off period by Scotland and England & Wales

<table>
<thead>
<tr>
<th>Length of wear-off period</th>
<th>Mortality periods</th>
<th>Number of deaths</th>
<th>England and Wales</th>
<th>Scotland</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 years</td>
<td>1991 - 1995</td>
<td>5,429</td>
<td>3,059</td>
<td></td>
</tr>
</tbody>
</table>

Source: Scottish Longitudinal Study

Our analysis focused on the working ages 35 up to state pension age of 64. Table 4.1 details the outcome variables that were constructed. Eight variables were constructed corresponding to different wear-off periods. The first involved 0 years of wear-off, and thus contained all mortality events that occurred within five years of the census (1991-1995). The second variable involved 1 year of wear-off and thus ignored all mortality events occurring one year after the census day but captured all mortality events in the subsequent 5 years (1992-1996). The third variable involved 2 years of wear-off and thus ignored all mortality events that occurred within 2 years of the census day, but captured mortality events in 1993-1997. The remaining variables were produced following the same procedure up to variable 8 which involved 7 years of wear-off and captured those deaths between 1998 and 2002. We decided to carry out 5 year follow up analyses, rather than 10 year, to allow for the maximum number of wear periods which also allowed us to extend this period two years beyond the traditional 5 year period that is often used whilst still allowing a substantial follow-up period.

4.2.2 Analysis

As the outcome was binary (whether the individual died or not), binomial logistic regression models were used for each of the different length wear-off periods beginning with 0 years through to 7 years successively. Two sets of two models were produced; a first set without direct control for LLTI and a second including this direct control. Within each set a base model including only age, age squared and sex was produced followed by a more complex full model including additional explanatory variables drawn from the 1991 census. The explanatory variables were chosen on the basis of known associations with mortality and included; age, age squared, sex, social class, marital status, educational attainment (1991),
Table 4.2 Frequencies and percentages for economic position, sex, social class, marital status, Carstairs deprivation, ethnicity and limiting long-term illness by Scotland and England & Wales.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Scotland</th>
<th>Frequency</th>
<th>%</th>
<th>England and Wales</th>
<th>Frequency</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Economic position</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In work</td>
<td>62,856</td>
<td>65.50</td>
<td></td>
<td>132,252</td>
<td>68.27</td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>5,600</td>
<td>5.84</td>
<td></td>
<td>10,841</td>
<td>5.60</td>
<td></td>
</tr>
<tr>
<td>Retired</td>
<td>6,487</td>
<td>6.76</td>
<td></td>
<td>13,104</td>
<td>6.76</td>
<td></td>
</tr>
<tr>
<td>Permanently Sick</td>
<td>8,726</td>
<td>9.09</td>
<td></td>
<td>12,705</td>
<td>6.56</td>
<td></td>
</tr>
<tr>
<td>Other Inactive</td>
<td>12,294</td>
<td>12.81</td>
<td></td>
<td>24,810</td>
<td>12.81</td>
<td></td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>46,747</td>
<td>48.71</td>
<td></td>
<td>96,522</td>
<td>49.83</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>49,216</td>
<td>51.29</td>
<td></td>
<td>97,190</td>
<td>50.17</td>
<td></td>
</tr>
<tr>
<td><strong>Social Class</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professional occupations</td>
<td>3,218</td>
<td>3.35</td>
<td></td>
<td>7,374</td>
<td>3.81</td>
<td></td>
</tr>
<tr>
<td>Managerial and technical occupations</td>
<td>22,331</td>
<td>23.27</td>
<td></td>
<td>48,496</td>
<td>25.04</td>
<td></td>
</tr>
<tr>
<td>Skilled non-manual occupations</td>
<td>15,961</td>
<td>16.63</td>
<td></td>
<td>34,377</td>
<td>17.75</td>
<td></td>
</tr>
<tr>
<td>Skilled manual occupations</td>
<td>17,179</td>
<td>17.90</td>
<td></td>
<td>34,501</td>
<td>17.81</td>
<td></td>
</tr>
<tr>
<td>Partly skilled occupations</td>
<td>12,599</td>
<td>13.13</td>
<td></td>
<td>26,461</td>
<td>13.66</td>
<td></td>
</tr>
<tr>
<td>Unskilled occupations</td>
<td>7,413</td>
<td>7.72</td>
<td></td>
<td>11,579</td>
<td>5.98</td>
<td></td>
</tr>
<tr>
<td>Armed forces</td>
<td>259</td>
<td>0.27</td>
<td></td>
<td>517</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td>No job in last ten years/not stated</td>
<td>17,003</td>
<td>17.72</td>
<td>30,407</td>
<td>15.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Marital Status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>8,639</td>
<td>9.00</td>
<td></td>
<td>15,766</td>
<td>8.14</td>
<td></td>
</tr>
<tr>
<td>Married (first marriage)</td>
<td>68,395</td>
<td>71.27</td>
<td></td>
<td>131,478</td>
<td>67.87</td>
<td></td>
</tr>
<tr>
<td>Remarried</td>
<td>7,350</td>
<td>7.66</td>
<td></td>
<td>21,159</td>
<td>10.92</td>
<td></td>
</tr>
<tr>
<td>Divorced</td>
<td>7,095</td>
<td>7.39</td>
<td></td>
<td>17,948</td>
<td>9.27</td>
<td></td>
</tr>
<tr>
<td>Widowed</td>
<td>4,484</td>
<td>4.67</td>
<td></td>
<td>7,361</td>
<td>3.80</td>
<td></td>
</tr>
<tr>
<td><strong>Carstairs quintiles</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One (least deprived)</td>
<td>19,116</td>
<td>19.94</td>
<td></td>
<td>31,001</td>
<td>16.00</td>
<td></td>
</tr>
<tr>
<td>Two</td>
<td>24,442</td>
<td>25.50</td>
<td></td>
<td>35,001</td>
<td>18.07</td>
<td></td>
</tr>
<tr>
<td>Three</td>
<td>20,187</td>
<td>21.06</td>
<td></td>
<td>37,065</td>
<td>19.13</td>
<td></td>
</tr>
<tr>
<td>Four</td>
<td>17,030</td>
<td>17.77</td>
<td></td>
<td>42,752</td>
<td>22.07</td>
<td></td>
</tr>
<tr>
<td>Five (most deprived)</td>
<td>15,087</td>
<td>15.74</td>
<td></td>
<td>47,884</td>
<td>24.72</td>
<td></td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>95,115</td>
<td>99.12</td>
<td></td>
<td>181,442</td>
<td>93.67</td>
<td></td>
</tr>
<tr>
<td>Non-white</td>
<td>848</td>
<td>0.88</td>
<td></td>
<td>12,270</td>
<td>6.33</td>
<td></td>
</tr>
<tr>
<td><strong>Limiting long-term illness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Has a health problem</td>
<td>14,885</td>
<td>15.51</td>
<td></td>
<td>25,996</td>
<td>13.42</td>
<td></td>
</tr>
<tr>
<td>Does not have a health problem</td>
<td>81,078</td>
<td>84.49</td>
<td>167,716</td>
<td>86.58</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Scottish Longitudinal Study

area deprivation quintile,4 ethnicity and finally, and most importantly, employment status. The frequencies and distribution of all of the explanatory variables is presented in Table 4.2.

---

4Deprivation was calculated using the Carstairs index. In Scotland this was calculated for 10,000 or so ‘Consistent Areas Through Time’ (CATTs); see Exeter et al. (2005) for further details. In England and Wales it was calculated for enumeration district.
Odds ratios associated with the different categories of employment status were then graphed to see patterns of change with different lengths of wear-off period. If selection into unemployment was an issue which could be properly addressed using a wear-off period, we would expect the odds of unemployment-related mortality to be higher in the initial period immediately after 1991 and lower in the later mortality periods, when we expect selection effects to have diminished.

4.3 Results

Figures 4.1 to 4.8 present the results from the analysis. In all of the graphs, and in line with previous studies, being out of work for whatever reason is associated with an increased risk of death relative to those in work. In the basic models in Figures 4.1 and 4.2, the unemployed are around two times more likely to die relative to those in work. Importantly, however, this relationship remains stable irrespective of the duration of wear-off that is used. However, for the permanently sick, there is a marked selection pattern, with higher relative mortality risk in the earlier periods shortly after the census followed by a steady decline with increased duration of wear-off. The addition of the social class, marital status, educational attainment, area deprivation and ethnicity variables (Figures 4.3 and 4.4) attenuated the relationship slightly but the relative relationships remained unchanged for unemployment and permanent sickness. The relationship between unemployment and mortality remained stable irrespective of the duration of the wear-off period that is used.

Controlling for LLTI attenuates the relationship for the unemployed but the relative pattern remains the same, with the mortality risk for the unemployed unchanged by the duration of wear-off (Figures 4.5 and 4.6). The selection shape for the permanently sick that is apparent in Figures 4.1 to 4.4 disappears with the addition of LLTI as a control variable. This has the effect of removing the excess risk of death for the permanently sick to the point that it closely matches the risk magnitude of the unemployed, a relationship that is mostly constant across all of the different wear-off periods. This relationship holds when we control for LLTI as well as the other explanatory variables (Figures 4.7 to 4.8).

5 Please note that all ratios are significantly different to full-time employees ($p < .05$) except for the ratio for period seven for the other inactive group in the LLTI controlled complex models.
Figure 4.1 Mortality odds ratios for levels of economic position within a given wear-off period with further adjustment for age, age^2 and sex in Scotland

Source: Scottish Longitudinal Study
Figure 4.2 Mortality odds ratios for levels of economic position within a given wear-off period with further adjustment for age, age\(^2\) and sex in England & Wales

Source: Scottish Longitudinal Study
Figure 4.3 Mortality odds ratios for levels of economic position within a given wear-off period with further adjustment for age, age^2, sex, social class, marital status, Carstairs deprivation and ethnicity in Scotland

Source: Scottish Longitudinal Study
Figure 4.4 Mortality odds ratios for levels of economic position within a given wear-off period with further adjustment for age, age², sex, social class, marital status, Carstairs deprivation and ethnicity in England & Wales

Source: Scottish Longitudinal Study
Figure 4.5 Mortality odds ratios for levels of economic position within a given wear-off period with further adjustment for age, age\(^2\) and sex and additionally limiting long-term illness in Scotland

Source: Scottish Longitudinal Study
Figure 4.6 Mortality odds ratios for levels of economic position within a given wear-off period with further adjustment for age, age$^2$ and sex and additionally limiting long-term illness in England and Wales.

Source: Scottish Longitudinal Study
Figure 4.7 Mortality odds ratios for levels of economic position within a given wear-off period with further adjustment for age, age², sex, social class, marital status, Carstairs deprivation and ethnicity and additionally limiting long-term illness in Scotland

Source: Scottish Longitudinal Study
Figure 4.8 Mortality odds ratios for levels of economic position within a given wear-off period with further adjustment for age, age2, sex, social class, marital status, Carstairs deprivation and ethnicity and additionally limiting long-term illness in England and Wales.

Source: Scottish Longitudinal Study
4.4 Discussion

In line with almost all work completed previously in this field, being unemployed is related to significantly higher odds of death relative to being employed in all of our models. This relationship is apparent both generally and after adjusting for individual socio-economic circumstances.

However, as described above, drawing causal conclusions about this relationship has been complicated by the potential for ill health to influence employment status. Thus, the strong associations between unemployment and mortality may reflect selection of the unhealthy into unemployment rather than a causal effect of unemployment on health per se. A popular technique in prospective studies has been to allow these potential selection effects to wear-off by ignoring all mortality events that occurred within the first five years. As Fox, Goldblatt & Adelstein (1982) suggest, this period allows for non-steady state unemployed individuals to either recover and return to the labour force or die.

We therefore explored whether it was possible to detect a change in the odds of death of the unemployed relative to the employed depending upon the length of the wear-off period. We were interested in both the effectiveness of this technique as well as potentially illustrating the effect and magnitude of these selection effects. Assuming that this wear-off period approach is appropriate, the evidence presented above suggests that the process of selection into unemployment of sick individuals is non-existent. If strong selection effects were present, we might expect the unemployed group to experience higher likelihoods of death initially with this figure declining steadily over time until reaching a steady state around the five year wear-off period. Indeed we might expect the distribution, at a lower magnitude, to match more closely the pattern exhibited by the permanently sick category in which we see a steady decline in the likelihood of death as those suffering acute illness either die or recover. This was not the case. Furthermore, direct control for LLTI had the effect of drastically reducing the apparent effects of selection in the permanently sick and we might argue that this control for baseline health may have isolated the effect of health-related selection out of work on health. However, it is important to note that nearly all individuals who are registered permanently sick in the census are also coded as having a LLTI which may complicate this conclusion.

Our results seem to accord with some suggestions made in the literature previously, although there seem to have been few rigorous tests of this effect. Thus, Moser, Fox & Jones (1984) used a “wear-off” period following the 1971 England and Wales census and suggested that the overall mortality risk associated with unemployment was little changed by excluding
deaths in the initial years. And Bartley’s (1994 pg. 335) review of unemployment and mortality studies, mainly conducted using 1971 and 1981 census data, concludes that “this [wear-off] pattern is seen in men who are ‘permanently sick’ but not in the unemployed.”

These results, particularly the apparent lack of any selection processes among the unemployed, leave two possible explanations. First, our sample of individuals (aged 35 – 64) may include a majority of steady-state employment status experiences. For example, it is plausible that a significant majority of our group of unemployed have been unemployed for at least five years (prior to 1991) thus removing the potential for them to be selected into unemployment through ill health in this period. As we saw, the unemployed maintained constant higher odds ratios relative to the employed group. This explanation was suggested as a possible reason accounting for the lack of apparent selection effects found in the work of Fox, Goldblatt & Adelstein (1982). However, it seems highly improbable that a majority of the unemployed group would have been in this steady state of unemployment for the requisite length of time to mask the effects of selection.

A second explanation could be simply that the role or magnitude of any selection effects for employment status and health are minimal in this sample. The odds ratios for the unemployed, whilst being significantly higher relative to the employed, remained mostly constant suggesting that unemployed individuals are dying at the same rates regardless of the wear-off period after the census. In the literature to date, a five year period has been widely adopted as it is deemed an appropriate timescale a) for unemployment to have had an impact on a person’s health if one were to exist, b) to allow for those individuals who became unemployed due to poor health to die or recover and return to good health and therefore c) to begin to rule out the possibility of reverse causality between unemployment and poor health. Our results perhaps suggest that the role of health in determining unemployment could be minimal in the first place. If our unemployed group had contained a significant number of recently unemployed individuals due to ill health, we would expect to see the patterns described by Fox and colleagues of an early peak in odds ratios followed by gradual decline.

In conclusion this study has found little evidence of a selection effect operating on the unemployed group. The use of a wear-off period following measurement of employment status may be an unnecessary and ineffective method for mitigating the effects of selection. Furthermore, we suggest that controlling for LLTI may go some way to isolating the effect of unemployment on health.
Chapter 5

THE CAUSAL EFFECT OF UNEMPLOYMENT ON ALL-CAUSE MORTALITY: AN INDIVIDUAL OBSERVATIONAL STUDY USING PROPENSITY SCORE ANALYSIS

5.1 Background

The association of unemployment with increased mortality risk has received extensive attention over the past 20-30 years with early studies finding a significant excess mortality risk attached to spells of unemployment (Moser, Fox & Jones 1984; Moser et al. 1986; Iversen et al. 1987; Moser et al. 1987; Martikainen 1990). However, the extent to which the association is causal has been questioned as prior poor health may cause unemployment as well as, of course, raising mortality risk (Bartley, Ferrie & Montgomery 2006). Such health selection is particularly difficult to assess due to the ethical and practical barriers that prevent a randomised experiment approach (Bonell et al. 2009). Furthermore, unemployment is not randomly distributed and is more common among certain groups (e.g. those from lower socio-economic backgrounds) and it may be that the deleterious health effects of these factors, rather than unemployment per se, is responsible for the increase in mortality risk. As a result, it is difficult to demonstrate a causal effect and the true influence of unemployment on mortality may be overestimated in observational studies (Bartley, Ferrie & Montgomery 2006).

More recent studies have attempted to mitigate the effects of health selection. Some have studied mortality rates for individuals made redundant following large-scale company
downsizing or collapse, as poor health would not be the reason for job loss (Martikainen, Maki & Jäntti 2007; Steenland and Pinkerton 2008). Similarly, periods of recession allow analyses of health selection because the pool of unemployed contains a lower prevalence of individuals forced out of work due to illness than during periods of lower overall unemployment (Martikainen and Valkonen 1996; Martikainen and Valkonen 1998). Furthermore, the availability of comprehensive data linkages between routine register based health and socio-economic data sources in the Nordic countries has allowed for direct health adjustment prior to unemployment (Lundin et al. 2010). Generally this more recent evidence concludes that selection bias rather than causality accounts for most of the excess mortality risk among the unemployed (Lundin and Hemmingsson 2009).

However, it may be a mistake to generalise from these studies. Firstly, we argue that some studies may have over-adjusted (Schisterman, Cole & Platt 2009) for certain nominated confounding factors that occurred subsequent to unemployment. For example, many studies adjust for confounding effects that are measured concurrently or cross-sectionally with unemployment (Nylen, Voss & Floderus 2001; Gerdtham and Johannesson 2003; Gardner and Oswald 2004) which does not allow examination of the timing of these events relative to unemployment. Such events could therefore be considered outcomes of unemployment rather than confounders on the causal pathway from unemployment to mortality. Secondly, many have been conducted in countries with relatively high spending on unemployment-related welfare and active labour market policies (Finland and Sweden in particular) which could act to protect individuals from the potential deleterious effects of unemployment. A recent comparative ecological study across European Union countries suggests this may be the case, particularly with regard to suicide rates (Stuckler et al. 2009). This study found that the national socio-political context may mediate the individual experience of unemployment to the extent that we might expect to see differences in the relative mortality risk between countries.

Despite this, adequate evidence that explicitly tests for health selection has yet to be gathered outside Scandinavia, mainly because of the limitations in available health data. Many studies rely on self-reported health measures that are recorded simultaneously with economic activity rather than more detailed historical health data allowing adjustment prior to unemployment. The timing of health events is particularly important to ensure that over-adjustment for poor health events that occur after unemployment can be avoided. In the absence of such linked data some studies employ the technique of ‘wear-off periods’ during which mortality events are ignored for a period subsequent to baseline observation to allow selection effects to diminish (Bethune 1996a; Bethune 1996b; Lewis and Sloggett 1998). However, the effectiveness of these approaches in mitigating health selection amongst the
unemployed has been questioned recently, perhaps because there is little health selection of the AU in the first place (Clemens, Boyle & Popham 2009). Consequently, there is a need to further explore the potential for unemployment to raise mortality risk in a context such as the UK where state expenditure for unemployment related welfare and active labour market programmes is relatively small compared to Scandinavian countries (Stuckler et al. 2009). Utilising both self-reported and hospital-based health data prior to unemployment coupled with a robust analytical approach, the present study addresses this need during a period of increasing concern over rising levels of national unemployment rates during the current economic recession.

5.2 Methods

5.2.1 Data and sample

The data was extracted from the Scottish Longitudinal Study (SLS). The SLS consists of linked 1991 and 2001 national census records for a 5.3% sample of the Scottish population (Boyle et al. 2009). Additional linkages have been made to vital events registry data, including death registration, from 2001 onwards and hospital admission records from the Scottish Morbidity Records (SMR) from 1980 onwards (Boyle et al. 2009). Initial sample selection from the SLS for the present study and corresponding sample sizes are detailed in Figure 5.1. The sample was restricted to males aged 35-54 in 1991. This age range was chosen in order to avoid problems associated with student related inactivity in the labour market. The economically inactive and unemployed in 1991 and those untraced at the 2001 census were excluded. Women were excluded from the analysis because of the often more complex ways in which they interact with the labour market (Gonzalo and Saarela 2000). All cases missing information for any of the baseline covariates (11) were excluded listwise from the sample. The resulting sample size was 17,810.

5.2.2 Study Design

The analysis was conducted using propensity score matching as detailed in panel one. The treatment group in this study was defined as those men who moved from employment in 1991 to unemployment in 2001 with the control group defined as those men who were in work in both 1991 and 2001. The propensity score (estimated probability of unemployment in 2001) was estimated for all individuals as a function of socio-demographic and health risk factors identified at baseline (1991 census) and from psychiatric and hospital admissions from 1980 until the year that each individual was last in work, or until 2001 for those who remained in work. The variables used to estimate the propensity score are detailed in Figure 5.1 and were chosen on the basis that they are factors known to influence the risk of
mortality and unemployment which thus might confound an observed unemployment-mortality relationship. The hospital admissions cases were coded using the international classification of diseases (ICD) versions nine and ten. Our analysis was specifically interested in health problems that could hinder attempts to work or that could cause unemployment and we thus adhered to a scheme suggested by Lynch, Holman & Moorin (2007) which identifies conditions that are activity limiting or disabling. Aggregation of the hospital admission information is detailed in Figure 5.1.

**Figure 5.1 Sample selection criteria (corresponding sample size in brackets) and outline of covariates used to estimate the propensity score for unemployment in 2001**

<table>
<thead>
<tr>
<th>Age restricted (35-54 in 1991) sample size: 67,846</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample selection criteria at baseline in 1991 census:</td>
</tr>
<tr>
<td>- Male (33,374)</td>
</tr>
<tr>
<td>- In work (27758)</td>
</tr>
<tr>
<td>- Full complement of covariate information (27739)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Socio-demographic and health Covariates from 1991 census:</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Age</td>
</tr>
<tr>
<td>- Housing tenure</td>
</tr>
<tr>
<td>- Educational attainment</td>
</tr>
<tr>
<td>- Ethnicity</td>
</tr>
<tr>
<td>- Marital status</td>
</tr>
<tr>
<td>- Area deprivation</td>
</tr>
<tr>
<td>- Social class</td>
</tr>
<tr>
<td>- Self-reported limiting long term illness</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ICD9 and ICD10 hospital admissions and cancer registrations from Scottish Morbidity Records:</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Illnesses grouped by ICD10 chapter scheme</td>
</tr>
<tr>
<td>- Each variable records at least one instance of an admission for an illness within that chapter</td>
</tr>
<tr>
<td>- Records are censored (i) at year of last employment prior to follow-up for the unemployed and (ii) at the start of follow-up for the in work</td>
</tr>
</tbody>
</table>

Final sample size of 17,810 for which propensity score for unemployment in 2001 was calculated
Using the values of the propensity score, we matched each treated unemployed case to a maximum of ten closest in work control cases ensuring that the propensity score values of matched cases were within an interval of width +/- 0.005 of the propensity score of the treated unemployed case. Unmatchable unemployed cases were excluded from the new sample due to the lack of a corresponding comparator case amongst the in work group. Once all of the treated cases had been matched any remaining cases amongst the in work group without an unemployed match were also dropped. The matching quality was evaluated with t-tests by comparing the prevalence of the variables used to estimate the propensity score by treatment and control groups pre- and post-matching. To determine the sensitivity of the results, we also estimated the mortality risk separately within quintiled subclasses of the propensity score across the entire sample (Rosenbaum and Rubin 1984). These estimates were then combined by taking a weighted average to produce an overall effect. The results from both approaches were compared.

The period of mortality follow-up started from the 2001 census (29th April 2001) to the end of 2006. Embarkations from the study during the follow-up period due to migration were identified and censored. Cox proportional hazards models were used to estimate the relative mortality risk of unemployment with a total of twelve models fitted to the mortality data. Of these twelve models, ten were fitted to the quintiled samples corresponding to one adjusted and one unadjusted model in each quintile and two were from the case matched sample corresponding to one adjusted and one unadjusted model. Adjusted models controlled for the variables that were used to construct the propensity score. Finally, when fitting models to the case matched sample it was necessary to apply a weight to those observations that were matched more than once.
5.3 Results

Table 5.1 is an overview of the sample of men drawn in 1991 and shows the number who became unemployed by 2001, the numbers who remained in employment, and the distribution of deaths among them in the follow-up period 2001-2006. This descriptive result based on unmatched samples demonstrates that the unemployed are 2.6 times more likely to die as the in work. This estimate ignores potential selection and confounding effects.

Table 5.1 Distribution of deaths during follow-up period (2001-2006) by unemployed and in work

<table>
<thead>
<tr>
<th>Employment status in 2001</th>
<th>Deaths</th>
<th>Died</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Survived</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In work (Row %)</td>
<td>16,608 (97·40)</td>
<td>444 (2·60)</td>
<td>17,052</td>
</tr>
<tr>
<td>Unemployed (Row %)</td>
<td>707 (93·27)</td>
<td>51 (6·73)</td>
<td>758</td>
</tr>
<tr>
<td>Total (Row %)</td>
<td>17,315 (97·22)</td>
<td>495 (2·78)</td>
<td>17,810</td>
</tr>
</tbody>
</table>

Source: Scottish Longitudinal Study

Table 5.2 and Table 5.3 compare the means and proportions of the covariates (dummy variables) used to predict the propensity score for unemployment in 2001 and notes significant differences across the in-work and the unemployed. Those who were unemployed in 2001 were, on average, significantly (p<0.05) more likely than the in work to have had a long term limiting illness in 1991 as well as being more likely to have been hospitalised prior to 2001 (when unemployment was recorded) for an infectious or parasitic disease, cancer (neoplasms), a mental or behavioural illness and factors influencing health status and contact with health services. They were also significantly older, less educated, more likely to reside in social housing and more deprived areas and more likely to be single or divorced and to occupy lower social class positions. Table 5.2 and 5.3 thus demonstrates the potential selection and confounding effects by illustrating the differences between the unemployed and the in work, both in relation to their socio-economic characteristics (confounding) and their prior health status (health selection).
Table 5.2 Characteristics of full sample in terms of socio-economic covariates predicting unemployment in 2001 and statistical tests of difference between 2001 unemployed and in work groups

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Unemployed</th>
<th>In work</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Age (in 1991)</td>
<td>43·9</td>
<td>42·8</td>
<td>&lt;0·0005</td>
</tr>
<tr>
<td>Housing Tenure (in 1991) %</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Owner Occupied</td>
<td>57·8</td>
<td>76·6</td>
<td>&lt;0·0005</td>
</tr>
<tr>
<td>Privately Rented</td>
<td>5·7</td>
<td>5·0</td>
<td>0·387</td>
</tr>
<tr>
<td>Social Housing</td>
<td>36·1</td>
<td>18·2</td>
<td>&lt;0·0005</td>
</tr>
<tr>
<td>Communal Establishment</td>
<td>1</td>
<td>1</td>
<td>0·399</td>
</tr>
<tr>
<td>Educational Attainment (in 1991) %</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>84·2</td>
<td>75·2</td>
<td>&lt;0·0005</td>
</tr>
<tr>
<td>Other Higher Qualifications (non-degree)</td>
<td>8·8</td>
<td>10·6</td>
<td>0·119</td>
</tr>
<tr>
<td>First Degree and Higher Degree</td>
<td>4·6</td>
<td>12·3</td>
<td>&lt;0·0005</td>
</tr>
<tr>
<td>Not Stated</td>
<td>2·4</td>
<td>1·9</td>
<td>0·313</td>
</tr>
<tr>
<td>Marital Status (in 1991) %</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married (first marriage)</td>
<td>66·6</td>
<td>79·6</td>
<td>&lt;0·0005</td>
</tr>
<tr>
<td>Single</td>
<td>15·6</td>
<td>7·5</td>
<td>&lt;0·0005</td>
</tr>
<tr>
<td>Remarried</td>
<td>9·5</td>
<td>8·1</td>
<td>0·176</td>
</tr>
<tr>
<td>Divorced</td>
<td>7·3</td>
<td>4·2</td>
<td>&lt;0·0005</td>
</tr>
<tr>
<td>Widowed</td>
<td>1·1</td>
<td>1</td>
<td>0·067</td>
</tr>
<tr>
<td>Deprivation Quintiles (in 1991) %</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Least Deprived Quintile</td>
<td>18·9</td>
<td>26·9</td>
<td>&lt;0·0005</td>
</tr>
<tr>
<td>2nd</td>
<td>22·3</td>
<td>29·9</td>
<td>&lt;0·0005</td>
</tr>
<tr>
<td>3rd</td>
<td>19·9</td>
<td>20·5</td>
<td>0·721</td>
</tr>
<tr>
<td>4th</td>
<td>22·2</td>
<td>14·5</td>
<td>&lt;0·0005</td>
</tr>
<tr>
<td>Most Deprived Quintile</td>
<td>16·8</td>
<td>8·3</td>
<td>&lt;0·0005</td>
</tr>
<tr>
<td>Social Class (in 1991) %</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professional Occupations</td>
<td>4·6</td>
<td>8·2</td>
<td>&lt;0·0005</td>
</tr>
<tr>
<td>Managerial and Technical Occupations</td>
<td>26·1</td>
<td>32·8</td>
<td>&lt;0·0005</td>
</tr>
<tr>
<td>Skilled Non-manual Occupations</td>
<td>9·1</td>
<td>9·1</td>
<td>0·983</td>
</tr>
<tr>
<td>Skilled Manual Occupations</td>
<td>35·8</td>
<td>33·1</td>
<td>0·134</td>
</tr>
<tr>
<td>Partly Skilled Occupations</td>
<td>17·8</td>
<td>12·7</td>
<td>&lt;0·0005</td>
</tr>
<tr>
<td>Unskilled Occupations</td>
<td>4·4</td>
<td>3·0</td>
<td>0·038</td>
</tr>
<tr>
<td>Armed Forces</td>
<td>1</td>
<td>1</td>
<td>0·38</td>
</tr>
<tr>
<td>No Job in Last 10 Years</td>
<td>1</td>
<td>1</td>
<td>&lt;0·0005</td>
</tr>
</tbody>
</table>

Source: Scottish Longitudinal Study

To avoid low numbers, cells with very low proportions are rounded up to one in all cases.
Table 5.3 Characteristics of full sample in terms of health covariates predicting unemployment in 2001 and statistical tests of difference between 2001 unemployed and in work groups

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Unemployed</th>
<th>In work</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-reported limiting long-term illness (in 1991) %</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>96.7</td>
<td>97.9</td>
<td>0.025</td>
</tr>
<tr>
<td>No</td>
<td>3.3</td>
<td>2.1</td>
<td>0.025</td>
</tr>
<tr>
<td>Hospital Admissions (during 1980 - 2001) %</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infectious and parasitic diseases</td>
<td>1.8</td>
<td>3.1</td>
<td>0.044</td>
</tr>
<tr>
<td>Neoplasms</td>
<td>4.7</td>
<td>7.2</td>
<td>0.009</td>
</tr>
<tr>
<td>Diseases of the blood and blood-forming organs disorders of the immune mechanism</td>
<td>1</td>
<td>1</td>
<td>0.774</td>
</tr>
<tr>
<td>Endocrine, nutritional and metabolic diseases</td>
<td>3.7</td>
<td>2.9</td>
<td>0.207</td>
</tr>
<tr>
<td>Mental and behavioural</td>
<td>6.9</td>
<td>2.6</td>
<td>&lt;0.0005</td>
</tr>
<tr>
<td>Diseases of the nervous system</td>
<td>2.6</td>
<td>2.5</td>
<td>0.785</td>
</tr>
<tr>
<td>Diseases of the eye and adnexa</td>
<td>1.8</td>
<td>2.1</td>
<td>0.642</td>
</tr>
<tr>
<td>Diseases of the ear and mastoid process</td>
<td>1.1</td>
<td>1.6</td>
<td>0.23</td>
</tr>
<tr>
<td>Diseases of the circulatory system</td>
<td>10.7</td>
<td>13</td>
<td>0.067</td>
</tr>
<tr>
<td>Diseases of the respiratory system</td>
<td>7.4</td>
<td>7.1</td>
<td>0.741</td>
</tr>
<tr>
<td>Diseases of the digestive system</td>
<td>18.9</td>
<td>20.3</td>
<td>0.343</td>
</tr>
<tr>
<td>Diseases of the skin and subcutaneous tissue</td>
<td>7.4</td>
<td>7.2</td>
<td>0.827</td>
</tr>
<tr>
<td>Diseases of the musculoskeletal system and connective tissue</td>
<td>9.5</td>
<td>10.5</td>
<td>0.363</td>
</tr>
<tr>
<td>Diseases of the genitourinary system</td>
<td>9</td>
<td>9.7</td>
<td>0.53</td>
</tr>
<tr>
<td>Congenital malformations, deformations and chromosomal abnormalities</td>
<td>1</td>
<td>1</td>
<td>0.052</td>
</tr>
<tr>
<td>Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified</td>
<td>17.5</td>
<td>15.6</td>
<td>0.149</td>
</tr>
<tr>
<td>Injury, poisoning and certain other consequences of external causes</td>
<td>15.7</td>
<td>13.6</td>
<td>0.096</td>
</tr>
<tr>
<td>External causes of morbidity and mortality</td>
<td>16.2</td>
<td>13.9</td>
<td>0.07</td>
</tr>
<tr>
<td>Factors influencing health status and contact with health services</td>
<td>15.6</td>
<td>20.4</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Source: Scottish Longitudinal Study
To avoid low numbers, cells with very low proportions are rounded up to one in all cases.

Figure 5.2 illustrates the distribution of the propensity score for unemployment in 2001 by unemployed and in work groups. Irrespective of actual movement into unemployment, the probability of becoming unemployed was very small with an unemployment rate of 4.3% in 2001. Furthermore, the minimum and maximum of the propensity score distribution for the in-work group lies outside the respective boundaries for the unemployed group, resulting in complete overlap of the unemployed propensity score distribution.

The sample that was derived from the case matching procedure is described in Tables 5.4 and 5.5. No significant differences between the in-work and unemployed are apparent in any of the pre-unemployment covariates used to predict the propensity score. The significant differences in the full sample that were observed in Tables 5.2 and 5.3 have been removed.
indicating that the new sample is balanced in the pre-unemployment covariates. This accounts for the selection and confounding problems identified above.

Figure 5.2 Distribution of the propensity score for unemployment in 2001 by in work and unemployed groups in 2001

Table 5.6 shows results from the mortality follow-up analyses, estimated from both the full sample sub-classification and the restricted case matched sample. A weighted average of the unadjusted and adjusted sub-class estimates produces overall effects (2.35 and 2.42 respectively) which are slightly higher than those acquired from the case matched sample (2.03 and 2.1 respectively). The estimates (adjusted and unadjusted) from the case matched sample are around 13% more conservative than those from the sub-class analysis. This is consistent with technical literature which suggests that the sub-class approach is effective at reducing up to 90% of the observed bias (Rosenbaum & Rubin 1984). The findings show a statistically significant (p > 0.01) doubling of the mortality risk in the period 2001-2006 for the unemployed relative to those in work.
Table 5.4 Characteristics of case-matched sample in terms of health covariates predicting unemployment in 2001 and statistical tests of difference between 2001 unemployed and in work groups

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Unemployed</th>
<th>In work</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean Age (in 1991)</strong></td>
<td>43.9</td>
<td>43.9</td>
<td>0.945</td>
</tr>
<tr>
<td><strong>Housing Tenure (in 1991) %</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Owner Occupied</td>
<td>58.2</td>
<td>58.2</td>
<td>0.988</td>
</tr>
<tr>
<td>Privately Rented</td>
<td>5.7</td>
<td>5.6</td>
<td>0.921</td>
</tr>
<tr>
<td>Social Housing</td>
<td>35.7</td>
<td>35.7</td>
<td>0.994</td>
</tr>
<tr>
<td>Communal Establishment</td>
<td>1</td>
<td>1</td>
<td>0.861</td>
</tr>
<tr>
<td><strong>Educational Attainment (in 1991) %</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>84.1</td>
<td>84.9</td>
<td>0.657</td>
</tr>
<tr>
<td>Other Higher Qualifications (non-degree)</td>
<td>8.9</td>
<td>9.5</td>
<td>0.706</td>
</tr>
<tr>
<td>First Degree and Higher Degree</td>
<td>4.6</td>
<td>3.4</td>
<td>0.218</td>
</tr>
<tr>
<td>Not Stated</td>
<td>2.4</td>
<td>2.2</td>
<td>0.852</td>
</tr>
<tr>
<td><strong>Marital Status (in 1991) %</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married (first marriage)</td>
<td>67.1</td>
<td>67.8</td>
<td>0.754</td>
</tr>
<tr>
<td>Single</td>
<td>15</td>
<td>14.5</td>
<td>0.792</td>
</tr>
<tr>
<td>Remarried</td>
<td>9.6</td>
<td>8.9</td>
<td>0.653</td>
</tr>
<tr>
<td>Divorced</td>
<td>7.3</td>
<td>7.8</td>
<td>0.726</td>
</tr>
<tr>
<td>Widowed</td>
<td>1.1</td>
<td>1</td>
<td>0.878</td>
</tr>
<tr>
<td><strong>Deprivation Quintiles (in 1991) %</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Least Deprived Quintile</td>
<td>19</td>
<td>19.3</td>
<td>0.896</td>
</tr>
<tr>
<td>2nd</td>
<td>22.4</td>
<td>22.6</td>
<td>0.944</td>
</tr>
<tr>
<td>3rd</td>
<td>19.9</td>
<td>20.3</td>
<td>0.86</td>
</tr>
<tr>
<td>4th</td>
<td>22.2</td>
<td>21.8</td>
<td>0.855</td>
</tr>
<tr>
<td>Most Deprived Quintile</td>
<td>16.5</td>
<td>16.1</td>
<td>0.837</td>
</tr>
<tr>
<td><strong>Social Class (in 1991) %</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professional Occupations</td>
<td>4.6</td>
<td>4.1</td>
<td>0.615</td>
</tr>
<tr>
<td>Managerial and Technical Occupations</td>
<td>26.3</td>
<td>27.7</td>
<td>0.537</td>
</tr>
<tr>
<td>Skilled Non-manual Occupations</td>
<td>9.2</td>
<td>8.1</td>
<td>0.455</td>
</tr>
<tr>
<td>Skilled Manual Occupations</td>
<td>36</td>
<td>35.8</td>
<td>0.929</td>
</tr>
<tr>
<td>Partly Skilled Occupations</td>
<td>17.9</td>
<td>18.5</td>
<td>0.757</td>
</tr>
<tr>
<td>Unskilled Occupations</td>
<td>4.3</td>
<td>4.5</td>
<td>0.821</td>
</tr>
<tr>
<td>Armed Forces</td>
<td>1</td>
<td>1</td>
<td>0.9</td>
</tr>
<tr>
<td>No Job in Last 10 Years</td>
<td>1.3</td>
<td>1</td>
<td>0.466</td>
</tr>
</tbody>
</table>

Source: Scottish Longitudinal Study

To avoid low numbers, cells with very low proportions are rounded up to one in all cases.
Table 5.5 Characteristics of case matched sample in terms of covariates predicting unemployment in 2001 and statistical tests of difference between 2001 unemployed and in work groups

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Unemployed</th>
<th>In work</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-reported limiting long-term illness (in 1991) %</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>96.7</td>
<td>97.3</td>
<td>0.479</td>
</tr>
<tr>
<td>No</td>
<td>3.3</td>
<td>2.7</td>
<td>0.479</td>
</tr>
<tr>
<td>Hospital Admissions (during 1980 - 2001) %</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infectious and parasitic diseases</td>
<td>1.9</td>
<td>1.8</td>
<td>0.91</td>
</tr>
<tr>
<td>Neoplasms</td>
<td>4.8</td>
<td>4.4</td>
<td>0.716</td>
</tr>
<tr>
<td>Diseases of the blood and blood-forming organs disorders of</td>
<td>1</td>
<td>1</td>
<td>0.746</td>
</tr>
<tr>
<td>the immune mechanism</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infective, nutritional and metabolic diseases</td>
<td>3.6</td>
<td>3.3</td>
<td>0.8</td>
</tr>
<tr>
<td>Mental and behaviour</td>
<td>6.8</td>
<td>6.9</td>
<td>0.938</td>
</tr>
<tr>
<td>Diseases of the nervous system</td>
<td>2.7</td>
<td>2.7</td>
<td>0.988</td>
</tr>
<tr>
<td>Diseases of the eye and adnexa</td>
<td>1.9</td>
<td>1.7</td>
<td>0.786</td>
</tr>
<tr>
<td>Diseases of the ear and mastoid process</td>
<td>1.1</td>
<td>1</td>
<td>0.592</td>
</tr>
<tr>
<td>Diseases of the circulatory system</td>
<td>10.8</td>
<td>10.8</td>
<td>0.971</td>
</tr>
<tr>
<td>Diseases of the respiratory system</td>
<td>7.4</td>
<td>7.4</td>
<td>0.973</td>
</tr>
<tr>
<td>Diseases of the digestive system</td>
<td>19</td>
<td>17.6</td>
<td>0.501</td>
</tr>
<tr>
<td>Diseases of the skin and subcutaneous tissue</td>
<td>7.3</td>
<td>6.6</td>
<td>0.606</td>
</tr>
<tr>
<td>Diseases of the musculoskeletal system and connective tissue</td>
<td>9.6</td>
<td>8.8</td>
<td>0.614</td>
</tr>
<tr>
<td>Diseases of the genitourinary system</td>
<td>9</td>
<td>8.2</td>
<td>0.552</td>
</tr>
<tr>
<td>Congenital malformations, deformations and chromosomal abnormalities</td>
<td>1</td>
<td>1</td>
<td>0.752</td>
</tr>
<tr>
<td>Symptoms, signs and abnormal clinical and laboratory findings, not</td>
<td>17.7</td>
<td>16.8</td>
<td>0.641</td>
</tr>
<tr>
<td>elsewhere classified</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Injury, poisoning and certain other consequences of external causes</td>
<td>15.7</td>
<td>15.1</td>
<td>0.759</td>
</tr>
<tr>
<td>External causes of morbidity and mortality</td>
<td>16.2</td>
<td>15.7</td>
<td>0.785</td>
</tr>
<tr>
<td>Factors influencing health status and contact with health services</td>
<td>15.7</td>
<td>15.1</td>
<td>0.761</td>
</tr>
</tbody>
</table>

Source: Scottish Longitudinal Study
To avoid low numbers, cells with very low proportions are rounded up to one in all cases.

5.4 Discussion

We have shown that, among Scottish men who were in work in 1991, movement into unemployment by 2001 doubled mortality risk in the period 2001 to 2006. It is a reasonable proposition that unemployment may have a causal effect on mortality. Potential mechanisms include the loss or decline of income and movement into poverty that can occur as a result of unemployment has been shown to increase inactivity and social isolation and quicken deterioration of physical health (Bartley, Ferrie & Montgomery 2006). Furthermore, there is evidence that unemployment may lead to poor health and mortality via deterioration in mental health resulting from feelings of social exclusion and a contraction of social networks and friendships (Bartley, Ferrie & Montgomery 2006) as well as marriage dissolution (Lampard 1994). Unemployment has also been linked to increased consumption of alcohol.
and tobacco, poorer diet and lower levels of exercise (Dooley, Fielding & Levi 1996). These linkages present plausible pathways between unemployment, health and subsequent mortality.

Table 5.6 Mortality risks of unemployment relative to employment during follow-up period 2001-2006

<table>
<thead>
<tr>
<th>Sample</th>
<th>Cox hazard ratio (C.I. &lt; .05)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unadjusted models</td>
</tr>
<tr>
<td>Quintile 1</td>
<td>1.51** (0.21-10.91)</td>
</tr>
<tr>
<td>Quintile 2</td>
<td>3.00*** (1.09-8.27)</td>
</tr>
<tr>
<td>Quintile 3</td>
<td>3.36*** (1.84-6.13)</td>
</tr>
<tr>
<td>Quintile 4</td>
<td>1.69** (0.86-3.33)</td>
</tr>
<tr>
<td>Quintile 5</td>
<td>2.16*** (1.40-3.33)</td>
</tr>
<tr>
<td>Weighted average effect estimate</td>
<td>2.35</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Case matched restricted sample</th>
<th>Cox hazard ratio (C.I. &lt; .05)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unadjusted Model</td>
<td>Adjusted model</td>
</tr>
<tr>
<td>2.03*** (1.266-3.264)</td>
<td>2.10*** (1.303-3.380)</td>
</tr>
</tbody>
</table>

ns (not significant)* (p<.10) ** (p<.05) *** (p<.01)

Hazard ratios show the mortality effect of unemployment relative to being in work. Unadjusted models contain no additional adjustment and adjusted models include adjustment for all of the covariates that were used to predict the propensity score.

Source: Scottish Longitudinal Study

There are limitations with this study. One shortcoming is the relatively small number of deaths in our sample which prevented an examination of the different causes of death. Future studies using the SLS will benefit from more years of follow-up to death. Secondly, although we are able to discern the year of last employment for those who became unemployed by 2001, we are unable to identify more detailed employment histories for those in employment in 2001, and particularly the possibility that some individuals may have experienced periods of short or medium term unemployment prior to the 2001 census. A third issue is the problem of how to define spells of prior health status for those who remained in work in 2001. For those who were unemployed in 2001, we know the year in which the unemployment spell began and we collected information from hospital episodes data to identify health problems prior to this date; those occurring after this date were ignored and considered as outcomes of unemployment. However, for those who were in work in 2001 we collected hospital admissions information for the entire period between 1980 and 2001. This
may give the appearance of a comparatively higher prevalence of sickness in some members of the in work group as a result of a lengthier ‘at risk’ period. These latter two design shortcomings are unlikely to affect our conclusions however, as they would serve to narrow rather than overestimate our risk estimates. A final shortcoming arises from the possibility of unobserved confounding from either factors that were unmeasured at 1991 or effects that may have occurred prior to 1991. These could include health behaviour variables (Jusot et al. 2008) or factors occurring earlier in the life-course that were unmeasured in 1991 (Smith et al. 1997). However, recent evidence suggests that these factors have a smaller attenuation effect for the mortality risk of unemployment than contemporaneous characteristics (Lundin et al. 2010).

There are a number of design features of the study that add robustness to the findings. Firstly we were able to overcome the problems associated with sample selection that have been highlighted in various commentaries and critiques of previous studies (Morris and Cook 1991; Bartley and Ferrie 2010). Some studies, particularly those using natural experiments and instances of company downsizing (Steenland and Pinkerton 2008) have been critiqued due to the selective nature of the sample and the problem of generalising to a wider population (Morris and Cook 1991). On this point our study benefits from the use of nationally representative census data.

Secondly, given the difficulties associated with internal validity in observational data generally, and unemployment and health studies specifically, another important design feature is the ability to adjust for various sources of selection and confounding, particularly related to health. In addition to adjusting for common confounders such as social class and deprivation, we were able to account for prior hospital admissions and self-reported limiting long-term illness. In particular, we were able to adjust for serious mental and psychiatric health problems which have often been highlighted as important pathways to unemployment and subsequent mortality in previous studies (Lewis and Sloggett 1998). By demonstrating statistical balance across these covariates we argue that unemployment status in 2001 was hypothetically exchangeable between our groups of unemployed and those in work. Although Tables 5.4 and 5.5 indicates some remaining residual differences amongst the covariates, none of these are significant and the similarities between the adjusted and unadjusted estimates in Table 5.6 suggest these differences have a minimal systematic effect. As a result we reduced the possibility of over-estimating the effect of unemployment on mortality due to direct health-related and other selection factors.

The potential for selection was further reduced by our baseline sample which was observed in 1991 during a period of recession, when selection amongst the wider population is suggested to be less likely (Martikainen and Valkonen 1996; Martikainen and Valkonen 1996).
1998). Also, the sample is restricted to men who were in employment in 1991, thus excluding the unemployed men who may still have been selected for poor health and other negative characteristics. This is particularly important as unemployment exhibits a strong tendency for state dependence (Heckman and Borjas 1980) as previous unemployment is a strong risk factor for later spells of unemployment (Lundin et al. 2010).

The combination of these features resulted in a baseline sample which, in a labour market sense, contained the “most advantaged” individuals in terms of, on average, a lack of risk factors for unemployment and consequently an increased ability to retain employment, even for those who were actually unemployed in 2001. Interestingly, a recent study in Sweden, which did not uncover evidence of a remaining relative risk after adjustment for health and other confounders (Lundin et al. 2010) was similarly interpreted as the result of a selective sample (Bartley, Ferrie & Montgomery 2006). The most vulnerable members of the population may have been excluded and the true effect of unemployment and mortality may not have been identified. The fact that our study finds a remaining relative risk, despite a similar sample, lends weight to the suggestion that we have identified a causal component of the association between unemployment and mortality in the UK context that may not be present in other countries.

In conclusion, this study provides strong evidence that unemployment is causally related to an elevated all-cause mortality risk. It is the only study of unemployment and mortality in the UK that has adjusted for health selection using hospital admissions and self-reported health status prior to becoming unemployed. Moreover, it benefits from a causal analytical framework combined with a well-defined measure of employment status which allowed us to differentiate and exclude the economically inactive out of work and those out of work due to sickness at baseline. The current climate of economic slowdown and the concomitant rise in the rate of unemployment, which has topped 10% across the EU recently, reiterates the importance of these findings. Recent ecological evidence found that levels of government social spending on active labour market programmes is a key mediator in the relationship between rises in unemployment rates and increases in suicide mortality rates (Stuckler et al. 2009). This suggests that government intervention in the form of active labour market policies may be important in reducing the deleterious health effects of unemployment that have been demonstrated in this study.
Chapter 6

ECONOMIC INACTIVITY DUE TO SICKNESS AND MORTALITY: EVIDENCE FOR A CAUSAL RELATIONSHIP?

6.1 Introduction

The relationship between labour market participation and health and mortality has been examined in some detail during the previous thirty years. Predominantly, these studies focus on the AU, that is, individuals who are unemployed but actively seeking work. Amongst these members of the workforce, studies report consistently higher mortality rates than their employed counterparts (Moser, Fox & Jones 1984; Moser et al. 1986; Iversen et al. 1987; Moser et al. 1987; Martikainen 1990; Bartley, Ferrie & Montgomery 2006). However, there have been relatively few studies examining the relationship between other forms of non-work and subsequent mortality, particularly individuals IDS (Bartley, Ferrie & Montgomery 2006). Those studies that do exist are mostly from Scandinavia (Quaade et al. 2002; Åhs and Westerling 2006; Wallman et al. 2006; Karlsson et al. 2007) or are restricted to older age groups (Akinwale et al. 2010).

This is important given that the numbers of individuals in the UK claiming incapacity benefit has increased dramatically during the 1980s and 1990s (Beatty and Fothergill 2003; Faggio and Nickell 2003). This growth has occurred against the backdrop of aggregate improvements in the level of health throughout the general population. These apparently paradoxical patterns can be explained, it has been argued, by the phenomenon of hidden unemployment whereby ill and vulnerable members of the employed workforce flow onto incapacity benefit during periods of higher unemployment (Beatty, Fothergill & MacMillan 2000). As both claimant count measures and measures using the international labour
organisation definition do not account for economically inactive numbers in their definition, these individuals are not included in official unemployment statistics. As a result many of the most vulnerable recently unemployed members of the workforce may not have been considered in studies of the AU. This form of selection bias in many studies is likely to underestimate the true effect of the unemployment experience in studies of mortality (Martikainen and Valkonen 1999).

The most likely explanation for this lack of research is the difficulties associated with health selection (poor health causing worklessness) (Clemens, Boyle & Popham 2009). Separating the effects of health selection from the potentially causal effects of labour market disadvantage is often difficult due to the lack of mortality linked longitudinal health and social data (Bartley and Ferrie 2001). Thus, despite the consistent finding that unemployment is related to mortality, relatively few studies have suggested that the link is a causal one (Martikainen and Valkonen 1996; Martikainen, Maki & Jäntti 2007; Steenland and Pinkerton 2008; Lundin and Hemmingsson 2009; Lundin et al. 2010). Amongst the IDS, the problem of health selection is particularly acute as this group is likely to be selective of former employees in manual labour industries who may be suffering work-related poorer health (Bartley and Owen 1996; Beatty and Fothergill 1996; Fieldhouse and Hollywood 1999; Nickell 2004). As a result isolating the independent effect on mortality of economic inactivity due to sickness from pre-existing health problems is very difficult without adequate adjustment for health prior to becoming inactive in the labour market. The present study utilises a unique data set which links socio-economic and hospital admissions data to subsequent mortality to address the lack of research into the mortality effects of economic inactivity due to sickness in the UK.

6.2 Method

6.2.1 Data and sample

The data was extracted from the Scottish Longitudinal Study (SLS) (Boyle et al. 2009). The SLS consists of 1991 and 2001 Scottish national census records for a 5.3% sample of the Scottish population. Individuals are linked between censuses with additional linkages made to vital events registry data, including death registration, and hospital admission records from the Scottish Morbidity Records (SMR) from 1980 onwards. Data for this study was extracted from the SLS according to the sample selection criteria detailed in Figure 6.1. The economically inactive or unemployed in 1991 and those untraced to 2001 were excluded and the sample was also restricted to males aged 35-54 to avoid problems with education related labour market inactivity. Women were excluded from the analysis because of the complex
Chapter 6: The causal effect of economic inactivity due to sickness for mortality

nature of their participation in the labour market which was beyond the scope of this paper. A very small number of cases were missing information for any of the baseline covariates and these were excluded from the sample. Our remaining baseline sample contained 19,165 individuals. To mimic the terminology of a randomised control trial the terms treatment and control will hereafter refer to the permanently sick and active in work categories within the census respectively.

6.2.2 Propensity Score matching

Given the high possibility for health-related and other confounding the analysis was conducted using a matching based causal effects framework which mimics randomisation in observational treatment and control studies. The first step is to estimate (via logistic regression) a probability of residing in the treatment group which is calculated as a function of a group of confounding variables. This probability is an estimate of the underlying risk of treatment assignment for each case within the sample expressed as a probability. Thus, every case in the sample, irrespective of the group they are actually observed in, is assigned a risk of treatment. The propensity score value for each case in the treatment and control groups is then used to match similar cases (and exclude where appropriate those that cannot be matched) between the two groups. Thus, with the incomparable cases removed, the distribution of these confounding factors (in terms of the means or proportions) is similar across treatment and control groups. This demonstrates that, in terms of the observed confounders, treatment assignment is equal between the two groups (Rosenbaum and Rubin 1983; Rosenbaum and Rubin 1985). Causal relative risks can then be estimated from the new samples using Cox proportional hazards modelling.

6.2.3 Study Design

The treatment group in this study was defined as those men who moved from employment in 1991 to inactively IDS in 2001 with the control group defined as those men who remained in employment in 1991 and 2001. The variables used to estimate the propensity score included socio-demographic and health risk factors identified at baseline (1991 census) and from psychiatric and hospital admissions from 1980 until the year that each individual was last employed or until 2001 for those who remained in employment. Details about the study design and variable coding is given in Figure 6.1.

Using the values of the propensity score, we matched one treatment case to the closest 10 control observations. In addition we also stipulated that the propensity score values of all matched pairs were within an interval (caliper) of 0.005. Those treated cases without a comparable case control were excluded from the resulting sample along with the unused
control cases. To evaluate the effectiveness of the matching we used t-tests to compare the prevalence of the variables used to estimate the propensity score by treatment and control groups before matching and compared these to the corresponding values in the matched sample. To determine the sensitivity of the results we also compared the findings with those from another approach in which we estimated separate effects within quintiles of the propensity score across the full unrestricted sample (Rosenbaum and Rubin 1984). The weighted average of these effects was then calculated and the two approaches compared.

Figure 6.1 Sample selection criteria, study design and details of the covariates used

6.2.4 Mortality Analysis

Mortality follow-up commenced from the 2001 census day (29th April 2001) to the end of 2006. Individuals who left the study during that period due to migration were censored at migration date. Hazard ratios associated with treatment were obtained from Cox proportional
hazards models. Given the susceptibility of the results to confounding a number of models were fitted to both the full sample as well as the various propensity score based samples (propensity score quintiles and case matched samples) in order to compare the estimates. When modelling the full sample we compared age adjusted estimates with those from fully adjusted models. When modelling the propensity score based samples, we first estimated effects without adjustment for the variables used to predict the propensity score and then compared these to effects estimated from models that were fully adjusted. This enabled us to correct for the effect of any residual confounding.

To investigate the potential for any remaining health-related selection specifically, we determined the stability of the results according to the theory of the ‘wearing off of selection’. Accordingly, we estimated mortality effects from the case matched sample in each of five lagged mortality follow-up periods beginning one, two, three, four and five years subsequent to baseline and compared the estimates. Previous evidence suggests that a pattern of higher estimates in models with a smaller lag period that decline with increasing lag would indicate residual health selection effects (Clemens, Boyle & Popham 2009).

6.3 Results

Table 6.1 describes the sample in terms of treatment and control numbers and the distribution of deaths throughout the two groups.

<table>
<thead>
<tr>
<th>Employment status in 2001</th>
<th>Deaths</th>
<th>Died</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>In work (Row %)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16,608 (97.40)</td>
<td>444</td>
<td>17,052</td>
<td></td>
</tr>
<tr>
<td>Inactive Sick (Row %)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,813 (85.80)</td>
<td>300</td>
<td>2,113</td>
<td></td>
</tr>
<tr>
<td>Total (Row %)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18,421 (96.12)</td>
<td>744</td>
<td>19,165</td>
<td></td>
</tr>
</tbody>
</table>

Source: Scottish Longitudinal Study

Additional sample description is provided in Tables 6.2 and 6.3 which compares the prevalence of the covariates used to predict the propensity score for treatment. The treatment group had significantly higher rates of hospital admissions for all disease types except neoplasms (cancers), digestive diseases, skin and tissue related illnesses and additional factors affecting contact with health services. Rates of hospital admissions for mental and behavioural illness, endocrine and nutritional diseases and circulatory problems were
particularly high in the treatment group compared to the in work controls. Amongst the socio-economic variables the treatment group were more likely to live in social housing, had lower educational attainment, and had higher rates of limiting long term illness. They were also less likely to be married and to occupy the higher social classes and more likely to live in deprived areas. Age adjusted cox regression estimates for the full sample were 4.14, p < 0.01, 95% CI 3.54-4.84 and adjusted estimates were 3.41, p < 0.01, 95% CI 2.88-4.03. The age adjusted model shows a strong and highly significant mortality risk associated with

Table 6.2 Socio-economic covariate comparisons (means and proportions) between treatment and in work groups across the full sample

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Inactive sick</th>
<th>In work</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (Mean)</td>
<td>46.84</td>
<td>42.81***</td>
</tr>
<tr>
<td>Housing Tenure (1991) %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Owner Occupied</td>
<td>54.52</td>
<td>76.61***</td>
</tr>
<tr>
<td>Privately Rented</td>
<td>3.69</td>
<td>4.97**</td>
</tr>
<tr>
<td>Social Housing</td>
<td>41.51</td>
<td>18.18***</td>
</tr>
<tr>
<td>Communal Establishment</td>
<td>0.28</td>
<td>0.24</td>
</tr>
<tr>
<td>Educational Attainment (1991) %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>90.91</td>
<td>75.22***</td>
</tr>
<tr>
<td>Other Higher Qualifications (non-degree)</td>
<td>4.59</td>
<td>10.62***</td>
</tr>
<tr>
<td>First Degree and Higher Degree</td>
<td>2.04</td>
<td>12.30*</td>
</tr>
<tr>
<td>Not Stated</td>
<td>2.46</td>
<td>1.87</td>
</tr>
<tr>
<td>Marital status (1991) %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married (first marriage)</td>
<td>72.79</td>
<td>79.62***</td>
</tr>
<tr>
<td>Single</td>
<td>10.03</td>
<td>7.49***</td>
</tr>
<tr>
<td>Remarried</td>
<td>10.18</td>
<td>8.12***</td>
</tr>
<tr>
<td>Divorced</td>
<td>6.39</td>
<td>4.22***</td>
</tr>
<tr>
<td>Widowed</td>
<td>0.62</td>
<td>0.55</td>
</tr>
<tr>
<td>Deprivation Quintiles (1991) %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Least Deprived Quintile</td>
<td>11.74</td>
<td>26.86***</td>
</tr>
<tr>
<td>2nd</td>
<td>20.59</td>
<td>29.86***</td>
</tr>
<tr>
<td>3rd</td>
<td>23.19</td>
<td>20.46***</td>
</tr>
<tr>
<td>4th</td>
<td>22.91</td>
<td>14.49***</td>
</tr>
<tr>
<td>Most Deprived Quintile</td>
<td>21.58</td>
<td>8.34***</td>
</tr>
<tr>
<td>Social Class (1991) %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professional Occupations</td>
<td>2.08</td>
<td>8.19***</td>
</tr>
<tr>
<td>Managerial and Technical Occupations</td>
<td>16.90</td>
<td>32.83***</td>
</tr>
<tr>
<td>Skilled Non-manual Occupations</td>
<td>8.61</td>
<td>9.13</td>
</tr>
<tr>
<td>Skilled Manual Occupations</td>
<td>43.30</td>
<td>33.13***</td>
</tr>
<tr>
<td>Partly Skilled Occupations</td>
<td>18.98</td>
<td>12.74***</td>
</tr>
<tr>
<td>Unskilled Occupations</td>
<td>7.62</td>
<td>3.02***</td>
</tr>
<tr>
<td>Armed Forces</td>
<td>0.14</td>
<td>0.66***</td>
</tr>
<tr>
<td>No Job in Last 10 years</td>
<td>2.37</td>
<td>0.31***</td>
</tr>
</tbody>
</table>

Source: Scottish Longitudinal Study
treatment relative to the controls which was attenuated slightly (18%) in the fully adjusted models (adjusted for all covariates). This indicates evidence of health selection and other confounding.

Figure 6.2 illustrates the distribution of the propensity score by treatment and control groups and shows the level of overlap and comparability of the two groups. Intuitively, the majority of the in work group had a low propensity score for treatment compared to the treatment group whose distribution was more even. The graph thus illustrates the significantly different composition of the group of treated compared to the controls. The effectiveness of the case matching approach can be seen in Table 6.4 and Table 6.5 which, when compared to Table 6.2 and Table 6.3, illustrates balance in the means and proportions of the covariates used to predict the propensity score in this restricted case matched sample.

The modelling results from the propensity score based samples are presented in Table 6.6. In all of the models the adjusted and age adjusted estimates did not differ significantly. The quintile estimates show that the mortality risk of treatment was significantly raised amongst those individuals in quintile one of the propensity score. This effect was reduced in the remaining quintiles. Weighted averages of these effects suggested a very strong mortality risk (4.53 and 4.47). In the case matched sample where incomparable cases were removed, the effect estimates were around 31% more conservative (3.11 CI 2.48-3.90 and 3.18 CI 2.53-3.98). The results of the sensitivity analysis show that the magnitude of the effect point estimates are stable with increasingly lagged follow-up periods and remains of a similar magnitude to the estimates obtained from the full-follow up analysis. This suggests that the findings are robust to the pattern of selection wear-off reported by Clemens, Boyle and Popham (2009).
Chapter 6: The causal effect of economic inactivity due to sickness for mortality

Table 6.3 Health covariate comparisons (means and proportions) between treatment and in work groups across the full sample

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Inactive sick (%)</th>
<th>In work (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-reported limiting long-term illness (1991)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>87.13</td>
<td>97.91***</td>
</tr>
<tr>
<td>Yes</td>
<td>12.87</td>
<td>2.09***</td>
</tr>
<tr>
<td>Hospital Admissions (1980 - 2001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infectious and parasitic diseases</td>
<td>3.88</td>
<td>3.14*</td>
</tr>
<tr>
<td>Neoplasms</td>
<td>7.38</td>
<td>7.25</td>
</tr>
<tr>
<td>Diseases of the blood and blood-forming organs disorders of the immune mechanism</td>
<td>1.37</td>
<td>1.00**</td>
</tr>
<tr>
<td>Endocrine, nutritional and metabolic diseases</td>
<td>5.96</td>
<td>2.90***</td>
</tr>
<tr>
<td>Mental and behavioural</td>
<td>8.28</td>
<td>2.55***</td>
</tr>
<tr>
<td>Diseases of the nervous system</td>
<td>5.35</td>
<td>2.48***</td>
</tr>
<tr>
<td>Diseases of the eye and adnexa</td>
<td>2.70</td>
<td>2.09*</td>
</tr>
<tr>
<td>Diseases of the ear and mastoid process</td>
<td>2.37</td>
<td>1.61**</td>
</tr>
<tr>
<td>Diseases of the circulatory system</td>
<td>26.50</td>
<td>12.97***</td>
</tr>
<tr>
<td>Diseases of the respiratory system</td>
<td>8.33</td>
<td>7.07**</td>
</tr>
<tr>
<td>Diseases of the digestive system</td>
<td>21.58</td>
<td>20.28</td>
</tr>
<tr>
<td>Diseases of the skin and subcutaneous tissue</td>
<td>6.48</td>
<td>7.18</td>
</tr>
<tr>
<td>Diseases of the musculoskeletal system and connective tissue</td>
<td>15.05</td>
<td>10.53**</td>
</tr>
<tr>
<td>Diseases of the genitourinary system</td>
<td>11.36</td>
<td>9.66**</td>
</tr>
<tr>
<td>Congenital malformations, deformations and chromosomal abnormalities</td>
<td>1.09</td>
<td>1.00*</td>
</tr>
<tr>
<td>Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified</td>
<td>23.28</td>
<td>15.60***</td>
</tr>
<tr>
<td>Injury, poisoning and certain other consequences of external causes</td>
<td>15.76</td>
<td>13.58***</td>
</tr>
<tr>
<td>External causes of morbidity and mortality</td>
<td>16.33</td>
<td>13.89***</td>
</tr>
<tr>
<td>Factors influencing health status and contact with health services</td>
<td>19.21</td>
<td>20.38</td>
</tr>
</tbody>
</table>

Source: Scottish Longitudinal Study

6.4 Discussion

6.4.1 Principal findings

It has long been established that labour market disadvantage is related to poorer health outcomes with much of this work focusing on the AU (Bartley, Ferrie & Montgomery 2006). However, work examining the health outcomes (particularly mortality) of individuals who are inactive in the labour market due to sickness in the UK is scarce despite growing numbers in this group during the 80s and 90s. This study has demonstrated that permanent sickness in the census is associated with a trebling of the risk of mortality relative to
individuals who are in work in a representative sample of Scottish men. This mortality effect is significantly higher than has been reported in recent studies of the AU.

**Figure 6.2 Distribution of the propensity score by treatment and in work groups in 2001**

![Distribution of the propensity score by treatment and in work groups in 2001](image)

Source: Scottish Longitudinal Study

### 6.4.2 Strengths and limitations

The study has some notable strengths that ensure robustness in the findings, particularly in terms of mitigating the effects of confounding and selection. Firstly, unlike many other studies it benefits from an initially representative sample (the SLS) and a large sample size which allows for subgroup analysis in conjunction with the propensity score approach. Secondly, we were able to account for prior health through the use of hospital admissions records which allowed us to control directly for the timing of these events prior to leaving the workforce. This is an important feature as it addresses the problem of health selection whilst avoiding the possibility of over-adjustment for non-antecedent health events that may lie on the causal pathway between labour market status and subsequent mortality. It is, as far as we are aware, the first study that has been able to do this. Thirdly, our baseline sample was restricted to individuals in employment in 1991 and excluded individuals who were either economically inactive or unemployed. The period around 1991 was characterised by economic recession and high unemployment rates which means that our baseline sample was
likely to be selectively advantaged in terms of labour market endowment and resilience to job loss and less likely to be health selected for economic inactivity in 2001 than the general population in 1991. Finally, our estimates benefit from a causal effects approach (propensity score matching) to identify the relative mortality risks. This approach explicitly addresses the issues of group non-comparability and confounding. Furthermore, we determined the sensitivity of the estimates to residual health selection using a commonly used mortality wear-off approach.

The study has limitations. Firstly, we are missing socio-economic information between the 1991 and 2001 censuses. As a result we are unable to account for changes in individual circumstances during this period, in particular the possibility of multiple changes in employment status. As a result, some individuals who were employed in both 1991 and 2001 may well have experienced periods of unemployment or inactivity during this ten year gap. However, the effect of this may reduce the estimate of the mortality effect of inactivity rather than overestimate the true effect. A second limitation is the use of hospital admission records for the in work controls. Although we are able to censor these records at the point at which the IDS were last in employment, comparable censoring for the in work is not possible and instead occurs on the day of the 2001 census. As a result, our design may over-represent the level of ill health amongst individuals in work due to the lack of a censoring point for their hospital records. However, this is unlikely to affect the substantive findings of the study as again the result would be to under-estimate the difference in mortality risk. Finally, partly as a result of the relatively small sample size and partly due to the short duration of follow-up, the number of deaths was relatively small and did not allow an examination of cause specific mortality. This would have been useful in terms of understanding the pathway or mechanism by which economic inactivity links to mortality. Future studies may well tackle these questions when more recent mortality data is subsequently linked to the SLS. Finally there remains the possibility of unobserved confounding. There is evidence for example that health behaviour variables often occur prior to as well as after employment status changes (Jusot et al. 2008; Roelfs et al. 2011). Furthermore, factors occurring earlier in the life-course that were unmeasured in 1991 may also confound the relationship (Smith et al. 1997).
Table 6.4 Socio-economic covariate comparisons (means and proportions) between treatment and in work groups across the propensity score case matched (nearest neighbour) restricted sample

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Inactive sick</th>
<th>In work</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (Mean)</td>
<td>46.84</td>
<td>46.81</td>
</tr>
<tr>
<td><strong>Housing Tenure (1991) %</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Owner Occupied</td>
<td>54.81</td>
<td>54.48</td>
</tr>
<tr>
<td>Privately Rented</td>
<td>3.72</td>
<td>3.66</td>
</tr>
<tr>
<td>Social Housing</td>
<td>41.18</td>
<td>41.74</td>
</tr>
<tr>
<td>Communal Establishment</td>
<td>0.29</td>
<td>0.13</td>
</tr>
<tr>
<td><strong>Educational Attainment (1991) %</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>90.90</td>
<td>92.01</td>
</tr>
<tr>
<td>Other Higher Qualifications (non-degree)</td>
<td>4.62</td>
<td>4.05</td>
</tr>
<tr>
<td>First Degree and Higher Degree</td>
<td>2.05</td>
<td>1.76</td>
</tr>
<tr>
<td>Not Stated</td>
<td>2.43</td>
<td>2.17</td>
</tr>
<tr>
<td><strong>Marital status (1991) %</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married (first marriage)</td>
<td>73.02</td>
<td>72.65</td>
</tr>
<tr>
<td>Single</td>
<td>9.96</td>
<td>9.67</td>
</tr>
<tr>
<td>Remarried</td>
<td>10.11</td>
<td>10.89</td>
</tr>
<tr>
<td>Divorced</td>
<td>6.29</td>
<td>6.16</td>
</tr>
<tr>
<td>Widowed</td>
<td>0.62</td>
<td>0.63</td>
</tr>
<tr>
<td><strong>Deprivation Quintiles (1991) %</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Least Deprived Quintile</td>
<td>11.77</td>
<td>10.73</td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>20.73</td>
<td>20.53</td>
</tr>
<tr>
<td>3&lt;sup&gt;rd&lt;/sup&gt;</td>
<td>23.31</td>
<td>24.12</td>
</tr>
<tr>
<td>4&lt;sup&gt;th&lt;/sup&gt;</td>
<td>23.02</td>
<td>23.38</td>
</tr>
<tr>
<td>Most Deprived Quintile</td>
<td>21.16</td>
<td>21.24</td>
</tr>
<tr>
<td><strong>Social Class (1991) %</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professional Occupations</td>
<td>2.10</td>
<td>1.79</td>
</tr>
<tr>
<td>Managerial and Technical Occupations</td>
<td>16.97</td>
<td>15.69</td>
</tr>
<tr>
<td>Skilled Non-manual Occupations</td>
<td>8.63</td>
<td>8.35</td>
</tr>
<tr>
<td>Skilled Manual Occupations</td>
<td>43.52</td>
<td>45.21</td>
</tr>
<tr>
<td>Partly Skilled Occupations</td>
<td>18.92</td>
<td>19.40</td>
</tr>
<tr>
<td>Unskilled Occupations</td>
<td>7.63</td>
<td>7.78</td>
</tr>
<tr>
<td>Armed Forces</td>
<td>0.14</td>
<td>0.14</td>
</tr>
<tr>
<td>No Job in Last 10 years</td>
<td>2.10</td>
<td>1.65</td>
</tr>
</tbody>
</table>

Source: Scottish Longitudinal Study
Table 6.5 Health covariate comparisons (means and proportions) between treatment and in work groups across the propensity score case matched (nearest neighbour) restricted sample

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Inactive sick</th>
<th>In work</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-reported limiting long-term illness (1991) %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>87.61</td>
<td>88.73</td>
</tr>
<tr>
<td>Yes</td>
<td>12.39</td>
<td>11.27</td>
</tr>
<tr>
<td>Hospital Admissions %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infectious and parasitic diseases</td>
<td>3.86</td>
<td>3.97</td>
</tr>
<tr>
<td>Neoplasms</td>
<td>7.39</td>
<td>7.14</td>
</tr>
<tr>
<td>Diseases of the blood and blood-forming organs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>disorders of the immune mechanism</td>
<td>1.38</td>
<td>1.49</td>
</tr>
<tr>
<td>Endocrine, nutritional and metabolic diseases</td>
<td>5.96</td>
<td>6.59</td>
</tr>
<tr>
<td>Mental and behavioural</td>
<td>7.87</td>
<td>8.09</td>
</tr>
<tr>
<td>Diseases of the nervous system</td>
<td>5.24</td>
<td>5.98</td>
</tr>
<tr>
<td>Diseases of the eye and adnexa</td>
<td>2.72</td>
<td>2.44</td>
</tr>
<tr>
<td>Diseases of the ear and mastoid process</td>
<td>2.34</td>
<td>2.18</td>
</tr>
<tr>
<td>Diseases of the circulatory system</td>
<td>26.22</td>
<td>26.72</td>
</tr>
<tr>
<td>Diseases of the respiratory system</td>
<td>8.34</td>
<td>8.99</td>
</tr>
<tr>
<td>Diseases of the digestive system</td>
<td>21.64</td>
<td>21.71</td>
</tr>
<tr>
<td>Diseases of the skin and subcutaneous tissue</td>
<td>6.44</td>
<td>6.87</td>
</tr>
<tr>
<td>Diseases of the musculoskeletal system and connective tissue</td>
<td>15.01</td>
<td>14.96</td>
</tr>
<tr>
<td>Diseases of the genitourinary system</td>
<td>11.44</td>
<td>11.47</td>
</tr>
<tr>
<td>Congenital malformations, deformations and chromosomal abnormalities</td>
<td>1.10</td>
<td>1.14</td>
</tr>
<tr>
<td>Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified</td>
<td>23.02</td>
<td>22.96</td>
</tr>
<tr>
<td>Injury, poisoning and certain other consequences of external causes</td>
<td>15.68</td>
<td>16.29</td>
</tr>
<tr>
<td>External causes of morbidity and mortality</td>
<td>16.25</td>
<td>16.86</td>
</tr>
<tr>
<td>Factors influencing health status and contact with health services</td>
<td>19.35</td>
<td>19.45</td>
</tr>
</tbody>
</table>

Source: Scottish Longitudinal Study

6.4.3 Interpretation

Given that the study accounts for health events that occurred prior to the unemployment event the findings are less likely to be mediated by higher levels of underlying poor health amongst the IDS compared to the active employed. Though we cannot conclude that we have removed all of the residual confounding by health our sensitivity analysis does appear to show little evidence of a large systematic health selection effect. As a result it is difficult to attribute the findings to confounding by health and we argue that our findings provide evidence of a causal effect.
**Table 6.6 Cox Hazard ratio estimates of treatment (compared to in work) and confidence intervals for propensity score matched samples and sensitivity analysis (wear-off selection) for follow-up period 2001-2006**

<table>
<thead>
<tr>
<th>Sub-classification on the propensity score (full sample)</th>
<th>Hazard ratio (C.I. &lt; .05)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unadjusted models</td>
<td>Adjusted models</td>
<td></td>
</tr>
<tr>
<td>Quintile 1</td>
<td>16.92*** (7.95-35.98)</td>
<td>15.60*** (6.24-38.96)</td>
<td></td>
</tr>
<tr>
<td>Quintile 2</td>
<td>5.74*** (2.84-11.62)</td>
<td>6.93*** (3.32-14.45)</td>
<td></td>
</tr>
<tr>
<td>Quintile 3</td>
<td>4.06*** (2.45-6.72)</td>
<td>4.44*** (2.61-7.53)</td>
<td></td>
</tr>
<tr>
<td>Quintile 4</td>
<td>3.76*** (2.78-5.10)</td>
<td>3.59*** (2.63-4.90)</td>
<td></td>
</tr>
<tr>
<td>Quintile 5</td>
<td>3.03*** (2.45-3.75)</td>
<td>2.86*** (2.29-3.58)</td>
<td></td>
</tr>
<tr>
<td>Weighted average effect estimate</td>
<td>4.53</td>
<td>4.47</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Case matched sample</th>
<th>Hazard ratio (C.I. &lt; .05)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unadjusted models</td>
<td>Adjusted model</td>
<td></td>
</tr>
<tr>
<td>1 Year</td>
<td>3.11*** (2.48-3.90)</td>
<td>3.18*** (2.53-3.98)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Selection wear-off sensitivity analysis (on the case matched sample)</th>
<th>Hazard ratio (C.I. &lt; .05)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Year</td>
<td>2.84*** (2.24-3.60)</td>
<td>2.90*** (2.28-3.68)</td>
<td></td>
</tr>
<tr>
<td>2 Years</td>
<td>2.95*** (2.25-3.86)</td>
<td>3.03*** (2.31-3.97)</td>
<td></td>
</tr>
<tr>
<td>3 Years</td>
<td>2.95*** (2.17-4.02)</td>
<td>3.05*** (2.24-4.17)</td>
<td></td>
</tr>
<tr>
<td>4 Years</td>
<td>2.93*** (2.01-4.27)</td>
<td>3.05*** (2.08-4.46)</td>
<td></td>
</tr>
<tr>
<td>5 Years</td>
<td>3.18*** (1.91-5.28)</td>
<td>3.31*** (1.98-5.53)</td>
<td></td>
</tr>
</tbody>
</table>

Source: Scottish Longitudinal Study

This permits two interpretations of our findings. The first of these is that previous studies of the AU may have underestimated the true effect of labour market disadvantage on mortality. Evidence suggests that ignoring the economically inactive excludes the most vulnerable members of the labour market (Martikainen and Valkonen 1999). We might argue that, without adequate adjustment for health selection, risk estimates of inactivity may be inflated by health selection so that the true underlying health effects of unemployment may be consistent between these two forms of labour market marginalisation. Evidence from previous studies would suggest that the health adjusted mortality risk of active unemployment ranges from zero to double relative to those in active employment (Clemens, Popham & Boyle 2011). As a result, this might suggest that comparisons using the AU risk missing the true effect of labour market disadvantage though there remains the possibility of residual confounding.

A second interpretation is that the experience of economic inactivity is independently worse for health than the experience of AU. A possible pathway in the relationship between labour market disadvantage and health is via deterioration in mental health which can often arise from the increase in feelings of social exclusion and a contraction of work related social
networks and friendships that accompany unemployment (Warr and Jackson 1987; Bartley, Ferrie & Montgomery 2006). Furthermore, it has been suggested that the working routine plays an important role in providing an individual with a reason to start and progress though each day which is lost during spells of unemployment or inactivity (Warr and Bryan 1987). It is quite plausible that these effects in particular may be more acute in the group of IDS specifically as evidence suggests that this group contains proportionally higher numbers of previously stably employed individuals who are more used to working (Beatty and Fothergill 2003). It is also possible that amongst the AU the process of active job seeking may replace some of the daily routine associated with employment, something which is less likely to occur amongst the economically inactive who are not actively seeking work.

6.5 Conclusions

This study has addressed the lack of research into the independent health effects of economic inactivity due to sickness. The findings, which are robust to both health selection and other confounding, suggest that the risk of mortality is higher in this group than has been reported in other studies which examine the effect of active unemployment. The findings lend weight to the argument that labour market disadvantage generally poses independently harmful consequences for health but also suggests that the magnitude of the effects may have been.
Chapter 7

DISCUSSION AND CONCLUSIONS

7.1 Introduction

The overarching objective of this thesis has been to examine and test for a causal relationship between worklessness and mortality amongst male individuals in Scotland using approaches and data that have been relatively underused in the literature to date. The project has examined three research questions which, collectively, have developed an innovative methodological framework to test for a causal relationship between forms of worklessness and mortality. These questions were examined in the form of three stand-alone journal articles (chapters four to six) which were written to be read in isolation but, when combined, offer a number of relevant and interesting implications for some of the wider debates and literature outlined in chapters two and three. The purpose of this final chapter is to collate the findings and to relate them back to these key debates and to evaluate the thesis in relation to the key aims that were discussed and laid out in chapters two and three.

The structure of the chapter is as follows. Firstly, the key contributions to the worklessness and mortality debate are highlighted. Secondly, a discussion of the limitations of each of the studies is provided, focusing primarily on the limitations of the data and issues of residual confounding. In the third and final section, the wider implications of the findings are examined in terms of the wider implications and applications of the methodological approach used in this study, and wider debates around welfare policy, active labour market policies and employment policies. It also includes a critique of the centrality of paid work in much of these discussions. The final section, by way of conclusion, summarises the main findings, outlines areas of future research and finishes with some final closing remarks of the thesis as a whole.
7.2 Key contributions to the worklessness and mortality debate

This thesis has made contributions to two key areas that were identified in chapter two; the selection and causation debate and the lack of research examining the relationship between IDS and mortality. In the following section these contributions are teased out further and in greater detail than was possible in the empirical chapters.

7.2.1 Selection versus causation

From the literature in chapter two, two explanations for the pervasive relationship between forms of worklessness and subsequent mortality were identified. On the one hand, the theory of direct health selection argues that it is driven by higher rates of pre-existing poor health amongst the workless that causes unemployment and increases the risk of mortality. Related to this, the theory of indirect selection argues that greater exposure to the social determinants of health in turn leads to the accumulation of poorer health, exit from paid employment and subsequently higher risk of mortality. Both of these explanations make it difficult to conclude that the relationship between worklessness and health is causal (Martikainen and Valkonen 1996). On the other hand, the theory of causality asserts that various aspects of the unemployment experience contribute towards poor health independently of pre-existing characteristics and circumstances and, therefore, provide a plausible causal link between worklessness and mortality (Bartley 1996). This theory highlights two components of the workless experience in particular. Firstly, the psychosocial distress associated with the stigma of being out of work, the removal of working routines and the decline in social interactions which may all lead to poorer health and health behaviours. The second component emphasises the role of financial disadvantage which often accompanies worklessness and has been linked to various health outcomes.

One of the most recent additions to the published evidence in this area, and one of great significance for this thesis, has suggested that the persistent association of unemployment with mortality may, in part, be due more to the higher prevalence of ill health and other socio-economic risk factors amongst the unemployed rather than a causal effect per se (Lundin et al. 2010). These findings are indicative of what remains a strong debate throughout the literature that was reviewed in chapter two (Roelfs et al. 2011). Broadly, whereas Lundin et al. (2010) conclude that much of the crude association between unemployment and mortality is related to confounding by other factors, the results from this thesis contribute different findings to the literature and point to markedly different conclusions. For the remaining sections, the nuances underpinning the strength of the findings in chapter’s four to six are explored in more detail.
Chapter 7: Discussion and Conclusions

7.2.1.1 Wearing off selection

In the UK, studies of the relationship between worklessness and mortality have been dominated by the Office for National Statistics Longitudinal Study due to its ability to link many years of mortality data to socio-economic information for a large population representative sample of individuals. The development of this technique originally reflected the inability of analyses to take account for health differences between the various forms of labour market disadvantage which was complicating interpretations of the observed association between different forms of worklessness and health. Originally, these studies noted that the mortality rates of the AU were actually lower when compared to the population as a whole (Moser, Fox & Jones 1984; Moser et al. 1986; Moser et al. 1987). In this work, the reference group to which the unemployed were compared included not just those in work, but also the economically inactive permanently sick and other inactive groups. This finding was interpreted as a healthy worker effect as this group were, by definition, sufficiently fit and healthy to be looking for work (Bartley and Ferrie 2010). In order to remove these health disparities, a wear-off period was inserted prior to follow-up to allow the ill members of the comparator group to die or recover which, theoretically, allowed the investigation of the relationship independently of health status (Bartley and Ferrie 2010). Thus, the use and development of selection wear-off periods in this way reflected the observation that the mortality rates of the unemployed were too low when compared to this reference group and would rise (relatively) as ill members of the comparator group died or recovered (Moser, Goldblatt & Fox 1987; Goldblatt, Fox & Leon 1991).

This suggestion, that the AU are selected for good health relative to the remaining population, has now largely been forgotten as the focus of many studies has been to differentiate between various forms of labour market disadvantage. However, whilst many studies have compared these different groups (Moser, Fox & Jones 1984; Moser et al. 1986; Moser et al. 1987), none have explicitly tested for a health selection wear-off pattern. Using many years of mortality data in combination with the limiting long term illness measure that became available in the 1991 census, the study in chapter four is the first in the UK to explicitly test the effectiveness of ignoring years of mortality data to remove these health differences. The findings illustrated in figures 4.1-4.8 do not reveal the wear-off pattern amongst the AU and instead show a consistent and raised mortality rate (relative to the in work) during the respective follow-up period. The findings therefore make an important contribution to the literature as they cast doubt on the effectiveness of the wear-off approach, particularly for studies interested in the association between AU and mortality, in terms of adjusting for un-measured health differences between the AU and in work. The findings imply that such an approach may be unnecessary and that health selection of the AU is not
Chapter 7: Discussion and Conclusions

responsible for their raised mortality. Findings from this study guided the choice of research design in chapter five which did not incorporate mortality wear-off periods and therefore allowed for analysis of seven years of mortality data rather than the two years that would have remained had the wear-off technique been used.

However, the findings in chapter four did confirm a strong wear-off pattern amongst the group of economically inactive permanently sick, which confirms earlier findings from the literature (Moser, Fox & Jones 1984; Moser et al. 1986; Moser et al. 1987). The odds ratios presented in figures 4.1-4.4 show a decline in risk for mortality that begins to level off at the five year wear-off period lending weight to the suggestion that a wear-off pattern does indeed exist for this group. This conclusion is not, however, supported by evidence from the sensitivity analysis in table 6.6 in which the study design from chapter four is replicated for the propensity score matched sample of IDS and in work. This analysis suggests that when adopting a causal methods framework and with good adjustment for prior health from hospital admissions data the strong selection effects for the inactive IDS can be mitigated without the use of wear-off periods. The central contribution of this project, with regard to wearing off selection, is to suggest that, if there are significant differences in the health of the AU and the in work, the use of mortality wear-off periods can be used to account for them. However, if the differences are negligible, then the observed mortality rate differentials can be interpreted as being free from confounding by health. For the IDS, the observed wear-off pattern in chapter four can be mitigated through the use of research designs that explicitly address problems of confounding and have access to prior health information. Given the significant amounts of lost data and the fact that wear-off periods continue to be used in more recent research (Akinwale et al. 2010; Goldring et al. 2011), these findings are an important addition to the wider literature.

7.2.1.2 Exchangeability of the active unemployed and permanently sick

Throughout the literature that was reviewed in chapter two, there was strong emphasis on the problems associated with confounding by health and other factors (termed direct and indirect selection) which often precluded a causal interpretation of the risk of active unemployment for mortality (Moser et al. 1986; Moser, Goldblatt & Fox 1987; Moser et al. 1987; Lundin et al. 2010). In the causal effects terminology outlined in chapter three, this presents a problem of non-exchangeability whereby AU individuals cannot be hypothetically substituted for corresponding individuals who were in work because of underlying differences in key confounding characteristics. In contrast to many of the studies outlined in chapter two, the propensity score approach allows for the direct examination of the degree of this non-exchangeability by calculating an estimated propensity score or probability of unemployment for both the actually unemployed and the in work. Comparing the distribution
of this score between workless and in work groups (Figures 5.2 and 6.2) gave valuable insights as to the degree to which the two groups were similar in the key confounding characteristics, including health.

Figure 5.2 illustrates a high degree of similarity in the distribution of the propensity scores of the AU and in work. This observation is very much in keeping with the argument that was put forward in chapter four which suggested that confounding of the AU of the sort that may be mitigated by the use of wear-off periods was overstated. Figure 5.2 thus illustrates that, rather than being what Oakes (2006) might describe as structurally confounded, the comparisons between the AU and the in work are actually comparisons between two very similar groups in terms of the potentially confounding characteristics identified from the literature in chapter two and observed in the census in 1991. This is intuitive since these individuals are actively seeking work and thus are likely to be similar in background health on average to individuals who are in work except for their status as unemployed. In Figure 4.3, this is perhaps illustrated by the lack of a wear-off shape for the various mortality risk estimates of the AU which thus remain constant irrespective of the length of wear-off. Table 5.3 illustrates the degree of similarity with respect to health conditions prior to unemployment in the AU and in work. Although episodes of infectious and parasitic diseases, neoplasms and mental and behavioural illness along with limiting long term illness showed significant differences they were extremely small in magnitude, and after matching, were attenuated and no longer statistically significant. This gives further explanation for the lack of a selection pattern in Figure 4.3.

The observations in Figure 6.2 which showed the distribution of the propensity scores for the permanently sick and the in work also shows the differing composition of this group when compared to the unemployed in Figure 5.2. It is clear that the permanently sick are a much less comparable group to the in work in terms of the characteristics that determine movement into ill health related detachment from the labour market. The pattern of health selection that is shown in Figure 4.3 is intuitive as it shows that the permanently sick are more likely to contain higher numbers of individuals with characteristics that might cause this selection wear-off shape. This is reflected in terms of the health characteristics in Table 6.3 where a much greater number of the different hospital episodes are significantly different between the permanently sick and the in work than was evident in the corresponding Table in chapter five (Table 5.3) for the unemployed. Furthermore, the magnitude of these differences is also greater. Given this strong possibility for confounding, a sensitivity analysis, presented in

---

In figure 4.5 where estimates are adjusted additionally for limiting long term illness it may appear that the selection pattern is removed because of the flat line for the permanently sick across wear-off periods. However, this is slightly misleading and reflects a census procedure in 1991 to hardcode all permanently sick as having a limiting long term illness irrespective of whether one was actually reported.
Table 6.6, was conducted based on the selection wear-off approach. Interestingly, and in contrast to Figure 4.4 which did not adjust for limiting long term illness or hospital episodes prior to unemployment, this analysis showed little evidence of a decline in the magnitude of the risks with increasing length of wear-off. This would suggest that, similarly to the AU, the use of wear-off periods for the permanently sick is also unnecessary when there has been adjustment for health prior to job loss.

These findings and their presentation within the framework of exchangeability of treatment (unemployment and IDS) and control (in work) make important contributions to the literature that was reviewed in chapter two. Though many of the studies, particularly those such as the Lundin et al. (2010) study, are able to adjust for a great number of characteristics, the studies in chapters five and six are the first to explicitly demonstrate the extent to which these characteristics are similar or different across groups of individuals who become unemployed or inactive having previously been in employment.

7.2.1.3 Statistical over-adjustment

Equally as important as the notion of exchangeability and comparability is the role of over-adjustment. Over-adjustment occurs when variables that are erroneously considered to be confounders or common causes of exposure and outcome are actually intermediary variables on the causal pathway between the exposure and outcome of interest (Schisterman, Cole & Platt 2009). It was noted that controlling or adjusting for these variables was likely to introduce a bias towards the null hypothesis rather than remove bias associated with confounding (Schisterman, Cole & Platt 2009). The notion of confounding, or selection as it is termed in the worklessness and mortality literature, is central to the debate around the effect of worklessness for mortality. However, a little discussed component of this is the potential for obscuring potentially harmful effects through spurious adjustment. Although the potential for over-adjustment was not explicitly tested within each of the studies, comparing the effect estimates from chapters five and six with those obtained in chapter four allows such an examination. Table 7.1 collates these effect estimates to aid comparison.

The differing magnitudes of the effects which differed by around 40% for both the unemployed and the permanently sick is likely to be the result of over-adjustment in chapter four. This was remedied by the specific analytical design of chapters five and six which reflected the temporal and longitudinal sequence of the various socio-economic confounding effects. Thus, rather than adjusting for characteristics observed concurrently with employment status, the adjustment for housing tenure, education, marital status, deprivation and social class was implemented based on their observation in 1991 when the sample of in work individuals was selected. The reasoning for this was to avoid over-adjustment for the
characteristics that may have occurred after a change in employment status which may themselves have been outcomes rather than determinants of becoming workless.

Table 7.1 Comparison of effect estimates from chapters four to six illustrating the potential over-adjustment

<table>
<thead>
<tr>
<th>Source</th>
<th>Effect estimate</th>
<th>Unemployment</th>
<th>Permanent sickness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter Four*</td>
<td>1.6***</td>
<td>2.25***</td>
<td></td>
</tr>
<tr>
<td>Chapter Five†</td>
<td>2.03***</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>Chapter Six‡</td>
<td>n/a</td>
<td>3.11***</td>
<td></td>
</tr>
</tbody>
</table>

*** (p<.01), *The effects presented from chapter four are those displayed in Figure 4.4 and are those gained from the five year wear-off period models, †effects from chapters five and six are from the analyses of unadjusted case matched samples. Source: Scottish Longitudinal Study

Whilst it may be argued that this may in fact constitute under-adjustment and, therefore, confounding on the basis that some of these effects may have caused worklessness, I argue that for all of the characteristics except marital status, this is unlikely to be true. For example, given the age of the sample (35-54), it is unlikely that education level may change from 1991 to the extent that it may jointly ‘cause’ (i.e. become a common cause) both worklessness and mortality. Similarly, housing tenure change (e.g. owner occupier to renter) or movement to a more deprived area is unlikely to precede a change in employment status, though both may well be an outcome of it. However, marital status is less easily dismissed as solely an intermediary variable. For example, there is some evidence to suggest that becoming workless is associated with marital breakdown (Lester 1996) and that marriage is also associated with longevity (Gardner and Oswald 2004). It is therefore possible that ignoring marital status change after 1991 may risk confounding through the pathway of changing marital status causing subsequent worklessness. However, it is also feasible to imagine that changes to employment status may itself cause marital breakdown and therefore could be considered an intermediary in the pathway between worklessness and mortality. In chapters five and six, both scenarios are possible but the lack of detailed longitudinal information between censuses on marriage precludes the investigation of such effects and there remains the possibility of residual confounding.

The issue of residual confounding by less serious health events may also explain the difference in the observed mortality effect for unemployment between this study and the study from Lundin et al. (2010). Adjustment for less serious types of health events in the Lundin et al. (2010) study is achieved through the use of sickness absence records prior to employment. They found that this variable held the single biggest attenuating effect on the
crude mortality hazard ratio of the unemployed, reducing it by around 40%. However, it could be argued that adjustment for sickness absence in this way also represents over-adjustment despite the fact that these events occur before unemployment. The root of this argument is that the prospect, or threat, of job loss in the future may hold negative effects for health (Ferrie 2001; Ferrie et al. 2001) and psychiatric well-being (Ferrie et al. 1995). Though this does not actually constitute unemployment in the sense of being out of work, the effect is likely to be the same in terms of health consequences. As a result, the anticipation of losing one’s job may induce poor health and raise the risk of sickness absence. It could be argued that adjusting for this sickness absence therefore runs the risk of adjusting for an unemployment related health effect. This might explain the strong attenuating effect observed by Lundin et al. (2010).

The recent meta-analysis by Roelfs et al. (2011) provides an interesting comparison with table 7.1, particularly for the AU. For example, they report that the average unadjusted risk estimate across all of the studies was around 2.08. This is very similar in magnitude to the estimate obtained in chapter five. Furthermore, the corresponding average risk estimate for adjusted estimates in Roelfs et al (2011) was 1.63 which is strikingly similar again to the estimate obtained from the adjusted models in chapter four. This suggests that many of the studies in the meta-analysis suffer from the effect of over-adjustment (Nylen, Voss & Floderus 2001; Gerdtham and Johannesson 2003; Gardner and Oswald 2004). The discussion in chapter three, in particular, reinforced the importance of the timing of confounding effects and consequently the difference between confounding and intermediary effects in studies of causal effects. However, it is clear, both from this meta-analysis and the individual studies included in it, that this is a relatively under-theorised area within the worklessness and mortality literature. The studies contained in this thesis and the particular attention paid to their design within a causal framework has, therefore, made a significant contribution to the literature.

7.2.2 Mortality and health related labour market inactivity

From chapter two, it was clear that the relationship between active unemployment and subsequent health and mortality has received a great deal of attention, particularly in terms of the competing explanations of selection and causality. However, it was noted that other forms of labour market disadvantage, such as economic inactivity due to sickness, had received far less attention. The handful of studies that had examined this had been conducted outside the UK (Quaade et al. 2002; Åhs and Westerling 2006; Wallman et al. 2006; Karlsson et al. 2007). The study presented in chapter six therefore makes an important contribution to our knowledge about the health effects of a previously under-researched
group. This gap, particularly within the UK context, is important, given the fact that, since the 1980s, the numbers of IDS sick in the UK have been rising considerably (Webster 2000). As Bambra (2011b) notes, this form of labour market disadvantage has traditionally been ‘framed as a social policy and labour market problem rather than a public health or medical concern’ (pg. 134). The study is especially timely given that the focus of welfare state reform in the UK has now switched to these individuals at a time of continuing entrenchment of discourses that emphasise the dependency culture amongst individuals on incapacity and other related benefits. The implications of the findings for wider debates around the provision and entitlement to welfare benefits and the discourses that underpin these will be discussed in subsequent sections of the thesis but, for the remainder of this section, the findings are discussed in the context of the literature examined in chapter two and in relation to the findings observed in chapters four and five. In particular, it will focus on the contribution of the findings to the theory of hidden unemployment before discussing the possible interpretations of the finding of higher excess mortality amongst the IDS.

7.2.2.1 Hidden unemployment

A key debate that was discussed in chapter two concerns the problem of health related worklessness. Of particular concern was the degree to which rises in the numbers of individuals claiming incapacity benefit reflected underlying poor health or whether it was a function of the labour market and the organisation and provision of welfare support. The theory of hidden unemployment and the related concept of hidden sickness provide a theoretical basis for the latter and the findings from the studies in both chapters five and six of this thesis appear to lend support for such an explanation. In particular, examination of tables 5.2 to 5.5 and 6.2 to 6.5 together with the propensity score approach, provide insights into the composition of the various groups in terms of socio-economic and health characteristics prior to changes in labour market status. The data in these tables therefore hold important wider implications for the current agenda of welfare reform which will be discussed in section 7.3.3. Furthermore, they hold implications for the conclusions of the study in chapter six.

The theory of hidden unemployment suggests that economic change associated with post-Fordism accelerated the decline in demand for manual unskilled labour and resulted in large numbers of individuals becoming unemployed during the 1980s. Generally, these individuals occupied lower social classes and were more likely to be suffering occupation related illnesses and general poor health. This was apparent in chapter six when examining table 6.3, which illustrates the health characteristics of the IDS compared to the in work and shows significantly higher proportions of almost all serious health outcomes and self-reported health when compared to the in work control group in the full sample. However, despite this,
Chapter 7: Discussion and Conclusions

the propensity score matching analysis was still able to produce a sample of working individuals to which the sick could be compared. This was achieved by removing the healthiest and most advantaged individuals from the sample to ensure overlap and comparability of the two groups in terms of their propensity for health related worklessness in the 2001 census. In doing so, these health and socio-economic differences were removed in tables 6.4 and 6.5. A useful way to illustrate this further is to compare the composition of the case matched control group in chapter five (tables 5.4 and 5.5) with that in chapter six (tables 6.4 and 6.5). For example, in chapter five, the proportions of the control group with a LLTI or who were hospitalised with a disease of the nervous system, the circulatory system or the musculo-skeletal system was around 3%, 3%, 11% and 9% respectively. The corresponding figures in the control group for the IDS analysis in chapter six were all considerably higher at around 11%, 6%, 27% and 15% respectively. Similar patterns are apparent in terms of many of the socio-economic characteristics. This indicates firstly that the IDS are considerably worse off than the AU in terms of health and socio-economic characteristics but, secondly, that the pool of in work in each case contained at least enough individuals of sufficient similarity to their respective treatment group to allow direct comparisons to be made. In other words, the case matched in work sample in chapter six contained individuals suffering equally high levels of both socio-economic disadvantage and poor health as their counterparts in the IDS group. Because the observed sickness that is apparent amongst the in work group is evidently not sufficient to cause job loss, it remains hidden when making comparisons between the individuals in this group and those who are categorised as out of work due to sickness.

These findings therefore lend weight to a crucial element of the hidden unemployment theory by uncovering strong evidence of ‘hidden sickness’. It is crucial because, as a result of this poor health which remains ‘hidden’ during employment, newly workless individuals qualify for sickness benefits and therefore flow onto incapacity benefit rather than job-seeking unemployment. This is because they are looking for work in an inevitably tight labour market, with fewer opportunities for re-employment (Bartley and Plewis 2002; Beatty and Fothergill 2005). This also explains the initially paradoxical observation of rising incapacity benefit claimants at a time of overall downward trends in mortality throughout the population (Beatty, Fothergill & MacMillan 2000). Crucially, in slack labour markets and despite poor health, ill individuals are able to maintain employment but are amongst the first to be released when the labour market tightens. Furthermore, in a tight labour market, these individuals are often at the back of the jobs queue as a result of their ill health with little prospect of re-employment (Beatty, Fothergill & MacMillan 2000). The findings in this thesis provide support for the theories of hidden unemployment and hidden sickness and
stress the importance of the labour market rather than health as the initial determinant of health related worklessness. The wider policy implications of this will be discussed in section 7.3.3. The remainder of this section will focus on the implications of the main finding in chapter six in the context of the preceding discussion.

7.2.2.2 Interpreting the relationship between mortality and health related worklessness

Given what was discussed in the previous section, any interpretations of the effect that was observed in chapter six are considerably more complex than for the findings in chapter five. This is true in terms of both the greater risk of confounding and in terms of the complex relationship between the labour market and hidden unemployment & sickness. As highlighted above, it is known that many of the individuals who make up the IDS population are more likely to be of lower social class, lower educational attainment and to come from particular employment sectors. Most crucially, and intuitively, they are almost certainly suffering an illness or poor health which might have contributed to their exit from paid employment. This pattern of confounding was evidenced by Figure 4.1-4.8 in chapter four whereby, in contrast to any of the remaining economic activity groups, the IDS showed considerably higher risk of mortality initially, which declined with increasing duration of follow-up. It is also evidenced in Figure 6.2, which showed significantly reduced overlap in the distribution of the propensity scores suggesting greater differences in terms of confounding effects than was in evidence in comparisons between the AU and the in work. Separating mortality effects that are due to this form of labour market disadvantage from those that occur as a result of these confounding effects is considerably more difficult.

Despite this, similar to the findings for active unemployment, the design of the studies allows greater confidence in the assertion that the observed effect is causal through a combination of sample selection, a causal effects methodological framework and control for a number of confounding variables including health and other factors that, crucially, are censored at the point of exit from the labour market. Arguably, the application of these techniques is more relevant in chapter six than it is in chapter five due to the marked differences that were apparent in comparisons between the IDS and the in work. Thus, and in contrast to the analysis of the AU, the matching approach removed a significant number of the ‘different’ cases from both the IDS group and the active employed group, ensuring that comparisons between the IDS and the in work are as un-confounded as possible with the data at hand. Collectively, the findings from chapters four and six suggest that the observation of a significantly greater risk of mortality amongst the IDS cannot be attributed solely to confounding effects.
On the basis of this conclusion, two theoretical interpretations can be offered. On the one hand, given the likelihood that many unemployed individuals may become hidden amongst the IDS, one interpretation is to suggest that estimates of the mortality effect of worklessness may be under-estimated in studies which are restricted to the AU. This idea is similar to that expressed by Martikainen and Valkonen (1999) who caution against the exclusion of economically inactive individuals when examining social class differentials in mortality. In their study, the authors go so far as to say that ‘social class differences in health based on the economically active population may give a partial view of the relevant causal factors and processes’ (Martikainen and Valkonen, 1999 pg. 903). From this perspective, it seems plausible that the effect estimate identified in chapter six may in fact represent a more accurate measure of the effect of worklessness for health than relying on effect measures derived from the group of AU alone. The findings make worthwhile contributions to the worklessness and mortality debate specifically but also to wider issues outlined by Martikainen and Valkonen (1999) who show that biases are introduced when ignoring the economically inactive in wider studies of health inequalities. In the context of the mortality effects of worklessness, comparing the effect estimates from chapters five and six suggests that the magnitude of this potential under-estimation would be considerable, raising the risk from a doubling to a trebling of the risk of mortality.

An alternative explanation is to suggest that the experience of IDS may be different to that experienced during AU and may hold greater risks for health and subsequent mortality. Although this theory has not been directly explored as it has for the AU, it is consistent with psychosocial explanations. These ideas suggest that much of the effect of worklessness that impacts on health occurs through the loss of Jahoda’s (1979) ‘latent functions’ of work which include factors such as daily time structures, shared experiences with colleagues and continued and constant activity which all help to provide self-esteem and feelings of respect from others. Thus, it is probable that the daily routine of worklessness associated with active unemployment is likely to be very different to that of inactivity due to sickness. It may also be the case that the higher levels of poor health may exacerbate the negative effects of unemployment. For example, mental health problems that precipitate an exit from the labour force into health related inactivity may be worsened by the lack of the ‘latent functions of work’ to a greater degree than the experience of active unemployment which may still provide a degree of routine and interaction with others. The implications of this explanation in terms of health inequalities may be crucial particularly as there is a great deal of evidence pointing to the unequal distribution of health related worklessness across socio-economic groups (Whitehead et al. 2009). In reality, aspects of both of these explanations are likely to be responsible for the findings observed in chapter six and the collective conclusion across
all of the studies is that worklessness of all forms is highly significant in terms of determining population level health and also in propagating inequalities in mortality.

### 7.2.3 Summary

The three studies combined present an overarching story of the limited role of confounding in the worklessness and mortality relationship which has rarely been explicitly developed in the literature previously. Using an innovative data source with linkages to other administrative data sources in combination with newly developing causal methods in social epidemiology, the studies have uncovered a number of findings which contribute to the wider debate around the causality of the association of worklessness with mortality. These contributions include a lack of support for the selection wear-off approach for the unemployed, a lack of evidence for large scale systematic confounding of the workless and active unemployment relationship and an important finding regarding the possibility for statistical over-adjustment which may have resulted in under-estimation of the effect of worklessness for mortality in previous studies. Collectively, these findings provide strong evidence that worklessness is causally related to mortality which runs counter to recent literature which has argued that the relationship may be mostly attributable to confounding.

The second key contribution of the thesis and one that is unique in the wider literature was to examine the independent effects of health related worklessness for mortality. The importance of this question was highlighted in chapter two in relation to the pattern of continued growth in numbers of individuals claiming incapacity benefit in the UK, as well as the increasing policy attention that has been paid to this group. Within current debates, the thesis made two key contributions. Firstly, the findings found support for the theories of hidden unemployment and hidden sickness which was discussed in chapter two and, secondly, they highlighted that the effects for mortality of this group appear to be worse than for active unemployment when applying the same methodology and adjusting for the same confounding effects. Two possible interpretations were offered. On the one hand this might indicate that the effect of health related worklessness is worse than the effect of active unemployment and on the other it might indicate that the true effect of active unemployment might have been underestimated in studies that restrict analysis solely to the AU. Both explanations represent important contributions to the wider literature.
7.3 Limitations of the Studies

7.3.1 Residual Confounding

The possibility for residual confounding remained a potential limitation in all of the empirical chapters. Although propensity score matching is a highly effective approach for achieving balance in terms of observed characteristics, it is unable to address the problems of unobserved confounding (Oakes and Johnson 2006). In the literature, there exists a sensitivity type approach which assesses the level of significance of effect estimates from propensity score matched samples against hypothetical levels of unobserved confounding. This approach essentially allows for an estimation of the level of unobserved confounding that would be necessary to render the estimated effect size statistically insignificant (Becker and Caliendo 2007). However, as of writing the technique has not been generalised to Cox proportional hazards and nor to caliper matching with replacement and is therefore not available for this study.

One source of residual confounding that may be important in these studies is differences in rates of poor health behaviours between the workless and the in work. In the recent meta-analysis by Roelfs et al. (2011) it was found that the average hazard ratios from studies that adjusted for health behaviours were reduced by around 24% compared to studies without adjustment for health behaviours. However, few of these studies were able to account for the timing of health behaviours in relation to employment status change and may run the risk of over-adjustment for health behaviours that are themselves caused by unemployment (Schisterman, Cole & Platt 2009). One study that was able to take account of this was the study by Lundin et al. (2010) who were able to adjust for smoking and alcohol use 32 years prior to baseline. Although this represents a significant lag, both measures had a modest attenuating effect on the mortality risk estimate of unemployment reducing it from 1.91 in the crude estimate to 1.83 and 1.85 respectively. This does suggest that the estimates in chapters five and six may suffer from unobserved confounding by health behaviours, though the magnitude of this unobserved effect is likely to be modest given the evidence presented by Lundin et al. (2010).

There is also the possibility that the workless may have experienced greater levels of social disadvantage, prior to becoming workless, throughout various stages of the life course relative to the in work. Much work has investigated the possibility of life course effects for health (Smith et al. 1997; Smith et al. 1998; Wadsworth, Montgomery & Bartley 1999) which opens the door for confounding when relying on measures of socio-economic position at one point in the life course. The recent study by Lundin et al. (2010) was able to account for these effects and measured factors such as crowded housing and childhood socio-
economic position 40 years prior to baseline. However, the effect of including these variables in their model was minimal in terms of attenuating the effect of unemployment for mortality. Thus, although these factors were predictive of mortality they did not seem to be related to later unemployment.

A final limitation in terms of the potential for residual confounding was the use of hospital admissions to adjust for health differences between the workless and in work in 2001 in chapters five and six. Though they afforded great flexibility with regard to the timing of health events relative to an individual becoming workless, they represent a serious health event and, as a result, may not capture all of the differences in less serious health problems. However, comparing the findings across chapters four, five and six, it is clear that confounding by health, in terms of the notion of a wearing off pattern, was markedly reduced within the study design framework that was used in both chapters five and six. This might suggest that the more serious health events are more likely to result in death than less serious ones which would indicate that the hospital admissions are likely to capture the majority of any potential health-related confounding effects.

### 7.3.2 Census and Mortality Data

All of the studies used data originating from the UK census. One of the limitations of such data (particularly in chapters five and six) is the lack of information during the inter-censal period. This meant the studies were unable to account for multiple transitions into and out of employment during the period between 1991 and 2001. It is, however, unlikely that this shortcoming would have much of a substantive impact on the findings as the main problem would be for individuals who were employed in 2001 but who had experienced significant periods of unemployment in the decade previous. If this were the case, the effect would be to reduce the magnitude of the observed effect and to underestimate the worklessness effect.

Another limitation in these studies was the numbers of deaths in the sample. In chapter four this restricted analysis to the 1991 census where it would have been useful to extend to 2001 in order to compare the potential change over time. In chapters five and six the small numbers of deaths prevented a cause specific mortality approach. This would have allowed greater understanding of the causal pathways between worklessness and health. In particular, examination of suicide deaths within the analytical framework of these studies would have been a useful and valuable addition.

The final limitation with regards to the use of census data in the studies was the decision to exclude women from the analyses in chapters five and six. As was noted in detail in section 3.6.2.2, the methodological focus of the thesis coupled with a lack of detail in the economic activity question in the census precluded an examination of the same questions amongst the
population of working age women. This represents a significant limitation of the research conducted in the thesis as it significantly restricts the generalisability of the findings. Future research should look to apply the methods and analytical techniques from the thesis to a study of the effect of worklessness for mortality amongst women.

7.4 Wider implications

The main overarching contribution of the thesis has been to test for a causal and independent relationship between both active unemployment and health related worklessness and subsequent mortality. It has achieved this through the application of a number of innovative methods with a unique dataset within the UK. These contributions have been discussed at length in relation to the immediate literature to which they have contributed. However, also in these chapters, a number of wider and broader debates were highlighted to which this thesis may offer important insights. These are outlined and discussed in the following section. It will begin with a discussion of the wider implications for public health research of the projects methodological innovations including both the application of propensity score matching and the use of linked administrative data. The second section will focus on the broader UK policy context and the implications of the findings in terms of welfare benefit payments, active labour market policies, and policies for those currently in employment before finishing with a broader critique of the central role of paid work in social policy and the problems this may hold in terms implementing appropriate polices to reduce the health effects of worklessness.

7.4.1 Data and methodological innovation

7.4.1.1 Accounting for structural confounding

The studies in chapters five and six were all conducted within a counterfactual methodological framework reflecting the extensive discussion in chapter three which examined the methodological debates around causal effects in epidemiology. The key tenet of the counterfactual framework is the importance that is placed on ensuring that the two groups under comparison are as comparable as possible with no cases in each group that cannot be matched with another in the corresponding group. In other words and to use the terminology of Oakes (2006), the comparisons are free from structural confounding and there is sufficient overlap in the propensity scores of both groups to enable these comparisons free from confounding and bias. By its nature, regression analysis obscures this problem in many research settings because any cases that do not compare to the reference group are extrapolated from the model itself rather than from the observed data at hand. Crucially, there are no diagnostic tools available in the general regression toolbox that
explicitly allows for the examination of structural confounding. The discussion in chapter three revealed substantial debate in the literature about the extent of this problem and drew on the example of research into the establishment of neighbourhood effects for health (Diez Roux 2004; Subramanian 2004; Chaix et al. 2006; Oakes 2004; Oakes 2006; Oakes and Johnson 2006; Messer, Oakes & Mason 2010). The analysis conducted in this thesis represents the first application of these ideas in the worklessness and mortality literature and in doing so has contributed a number of important findings to the literature in relation to the competing explanations of confounding and causality. As a result, this thesis endorses the use of techniques such as propensity score matching and argues that similar approaches could well make important contributions to wider research questions in observational social epidemiology.

More specifically, the analysis presents challenges to the various other methods that have been applied in previous studies of the relationship between worklessness and mortality that were discussed in chapter two. These include studies that rely on wear-off periods (Akinwale et al. 2010; Goldring et al. 2011), those that rely on traditional regression analysis (e.g. Lundin et al. 2010) or those that use a natural experiment type approach (Martikainen and Valkonen 1996; Steenland and Pinkerton 2008). In the case of natural experiments for example, there is likely to be problems in terms of the generalizability of the findings when the sample is restricted to individuals in one factory or restricted to periods of recession when unemployment rates are higher. The problems of data wastage allied to the questions over the effectiveness of wear-off period analyses were well rehearsed in chapter four but both problems were mitigated by the use of the propensity score approach. In all of these cases, it would provide a flexible and intuitive approach to understanding structural confounding and the extent to which it undermines the analysis in a way that other approaches do not.

However, a major limitation is that it does not provide safeguards against the possibility of unobserved or residual confounding and requires significant amounts of data on a variety of confounding effects in order to be effective. In this thesis, data of this nature was able to be provided by the Scottish Longitudinal Study which is relatively unique internationally and certainly so in the UK context.

7.4.1.2 Data

The unique nature of the Scottish Longitudinal Study, in terms of its linkages to substantial amounts of health data from hospital admissions records and linkages to death records from vital events registers, was crucial for the completion of the analyses in the thesis. Importantly, it allowed the studies to adjust for a large amount of potentially confounding effects in the propensity score matching approach. Crucially, the data also allowed for the
incorporation of timing in the study designs which ensured that these various confounding effects were censored at the point of exit from employment. This was a crucial component in terms of mitigating the possibility of over-adjustment which might spuriously obscure potentially important effects. The ability to incorporate all of this information in a sample of sufficient size to observe a mortality effect is entirely unique in the UK.

Studies such as those presented in this thesis reiterate the importance of continued investment and development of opportunities for data linkage, including, in particular, continued efforts to make as much of the routinely collected administrative data available for research purposes as possible. In this regard, Scotland leads the way, particularly in terms of health data, but still lags behind many of the Scandinavian countries. The research in this thesis illustrates the value and the power of such data.

7.4.2 Implications for theory and policy

7.4.2.1 Welfare benefit payments

One of the key policy approaches to reduce the negative effects of worklessness is through the payment of social security benefits. These can begin to alleviate some of the financial penalties associated with worklessness and reduce the risk of individuals falling into poverty (Bambra 2011b). One approach for achieving this is to implement what is known as a minimum income for healthy living (MIHL) which is calculated on the basis of meeting basic needs in terms of nutrition, physical activity, housing, psychosocial interactions, transport, medical care and hygiene (Morris et al. 2000). The initial recommendation from this analysis, which was calculated for single British men during 1999, was that such needs could be met on a minimum amount of £131.86 per week (Morris et al. 2000). This figure was considerably greater than contemporary job seekers allowance levels which were between £40.70 and £51.40 (Morris et al. 2000). While a basic level of subsistence living is possible at such financial levels such is the level of disparity that there exists genuine concern about the ability individuals to live healthily on these levels of financial support (Deeming 2005). The implementation of such a minimum payment system for both the unemployed and the IDS is likely to represent an important approach in reducing some of the potential health impacts of worklessness (Morris et al. 2010).

In addition to the payment of welfare benefits for the workless, recent evidence also appears to suggest that the manner in which these payments are provided may also determine the severity of the impact of worklessness for health. In particular, it appears that means tested welfare benefits are less protective than non-means tested payments. For example, evidence from the US appears to suggest that receipt of means tested welfare is more likely to result in depressive symptoms than those in receipt of non-means-tested support (Rodriguez,
Chapter 7: Discussion and Conclusions

Frongillo & Chandra 2001). A further study examining data from the UK, the US and Germany found similar results for self-perceived health and included adjustment for a range of factors that might confound the relationship. It found that despite significant differences in the degree of means testing across the three countries, the effects remained across all of the countries suggesting that in each case the provision of such benefits is not sufficient to reduce the negative effects of worklessness (Rodriguez 2001). These findings are also supported by more recent comparative public health research which draws on welfare state regime theory to make comparisons between countries with different systems of welfare support. This literature was examined in chapter two and suggested that countries with a greater degree of labour commodification (in terms of means testing and associated levels of stigma that are attached to these benefits) generally suffer worse population level health outcomes (Coburn 2004; Eikemo et al. 2008b). This would appear to suggest that policies intended to improve the health of the workless in terms of welfare benefit payments should look to reduce the degree of means testing and look to provide a greater level of financial support in order to meet the basic income level suggested by Morris et al (2000).

The payment of welfare state benefits to the AU in the UK has long been conditional on individuals demonstrating active job seeking behaviour since reforms implemented during the 1980s and 1990s (Bambra 2011b). However, as was outlined in chapter two, the UK has implemented marked changes to the system of benefit receipt for the IDS through the introduction of the Employment and Support Allowance (ESA) in 2008 to replace incapacity benefit. This welfare reform agenda arose out of the continued and longstanding trend of growth in the numbers of individuals claiming incapacity benefit in the UK which was explained, in chapter two, in terms of the theory of ‘hidden unemployment’ (Beatty, Fothergill & MacMillan 2000; Lindsay and Houston 2011). The findings in this thesis, as discussed in section 7.2.2.1, provide support for the theory of both hidden unemployment and sickness and, therefore, of the need for reform of the system of social security for the IDS more generally. Furthermore, in light of the discussion in section 7.2.2.1, the findings in chapter six might be interpreted as providing support for one of the key assumptions underpinning the implementation of the ESA. This assumption posits that there is a clear division within the population of health related workless; those who are able to work and those who are not and therefore those who may have become ‘dependent’ on the passive welfare provided by incapacity benefit and those for whom the degree of illness or disability is too severe for work.

Although the findings in this thesis, particularly those in chapter six, provide support for the need for policy intervention, the implementation of the ESA would appear unlikely to address the poor health and wellbeing of these individuals. In line with predominant
discourses of dependency, in effect the ESA legitimates a process of separating the deserving sick from the undeserving sick in terms of who is entitled to the previously universal benefit for all sick individuals (Bambra 2008; Bambra and Smith 2010). Of the ‘undeserving’, who are deemed “sick but able to work”, the transfer from incapacity benefit onto employment and support allowance will represent a considerable drop in income to levels even further below the minimum level suggested by Morris et al (2001). The potential result of this policy therefore, is to reduce the real incomes of some of the poorest and most vulnerable members of the population. Though there is likely to be a wide spectrum in terms of the severity of the different health conditions amongst this group, all individuals experiencing this transition will be suffering illness and disability to some degree and are likely to be particularly vulnerable to changes in financial circumstances.

It is difficult, therefore, to see the health effects of this form of labour market disadvantage and the individuals within it improving under the new policy regime. Moving onto lower paying benefits that come with the expectation of taking on work when it becomes available may result in many sick individuals being compelled to take on temporary and insecure employment or jobs that may not be suitable in terms of expectations or physical ability. The potentially negative health effects of such forms of employment have been widely noted in the literature (Winefield et al. 1991; Graetz 1993; Ferrie 2001; Kivimäki et al. 2003c; Virtanen et al. 2005a; Virtanen et al. 2005b; Virtanen et al. 2005c; Waddell and Burton 2006) with many of these types of employment opportunities a characteristic of the flexible labour markets of post-Fordist and service sector dominated economies that predominate in developed Western economies (Dicken 2003; Bartley, Ferrie & Montgomery 2006; Bambra 2011b). This recent reform agenda is the continuation of a trend of decreasing real levels of social security in the UK, a trend which looks set to continue under current coalition government policy which aims to restrict access to benefit payments to an even greater degree (Lindsay and Houston 2011). This trend may need to be broken and replaced with the MIHL approach suggested by Morris et al. (2000) and others if there is to be a real commitment towards reducing the health impacts of worklessness that have been identified in this thesis.

7.4.2.2 Active labour market policies
As was discussed in the previous section, levels of social security benefits in the UK have continued to decline reflecting the continued power of discourses around welfare dependency and the benefits of paid work for reducing poverty and social exclusion. As a result, in addition to social security payments for the workless, another policy approach that has developed since the 1980s is that of active labour market policies (ALMPs) (Bambra 2011b). The focus of these types of policies is to reduce the level of unemployment through
interventions in both the supply (individual) and demand (employers) sides of the labour market. These ALMPS often take the form of measures that aim to improve the chances of successful job search, or that improve the skillsets of workless individuals through training and education, for example. They may also encourage job growth through wage subsidies or public employment (Bambra 2011b). Thus, these policies differ compared to the provision of social security payments as they provide the policy means by which to actively begin to move individuals back into employment. The central tenet of these types of policy interventions from a public health perspective is that employment provides great benefits in terms of health and wellbeing relative to being out of work.

The degree to which the health of AU participants improves through participation in individual based ALMPs remains an active area of research and published literature to date returns mixed findings. The main pathway through which ALMPs are expected to improve health and wellbeing is through improvements to psychosocial functioning and psychological health (Coutts 2005). These ideas very much follow the theories developed by Jahoda (1979; 2002), Warr (1987) and Fryer (1987) that were discussed in chapter two. These explain the negative effects of unemployment through the lack of both the latent functions of work and the psychosocial experiences associated with the working activities. In line with this, some evidence suggests that ALMPs bring about a reduction in psychological distress and depression, increased self-reported well-being, higher levels of control or mastery, motivation and self-esteem and increased opportunities to meet and socialize with others. Furthermore, they may also help to lessen feelings of stigmatization that often accompanies unemployment. Crucially, it has also been suggested that not only do these policies help to improve the unemployment experience; they may also serve to improve the chances of finding suitable paid employment quickly so as to outweigh the financial and mental health penalties that are associated with unemployment (Coutts 2009).

The evidence is not, however, entirely unequivocal and, in particular, some studies suggest that the effects of participation in these programs does not last indefinitely and in some cases the effects have been seen to disappear within six months (Creed, Hicks & Machin 1998). Other studies suggest the positive effects can continue for up to two years (Vinokur et al. 2000). Furthermore, the degree to which ALMPS may hold positive effects for health appears to be dependent on individual characteristics prior to participation in ALMPS (Coutts 2009). In particular, evidence suggests that individuals with depression or a low sense of motivation, self-efficacy and mastery stand to gain the most whereas the benefits for individuals who are perhaps closer to the labour market in terms of occupational skills and experience may be less pronounced. However, for individuals who may be considered to have ‘adapted’ to the unemployment experience, the transition to an ALMP programme may
be associated with a net negative effect for health, particularly if new employment is characterized by insecurity and poor pay which may reverse the effects of this adaptation (Creed, Machin & Hicks 1999; Creed, Muller & Machin 2001; Westerlund, Theorell & Bergström 2001).

Though the development of ALMPs for the AU has a long history in the UK, their extension into policy discussions around the IDS is relatively recent beginning in the 1990s and continuing into the 2000s (Bambra 2011a). If the findings in chapter six are interpreted as the effect of IDS being worse for mortality than active unemployment, then policy responses that focus on return to work strategies for this group may be seen as crucial. Furthermore, chapter two noted that the strength of discourses that characterize the IDS as ‘dependent’ on state benefits, as well as notions of hidden unemployment and sickness, has resulted in an increase in the use of ALMP type policies in attempts to reduce the numbers of individuals in this labour market position. However, in contrast to the AU, the population of IDS contains a greater number of individuals with significant health problems and lower skillsets and qualifications. As a result, it is much less likely that the positive benefits of ALMP in relation to the AU will accrue to the same degree in interventions of a similar nature aimed at the IDS.

There is some evidence to suggest that these policies have been effective in terms of improving the employment of people with a disability or health problem with a number showing higher rates of employment amongst participants (Bambra, Whitehead & Hamilton 2005). However, studies reviewing this evidence note that in many cases it is difficult to determine whether these improvements could be attributable to the policies themselves or whether they reflect wider contextual and temporal factors. For example, many of the policies were implemented during a period when UK rates of employment were experiencing an upturn (Bambra 2011a). Similarly to the AU, many of the effects appeared to be dependent on a number of individual characteristics such as job-readiness and the type and the context of the local labour market (Clayton et al. 2011). Although little evidence exists describing the health effects of such policies for the IDS as it does for the AU, the reverse adaptation effect that was observed in the latter may be particularly important in terms of the health effects of ALMP policy interventions for the health related workless. For example, there is strong evidence that suggests that a large proportion of the IDS have been out of work for longer than five years (Bambra, Whitehead & Hamilton 2005). This might indicate a greater level of adaptation to the workless experience than might be observed in the AU and therefore might lead to greater negative health effects when beginning participation in ALMPs. Furthermore, Bambra (2011b) notes that many ALMPs in the UK do not take any account of sickness as a barrier to employment compared to corresponding policies in the
Nordic countries for example. Recognition of this in policy decisions might serve to improve the effectiveness of ALMPs in terms of returning individuals to work (Bambra 2011b) as well as, perhaps, reducing the independent health effects of health related worklessness that have been observed in this thesis.

Thus far the discussion has focused on supply side policies that focus on individual based approaches towards moving people into work. Since the 1980s, UK labour market policy has traditionally focused almost entirely on neo-liberal supply side policies and there has been little room for policies that recognize demand deficiencies in local labour markets (Lindsay and Houston 2011; Bambra 2011b). As discussed in chapter two, in the past 10 to 20 years, this has culminated in considerable policy focus on the growing numbers of health related workless. This includes the most recent round of welfare and labour market reforms such as firstly the ESA and latterly the universal working age benefit which includes, amongst other things, a compulsory work for benefit component which is due to be implemented across the UK by 2015 (Bambra 2011b). Lindsay and Houston (2011), highlight a number of assumptions that underpin these recent policy responses to the IB problem including the unemployability of individuals, the presence of a culture of dependency or a loss of job-seeking confidence amongst IB claimants. As the authors highlight, such accounts fail to recognize that, for example, rates of the old incapacity benefit were strongly correlated with local labour market conditions and disproportionately affected older industrial areas of Britain. These areas saw a collapse in demand for unskilled and manual labour following deindustrialisation which was not matched with suitable new employment opportunities.

The result is that spatial patterns of claimant numbers of the new benefit system correlate very strongly with the spatial patterns of incapacity benefit. This, coupled with very little evidence of a dependency culture amongst the IDS (Kemp and Davidson 2010), appears to provide support for a demand based component to labour market policy intervention. Though previous demand based interventions have been shown to have had a limited impact (Bambra 2011b), this may be due to the lack of engagement with regional labour market patterns that were outlined in chapter two. Thus, future strategies may look to incorporate supply-side policies to aid employability and develop skills of individuals in conjunction with area based initiatives to address the specific contextual demand based factors that are at play in these different regions (Lindsay and Houston 2011).

7.4.2.3 Employment policies

A consistent theme that has run throughout the thesis has been the fact that poor health is a significant risk factor for unemployment and worklessness. This was in evidence both from the literature that was discussed in section 2.5.1 as well as in the full unmatched sample comparisons (Table 5.3 & Table 6.3) in chapters five and six of the empirical findings. All of
these showed significantly higher rates of self-rated poor health and greater incidences of admissions to hospital for a range of conditions. A proportion of this sickness is likely to have arisen from exposure to health damaging dimensions of certain physical work environments. Bambra (2011b) notes three key sources of such effects; chemical hazards such as asbestos and lead, environmental factors such as noise, vibration and injuries and ergonomic hazards such as repetitive strain injury, physical load and shift work outside of normal business hours. These factors remain an important public health issue in their own right as well as being important in terms of increasing the risk of unemployment or, more likely, inactivity due to sickness or disability in the labour market.

Appropriate policies that reduce these health effects in the workplace are, therefore, an important component of the policy responses to the evidence that is presented in this thesis. According to Bambra (2011b), examples of such interventions include health and safety legislation which incorporate provision for measures to ensure that such legislation is adhered to. Other policies can be implemented to ensure reductions in the exposure to harmful chemicals and airborne particles including the use of respiratory equipment and the reduction of noise levels to acceptable levels (Bambra 2011b). Furthermore, measures to reduce the duration of repetitive tasks at work is an important part of reducing potentially serious musculoskeletal problems, which, from the analysis in chapter six, represented a key risk factor for health related worklessness.

In addition to this, there is strong evidence linking psychosocial aspects of the working environment to various health outcomes through the pathway of elevated levels of stress. Other psychosocial theories include the demand-control-support model, the effort-reward imbalance model and the organisational injustice model. All of these models emphasise the non-physical characteristics of work that include levels of responsibility, psychological demand and the degree or perceptions of equity between workers. Such theories are given added weight in tables 5.3 and 6.3 which show that hospitalisation for mental and behavioural illness and circulatory problems, which can all be related to stress, were all significantly higher in individuals who subsequently became unemployed or workless. At the micro scale, it has been proposed that such psycho-social problems can be mediated by, for example, job enrichment and enlargement, collective coping and decision making and the use of autonomous production groups in factories to widen skill sets and reduce the level and frequency of repetitive tasks (Karasek 1992).

7.4.2.4 A wider critique of paid work in social policy

Chapter two highlighted that the overarching aim of labour market policy in the neo-liberal era has been to encourage workless individuals back into paid work rather than addressing some of the aspects of the unemployment and worklessness experience that may be negative
for health and wellbeing. Much public health research in the area of worklessness and health has argued for similar policy recommendations that centre on paid work as the crucial pathway back to health and wellbeing (Coutts 2009; Stuckler et al. 2009; Stuckler, Basu & Mckee 2010; Stuckler et al. 2011). Whether this is through demand and more commonly supply ALMPs for example, or the reduction of passive social security payments to reduce potential ‘dependency’, the dominant discourse around labour markets, the economy and social policy interventions reflects unerring faith in the benefits of paid work for health and wellbeing. However, this faith overlooks the potentially harmful effects of employment in jobs towards the bottom of the labour market which can often be characterised by insecurity, poor conditions, repetitiveness and low pay (Bartley, Ferrie & Montgomery 2006; Waddell and Burton 2006; Lindsay and Houston 2011; Bambra 2011b). In these types of jobs, there is likely to be a significant health penalty compared to more favourable employment which might be particularly severe for sick or disabled individuals who have been moved from health related worklessness into these forms of employment. Thus, though the independent effect of worklessness that has been identified in this thesis may be mitigated through the transition to paid employment, this may not translate into universal net gains in health across all individuals, some of whom may be more vulnerable or susceptible to work related determinants of health.

More fundamentally, there is evidence, which was touched upon in section 7.4.2.2, that paid work or participation in temporary alternative employment may interrupt health protective adaptation to the workless experience. Westerlund et al. (2001), for example, found that levels of a particular hormone associated with helplessness were found to be higher compared with baseline measurements after participation in a program which was designed to ‘create supportive and empowering environments’ for the workless (pg. 407). This finding was in contrast to the hypotheses of the paper and previous evidence from self-reported data which suggested that the effects of participation in the scheme would reduce the feelings of helplessness associated with unemployment and worklessness (Westerlund, Theorell & Bergström 2001). Interestingly, the participants in the study were described as having been outside the labour market for an extended period of time (around 2.5 years) and were identified as eligible for the studied program on the basis of difficulties in terms of gaining employment. As a result, the findings were interpreted as evidence of adaptation to the workless lifestyle which was acting to lower levels of helplessness in these individuals. Crucially, as the authors note, this adaptation effect is unlikely to be uniform across workless individuals (Westerlund, Theorell & Bergström 2001). For example, although the ‘helplessness’ effect of participation was high amongst highly adapted or ‘weak’ individuals in terms of labour market potential, individuals with greater potential for employment in the
labour market may experience greater and more severe impacts of unemployment related helplessness (Westerlund, Theorell & Bergström 2001). From a public health policy perspective, this suggests that return to work based policy frameworks may not be beneficial across, particularly, individuals in the heterogeneous group of IDS who will comprise individuals who are ‘adapted’ to their labour market environment to greater or lesser degrees. In the current policy environment, particularly for the IDS described in section 7.4.2.2, there is a very real risk that ‘substituting dependency on state benefits for dependency on jobs at the bottom end of the labour market’ might therefore result in worsening health for these individuals (Waddell and Burton 2006).

The review in section 2.2.2 noted that the relationship between social policy and paid work can be traced back to the transition from Fordism to post-Fordism following the inflationary crises of the 1970s and the concurrent growth in neo-liberal economic policies. From a political economy perspective, this gradual transition resulted in marked changes to the mutual and reinforcing relationships between the state, the labour market and the economy. During the post-war settlement era, increasing the spending power of citizens through full employment policies and a wide reaching and comprehensive social security system was essential for Fordist capital accumulation which relied on domestic demand for produced goods. In the de-regulated and export orientated post-Fordist system, this demand is met from international markets reducing the need to maintain spending power of citizens through social security. Conversely, as markets de-regulated and capital became more footloose, the demand for flexible and competitive labour markets grew, necessitating state intervention to reduce inflation. This was achieved through the maintenance of a ‘reserve army’ of unemployed work-focused labour which is conditioned, through both neoliberal social policy and social security, to continually search for opportunities to re-enter paid employment. Thus, discourses of dependency develop alongside the central role of paid work to counter the potential for this reserve army of labour to subvert their role as providing competition in the labour market.

Radical critiques that counter the centrality of paid work in social policy discourses are plenty (Rifkin 1995; Gorz 1999; Cole 2007; Cole 2008). Drawing on critical social theory, these accounts strike at the heart of the notion of paid work as a central source of health and wellbeing and permit a worthy critique of the wider worklessness and mortality research that was reviewed in chapter two. Amongst these critiques, Cole (2008) suggests that, coinciding with the period of the 1980s, there was a growth in sociological literature that sought to defend the unemployed from stigmatic perceptions of a lack of work ethic. On one level, these accounts were clearly justified in their critique of the damaging effect of supply based economic policy, but on another, they indirectly served to legitimize a ‘moral discourse’
which underlined the centrality of paid work to unemployment policy thereby restricting ‘deeper critiques of the meaning and purpose of work’ (pg. 1). Though Cole (2008) restricts this critique to the sociological literature of the period during the 1980s, the same ideas can be pointed at much of the epidemiological literature examining the relationship between worklessness and health that was reviewed in chapter two. Although much of this literature stems from a social inclusion discourse (e.g. Jahoda 1979; Jahoda, Lazarsfeld & Zeisel 2002) it could be argued that they, perhaps inadvertently, serve to reproduce the very system of capitalist accumulation that is at the root of the worklessness and health relationship (Cole 2007).

Critical engagement with these powerful discourses should, therefore, become a central, if challenging, part of epidemiological research more widely but particularly for research into the effects of worklessness for health. Fundamentally, critical engagement with this underlying notion of paid work could help to reduce the degree of stigma attached to worklessness and begin to undermine accounts of a dependency culture or a lack of ‘work ethic’ (Kemp and Davidson 2010; Lindsay and Houston 2011). In combination, these ideas could start to reduce the psychosocial effects of worklessness that were noted in chapter two and may well be a component of the findings in this thesis. Equally, it might also begin to address the lack of public, and therefore political, support for unconditional social security payments (Dwyer 1998; Dwyer 2002). This would have the benefit of potentially removing barriers that are hindering the implementation of policy alternatives such as the MIHL which has yet to be implemented since its inception in 2000. The accounts provided by Cole (2007 & 2008) and others surrounding paid work in modern society are, necessarily, radical and progressive in nature. However, engagement with them by public health and epidemiology researchers perhaps represents an opportunity to make real progress in reducing health inequalities. As doubts remain about the ability of these disciplines to effect real change (Bambra et al., 2011; Mackenbach, 2010b) now is perhaps as good a time as ever.

7.4.3 Summary

This section has covered significant ground. It has examined the findings in the thesis in terms of wider implications for analytical methods in public health research and the potential policy implications of the three studies. In terms of methods, the particular contribution of the studies in terms of wider methodological issues such as confounding in statistical analyses of observational data was noted. The approach used in the thesis is likely to be of great benefit to other studies by explicitly addressing problems associated with structural confounding. It was noted that standard regression analysis tools do not explicitly provide for an examination of the degree of structural confounding and therefore do not illustrate the
degree to which model parameters are based on actual data points or are extrapolated to overcome the lack of a counterfactual observation. In chapters five and six this was illustrated by comparing the characteristics of both workless and in work groups and ensuring comparability by matching on the propensity score.

In the second section, the wider policy implications of the studies were examined in relation to the political economy explanations for labour market and social policy patterns in the UK over the past 70-80 years that was developed in chapter two. In terms of social security policies, a minimum income for healthy living approach was recommended in conjunction with a reduction in the degree of means testing of benefit entitlement. This approach would help to reduce the financial burden of worklessness and reduce the potentially harmful effects for health. A critique of the current round of welfare reform was also developed, including the ESA and newly announced universal credit system, arguing that such policies might actually worsen the health of workless individuals, particularly the IDS.

The discussion then examined the role of both supply and demand side active labour market policies in terms of their effects for the health of the workless. It was noted that in the former, though they may have some positive effects for health, these effects may not last beyond six months and very much depend on the nature of the specific implementation of the programme. It was also noted that their successful application to the IDS would likely depend on the degree to which the underlying health issues afflicting individuals were accounted for in the job search programme. Furthermore, it was noted that a greater commitment to incorporating demand side policies was needed. This represents an important component of successful activation interventions, particularly regarding the provision of suitable jobs for the newly job seeking IDS in conjunction with these individual based activation policies. In addition to these policies which focus explicitly on the workless, a number of other policies relating to the working environment were also examined including health & safety legislation and improvements to the psychosocial working environment. All of these might reduce the likelihood of work-related ill health which, in turn, would decrease the risk of health related exit from employment.

Finally, a more widespread critique of the role of paid work in social policy was developed to counter the predominantly work based focus of the preceding policy recommendations. This focused on the underlying issue that in many cases paid work might be as bad for health as being in work in certain types of jobs, particularly in individuals who are further from employment and who may have become adapted to their workless situation. Engagement with more radical and progressive sociology theory allowed this argument to extend into a wider and more fundamental critique of the centrality of paid work in social policy and public health research. It argued that these disciplines could, perhaps, do more to recognise
their own role in reproducing neo-liberal discourses of paid work and, in doing so, might begin to develop the tools with which to break the longstanding pattern of health inequalities that have persisted since the publication of the Black report in 1982.

7.5 Conclusions

7.5.1 Summary of main findings and key contributions

The overarching objective of this thesis has been to examine and test for a causal relationship between worklessness and health amongst male individuals in Scotland using approaches and data that have been relatively underused in the literature to date. The project has examined three research questions and, to conclude the thesis, the findings are summarised and the overall contribution of the thesis reiterated.

The first question examined the concept of ‘health related selection’ amongst the unemployed and economically inactive permanently sick in both the Scottish and England & Wales 1991 censuses. This question arose out of the observation that many studies of the relationship between worklessness and mortality employed the technique of selection wear-off periods when following up individuals. Wear-off periods involve ignoring deaths in a period (normally five years) following the commencement of follow-up in order that individuals ‘selected’ for poor health will either die or recover. The theory suggests that ignoring these individuals reduces the potential for selection and confounding. However, the findings in chapter four found little evidence of such a wear-off effect for the AU. On the contrary, the risk of mortality remained relatively constant irrespective of the length of wear-off period. However, for the economically inactive permanently sick, the mortality risk was found to decline with increasing length of wear-off period suggesting a wearing off effect amongst this group. The main conclusions from this study were that the use of wear-off periods for the AU in mortality follow-up studies was unnecessary.

The second research question, addressed in chapter four, tested for a causal relationship between male AU and mortality using linked data and a causal effects modelling framework. Instead of relying on wear-off periods, chapter three adopted a propensity score matching design. It adjusted for health differences between the AU and in work using hospital admissions records prior to unemployment and socio-economic differences using information from the census at baseline. Furthermore, rather than adjusting for confounding characteristics that were contemporaneous to the observation of labour market status in 2001, the risk estimates were adjusted for characteristics that were observed in 1991 when the sample was selected. This was an important feature of the design because of the potential for over-adjustment which, as outlined in chapter three, may result in a bias towards the null
hypothesis. In contrast to recent evidence, the findings from the study illustrated that the unemployed were at around double the risk of mortality compared to the in work a relationship which was interpreted as containing a significant causal element. Two broader implications were noted. Firstly, the potential for over-adjustment in studies that adjust for baseline confounding effects that are measured contemporaneously with employment status and, secondly, that selection and confounding of the AU when compared to the in work may be overstated.

In chapter five, these ideas were extended to the permanently sick group in the census (which was used to proxy incapacity related detachment from the labour market) and compared their risk of mortality to those individuals in work. The results showed around a trebling of the risk of mortality for the permanently sick. The potentially health-related confounding effects operating on this group that were identified in chapter four were examined with a sensitivity analysis using wear-off periods. This analysis showed that, in contrast to the pattern identified in chapter four, the risks remained mostly constant throughout the wear-off periods suggesting it unlikely that the findings could be attributed solely to health selection. Examination of the common support of the two groups in Figure 6.2 revealed the region of support to be smaller than that observed in Figure 5.2.

Collectively, the findings build towards the key contribution of the thesis as a whole which is to reassert the importance of labour market participation as a key determinant of individual male mortality in the UK. The studies represent a significant contribution in the UK context as they are the first that have been able to utilise health information to adjust for health differences between the various labour market groups prior to mortality follow up and the first to study the effects of inactivity due to sickness. More widely, the studies contribute contrasting findings to a growing international literature which has cast doubt on the potentially causal nature of the relationship between worklessness and mortality.

7.5.2 Future research

7.5.2.1 Cause specific mortality

As has been noted in each of the empirical studies and in section 7.3, one of the key limitations of the research is the lack of cause specific mortality analyses. Such an analysis, using the same analytical and methodological framework as the studies in this project, would enable a more detailed understanding of the potential pathways between worklessness, poor health and subsequent mortality. This analysis would be particularly beneficial for the study of the IDS as it would enable a comparison with the in work in terms of the types of illness that are experienced and the relationships between these disabling and long term illnesses
and subsequent mortality. The continued addition of death events to the SLS as they become available will allow such an analysis in the near future.

7.5.2.2 Youth unemployment and mortality
The current recession has been characterised by very high levels of youth unemployment and worklessness. The very latest figures released in April 2012 show that unemployment amongst 16-24 year olds was over 1 in five at around 22.2%. This should be significant cause for concern given evidence that, for example, unemployment seems to hold more serious consequences in young people in terms of psychological health and smoking than in older individuals (Reine, Novo & Hammarström 2004). The studies in this thesis were restricted to middle-aged men and did not consider younger ages due to problems with disentangling the labour market position of younger individuals from student related inactivity in the census. However, this area remains one of great concern for future research particularly in light of the persistence of low unemployment in the UK following the recent economic crisis.

7.5.2.3 Worklessness, employment and mortality at the bottom of the labour market
The policy implications of the findings of this thesis were discussed at length in section 7.4.2. One of the recurring themes throughout this discussion was the assumption that paid work is naturally beneficial for health and wellbeing. This assumption underpins most if not all of the policy interventions of the last 30 years or so in the UK which have focused on the ‘activation’ of individuals through supply side active labour market policies. Thus, implicit in these policies is the assertion that any work is better than no work. Though this has been a feature of policies for active unemployment, the development of similar policies for the IDS is more problematic and there is a need for research to keep pace with the changes to these policy interventions. The findings in this thesis have demonstrated robust evidence for a strong and causally negative effect of health related labour market inactivity for the risk of mortality amongst men, which, to an extent, adds weight to the need for policy interventions to continue to target this group. However, as discussed in section 7.4.2, the advent of the ESA, and, more recently, the labour market policies of the coalition government, have ushered in a period when increasing numbers of sick and disabled individuals will become dependent on poorer quality and less secure employment at the bottom of the labour market. While there is little doubt that secure, fulfilling and suitable employment would provide great benefits for health and wellbeing for many of these individuals, it is unlikely that these sorts of jobs will become available without government intervention to stimulate job growth in local, demand deficient, labour markets. Therefore, future research should look to focus on the health effect of employment in these lower quality manual and unskilled occupations, particularly for individuals previously out of work due to sickness and disability.
7.5.2.4 Incorporating benefits data

It would have been useful to have a more objective measure of employment status in order to examine the complex subgroups that exist amongst the IDS. This could perhaps be achieved in the future through the linkage of benefit records data to the SLS. Though not all sick or disabled inactive individuals will claim benefit payments, this information could be used in conjunction with census information to validate the census measures used in the analysis and reduce potential miss-classification around the census period as well as providing interesting avenues for future research. Furthermore, such a linkage would also mitigate some of the obstacles highlighted above regarding multiple transitions into and out of unemployment or inactivity during the intercensal period. More interestingly, a linkage to benefits data would enable extensions to the research contained in this project through examination of these different patterns of transition and their effects for health and mortality. It might also make possible studies that examine the health effects of different policy interventions in the labour market which is especially prescient given the recent reforms to the welfare state that have been outlined at various points in this thesis.

7.5.3 Concluding remarks

From the review of the literature presented in chapter two it was noted that there remains significant debate surrounding the nature of the relationship between unemployment and mortality. On the one hand, the observed association between worklessness and mortality is the result of poorer underlying health amongst workless individuals and, on the other, it is the result of a causal effect of worklessness that raises the risk of mortality independently of other characteristics. The main contribution of the empirical findings in this thesis is to reaffirm the causal explanation for the worklessness and mortality association and to highlight the importance of understanding the notions of confounding and selection when examining these relationships. In addition to this, the thesis also uncovered a number of other important additions to our knowledge of the roles of confounding in the worklessness and mortality relationship, particularly the limited role of health differences between the AU and the in work. Furthermore, important contributions to the debate surrounding health related worklessness were also made in relation to the theory of hidden unemployment and the socioeconomic and health precursors to health related labour market inactivity. All of these observations were made possible through the application of a causal analytical framework which illustrated the potential dangers of over-adjustment and allowed for a greater illustration and visualisation of structural confounding when compared to a standard regression approach. The thesis therefore makes a worthwhile contribution to
methodological issues in the discipline of public health and epidemiology which hold potential implications for many other questions in these disciplines.

The overall theme that has been developed throughout this project is the central importance of labour market marginalisation as a determinant of health. Since the publication of the Black report in 1980 which highlighted growing social disparities in health in the UK (Black 1982), the potential for various forms of socio-economic disadvantage to determine individual health has seen a great deal of research interest. Motivated in a large part by the shifts in labour market dynamics and economic & social security policies during the transition to post-Fordist modes of production in the 1980s and early 1990s, the role of labour market marginalisation has been of particular interest. The collective findings from this study suggest that amongst males in Scotland, there does indeed appear to be a causal pathway from worklessness to mortality and that, from a public health perspective, worklessness continues to be a price not worth paying for economic growth.

However, implementing policies that address these issues remains politically unpalatable. From political economy theory, the role of the state, labour market policy and the social security system becomes one of maximising the accumulation of capital and economic growth. The role of unemployment and worklessness more generally is to regulate and ‘discipline’ workers through competition and wage control. Thus, the dominant policy approach to addressing unemployment and worklessness is through ‘activation’ interventions for the workless to ensure the supply of work-ready labour from a competitive and flexible labour market. These interventions themselves can be implicated in various ways in the mechanism underpinning the health effects of worklessness. Yet, undermining the centrality of paid work in policy discourses requires transcending essential and hegemonic discourses of the relationship between labour and the accumulation of capital. If research in public health and social epidemiology (including the research in this thesis) is committed to the project of reducing health inequalities generally and the causality of worklessness and health specifically, engagement with more radical and progressive critical perspectives is essential. As much of the west progresses through what is feared to be a crisis of capitalism on the scale of those experienced in the 1930s and 1970s, the imperative for such work is as great as ever.
References


BBC. (2012). The economic downturn has taken its toll on the UK's workforce.


References


Brenner, M.H. (1976). *Estimating the social costs of national economic policy: Implications for mental and physical health, and criminal aggression*: Johns Hopkins University, School of Hygiene and Public Health and Department of Social Relations.


References


Ellaway, A., & Macintyre, S. (1998). Does housing tenure predict health in the UK because it exposes people to different levels of housing related hazards in the home or its surroundings? *Health & Place, 4*, 141-150.


References


a 10 year follow up using the health and retirement survey. *Occupational and Environmental Medicine*, 63, 683.


Heron, E., & Dwyer, P. (1999). Doing the right thing: Labour’s attempt to forge a new welfare deal between the individual and the state. *Social Policy & Administration*, 33, 91-104.


References


References


McKeown, T. (1976b). *The role of medicine: dream, mirage or nemesis*.


ONS. (2010). How exactly is unemployment measured?


References


References


alcohol consumption on a typical occasion*. British Journal of Addiction, 86, 1269-1281.


193


