Promotion of Health and Prevention of Ill-Health for Camden's Older Citizens:

The systematic use of existing Administrative Data to examine the relationship between Health, Contact with Social Services and Socio-economic Characteristics

Behrooz Tavakoly

Thesis submitted for the degree of
Doctor of Philosophy

City University
Department of Sociology

June 2008

THE FOLLOWING HAVE BEEN REDACTED AT THE REQUEST OF THE UNIVERSITY

MAPS ON PAGE 5 AND 6

Dedicated to my mother and father who passed away during this study

TABLE OF CONTENTS

PART I - INTRODUCTION, BACKGROUND, DEFINITIONS AND DATA PREPARATION

PART III - POLICY IMPLICATIONS, CONCLUSIONS & DISCUSSION

LIST OF TABLES

LIST OF FIGURES

Acknowledgements

I would like to thank my supervisor Dick Wiggins, who provided invaluable support during this study and significantly influenced my knowledge of statistics, research methods and planning the research. I am also grateful to my co-supervisor Les Mayhew, who supported me with methodological and analytical experience in analysis of health data. Thanks also to my Camden Primary Care Trust supervisor Natasha Roberts and all members of the study advisory board for their valuable contributions; Adele Yemm, Elizabeth Breeze, George Magoulas, Laidon Shapo, Mousumi Basu-doyle, Ian Rees Jones and Suzanne Lutchmun. I would also like to add special thanks to Barry Kelly for his support. I must thank the Economic and Social Research Council together with Camden Primary Care Trust who provided the funding to make this study possible.

Abstract

This study investigates the relationship between mortality and a number of factors drawn from existing administrative databases including gender, housing tenure, council tax bands (a proxy of wealth) and three popular causes of hospital admission (falls, strokes and ischemic-heart disease) for Camden residents aged 50 years and older. The study also includes an assessment of information on social service contact in order to identify the potential and/or the effectiveness of service delivery.

Existing data sources are merged using a relational database management systems approach. Risks of mortality are examined for different combinations of factors (Risk Ladders). The relative importance of risk factors are assessed by logistic regression and the model's ability to discriminate between 'those subjects who experience the outcome of interest versus those who do not', are also evaluated by use of Receiver Operating Characteristic (ROC) Curves.

The risk of mortality is more likely to occur for people living in social housing and lower council tax bands (A-C) than private housing and higher tax bands (D-H) and for men rather than women. However, the effect of tenure varies for different age groups, gender and tax band. The risk of mortality significantly increases for those groups of individuals who had at least one hospital admission for any of the three causes during 2002-04. Our results show the extent to which contact with social services is aligned with mortality risk among this age group with consequent implications for how social services are organised and delivered.

Abbreviations

AUC	Area Under Curve
CHD	Coronary Heart Disease
EHCS	English House Condition Survey
FIS	Falls, Ischemic-Heart Disease and Strokes
GIS	Geographical Information System
IMD	Index of Multiple Deprivation
LASIR	Local Area Shared Information Resource
LHO	London Health Observatory
LLTI	Limiting Long Term Illness
LPG	Local Property Gazetteer
NHS	National Health Services
NSF	National Service Framework
ONS	Office for National Statistics
OP	Older People
OR	Odds Ratio
PCT	Primary Care Trust
RDBMS	Relational Database Management System
ROC	Receiver Operating Characteristic Curves
RSL	Registered Social Landlords
SEP	Socio-Economic Position
SES	Socio-Economic Status
SOA	Super Output Area
SQL	Structured Query Language
SS	social services/being know or being in contact with social services
UPRN	Unique Property Reference Number
WHO	World Health Organization

PART-I:

INTRODUCTION, BACKGROUND,

DEFINITIONS AND

DATA PREPARATION

1 Introduction and background

1.1 Introduction

Health is defined in the World Health Organization (WHO) constitution of 1948 as: a state of complete physical, social and mental well-being, and not merely the absence of disease or infirmity (WHO, 1998). Robertson & Minkler (1994) take a broader definition and define health as: "a complete state of physical, mental and social well being and not merely the absence of disease, and focused on the social, political and economic determinants of health not amenable to improvement by medical care". This broader definition of health is an alternative to medicalized notion of health as it focuses on individual lifestyles (Oliver & Peersman, 2001).

Acheson (1998) in his report of *'independent inquiry into inequalities in health'* states: "Inequalities in health exist, whether measured in terms of mortality, life expectancy or health status; whether categorised by socioeconomic measures or by ethnic group or gender". Factors related to the socio-economic status of individuals always have been a central issue in the debate on social determinants of health. Indeed as Krieger (2001) suggests "...poverty has a direct effect on mortality rates. In general, people of higher socioeconomic position (SEP) ... enjoy better health. SEP is thought to affect health through a multileveled set of mechanisms".

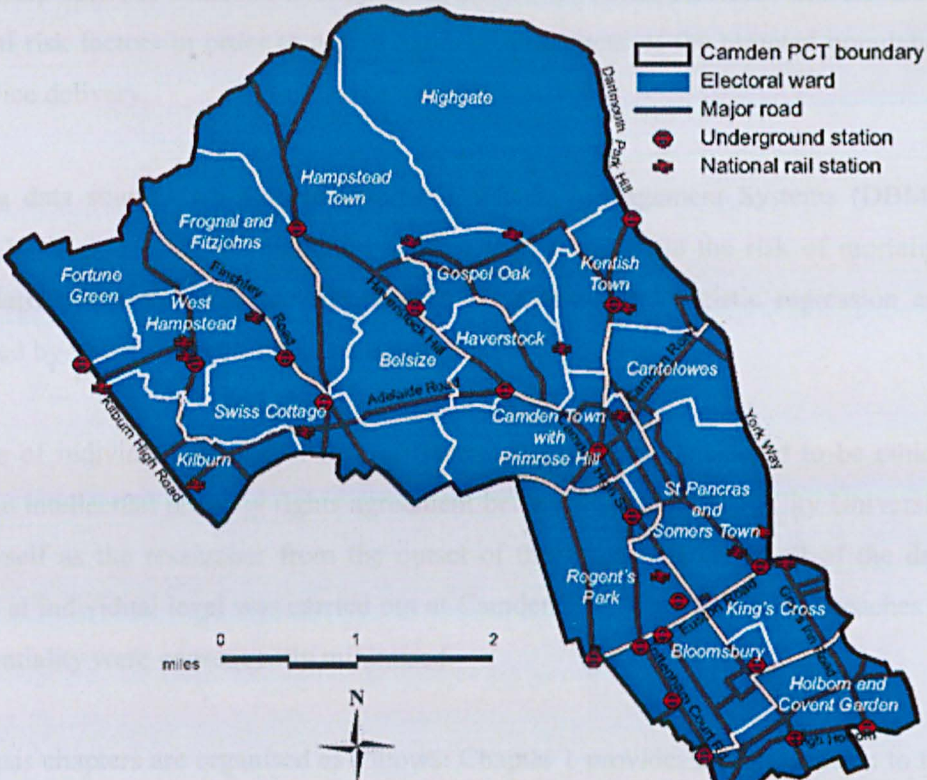
In the UK, heart disease and stroke are two of the top three most likely causes of death (Philp, 2004), which also mirrors the situation in the USA (American Heart Association's Heart Disease and Stroke Statistics, 2004). In addition, in the UK falls and fractures are also common causes of death (Philp, 2004). It is also important to note that people that belong to a lower socioeconomic status have higher mortality, morbidity, and risk factor levels for heart disease and stroke than persons of higher socioeconomic status (Centre for Disease Control and Prevention, 1999).

In order to address these issues the Health Act 1999 was implemented, which introduced partnerships to improve local services at the intersection of health and local services. The 2001 National Service Framework (NSF) in England established national standards and a national program of reform for older people's services. NSF Standard-8 is defined as: "the health and well-being of older people is promoted through a co-ordinated programme of action led by the NHS with support from the councils' (Department of Health, 2001). It emphasised that: "The aim of the integrated strategies for older people has been declared as promotion of good health and quality of life, and prevention or delay of frailty and disability which can have significant benefits for the individual and society" (Department of Health, 2001).

This research project conducted in the context of this policy setting aims to enhance Camden Primary Care Trust's strategy to specifically characterize the needs of the older people in response to NSF Standard 8. Camden is one of the most appropriate boroughs in London to undertake such research, as it is an inner city borough with a large population over 50 years old, with a wide range of socio-economic groups and variety of ethnicities. Figure 1.1 bellow illustrates map of greater London with borough's boundary and Figure 1.2 shows London Borough of Camden with electoral wards boundary.

Figure 1.1 Map of London boroughs

Figure 1.2 London Borough of Camden, Electoral ward boundaries



In Camden tackling health inequalities is a key local objective, and the 2001 Annual Public Health Report prepared by Camden and Islington Health Authority (2001), provided a detailed examination of the key adverse social, economic, environmental and lifestyle factors which drive local patterns of ill-health. The findings of the current research will help the Trust identify where to target future services to meet its priorities for older citizens.

The specific aims of this research are to examine the patterns of mortality among 'Older People' in the Borough of Camden for 2002-04, in order to target population disease prevention and health promotion interventions.

Furthermore, the research will examine the relationship between mortality rates and a number of explanatory factors (potential risk factors) drawn from existing databases in the London Borough of Camden. These factors include; age, gender, housing tenure, council tax banding (a proxy of wealth) and three popular causes of hospital admission (including falls, ischemic-heart disease and strokes). The study will also examine the relationship between someone already being known to 'social services' and the above potential risk factors in order to assess the appropriateness of the targeted population for service delivery.

Existing data sources are merged using a Database Management Systems (DBMS) approach. 'Risk Ladder' methodology is then used to estimate the risk of mortality. The relative importance of the risk factors are assessed by logistic regression and evaluated by Receiver Operating Characteristic (ROC) Curves.

Linkage of individual level data from different sources was considered to be ethical under an intellectual property rights agreement between Camden PCT, City University and myself as the researcher from the outset of the project. Because all of the data linkage at individual level was carried out at Camden PCT site, the risk of breaches of confidentiality were consequently minimised.

The thesis chapters are organised as follows: Chapter 1 provides an introduction to the study and the background of the research. It also includes a literature review on the wider determinants of health and develops the local and national policy context. Chapter 2 provides an overview of definitions and terminologies related to risk and its prediction. Chapter 2 also discusses a number of methodological issues and their definition. Chapter 3 covers the process of data management including data cleaning, data integration, and variable creation for the purpose of analysis. Chapters 4, 5, 6 and 7 will focus specifically on data analysis. Chapter 4 concentrates on the risk ladder approach and the observed risk of mortality for different groups of people with similar characteristics will be estimated. Chapter 4 will also include the mapping of observed risk based on analysis presented in the first part of the Chapter. The relative importance of risk factors will be assessed using logistic regression in Chapter 5 and its results will be evaluated in Chapter 6. Chapter 7 discusses 'policy implications' and finally Chapter 8 will provide discussion, conclusions and recommendations for future work.

1.2 Background

Inequalities in health and their relationship to poverty have been a well known theme among public health researchers for centuries. The French physician Louis René Villermé as early as 1826, proved that poorer neighbourhoods in Paris had higher mortality rates. Edgar Sydenstricker, an American epidemiologist in the 1930s, showed how the Depression impacted upon people's health (Krieger, 2000). Drever & Whitehead (1997), also in relation to the record of health inequalities, state "There is relatively firm evidence of substantial social inequalities in mortality in 17th century Geneva, and other parts of Europe and Britain in the 18th century. Throughout the 19th and 20th centuries evidence has continued to emerge of differentials in health between different population groups".

An example of health inequalities in UK in 19th century is illustrated in Table1.1. The longevity of families[1] by class and area of residence between 1838 and 1841 is shown in Table1.1.

Table1.1 Longevity of families, by class and area of residence, 1838–41

District	Gentry and professional	Farmers and tradesmen	Labourers and artisans
Rural			
Rutland	52	41	38
Urban			
Bath	55	37	25
Leeds	44	27	19
Bethnal Green	45	26	16
Manchester	38	20	17
Liverpool	35	22	15

Source: Drever and Whitehead (1997) adapted from Lancet 1843, Office for National Statistics

The above table not only shows much higher mortality in urban area than rural districts, but also illustrates the huge gap in average age of mortality between

[1] Longevity of families (assumed) = Average life of family members

different socio-economic groups of people. The gap between Gentry and Labourers in Bath is 30 years and the average mortality age of 15 for labourers in Liverpool is extreme. Comparing the longevity of families of labourers in Liverpool and Gentry in Bath shows the longevity of families for the latter group to be nearly 4 times longer than for the former.

Drever & Whitehead (1997) claim, in relation to the recent context of influence of socio-economic status on health, that "Health in the late 20th century is still greatly influenced by the prevailing social and economic conditions, and there remain large differentials in the health of different groups of the population". Graham (2001) also highlights that the long tradition of health inequalities research in the UK makes it well placed to unravel the links between inequalities generally and health inequalities in particular.

Promotion of health and prevention of ill health has become a key theme in social and national policy across health and social care for older people in recent years. It has been the central issue in many policy debates related to the public health, including: the white paper 'Caring for People' (Department of Health, 1989), 'Better Services for Vulnerable People' (Department of Health, 1997), 'Modernising Health and Social Services' (Department of Health, 1998), the white paper 'Saving lives: Our Healthier Nation' (Department of Health, 1999), the Cabinet Office initiative on 'Better Government for Older People' (Report of the Steering Committee of the Better Government for Older People Programme, 2000) and 'National Service Framework (NSF) for older people' (Department of Health, 2001).

The white paper 'Saving lives: Our Healthier Nation' (Department of Health, 1999) and the Cabinet Office initiative on 'Better Government for Older People' (Report of the Steering Committee of the Better Government for Older People Programme, 2000) are identified by Godfrey (2001) as two important policy strands that have relevance to how prevention is conceptualised in official discourse by focusing on reducing health inequalities. Godfrey states that: "In addition to the long standing public health concern with individuals assuming responsibility for their own health, including changing their lifestyle to reduce risk of chronic illness and disease, there was a new emphasis on government action to improve living conditions and secure healthier living". In the white paper of Department of Health (1999) it has been suggested that: "...there are powerful factors beyond the control of the individual which can harm health. The Government has a clear responsibility to address these problems. Striking a new balance – a third way – linking individual and wider action is at the heart of our new approach". Godfrey (2001), in reference to health promotion within the above framework, states: "...health promotion is conceived of as part of a wider strategy to reduce social and economic inequalities that impact on illness and disability".

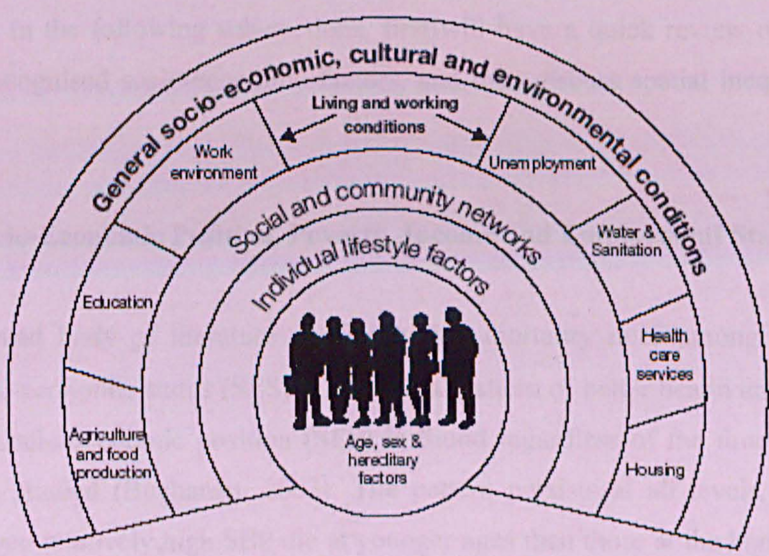
In the following section,Iwill review some of the most recent research and literature to examine the relationship between health and some of the major risk and protective factors. FirstIwill have a quick look at the major influences on health and a more detailed review of the social determinants of health.

1.3 Determinants of health

1.3.1 Major influences on health

Drever and Whitehead (1997) highlight that social and environmental issues have an influence on any population's health and well-being. They of course indicate that a number of factors are fixed such as gender, genetic make up and age but that there are also external influences on health, community living, working conditions and environmental factors. Figure 1.3 illustrates the many determinants of health introduced by Dahlgren & Whitehead, 1991. The key aspect of Dahlgren and Whitehead's model of health is that all of these factors interrelate to form a very complex relationship. In short, these factors do not exist in isolation.

Figure1.3 the main determinants of health



Source: Drever & Whitehead (1997) adapted from Dahlgren & Whitehead, 1991

While the Dahlgren & Whitehead model of health determinants focus on a wide range of health determinants including general, cultural or environmental factors, the particular area of interest for this research lies in the social determinants of health. In Dahlgren & Whitehead's model social determinants include factors such as housing, employment, care and services etc. The social determinants of health are represented on the second outer ring on the model illustrated in Figure 1.3.

1.3.2 Social determinants of health

The relationship between poor health and low socioeconomic position in Britain is well recognised but its origins are complex. Lower socioeconomic groups are seen to have a greater incidence of health disorders such as heart disease, stroke, and some cancers in adults. Risk factors including smoking, physical inactivity, obesity, hypertension, and poor diet are also clustered in the lower socioeconomic groups (James et al., 1997).

It is difficult to separate out and to assess the importance of impact of each socio-economic factor on health individually as most of the factors interrelate with each other. For example poverty, income, social class, education and tenure are all correlated. In the following sub-sections, firstIwill have a quick review of some of the well recognised socio-economic factors, and then discuss spatial inequalities in health.

1.3.2.1 Socio-Economic Position, Poverty, Income and Employment Status

A widespread body of literature verifies higher mortality rates among people of lower socio-economic status (SES). The general pattern of better health among those of higher socio-economic position (SEP) is found regardless of the time period or population studied (Buchanan, 2003). The pattern persists at all levels, such that those of even relatively high SEP die at younger ages than those at the highest levels (Marmot et al., 1997).

In the USA the interrelationship between SES and health and subsequently mortality has been observed many times in health research. Death rates for poor groups of people is generally two to three times higher than the death rates for rich ones (Auerbach & Krimgold, 2001). In the 1980s, white men in the USA with a family income lower than $10.000 had a 6.6 years lower life expectancy than those with an income higher than $25.000 (Smith, 1999).

Bowling (2004) and Mackenbach et al. (1997) also state that people in lower socioeconomic status groups experience poorer health and live shorter lives than those in higher status groups. "It has been consistently found that among adult populations, mortality at the lower end of the socioeconomic scale is higher than mortality at the higher end. Also among elderly populations, socioeconomic mortality inequalities are found" (Huisman et al., 2004). They argue these inequalities often decrease with increasing age. Research by Mackenbach et al. (1997) also shows that in all Western European countries the risks of morbidity and mortality were higher in the lower socioeconomic groups.

Subject to the relationship between age and income and their combined effect on mortality, Wolfson et al. (1993) cited by Gardner & Oswald (2004), analysed nearly 550,000 administrative records from the Canadian Pension Plan in a longitudinal analysis of male mortality after the age 65. They found that higher earnings in late middle age (age 45–64) were associated with significantly lower mortality at older ages (65–74)". The effect of constant low income on mortality was assessed by McDonough et al. (1997). They found that persistent low income was a good predictor of early mortality.

In relation to the effect of employment status on mortality; Iversen et al. (1987), Moser et al. (1984), Morris et al. (1994), Martikainen and Valkonen (1996) and Gardner & Oswald (2004) all demonstrate that individuals who experience unemployment are more likely to experience reduced longevity than comparable individuals who are continuously employed.

1.3.2.2 Social Class

The relationship between health inequality and social class has been examined in many studies. For example, in the study of the 'magnitude of social inequalities in Coronary Heart Disease' (CHD) by Marmot (1998), this relationship was described as, "Among men, death rates from CHD are about 40 percent higher among manual workers than among non-manual workers; the death rate for wives of manual workers is about twice the rate of wives of non-manual workers".

White et al. (2006) provide evidence for the UK that men in semi-skilled and unskilled social classes (Social Classes IV and V combined) had odds of death 1.54 times that of men in the professional classes.

1.3.2.3 Relationship between Income and Education with Lifestyle

A report by Washington State Department of Health (2002a) shows that Washington residents with lower incomes are more likely to smoke and also that women with lower incomes report more obesity compared with those people in higher income groups. Washington residents with lower levels of formal education report more smoking and obesity and lower consumption of fruit and vegetables than those with higher education. A similar study of Dutch men also shows the difference in life expectancy for Dutch men between the highest and lowest educational group is 4 years (Hoffmann, 2005).

1.3.2.4 Relationship between Socio-Economic Status and Ethnicity

Inequalities in social position have a substantial impact on the health experience of ethnic minority groups in terms of socio-economic disadvantage and discrimination. Modood et al. (1997) claims that the measures of SES have been developed for the White population and explains, "...ethnicity mediates access to the domains which these measures are designed to capture. For example, people from minority ethnic groups face higher rates of unemployment and of employment in low-skilled jobs than similarly qualified Whites". Graham (2001) also argues: "... measures of SES may have a variable - rather than consistent - relationship to life chances and living standards in different ethnic groups".

1.3.2.5 Tenure

Tenure has been considered as an important factor in determination of health for a long time. Chadwick (1842) states that in the nineteenth century: "... public health practitioners and theorists regarded housing conditions as a major determinant of population health and of the differences in health between social groups".

Much research has been conducted to examine the relationship between housing tenure and wealth. Tenure could be a good marker of income and socio-economic status which are difficult to measure with other methods. Macintyre et. al 2001 state that, "the frequent, but usually implicit, hypothesis underlying the use of housing tenure in planning and in social epidemiology is that it is simply a marker of income or social class, both of which are major determinants of health but are difficult to collect in surveys or are inappropriate for some groups" (Macintyre et al., 2001; Macintyre et al., 2001). Huisman *et al* (2004) used housing tenure and level of education as their indicators of SES for a target population of people aged 65 and over who were retired and occupation as a measure of SES for this group is less relevant. Dorling et al.(2001) in their study of *'Housing wealth and community health'* found out the owner-occupiers were, on aggregate, the tenure group with the greatest financial resources, in terms of both wealth and income. They also claim, in order to become owner-occupiers, and live in the types of areas where this tenure is concentrated, requires a certain level of capital and/or income.

A report by the ONS (1998) based on the result of 1996 General Household Survey (cited in Graham (2001)) shows that education is increasingly determining access to employment and employment is in turn increasingly determining access to housing. The report emphasises, "...since the 1970s, there has been a rapid rise in owner occupation and a sharp decline in the availability of social housing (homes to rent from local authorities and housing associations)". It also states, "...the neighbourhoods in which tenants and owner occupier live, eight in ten heads of household are in paid employment; in the social housing sector, six in ten are economically inactive".

Macintyre et al. (2001) suggest that owner occupiers have significantly greater monthly household income adjusted for family size and are much less likely to receive all the household income from benefits. They also claim that owner occupiers were more than twice as likely as to be in paid employment than those renting and were more likely to be in non-manual occupations. Their findings also suggest that various socially desirable features of the home (such as the dwelling being a house rather than a flat, having more rooms, the presence of a garden and the main accommodation being on the ground floor rather than in the basement or above the fifth floor) are more commonly found in owner-occupied properties. Those renting accommodation in this study were found to be more likely to struggle with a range of stressors (with the exception of burglaries).

The direct relationship between housing tenure and health has also been assessed in a number of research projects. Breeze et al. (2004) indicate "In Britain people in rented homes in old age-whether living independently or with relatives-were more likely to have poor health related quality of life than those in owner occupied homes". Macintyre et al. (2001) also argue that the owner-occupiers of all ages have lower risk of death and better health than people who rent their homes. Their justification for this claim is that the housing tenure is acting as a marker for social class or for income and wealth.

A study based on the Office for National Statistics Longitudinal Study by Filakti and Fox (1995) shows between 1971 and 1981, age standardized mortality rates were around 25% higher for social tenants than for owner-occupiers. Moreover, although death rates have declined since that time, the decrease has been larger among owner-occupiers (Harding et al., 1997). According to the study based on the 1991 UK census, housing tenure is associated with a range of health measures, including higher rates of long-term illness (Gould & Jones, 1996) and psychosocial problems (Lewis et al., 2003) among social renters.

White et al. (2006) claim male residents in social housing in 1991 had an odds of death 1.41 times that of men in owner occupation. "Both private rented and social housing tenures increased the risk of death compared to men in owner occupied tenure… social housing in particular tends to be associated with social disadvantage" (White et al., 2006).

However, there are some exceptions in the relationship between housing tenure and socio-economic status, suggesting that tenure cannot be used as a precise marker of socio-economic status and material resources. The studies by Danesh et al. (1999), McLoone and Ellaway (1999) and Macintyre et al. (2001) show 7% of the owner occupier were indeed in the lowest income and 13% in lowest social class groups, while 9% of the social renters were in the highest income and 15% of social renters in highest social class groups.

1.3.3 Spatial inequalities

Spatial inequalities in health or, in other words, the health gap, in Britain between those with poor health and the healthy is wider now than ever (Dorling, 1997a; Shaw et al., 2000). A study conducted by Shaw et al. (1998) provides an example of spatial inequalities, with a concentration of premature deaths in areas of high deprivation. It states: "Poorer areas which had mortality rates 20 percent above the national average in the 1950s, like Oldham, Salford and Greenock, had mortality rates 30 percent above the national rate by the 1990s".

Hattersley (1999) in study of mortality by social class uncovered similar spatial inequalities. "Between 1980 and 1992, the life expectancy has continued to rise for men and women in all socio-economic groups, but the differential has become more pronounced. Between 1972 and 1996, life expectancy for men in social class 1 increased by 5.7 years: among men in social class 5, the gain was 1.7 years".

Dorling et al. (2001; Dorling et al., 2001) argue that spatial inequalities are very much consequence of 'social policy'. They clarify: "If fiscal policies continue to lead to increased income inequality, we can expect to see the spatial polarization of mortality continuing. Wealthy areas will get wealthier and healthier, and poor people (with poor health) will be left behind in those areas which are considered undesirable and where opportunities are sparse".

Another factor which influences the spatial inequalities is social capital. Social capital as Putnam (1995) states, refers to "... connections among individuals, social networks and the norms of reciprocity and trustworthiness that arise from them". Low levels of social capital have been associated with higher mortality rates (Kawachi et al., 1997). In areas with high income inequality, social trust is low, in part because, as Wilkinson (1999) notes, friendship and inequality are not compatible. Friendship includes the concepts of acceptance, appreciation, and reciprocity, while social hierarchy involves dominance and subordination, competition, and social comparison. In communities in which most people are social equals, levels of friendship and social trust and hence, social capital will be relatively high (Washington State Department of Health, 2002a).

So far some of the most influential socio-economic factors on health and their relationship with geography have been discussed. Given the continuing awareness of the relationship between the stated socio-economic factors and ill-health, the various relevant authorities have sought to improve the health through attempting to impact on these socio-economic factors. Thus, in the following sections the relevant policy context at national and local level will be reviewed.

Dorling (1997b) argues that nationally, mortality rates are higher in the north and in urban areas. In 1990s, a person living in Glasgow was 66% more likely to die in any given year than someone living in the districts of rural Dorset and 31% more likely than a resident of Bristol. While at the end of the 1960s the excess chance of dying in Glasgow, relative to rural Dorset and Bristol, was much lower, at 42% and 21% respectively.

In order to produce a fair measure of geographical inequality in mortality, Dorling (1997) included all parts of Britain by dividing the population into ten group of equal size (deciles). However, his analysis rested upon the population aged less than 65 years old. Dorling (1997) states: "...Historical records do not provide enough detail to look at variations in mortality over the age of 65". Table 1.2 adapted from Dorling (1997b) shows the age and sex standardised mortality ratio of the under-65 population of Britain living in each decile group of areas in 1950-53 and 1990-92. At the start of the 1950s people in the worst decile areas were 31% more likely to die than average. By the early 1990s that differential had grown to 42% (the largest divergence ever recorded). However, people living in the areas with least deaths were 18% less likely to die than average in 1950-53. By 1990-92 this differential had grown to 24% less likelihood of death in that period (Dorling, 1997b).

Table 1.2 Relationship between health inequalities and poverty adapted from Dorling (1997b)

Health decile	Standardised mortality ratio 1950-53	1990-92	Current poverty indicator: residents in households with no car	Children in households with no work	65 with a long term illness
Worst	131	142	40.8%	33.2%	9.7%
2	118	121	31.4%	24.2%	8.4%
3	112	111	30.8%	21.0%	8.0%
4	107	105	26.2%	19.9%	8.3%
5	103	99	23.1%	15.2%	6.9%
6	99	94	22.3%	15.7%	6.4%
7	93	91	19.7%	14.1%	6.0%
8	89	86	17.0%	11.6%	5.6%
9	86	80	13.0%	9.6%	4.9%
Best	82	76	10.9%	7.9%	4.5%
Britain	**100**	**100**	**23.6%**	**17.4%**	**6.9%**

Curtis & Jones (1998) argue that health inequalities are influenced both by the characteristics of individuals, and the context or setting (landscape) in which they are situated. They also claim that there is theoretical and empirical evidence that health disadvantage may be experienced differently by socially disadvantaged individuals according to their geographical setting.

Power (2000) also emphasizes that social exclusion is almost entirely an urban problem. She adds: "Council estates have become increasingly unpopular and stigmatised as they became tied to slum rehousing, then became housing of last resort for people who might otherwise become homeless". In the UK, in a descriptive sense a strong link is often made, between social housing and the notion of social exclusion. Ratcliff (1998) also confirms the relationship between social exclusion and social housing. From this perspective, poverty and disadvantaged places are implicated in social exclusion and therefore are clustered spatially.

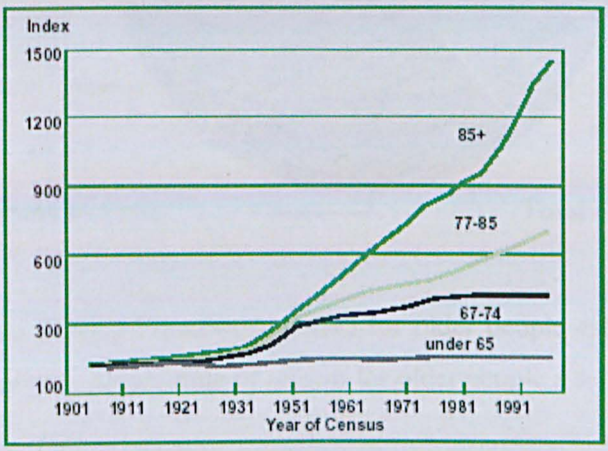
Provided this evidence holds for an older population then those older people living in deprived areas are likely to be at greater disadvantage or risk of premature mortality than those living in more affluent areas. This will be typically the case in urban areas where older (more affluent) residents migrate out in retirement leaving the less advantaged remaining in the borough (London Borough of Camden, 2007a).

1.4 Policy context

1.4.1 National context (Review of NSF)

England is an ageing society. Since the early 1930s the number of people aged over 65 has more than doubled and today a fifth of the population is over 60. Between 1995 and 2025 the number of people over the age of 80 is set to increase by almost a half and the number of people over 90 will double (Department of Health, 2001). Figure1.4 illustrates the increase of the number of older people in England from 1901 to 1991.

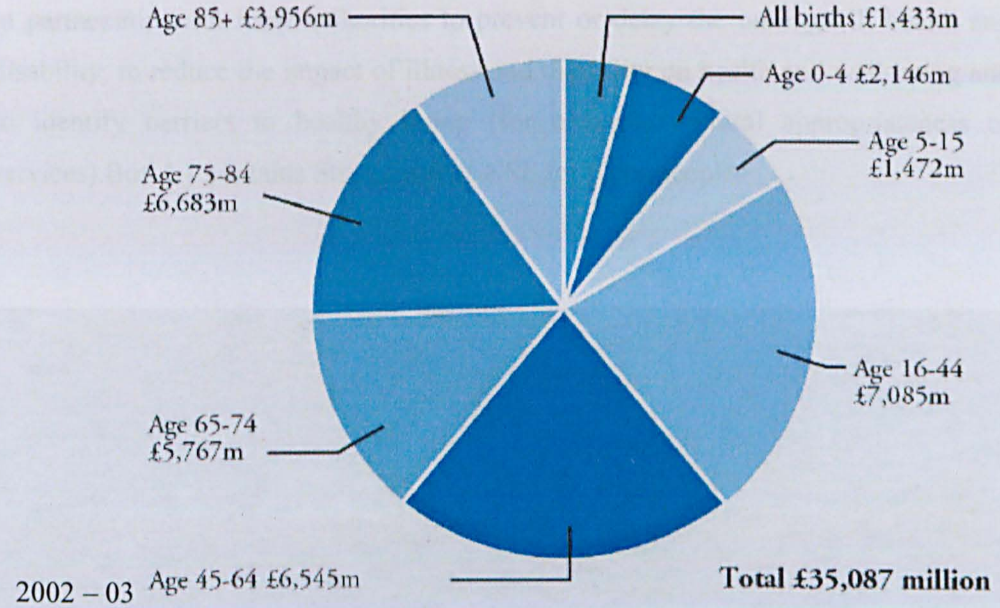
Figure1.4 Illustration of increase of number of older people in England 1901-1991(Department of Health, 2001)



The NHS spent around 40% of its budget - £10 billion – on people over the age of 65 in 1998/99. In the same year social services spent nearly 50% of their budget on the over 65s, some £5.2 billion(Department of Health, 2001). In 2002-03 people aged 65 and over accounted for approximately 47% of total expenditure, a group however, that comprises around 16% of the population (Department of Health, 2006). Figure 1.5 shows the hospital and community health services gross current expenditure by age for 2002-03.

Older people tend to have a much greater need for health and social services than the young, so the bulk of health and social care resources are directed at their needs. For example, almost two thirds of general and acute hospital beds are used by people over 65 (Cowan, 2003).

Figure1.5 Hospital and community health services gross current expenditure by age (Department of Health, 2006)



The 2001 National Service Framework (NSF) for older people established national standards and a national programme of reform for older people's services.

The NSF-2001 for older people emphasises that in both social care and in health care there are many examples of excellent service provision for older people. It continues, "However, there have been reports of poor, unresponsive, insensitive, and in the worst cases, discriminatory, services. Instances of adverse discrimination have usually been inadvertent, a result of the survival of old systems and practices that have failed to keep pace with changing attitudes or advances in the capacity of professionals to intervene successfully. Health and social care staff have been at the forefront of efforts to secure a better deal for older people, but too often the structures and practices that they have had to work with have frustrated these efforts" (Department of Health, 2001).

The NSF for older people set eight standards for the care of older people across health and social services and focuses on: rooting out age discrimination, providing person-centred care, promoting older people's health and independence and fitting services around people's needs (Department of Health, 2001). Appendix-A includes a summary of the eight standards of NSF for older people.

In Standard-8 of the NSF it has been highlighted that action can be taken by the NHS in partnership with local authorities to prevent or delay the onset of ill health and disability, to reduce the impact of illness and disability on health and well-being and to identify barriers to healthy living (for example cultural appropriateness of services).Box 1.1 contains Standard-8 of NSF for older people.

Box1.1 Standard Eight: The promotion of health and active life in older age

Aim: To extend the healthy life expectancy of older people.

Standard: The health and well-being of older people is promoted through a coordinated programme of action led by the NHS with support from councils.

Rationale: Growing old has been seen to represent a period of increasing dependency, as physical strength, stamina and suppleness decline, and the individual has to cope with chronic or long-term conditions. But chronic degenerative disease, disability and ill health are not an inevitable consequence of ageing.

There is a growing body of evidence to suggest that the modification of risk factors for disease even late in life can have health benefits for the individual; longer life, increased or maintained levels of functional ability, disease prevention and an improved sense of wellbeing.
Integrated strategies for older people aimed at promoting good health and quality of life, and to prevent or delay frailty and disability can have significant benefits for the individual and society.

This thesis will explore the use of existing data in the inner London Borough of Camden, in order to examine the potential for the council and the PCT to make suitable policy interventions at a local level.

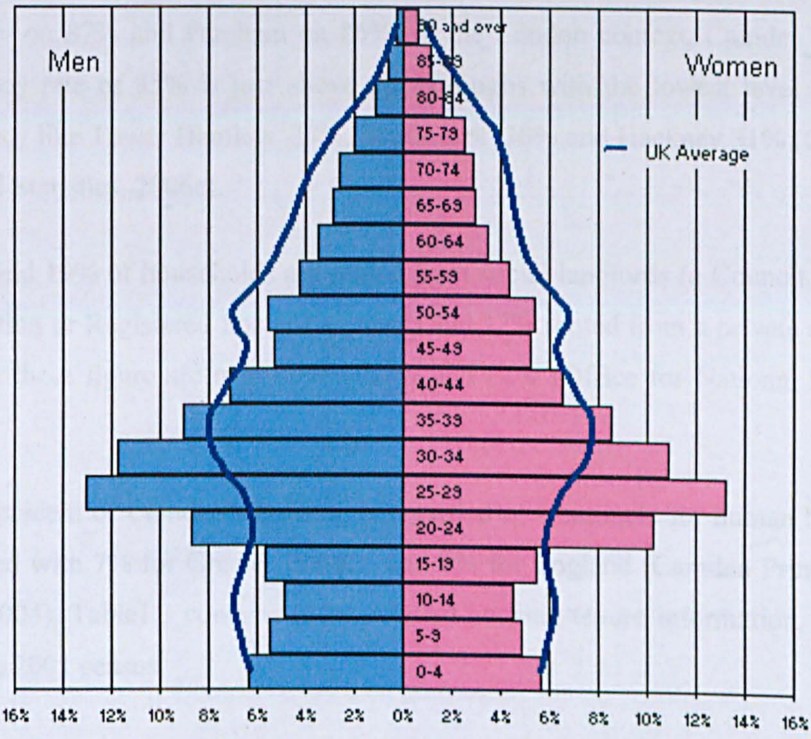
1.4.2 Local Context

The following section provides a socio-economic picture of Camden including: Camden's population structure and distributions of ethnicity, housing tenure, deprivation, mortality and life expectancy.

1.4.2.1 Camden's People

The population of Camden in mid-year 2003 were estimated at 210,700 people by the Office for National Statistics (Camden Primary Care Trust, 2005; Cowan, 2003). The population is comparatively young: 72% of the population are under 45 years old, 22% are aged 20-29 years (12% for England) and 10% over 65 years old (16% for England). Twenty seven percent of Camden's population are from ethnic minority communities (Camden Primary Care Trust, 2005). The structure of the population in Camden, compared with UK, is illustrated by means of a 'population pyramid' (Office for National Statistics, 2006b) in Figure1.6.

Figure1.6 Population pyramid for Camden's population structure by 5 years age group, compared to the UK.



1.4.2.2 Ethnicity

The Camden Primary Care Trust (2005) annual report states: "Camden's population is culturally and ethnically diverse. Culture and ethnicity may affect health beliefs and behaviours, and can therefore be important influences on health and wellbeing". Based on the 2001 Census, 26.8% of the Camden's population is from black and minority ethnic (BME) groups. The largest BME groups in Camden are Bangladeshi (6.4%), Black African (6%) and Irish (4.6%). In fact, 8% of all Bangladeshi and 4% of all Irish people in London live in Camden.

The report shows that some communities are concentrated in particular neighbourhoods. For example, Kings Cross has a high proportion of Bangladeshi people and Kilburn has a high proportion of Irish people (Camden Primary Care Trust, 2005).

1.4.2.3 Housing in Camden

Based on data from the 2001 census there are 91,603 households in Camden. For England and Wales more than two-thirds of homes are owner occupied and 31% rented. Castlepoint has the highest percentage of owner occupancy at 88%, followed by Blaby on 87% and Fareham on 86%. In the London context, Camden's owner occupancy rate of 35% is just above the boroughs with the lowest level of owner occupancy like Tower Hamlets 28%, Southwark 30% and Hackney 31% (Office for National Statistics, 2006c).

In England 19% of households are rented from social landlords (a Council, Housing Association or Registered Social Landlord) and 12% rented from a private owner. In Camden these figure are respectively 37% and 28% (Office for National Statistics, 2006a).

Eleven percent of Camden's housing is regarded as unsuitable for human habitation compared with 7% for Greater London and 6% for England (Camden Primary Care Trust, 2004). Table1.3 contains more detailed housing tenure information, extracted from the 2001 census.

Twenty nine percent of the Camden population is receiving housing benefit, while this figure for Greater London is 26% and for England is 15%. In Camden 22% of the population receive 'Housing Benefit' with 'Income Support' or 'Job Seeker Allowance' compared with 19% for Greater London and 11% for England (Camden Primary Care Trust, 2004).

Table 1.3 The distribution of housing tenure in Camden compared to England and Wales (Office for National Statistics, 2006b)

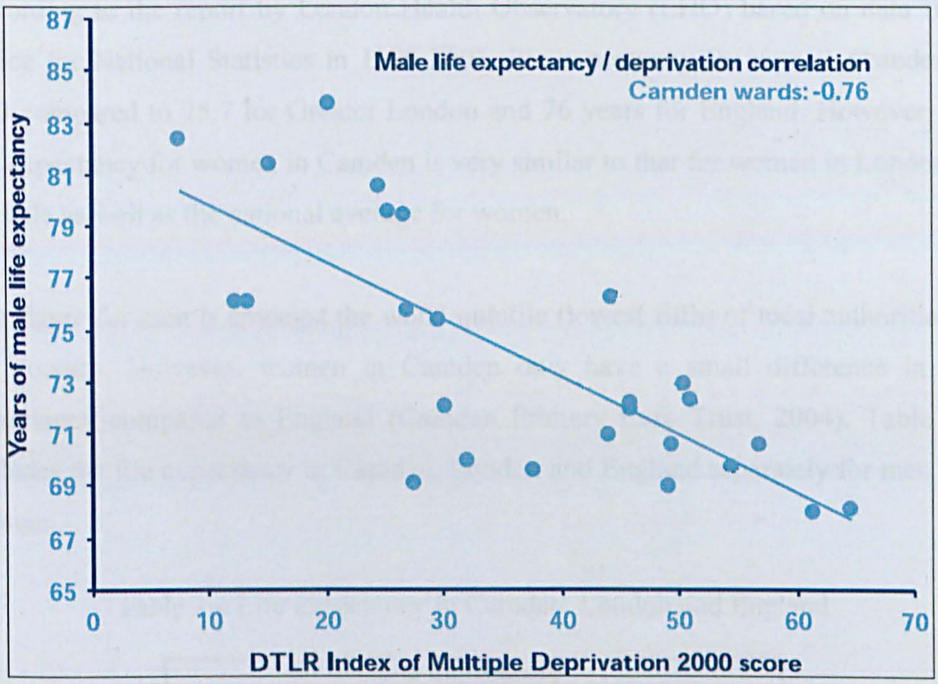
Tenure	Camden %	England and Wales %
Owner occupied	34.9	68.9
Rented from Council	26.0	13.2
Rented from Housing Association or Registered Social Landlord	11.4	6.0
Private rented or lived rent free	27.7	11.9

1.4.2.4 Deprivation in Camden

The Index of Multiple Deprivation (IMD) 2004 provides an overall deprivation score for local authorities and smaller areas known as 'lower layer Super Output Areas' (Camden Primary Care Trust, 2005). Lower layer super output areas consist of approximately 1,500 residents (Office for National Statistics, 2006d). The index incorporates the following seven elements: income, employment, health and disability, education, skills and training, barriers to housing and services, living environment and Crime.

According to the IMD, Camden is the 19th most deprived local authority of the 354 in England. Within Camden, 84% of the 133 Super Output Areas (SOAs) are more deprived than the national average, and almost a quarter of SOAs are among the 10% that are the most deprived in the country (31 SOAs). None of Camden's SOAs are in the least deprived 20% in England (Camden Primary Care Trust, 2005). Figure1.7 illustrates the association between deprivation and life expectancy for men living in Camden.

Figure1.7 The relationship between life expectancy and deprivation for males living in Camden wards 2000 (Camden and Islington Health Authority, 2001)



1.4.2.5 Mortality

The mortality rate in Camden is about 5% more than would be expected given the age and sex structure of the population. While in some wards (Kentish Town, Kilburn and St Pancras & Somers Town) the death rate is up to 30% more than expected, in some other wards (Belsize ward) it is 20% below expected death rate. The death from coronary heart disease accounts for nearly a fifth (18%) of the total death in Camden and is 6% (in some wards up to 35%) more than expected. Death as a result of mental health is 21% above the England rate and the deaths related to suicide or undetermined injury is 60% more than national average. The number of deaths caused by cancer is also 12% higher than expected (Camden Primary Care Trust, 2004).

1.4.2.6 Life expectancy in Camden

According to the report by London Health Observatory (LHO) based on data from Office for National Statistics in 1999-2003, life expectancy for men in Camden is 74.3 compared to 75.7 for Greater London and 76 years for England. However; the life expectancy for women in Camden is very similar to that for women in London as a whole as well as the national average for women.

The figure for men is amongst the worst quintile (lowest fifth) of local authorities in the country. However, women in Camden only have a small difference in life expectancy compared to England (Camden Primary Care Trust, 2004). Table 1.4 includes the life expectancy in Camden, London and England separately for men and women.

Table 1.4 Life expectancy in Camden, London and England

	Life Expectancy	
	Male	Female
Camden	74.3	80.6
Inner London	74.3	79.9
Outer London	76.5	80.9
London	75.7	80.6
England	76.0	80.6

Source: ONS and GLA Analyses by LHO

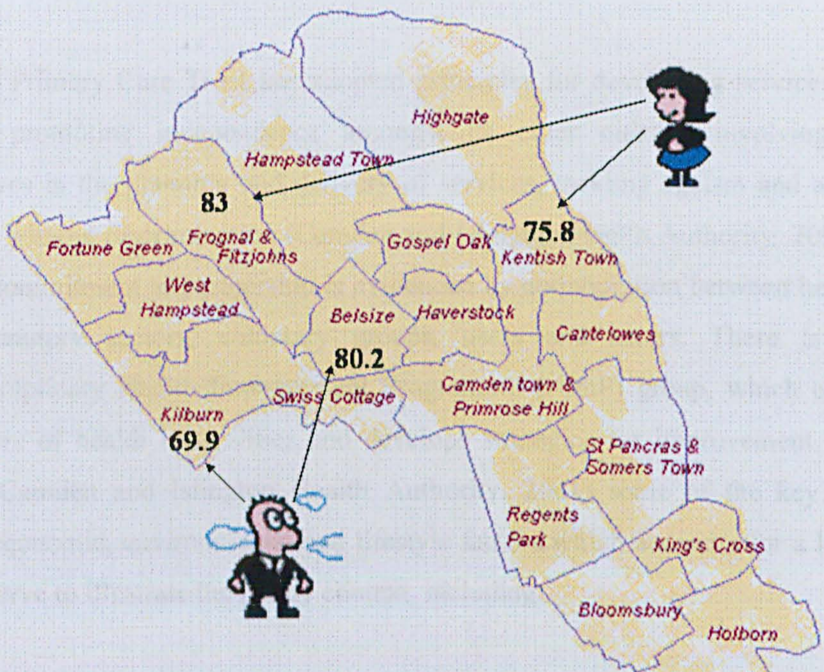
1.4.2.7 Inequalities in life expectancy in Camden

For men there is more than 10 years difference in life expectancy between the best and worst ranking wards: in Kilburn it is 69.9 years (ranked 620th of the 624 electoral wards in London) while in Belsize it is 80.2 years (ranked 30^{th} in London).

The pattern for female life expectancy within Camden is similar to that for men. For women, life expectancy is lowest in Kentish Town at 75.8 years (ranked 620^{th} of electoral wards in London) and highest in Frognal & Fitzjohns at 83 (ranked 94th).

Figure1.8 illustrates the inequalities in life expectancy for both male and female in Camden (Camden Primary Care Trust, 2004).

Figure1.8 Difference in life expectancy between the best and worst ranking Electoral wards (from the total of 18 Electoral wards) in Camden for both male and female



We will now turn our attention on Camden's older citizens.

1.5 Older People in Camden

There is no an agreed definition of older people. However Camden's definition of older age as indicated in their Community Strategy to enhance the 'Quality of Life (QoL) of Camden's older citizens' is the same as NSF for older people which includes people as young as 50 (London Borough of Camden, 2002).

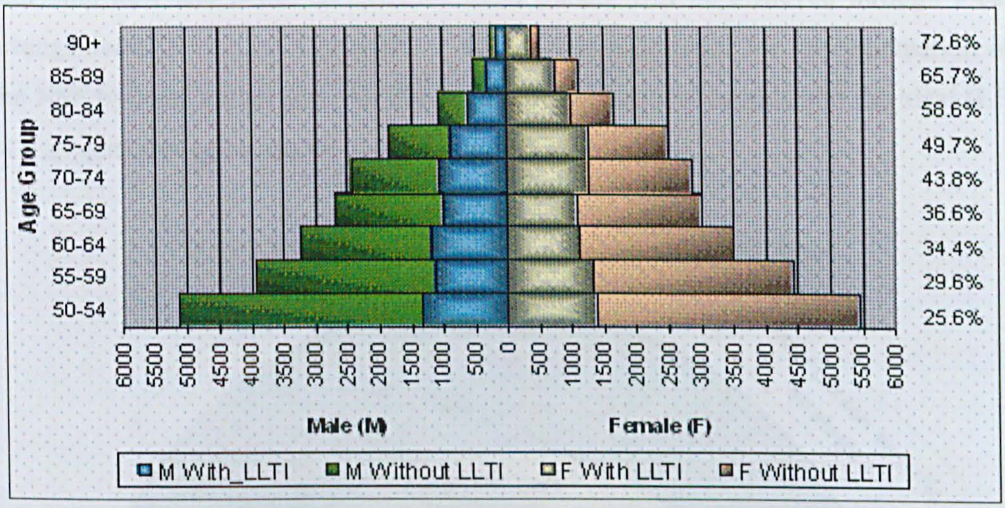
Camden Primary Care Trust has adopted principles for developing services. These include promoting independence amongst its elder citizens involving elders themselves in the planning and delivery of services, tackling ageism and access to services without discrimination (Camden and Islington Health Authority, 2001). The Trust's commitment to partnership is evidenced by collaboration between health and local statutory sectors, voluntary groups, users and carers. There is also a multidisciplinary Health Improvement Programme (HImP) group, which examines the nature of health inequalities and develops strategies for improvement. In their report (Camden and Islington Health Authority, 2001) some of the key adverse social, economic, environmental and lifestyle factors which accounts for a lot of ill-health serve to illustrate the policy context, including:

Twenty percent of persons aged over 65 years in Camden receive income support. One in four may not be claiming benefits for which they are eligible. Around 7,000 dwellings in Camden are unfit for habitation and there are no specific services for homeless elders. Seventeen percent of the elderly in one part of the district suffered clinical depression.

1.5.1 Limiting Long Term Illness (LLTI)

Figure1.9 shows the population of Camden (over 50 years old) with LLTI for every five years age group and for both male and female. The figure also illustrates the percentage of people with LLTI in each age group. This chart has been produced based on information from the 2001 census. The percentage of people with LLTI for age group 50-54 increases from 25.6% to 72.6% for those over 90 years old.

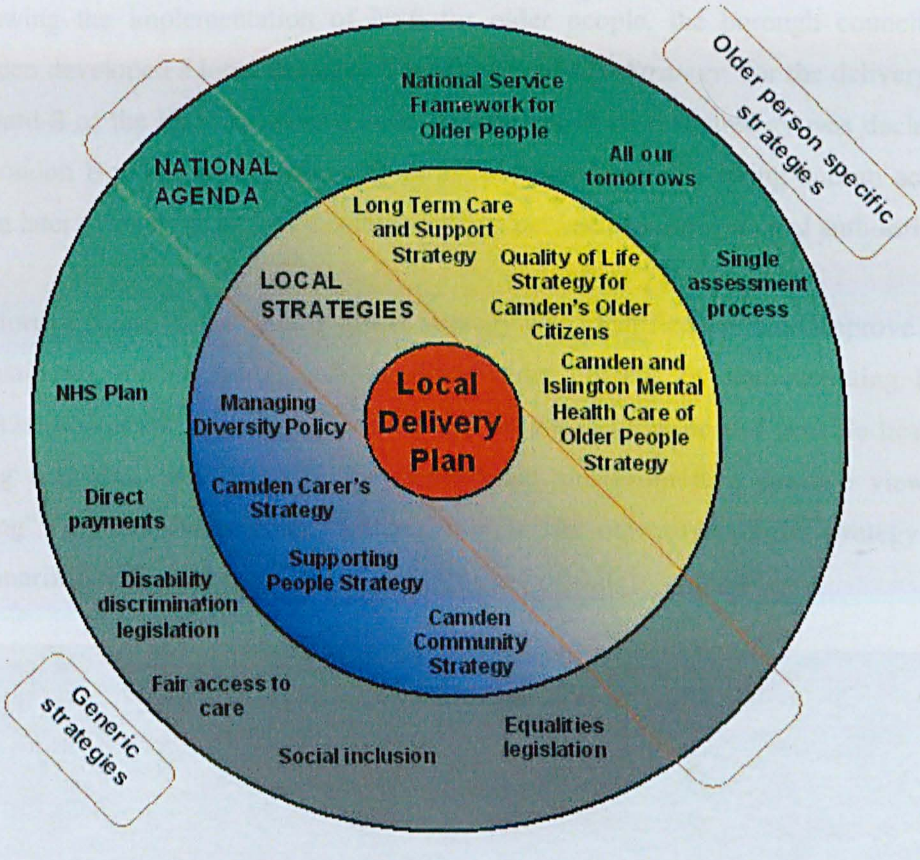
Figure1.9 Illustration of the distribution of older people with limiting long term illness in Camden



1.6 Local Implementation of NSF for older people in Camden

The NSF for older people highlights the translations of the national standards of the NSF into new and better services for older people will be achieved through local arrangements. It also set a program of actions and milestones for implementation of the NSF by local authorities. An illustration of the national agenda and the local (LB of Camden) strategies can be found below in Figure1.10.

Figure1.10 Health strategies at National and Local level



Source: London Borough of Camden Social Services Department & Camden Primary Care Trust (2004), Serving older people

In response to the implementation of NSF for older people, local authorities in Camden developed several strategies and many of the above needs are being met or plans are being developed. Three principal strategies developed by the local authority in Camden include:

- The Quality of Life Strategy for Camden's older citizens, implemented by Camden council.
- Long Term Care and Support Strategy implemented by London Borough of Camden Social Services Department and Camden Primary Care Trust
- Mental Health Care of Older People Strategy (MHCOP) implemented by 'Steering Group of the Camden Mental Health Joint Commissioning Group'.

1.6.1 The Quality of Life Strategy for Camden's older citizens

Following the implementation of NSF for older people, the borough council in Camden developed a local strategy, *'The Quality of Life Strategy'* for the delivery of standard-8 of the NSF for older people. The objective of this strategy was declared by London Borough of Camden (2002) as: promoting healthy living and an active life in later years by closer partnership working between health and local authorities.

The foremost aim of the quality of life strategy was identified as "...to improve and maintain the quality of life of Camden's older citizens by demonstrating how agencies will work together and with older people to promote and provide healthy living activities, sustain people's independence and promote a positive view of ageing" (London Borough of Camden, 2002). The objectives of the strategy are summarized in Box1.2.

Box1.2 The Objectives of The Quality of Life Strategy for Camden's older citizens

1. To promote the principles of *active engagement*.
2. To make older people aware of a *range of opportunities and activities* that facilitates health and well-being.
3. To ensure older people have *equal access to* both statutory and non-statutory *services*.
4. To ensure older people have access to a range of services that can help to *maximise their income*.
5. To ensure older people *feel safe and secure* in their homes and in the community.
6. To promote a positive approach to the experience of ageing *(campaign against ageism)* through older people working together and with others.
7. To *challenge assumptions about ageing* by promoting links and activities across generations.
8. To *prepare generations* for the opportunities and challenges of later years.

The following initiatives and programmes were established:

The Health Improvement Programme for older people: to reduce inequality in accessing services.
Joint Investment Plan for older people: To enable the older people to stay at home as long as possible.
Well and Wise Healthy Living Network: Reducing social exclusion and poverty.
Camden Gold: Focus on minority ethnic groups.
Camden's Champion for Older People: Engaging older people in community planning & local political process.

1.6.2 Long-term Care and Support Strategy

The aim of this strategy as it has been stated in the report is to "... provide the London Borough of Camden, Camden Primary Care Trust (PCT) and other partner agencies with a framework for the future planning of accommodation and related services for older people" (Camden Primary Care Trust & Camden Council, 2003). The report also states that it focuses on long-term care and support for older people and not only will it address the minority of older people with explicit health and social care needs but also the majority of older people who are not major users of health and social care services. The strategy will be implemented within the overall framework of Camden PCT's Local Delivery Plan, which is produced every three years (currently 2003 – 2006) and updated annually.

The strategy puts emphasis on three outcomes and the key action points for achieving those outcomes were also been identified. Box1.3 includes the three key action points.

1.6.3 Mental Health Care of Older People Strategy

As yet there is not any formally published strategy in relation to the mental health care of older people in Camden. The framework of specialist 'Mental Health Care of Older People' (MHCOP) includes the needs of people with severe functional mental illness and of people with dementias. The MHCOP services have very close links with the 'General Adult Psychiatry Services' (GAPS) and with 'Services for Frail Older People'.

Box1.3 Key action points for Long Term Care and Support Strategy	
Key action point-1: To ensure that older people are able to live at home for as long as they can Developing extra-care sheltered housing Undertaking a review of sheltered housing Developing an intensive *home-care support* Preventing unnecessary hospital admission Improving the community meals service Reviewing frequent emergency admissions Supporting the Quality of Life strategy Supporting black and minority ethnic communities Keeping People Independent with Assistive Technology Consultation questions **Key outcome 2 : Max benefit From their stay in Hospital** Improving "*patient-centred*" care	Improving admission and discharge procedures. Providing prompt access to integrated *stroke* care services Establishing an integrated *falls* service **Key outcome 3: Access to accommodation, appropriate to their needs** Ensuring that an appropriate *mix of accommodation* is available *Increasing* the provision of general care *beds in nursing homes* Ensuring the provision of *specialist beds* for older people with *mental health* problems Providing choice in the delivery of *continuing care* *Tackling inequalities* in access to specialist palliative care Improving support for *primary care* Supporting black and *minority ethnic* communities

1.7 The factors to be considered in this study

The factors 'age' and 'gender' are two fundamental factors in health-related research and included in this study on both substantive and practical grounds (readily available). However, access to the appropriate factors that could represent the socio-economic status of a person (such as their occupational status or educational attainment etc) for every individual from available administrative records is not always possible. This difficulty will be more complex when the majority of the population under investigation are older people in their retirement. Thus, after careful consideration it was decided to use 'housing tenure' and 'council tax banding' as proxies for wealth and material circumstances. Whilst it is possible to justify the use of theses proxies on theoretical grounds it so happens that these items were the only ones readily available for extraction from the administrative data source.

It is also necessary to state that the focus of this study is on promotion of health and prevention of ill-health in old age. For this reason I have included all people aged over 50 which covers those entering old age (the 'young-old') as well as the 'old-old' who are likely to be frail and vulnerable. There is no agreed chronological definition of what ages exactly define these boundaries (Laslett, 1996) but it was felt that this population would provide enough variation in individual circumstances to capture the changing demography of ageing in Camden.

In addition to the broad socio-economic factors, discussed in the previous sections, there are some health related factors such as falls and strokes etc which directly increase the risk of ill-health and consequently the risk of hospital admissions or death. However, as noted earlier, there is a correlation between these factors and the socio-economic status of a person. To put it another way, at some stage, the above health related factors are influenced by the socio-economic factors. For the purpose of this study three popular causes of hospital admissions including falls, stroke and ischemic heart disease are chosen for further analysis.

In order to examine the effect of each of the above socio-economic and health related factors on each individual's health, and also on policy of service delivery, variables 'mortality' and 'contact with social services' were used as outcome variables.

As it has already been stated, there are some local or national service targets for older people, which are expressly designed to monitor aspects of older person's health and services. The findings of this research will help Camden's authorities to identify where to target future services to meet its priorities for the elderly. In another words it will allow older people's needs in ill-health prevention and health promotion in specific geographical locations to be more clearly identified. It will also look at whether timelier, accurate and possibly better measures can be identified which provide a better basis for service development and older people's policy.

A project of this nature requires a strong evidence base but any analysis using official statistical sources is heavily constrained by poor quality and coverage of data for micro spatial scales. This research is based on existing administrative records which are completely anonymised. Further details are provided in Chapter 3.

In the next chapter I will define the terminology used in this study in order to justify our understanding through whole study. The methodologies to be employed in this thesis research will also be discussed and defined in Chapter 2.

2 Terminology and Methodology

This chapter includes two sections: Section 1 contains the definition and description of some of the terminology used in this study and Section 2 contains a brief introduction of the methodologies employed in this research.

2.1 Predicting risk

The concept of 'risk' is central to the analysis that follows. It is therefore necessary to define what is meant by 'risk' and other relevant terminologies. It is also necessary to review the importance of identifying the risk factors and predictive modelling.

2.1.1 Risk

The Royal Society (1983) defines risk as: "...a particular adverse event which occurs during a stated period of time, or results from a particular challenge." The Royal Society continues: "...as a probability in the sense of statistical theory risk obeys all the formal laws of combining probabilities".

Risk by Holton (2004) is defined as: "...exposure to a proposition of which one is uncertain". Another definition of risk by Moreau & Jordan (2005) is: "... the likelihood of the occurrence and the magnitude of the consequences of an adverse event: a measure of the probability of harm and the severity of the impact of a hazard". Simply defined; Risk = Hazard × Exposure. Hazard also has been defined by them as "...the way in which a thing or situation can cause harm," and exposure as "The extent to which the likely recipient of the harm can be influenced by the hazard".

The analysis presented in this thesis is based on two binary outcomes; death or 'being known to social services'. The second item is also used as a predictor in models of mortality. Essentially, all models predict an overall probability of an event occurring.

Whenever this event is judged to be adverse (like death itself) the term 'risk' is adopted. Thus the main focus is the extent to which the observed or predicted risk of death is higher or lower for certain subgroups in the population. Equally, there is also a policy dimension in knowing the extent to which the overall probability of being known to social services varies by socio-economic characteristics. It is arguable as to whether being known to social services is adverse or not so I prefer to use the term probability when referring to the chances of being known to social services. A predictive factor is described as a 'risk factor' whenever the estimated odds increase the overall probability of death occurring and a 'protective factor' whenever the estimated odds decrease the overall probability of death occurring.

2.1.2 Risk factor

A useful working definition of a risk factor is provided by Mayhew (Mayhew, 2004); "A situation or an event that could increase or be associated with the probability of occurrence of an adverse event".

2.1.3 Risk assessment

Risk assessment is defined in the Encyclopedia of Public Health (Encyclopedia of Public Health, 2002) as: "A report that shows assets, vulnerabilities, likelihood of damage, estimates of the costs of recovery, summaries of possible defensive measures and their costs and estimated probable savings from better protection".

2.1.4 Predictive Modeling

Predictive modeling has been defined by Cousins et al (2002) as: "...a set of tools used to stratify a population according to its risk of nearly any outcome...ideally, patients are risk-stratified to identify opportunities for intervention before the occurrence of adverse outcomes that result in increased medical costs".

2.1.5 Why predict risk?

There is a direct relationship between predictions, planning, prevention and promotion. In order to promote quality of life and to improve the health, ultimately, we need to identify the risk factors which drive poor health and well being. Identifying risk factors is also important for the fair and efficient allocation of limited financial resources.

By identifying risk factors, we will be able to design a predictive model to assist us in planning. Axelrod & Vogel (2003) have argued that: "Over the last few years, an increasingly higher degree of interest has focused on the process of predictive modelling in healthcare. While risk assessment is embedded within most industries, post industrial revolution, the process of modelling prediction using advanced mathematical models is relatively new. Within the healthcare industry, multiple constituencies operate under the principle of risk and risk assessment".

2.1.6 The relationship between risk assessment and health promotion

Evidence shows that identifying health risk factors are an important step for the promotion of health. In other words, it was found that in order for health promotion programs to be effective, it is incumbent upon researchers to identify explicit health risk factors in order to promote health or prevent ill-health.

In a report by the World Health Organization (2001) it was argued that much of the progress in their health promotion programme has been achieved through the application of health promotion principles to specific risk factors and diseases in particular populations and settings, and the generation of an evidence base of effective practice. WHO also underlines that after 25 years of effort, community-based health promotion activities in North Karelia, Finland, have reduced age-adjusted mortality due to heart disease among men by 73% and cut 44% of all causes of mortality for men. The report also continues that over a 10-year period in California, United States of America, a comprehensive tobacco control programme has helped to prevent 33,000 heart disease deaths and reduced the incidence of lung cancer by 14%, compared to a reduction of 3% in the rest of the United States.

Standard-8 of the 2001 National Service Framework (NSF) for Older People relating to the importance of the modification of risk factors highlights: "There is a growing body of evidence to suggest that the modification of risk factors for disease even late in life can have health benefits for the individual; longer life, increased or maintained levels of functional ability, disease prevention and an improved sense of well-being." (Department of Health, 2001).

2.1.7 The importance of updating risk

The health of people changes over the course of time; while some healthy people get sicker, some un-well people become healthier and need less care. Adams (1995) states: "records of past risks are not an accurate guide to the future because people respond to risk, thereby changing it. For example, insurance companies consult past claim experiences in calculating premiums they charge to cover future risks. This in turn affects people's risk taking behaviour". Therefore, risk in general and particularly the risk related to health need to be updated more frequently.

In the following section, the methodological approach adopted in this study, will be discussed and a brief description of each of the methods will be provided.

2.2 Methodology

There were several distinct components to the research. For combining and enhancing various sources of data, database management system (DBMS) was used. A 'risk ladder' approach was utilised for the initial data analysis and following this the relative important of risk factors were assessed by logistic regression (Hosmer & Lemeshow, 2000). Finally, the results of logistic regression models were evaluated by means of Receiver Operating Characteristic (ROC) Curves.

The main reason for using risk ladder analysis in this study is its simplicity in showing the varying observed probabilities of mortality for different groups of people with similar characteristics (clusters). For the purpose of prediction and the estimation of the relative importance of each risk factor, logistic regression modeling is used. Alternative approaches include probit or linear probability models (LPM). LPM places no restrictions on the values that the independent variables (IVs) take on. They may be continuous (interval/ratio) or they may be dichotomous (dummy) variable. The dependent variable (DV), however, is assumed to be continuous. Because there are no restrictions on the IVs, the DVs must be free to range in value from negative infinity to positive infinity. The LPM predicts the probability of an event occurring, and, like other linear models, says that the effects of IVs on the probabilities are linear (Aldrich & Nelson, 1985).

Logistic regression models are based on the assumption that the categorical dependent variable reflects an underlying qualitative variable and uses the binomial distribution, whilst probit regression assumes the categorical dependent reflects an underlying quantitative variable and it uses the cumulative normal distribution. As with logistic regression, there are oprobit (ordinal probit) and mprobit (multinomial probit) options. In practice logit and probit analyses provide similar results. The preference for logistic modeling was also influenced by its dominant application in social epidemiology (Altman, 1999; Barros & Hirakata, 2003; Hosmer & Lemeshow, 2000). Both the cumulative standard normal curve used by probit as a transform and the logistic (log odds) curve used in logistic regression display an S-shaped curve. Though the probit curve is slightly steeper, differences are small. Because of its reliance on the standard normal curve, probit is not recommended when there are many cases in one tail or the other of a distribution (Pampel, 2000).

The following sub-sections provide a more detailed summary of each component of the overall methodological approach.

2.2.1 Relational Database Management System (RDBMS)

Database Management System is a collection of programs that enables one to store, modify, and extract information from a database. A Relational Database Management System (RDBMS) is a type of DBMS that stores data in the form of related tables and is based on the 'relational model' introduced by Codd (1990). Codd (1979) defines the relational model as: "...a time-varying collection of data, all of which can be accessed and updated as if they were organized as a collection of tabular time-varying tabular (nonhierarchic) relations of assorted degrees defined on a given set of simple domains".

The DBMS approach with Structured Query Language (Gennick, 1999) was used to join and integrate different data sources. Five fundamental operations in relational algebra, selection, projection, Cartesian product, union and set difference were employed to retrieve the data from different data sources (Connolly & Begg, 2001). The software packages used for data manipulation include Microsoft Excel (Microsoft Corporation, 2003b), Access (Microsoft Corporation, 2003a) and SQL (Microsoft Corporation, 2003c) . The next chapter (Chapter 3) provides a full account of data management and preparation.

2.2.2 Risk Ladder

"A risk ladder is an analytical tool to assist in the analysis of the risk or probability of an event and is based on the complete decomposition of a population according to selected risk factors" (Alder et al., 2005; Mayhew, 2004). A risk ladder in the context of this research is an exhaustive tabulation of all individuals over 50 years old in Camden according to the presence or absence of every possible combination of risk factors. The number of different possible combinations for '*N*' risk factors is equal to 2^N. For a model with 4 risk factors there are 16 possible risk categories and so the table has 16 rows in it; with 5 factors this increases to 32 and so on. Each row has an entry for the observed risk. The rows are arranged in ascending/descending order of 'risk'; thereby defining the risk ladder.

The risk ladder was used to cluster groups of the population with similar characteristics and accordingly to assess the probability of an adverse event (or risk) for each group.

2.2.3 Regression; Multiple Linear Regressions, Logistic Regression

Multiple linear regression is a technique used to estimate a statistical relationship between independent predictor variables and a dependent predictand variable (Tabachnick & Fidel, 2001).

A special case of the general model common in epidemiology is where the outcome of the dependent variable is binary (Jewell, 2004) and referred to here as binary logistic regression.

Binary Logistic Regression: Logistic regression allows one to predict a discrete binary outcome, such as group membership, from a set of variables that may be continuous, discrete, dichotomous, or a combination of any of these (Tabachnick & Fidel, 2001). In logistic regression.the dependent variable is always dichotomous, that is, the dependent variable can take the value '1' with a probability of occurring (*p*) or the value '0' with probability of not occurring (1-*p)*. Binary logistic regression is a special case of the general form of the model where all of the predictors are also binary.

Peat & Barton (2005) define binary logistic regression as: "...a mathematical method to measure the effects of binary risk factors on a binary outcome variable whilst adjusting for interrelationship between them". In other words it is primarily used to determine which binary explanatory variables independently predict a binary outcome (Wright 1995, Logistic Regression, cited in Peat & Barton, (2005). The outcome variable typically reflects the presence or absence of a condition or a disease.

In binary logistic regression, the variables that predict the probability of the outcome are measured as odds ratios. Therefore the interpretation of any fitted model relies on an understanding of odds and odds ratios. 'Odds' is simply the ratio of the probability of an event occurring (p) to the probability of its not occurring (1-p) which can be simplified to:

$$\text{Odds} = p/(1\text{-}p)$$

For example in mortality analysis, an odds of 2 (2/1) for a particular cell in a multiway table defined by a combination of factors would imply that two deaths occur for every survivor. Whereas an odds of 0.20 (1/5) would indicate that there was one recorded death for every 5 survivors. When probabilities are small, p/(1 - p) approximately equals p because 1 - p is approximately 1 (Gould, 2000).

An Odds Ratio (OR)[2] is the ratio of the <u>odds</u> of an event occurring in one group (p/1-p) to the odds of it occurring in another group (q/1-q) and can be written as:

$$\text{Odds Ratio} = \frac{p/(1-p)}{q/(1-q)} \qquad (2.1)$$

In logistic regression the outcome or dependent variable is the log (odds) and the transformation used is called logit transformation, written logit (p) (Altman, 1999) and expressed as:

$$\text{logit (p)} = \log_e\left(p/_{1-p}\right) \qquad (2.2)^3$$

Thus, the estimate of p can be written:

$$\hat{p}_i = \frac{e^u}{1+e^u} \qquad (2.3)^4$$

[2] Altman (1999, pp266-268)
[3] Altman (1999, p. 352)
[4] Tabachnick & Fidel (2001, p.518)

Where '$\hat{p}_i$' is the estimated probability of occurring for ith case (i = 1,...,n), 'e' is the base of the natural logarithm (about 2.718) and u is the usual linear regression equation

$$u = A + B_1X_1 + B_2X_2 + \ldots B_kX_k \qquad (2.4)^5$$

With constant A, coefficients B_j, and predictors, X_j for k predictors (j = 1,..., k). Thus, the linear regression equation which creates the logit or log of the odds can be written:

$$\ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = A + \sum_{j=1}^{k} B_j X_{ij} \qquad (2.5)^6$$

Hence in the absence of any of the factors 1 to k, the equation (2.5) becomes:

$$\ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = A$$

So, the odds is:

$$= \hat{p}/(1-\hat{p}) = \exp(A)$$

And the estimated probability / risk is:

$$\hat{p} = \exp(A)/(1+\exp(A))$$

(Multiply this by 100 and you get the predicted % risk for those without any risk factors present in the risk ladder which can be compared with the observed risk).

---

[5] Tabachnick & Fidel (2001, p.518)
[6] Tabachnick & Fidel (2001, p.518)

In general the predicted odds (probabilities) can be estimated in this manner for any combination of risk factors. Using equation 2.5 above the exponent associated with each risk factor ($\exp(B_j)$) represents a multiplicative factor that increases or decreases the odds of an event occurring. For example if risk factor X_1 were the only risk factor present the predicted odds would be $\exp^{(A+B_1)}$ written as a product: $\exp^{A} . \exp^{B_1}$.

Gould (2000) interprets the exponentiated coefficient in logistic regression as:

$$\exp(B_j) = \frac{odds(if\ the\ corresponding\ variable\ is\ incremented\ by\ 1)}{odds(if\ variable\ not\ incremented)}$$

Or, equivalently,

$$\exp(B_j) = \frac{(p(event \mid X_j + 1) / (1 - p(event \mid X_j + 1))}{(p(event \mid X_j) / (1 - p(event \mid X_j))} \qquad (2.6)^7$$

In other words, $\exp(B_j)$ is the ratio of odds for two groups where each group has a values of X_j and is one unit apart from the values of X_j in the other group (e.g. 'X_j' and 'X_j+1'). $\text{Exp}(B_j) > 1$ means the independent variable increases the logit and therefore increases odds(event). If $\exp(B_j) = 1$, the independent variable has no effect and if $\exp(B_j) < 1$, then the independent variable decreases the logit and therefore decreases the odds(event)(Hosmer & Lemeshow, 2000) .[8]

2.2.4 Receiver Operating Characteristic (ROC) Curve

ROC curves are used as a tool to evaluate the results of the prediction by logistic regression. A ROC curve can be represented equivalently by plotting the fraction of true positives (TP) or 'sensitivity' versus the fraction of false positives (FP) or '1-Specificity'.

---

[7] Gould (2000, p.20)

[8] 2.6 has generality for binary predictors and/or continuous variables which incremented by whole units. A categorical variable would therefore be represented by a set of 0/1 dummy variables.

For example, in a health setting, sensitivity refers to the people with disease who have a positive test result (True Positive or TP) and specificity refers to the people without disease who have a negative test result (True Negative or TN). Subsequently '1-Specificity' refers to the people without disease who have a positive test result (FP). In short, sensitivity indicates how likely the outcome of a test will be positive for actual positive cases and specificity indicates how likely the outcome of a test will be negative for actual negative cases (Peat & Barton, 2005). Detailed information on usefulness of the ROC curves will be discussed in Chapter 6.

2.3 Ethnicity

A key risk factor in the study of health inequalities is 'ethnicity' (Ward, 2003; Dressler et al., 2005; Pearce et al., 2004; Carter-Pokras et al., 2004). However, it was not possible to include this factor with any confidence as the available administrative data source only contained an ethnicity code for less than one third of the study population. Table 2.1 below shows a comparison of the marginal distribution for ethnicity using the 2003 mid-year estimates of population by ONS (Office for National Statistics, 2004b) for Camden older citizens with the percentage of those with a recorded ethnicity in this study. Using the 2003 mid-year estimates of population as a benchmark gives a chi-square of 9248, 4 df ($p<.001$), signifying that the sample (those with a recorded ethnicity in this study) is unlikely to have been selected randomly from the population. That is, reporting of ethnicity is non-random in the population of Camden residents (White & Mixed people were less likely to have their ethnicity recorded than the others).

Table 2.1 Comparison of percentage of ethnic groups living in Camden for 2003 ONS mid-year estimates of population (aged 50 years and older) by ONS to those with a recorded ethnicity in the administrative data available to this study

Ethnic-Group	Mid Year 2003 Estimates of Population	Those with a recorded ethnicity in this study
White	86.50%	66.55%
Mixed	1.30%	0.07%
Asian or Asian British	5.40%	17.63%
Black or Black British	4.00%	10.93%
Chinese or Other Eth-G	2.80%	4.82%

The interpretation of the health data by race and ethnic group, can also very much depend on the quality of data and the meaning assigned to the terms 'race' and 'ethnicity'. The quality of the data on race and ethnicity also depends greatly on how it has been collected and compiled, and there is often variation between organisations and departments.

In addition, the concept of race and ethnicity has developed over time (Washington State Department of Health, 2004). Debate on the notion of race and ethnicity is an on going subject and the concepts are continually being updated and re-defined owing to a number of factors (Bhopal, 1997). From the biological viewpoint, there are some arguments which suggest that races are not biologically distinct (Kuper, 1975). As Bhopal (1997) also states: "...the physical characteristics distinguishing races result from a small number of genes that do not relate closely to either behaviours or disease".

From the socio-economical perspective, Senior & Bhopal (1994) state that "...ethnicity is a fluid concept and depends on context. At Ellis Island millions of Europeans swapped European identities for American ones... ethnicity is not measurable with accuracy or validity".

There have always been difficulties in using the collected ethnicity data. In the 1991 UK census the question on ethnicity was responded to by those people who were willing to answer it, and the classification was arbitrary (Senior & Bhopal, 1994). The constraining factor in relation to the limited use of the collected data by the Department of Health in the UK by Jacobson & Aspinall (2006) are defined as the 'poor quality of much of this data', 'concerns over low rates of completeness' and the use of 'non standard classifications and questionable methods of ascertainment of ethnic group'.

At the beginning of this study it was decided to use the variable 'ethnicity' as one of the factors in the process of analysis. In the first instance it was attempted to find an ethnicity match for each of the 3188 records in the mortality list. In order to do this several data sources including 'hospital admissions', 'social services' and 'school pupil roll' were used to find an ethnicity match for each individual. For approximately 50% of the records (around 1600 records) ethnicity was present. For the remaining 50% the information including place of birth, surname, first name or a combination of two or more of these entities were used to extract an ethnicity (Lauderdalei & Kestenbaum, 2000; Mateos, 2007; Research and Development DH & NHS, 1998). This process took approximately two weeks but the allocated values could still not said to represent every individual's ethnic group correctly. Appendix-B includes the steps that were taken to find/allocate an ethnicity to each record in the mortality list. However, using this exhaustive and manual approach is time consuming and it was not possible to do it for whole population in this study.

In the process of finding an ethnicity for each record some problems were discovered. These problems included a large number of cases where the place of birth, the first name and surname did not match with the ethnicity extracted from administrative data sources like hospital admissions, social services and schools. For example, in some cases the name and surname of a person is South Asian in origin and the place of birth is listed as Bangladesh but ethnicity were recorded 'White'. There were also a considerable number of cases where the place of birth was recorded as 'Ireland' or 'Eastern European countries' with Irish or Eastern European names but ethnicity was listed as 'White English'.

Given the above problems, it was decided not to continue attempting to match an ethnicity for all of the 43472 records of population. Therefore the variable ethnicity as a parameter for further analysis had to be excluded from this study.

It is now important to continue with a detailed account of the stages involved in the data management and preparation.

3 Data management and preparation

3.1 Introduction

Adopting the right policy by any government is highly dependant on the periodic updating of information, accuracy of the information and cost of this information. It is also recognized that providing high quality health services requires comprehensive information to be collected on the health status of the population. Data sharing between health and local authorities is necessary to ensure comprehensive information is at hand (Mayhew & Harper, 2006; Raine et al., 2006). Deficiencies in the quality of this information as Raine et al. (2006) state: "...making it impossible to track, at local level, trends in major risk factors and in patterns of diseases". Therefore, the Statistics Commission welcomed the announcement in the Queen's Speech 2006, of plans for legislation governing UK official statistics, which was followed by Commission Chairman Professor David Rhind on 'enhancing the ONS's access to administrative data from across the public sector for valid statistical purposes' (Statistics Commission, 2006).

Owing to the lack of a Relational Database Management System (RDBMS) between administrative data available for this analysis, it is necessary to consider ways of combining information from several different data sources. The data sources used in this research were originally created to meet the specific needs of the particular agencies/bodies that were involved in their design. These data sources have varied formats and styles, which makes it difficult to analyse and compare the data sets.

In order to make best use of these data sources it is essential that comprehensive queries are undertaken to extract the necessary risk and predictive factors enabling them to be integrated into single comprehensive matrix. This matrix will ensure that all factors are included in one single source and will also help facilitate the import and export of data in different formats including Microsoft Excel (Microsoft Corporation, 2003b), Microsoft Access (Microsoft Corporation, 2003a), Microsoft SQL (Microsoft

Corporation, 2003c), Software Package for Social Science (SPSS, 2003) and Statistical Software for Professional, Stata (StataCorp, 2005).

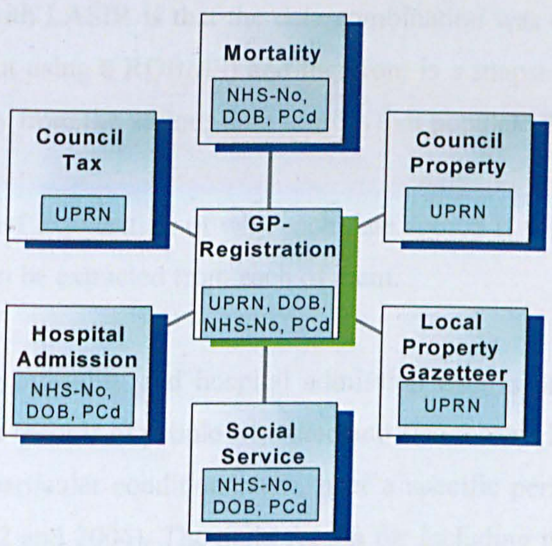
The following sections provide more detail the key data sources used in this research and outline the processes of data preparation.

3.2 Data Sources

Altogether seven distinct data sources were used for this research and these are shown in Figure 3.1. Specifically the boxes represent seven distinct sources:

1. Mortality data for three years (2002-2004)
2. Hospital admission records (2002-2004)
3. GP-Registration (October-2003)
4. Council tax (2003)
5. Council property (2003)
6. Local Property Gazetteer (LPG)
7. Social service data (2002- 2004)

Figure 3.1 Data sources with unique identifier(s) in each source.



Legend	
Variable Name	**Definition**
UPRN	Unique Property Reference Number
NHS-No	National Health Service Number
DOB	Date of Birth
PCd	Postcode

The data sources are integrated for the purpose of identifying risk factors. Also represented in figure 3.1 are the key variables (contained in the small internal boxes). These variables either individually or in conjunction with other variables form the 'primary key' (unique identifier of each record). Consequently the primary key has been defined in database design as a value that uniquely identifies each row (record) in a table (Date, 2000).

Some of the data sources included in Figure 3.1 utilized Camden Local Area Shared Information Resource (LASIR) which was introduced in April 2003 and includes a wide range of data sources, including; police, health and council data. The purpose of LASIR is to provide the means to join up information from different partners in order to enable the research to focus on local areas and neighbourhoods and therefore target resources more effectively (Mayhew Associates Ltd & Camden PCT, 2005). With LASIR the single records from each table have been matched to the Local Property Gazetteer (LPG), which includes the list of all properties in the Borough with a Unique Property Reference Number (UPRN). It should also be noted that one of the problematic issues with LASIR is that the data combination was completed manually (which means without using a RDBMS) and therefore is a snapshot which can not be updated automatically from the various data sources that populate the database.

What follows is a brief explanation of why each data source is useful and the type of information which can be extracted from each of them.

The reason for using mortality and hospital admission data is an obvious one. Put simply, we require the records of people who died and also those who were admitted to hospital owing to a particular condition/cause, over a specific period of time (in this instance between 2002 and 2004). The main reason for including the GP-register data is that the GP register provides a very extensive list of almost all of the people living in the Borough and has been used to extract the target population for this work. Also included in the data sources is 'Council tax' and 'Council property' which have been utilised for the extraction of council tax band and housing tenure for each individual in the target population.

In addition, the Local Property Gazetteer (LPG) includes the appropriate address of all properties in the Borough, including a UPRN. UPRN represents the same address in each different data sources. The UPRN provides a link between an individual to several data sources, in order to extract different variables related to the same individual. In other words UPRN by itself is an appropriate primary key.

The records from LB Camden's social services data have also been used in this study. The rationale behind using the social services Department data is to examine the relationship between the services provided by the department of Social Services and the mortality rates of older people in Camden. In other words, it has been used to test whether being at higher risk of ill-health and consequently mortality will increase the chance of being in contact with social services or not. It has also been used as a complementary source for extraction of the ethnicity codes for some records in order to undertake further analysis. It needs to be declared that the social services data will be limited in scope since it will only contain individual records for persons known to social services.

3.3 Data preparation

Data preparation is an important and critical preliminary step in any data analysis. In order to maximise the coverage of the information, it is very important to adopt procedures which maximise the accuracy of individual records. It is very difficult or sometimes impossible to go back to manipulate the data once records are extracted or merged. Therefore it is a time consuming stage of the work. The data preparation in this research involves five critical steps.

In the first step (data collection), the individual data sources are collected from different databases. The second step involves the 'data cleaning' in order to extract the appropriate records or variables, required for the research. In step three various data was integrated. In this stage the data was combined together via appropriate queries to create a comprehensive source. Step four involves the variable creation which the target variables are either extracted from the existing data sources or have been derived by manipulating the available data. A simple example of this would be the extraction of age from date of birth and the date of data entry. During this stage categorical or dichotomous variables were created too. Stage five focuses upon transferring the created data source to a software package such as 'SPSS' (SPSS, 2003) or 'Stata' (StataCorp, 2005) for statistical analysis.

To predict all required variables for the later stages of the research is not always possible. Therefore there are various examples of returning to the previous stages of data preparation to repeat some steps in order to retrieve a new variable.

Before going through the detailed explanation of the process of data preparation for each data source in the following sections, it is necessary to declare that on balance it was decided to focus on the non-institutionalised population of Camden aged 50 years and above given that the broad policy aim of the project was to increase the chance of this group remaining independent and healthy in their homes. Therefore any institutionalised people in this age group were excluded from the list of population. A more comprehensive explanation for the exclusion of institutionalized people from the analysis can be found in Section 3.4.

3.3.1 Mortality data

The mortality data includes information such as: person's name, date of birth, date of death, extracted age (at the time of death), gender, occupation, address, causes of death and many other attributes that are not relevant at this stage of the work.

In order to find the relation between mortality and socio-economic status of a person, the tax band and tenure have been used as a proxy for 'wealth' (Mayhew Associates Ltd & Camden PCT, 2005). The starting point of tax bands is the valuation of the property: "The basis of valuation for a dwelling...is the amount which, subject to certain assumptions, it would have sold for on the 'open market' by a 'willing vendor' on 1 April 1991" (Valuation Office Agency, 2006). Further information on tax bands for England and Wales can be found in Appendix-C.

Therefore, the council tax and council property data sources which include the tax band and tenure, have been linked to the mortality records by using the UPRN as a primary key (unique identifier), in order to extract more comprehensive information.

At every stage it was important to maximise the number of records. This has obvious benefit for the precision of any estimates. After careful consideration of the quality of the available mortality data in the borough of Camden, three years mortality records (2002-04) were selected for further analysis.

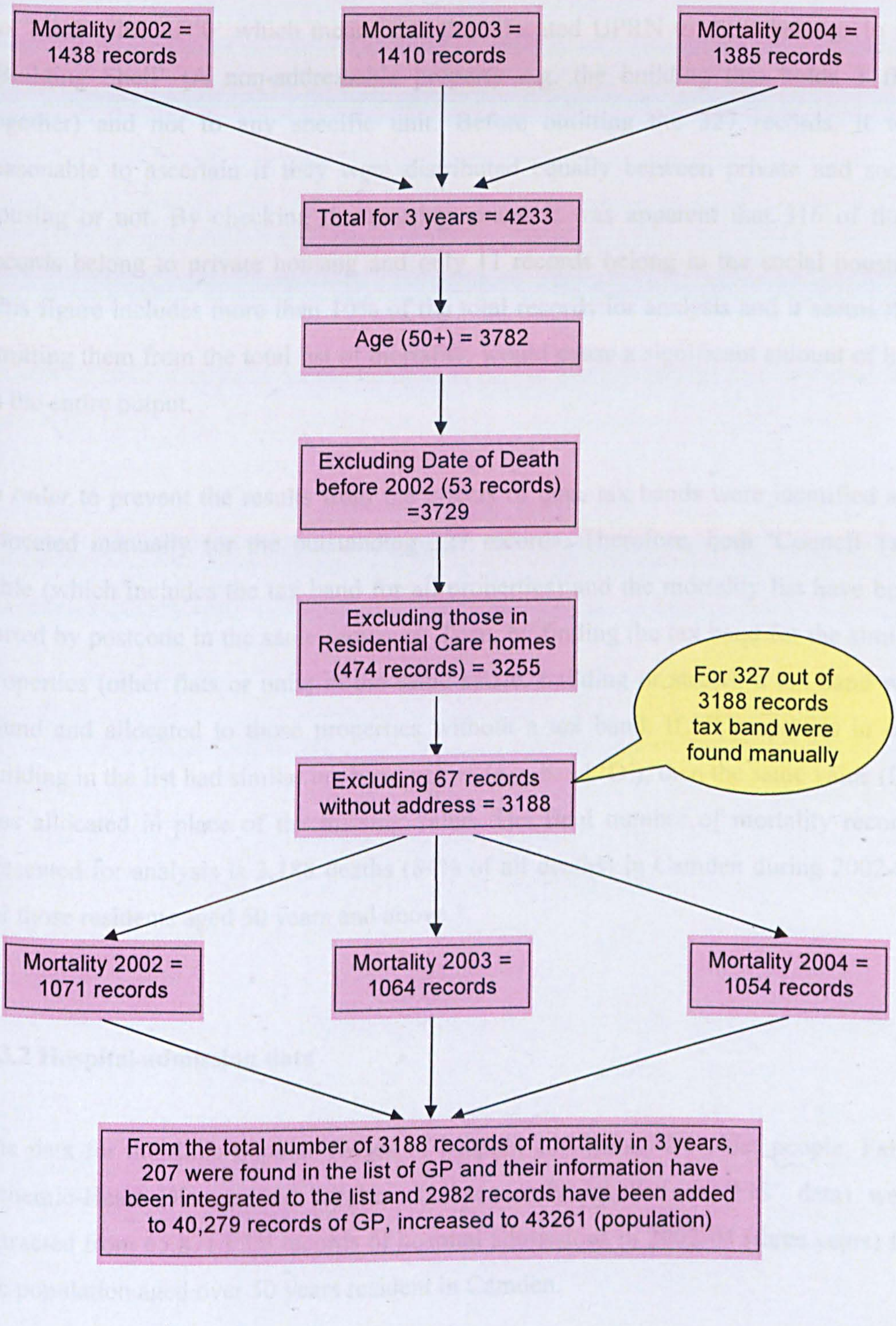
Preparing Mortality Data

The process of mortality data preparation is shown in Figure 3.2. The following steps are the detailed explanation of the process of mortality data preparation.

i) Integrating the 3 years mortality data: the total numbers of deaths for the period of 2002-2004 were 4,223 and the total number of deaths for people age over 50 was 3782.

By excluding those who died in residential care homes the number of records reduces to 3,255.

Figure 3.2 Mortality data preparation flowchart.



ii) Allocating a UPRN to each record by matching with LPG: there are 67 records that neither have an address in the address field nor a UPRN. The total number of records with a UPRN is 3,188.

iii) Council Tax Band: 327 records out of the total 3,188 records with a UPRN had a tax band value of '0' which means that the allocated UPRN to them belongs to the 'Building Shell' (A non-addressable property e.g. the building that holds 3 flats together) and not to any specific unit. Before omitting the 327 records, it was reasonable to ascertain if they were distributed equally between private and social housing or not. By checking the housing tenure it was apparent that 316 of these records belong to private housing and only 11 records belong to the social housing. This figure includes more than 10% of the total records for analysis and it seems that omitting them from the total list of mortality, would cause a significant amount of bias in the entire output.

In order to prevent the results from the effects of bias, tax bands were identified and allocated manually for the outstanding 327 records. Therefore, both 'Council Tax' table (which includes the tax band for all properties) and the mortality list have been sorted by postcode in the same sequence. Then, by finding the tax band for the similar properties (other flats or units in the same estate, building or street), a tax band was found and allocated to those properties without a tax band. If all properties in one building in the list had similar tax band values (e.g. band 'D'), then the same value (D), was allocated in place of the missing value. The final number of mortality records presented for analysis is 3,188 deaths (84% of all deaths) in Camden during 2002-04 for those residents aged 50 years and above.

3.3.2 Hospital admission data

The data for the three popular causes of hospital admissions for older people; Falls, Ischemic-Heart Disease and Strokes, (subsequently labelled as 'FIS' data) were extracted from 65,871 total records of hospital admissions in 2002-04 (three years) for the population aged over 50 years resident in Camden.

The records of the three causes of hospital admission (FIS) were extracted based on the codes of the 'Diagnosis' columns such as 'Prim (ary) Diagnosis', 'Sub-Diagnosis' or 'Second-Diagnosis' (columns 1 to 5). All hospital admission data that were used were those coded by ICD-10 (International Statistical Classification of Diseases and Related Health Problems-Tenth Revision). The outputs for the three causes of hospital admission were; 2,060 records of falls, 1,781 records of strokes and 2,475 records of ischemic heart disease (6316 in total).

Preparing hospital admission data

For the purpose of further analysis based on tenure and tax bands, a primary key such as UPRN, NHS-number or a combination of two or more variables (e.g. postcode and date of birth) is required. Therefore we need to look for some other attributes that can be found in other administrative data sources with UPRN, to use as a bridge between Hospital Admission and Council Tax or property tables (e.g. GP-Registration, Mortality etc). All three data sources (hospital admission records, GP-Register and Mortality records) have a NHS-Number. Therefore person's NHS number is the best attribute for the above purpose.

However, in some records in the hospital admission data, the NHS-Number is missing but the person's name and address is present. These records are more likely to be found in places such as a residential care home or hospital, which are excluded from this analysis. There are 1,097 records without NHS-Numbers from the total number of 6,316 FIS records and the reminder of the records (5,219), have a NHS-Number.

There is large number of cases where one person was admitted to the hospital on several occasions. Therefore by using a 'Distinct' query, all repeated records have been omitted and for each person there is only one record left. A column for entering the number of hospital admissions for each cause of hospital admission has been added in the table. Thus, the 5,219 admission records were reduced to 2,847 individuals (unique records).

In the next stage of work these records were cross checked with the main (population) list for matching (to find if the person in the hospital list is also included in the population list or not). For 1,846 out of 2,847 individuals in the FIS list, a match was found with the NHS-number and for 182 records a match was found by combination of 'date of birth' and 'postcode'. 819 records (individuals) were discarded because no match of any kind could be found for them in the population. The lack of sufficient information related to their addresses, which could have helped extracting other factors for them, also precluded their inclusion in the population. 1,097 records without a NHS-Number were cross checked with the population by using combination of 'DOB' and 'Postcode'. For 702 out of 1,097, neither any match was found nor did they have enough information to be added to the population list independently. The reminder of the records (395 records belong to 225 individuals) did not have any match in the population but they could be added to the population independently. Therefore after the distinction process they were added to the target population.

The percentage and distribution of the 2,399 individuals and causes are as follows:

From the total of 2,060 records of falls, a match was found for 1,266 records (61%). These 1,266 records belong to 885 distinct individuals.

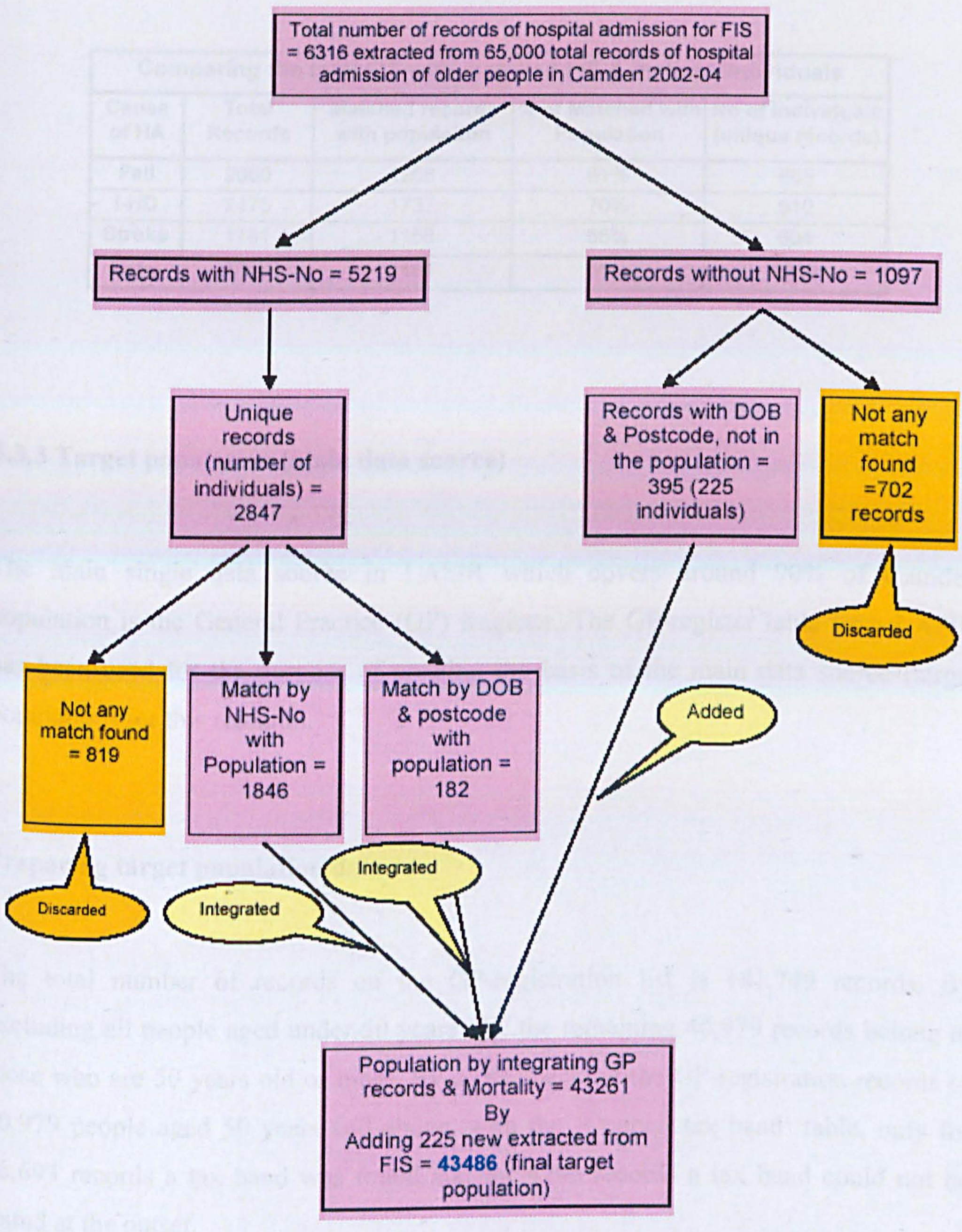
From the total of 2,475 records of ischemic heart disease, a match was found for 1,737 records (70%). The 1,737 records belong to 910 distinct individuals

From the total of 1,781 records of stroke, a match was found for 1,158 records (65%). The 1,158 records belong to 604 distinct individuals

By adding all causes of hospital admission, 66% of records have been integrated into the target population.

Finally, from the total of 6,316 records of hospital admissions as a result of the 3 causes (FIS) for 4,161 a match was found. These 4,161 matched records belong to 2,253 (1,846 + 182 + 225) individuals. Figure 3.3 illustrates the process of hospital admission data preparation.

Figure 3.3 process of data preparation for hospital admissions as a result of three causes



From the total of 6,316 records of hospital admissions as a result of the 3 causes (FIS) for 4,161 (66%) a match was found from the records of Camden population aged 50 years and above. Table 3.1 shows the percentage of matched records and the total number of individuals for each cause of hospital admission

Table 3.1 Matching 2002-04 hospital admission with population records for those aged over 50 years in Camden

Comparing the total FIS with matched FIS & distinct individuals				
Cause of HA	**Total Records**	**Matched records with population**	**% of Matched with Population**	**No of Individuals (unique records)**
Fall	2060	1266	61%	885
I-HD	2475	1737	70%	910
Stroke	1781	1158	65%	604
Total	**6316**	**4161**	**66%**	**2399**

3.3.3 Target population (main data source)

The main single data source in LASIR which covers around 90% of Camden population is the General Practice (GP) Register. The GP-register table from LASIR has been used for the purpose of creating the basis of the main data source (target population) for this research.

Preparing target population data

The total number of records on the GP-registration list is 181,749 records. By excluding all people aged under 50 years old, the remaining 40,979 records belong to those who are 50 years old or more. By cross checking the GP-registration records of 40,979 people aged 50 years and above, with the 'Council tax band' table, only for 36,693 records a tax band was found and for 4286 records a tax band could not be found at the outset.

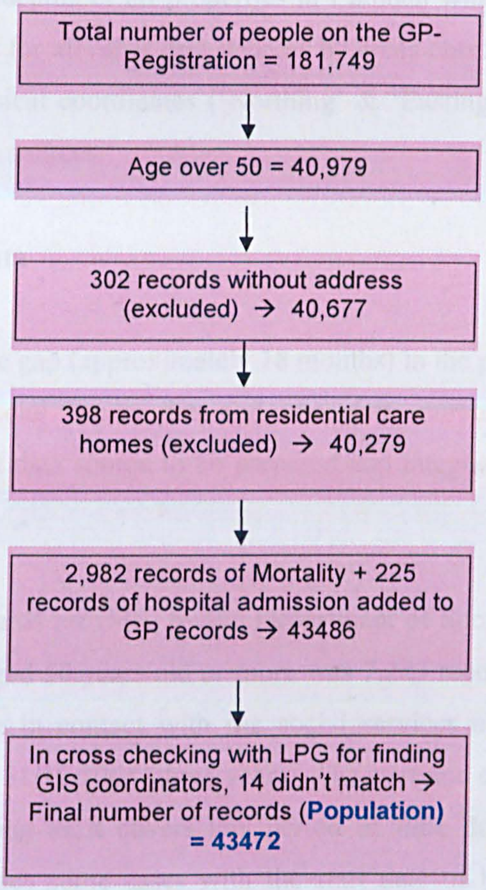
In order to extract a tax band for the above 4286 records, a similar technique to the one used for the mortality list, was employed. A tax band could not be found at all for 302 records because the address was un-identifiable. There were also 398 records for those living in residential care homes. These two groups (302 + 398 = 700 records) were excluded from the 4286 records and for the reminder of the records (3586 records), a tax band was found. The final number of records for people aged 50 years old or more with a council tax band value and not living in a residential care homes turned into: 36,693 + 3586 = 40,279 records.

The reason for excluding those records without a tax band is because tax bands will operate as a 'proxy for wealth' (as mentioned earlier). In the case of those people residing in residential care the tax band is misleading as the bands tend to be allocated to the institution. Therefore, the tax band in institutions does not represent the resident's socio-economic status.

There is still a noticeable difference between this figure and the estimated population of 50 years old or more in Camden for mid-2003 by the Office for National Statistics (2004b) which is approximately 47,700. To reduce the gap the records of three years mortality data (2002-04) were integrated into the above records (the duplicate records were omitted) and the total number of records increased to 43,261. The population records were increased to 43,486 by adding a further 225 unique records of hospital admission data (not duplicate by mortality or GP registration) for three years (2002-04).

For the purpose of mapping the outcome in the later stage, we need to extract the geographical coordinators data for all records. Finally by cross checking the 43,486 records with Camden Local Property Gazetteer (LPG) for allocating a 'North' and 'East' geographical coordinator to each record, the number of records decreased to 43,472 records. This figure is the final number of the target population, and has been used for all stages of this research. Figure 3.4 shows an illustration of the process of data preparation of the population aged 50 years and above in Camden.

Figure 3.4 process of data preparation of the population over 50 years old in Camden using the GP-registration list



3.3.4 Council tax

The Council tax table includes more than 95,000 records belonging to all residential properties with complete address, postcode, tax band and UPRN. This table has been used to extract a tax band for the records in population, mortality and hospital admission tables. Once the UPRN was allocated, linking them to the council tax table was achieved without any problems.

3.3.5 Council property

This table includes more than 56,000 properties. It has been used to distinguish the council property from other type of properties.

3.3.6 Local Property Gazetteer (LPG)

This table includes the details of all properties in Camden with a UPRN. It has been used to extract a UPRN for all other data sources by cross checking using address and postcode. The geographical coordinates ('Northing' & 'Easting') were also extracted from LPG for other data sources.

3.3.7 Social services Data

There was a considerable gap (approximately 18 months) in the period of time between gaining access to the Social Service data and other data sources, as described above. Therefore it was the last data source to be prepared and integrated into the main data source (population).

The total number of records provided by the Department of Social Services in the LB of Camden for people aged 50 years old or more was 7,223 records. This list includes everyone who had been in contact with the social services at any period of time between 01/01/2002 and 31/12/2004 (three years). The start and end date of the service could be any date as long as it covers any period of time during the three years. Therefore this list includes some cases with the start date of being in contact with social services as early as 1970 and many cases who were in contact by the time the list was provided for this study (31st of March 2006).

Preparing social services data

In order to integrate the social services data with the population data for further analysis, the following steps were taken:
i) The total list of records was filtered by 'Date of Birth' and 'Postcode' in order to identify any duplicate records. After filtering the records by 'DOB' & 'Postcode', 7,188 records were returned.

ii) The postcodes of all records extracted in step '*i*' were cross checked with the Camden postcodes. The postcodes for around 900 records of social services could not be matched with the Camden postcodes, (either belonging to other boroughs or the postcode was wrongly entered), and as a result these records were dropped. Therefore the total number of records with a Camden postcode was reduced to 6,280.

iii) The 6,280 records with a Camden postcode were cross checked with the list of the entire population of Camden aged 50 years and above (43,472) in several steps, as follows:

a) First it was cross checked across four variables: First Name, Surname, DOB and Postcodes. It returned 2,439 matched records of individuals.

b) After excluding 2,439 matched records extracted in the previous stage from the total 7188 records of Social Service, the reminder of the records (4,749 records) were cross checked with the population list, but this time with three variables: Surname, DOB and Postcode. The number of the matched records at this stage was 573 records.

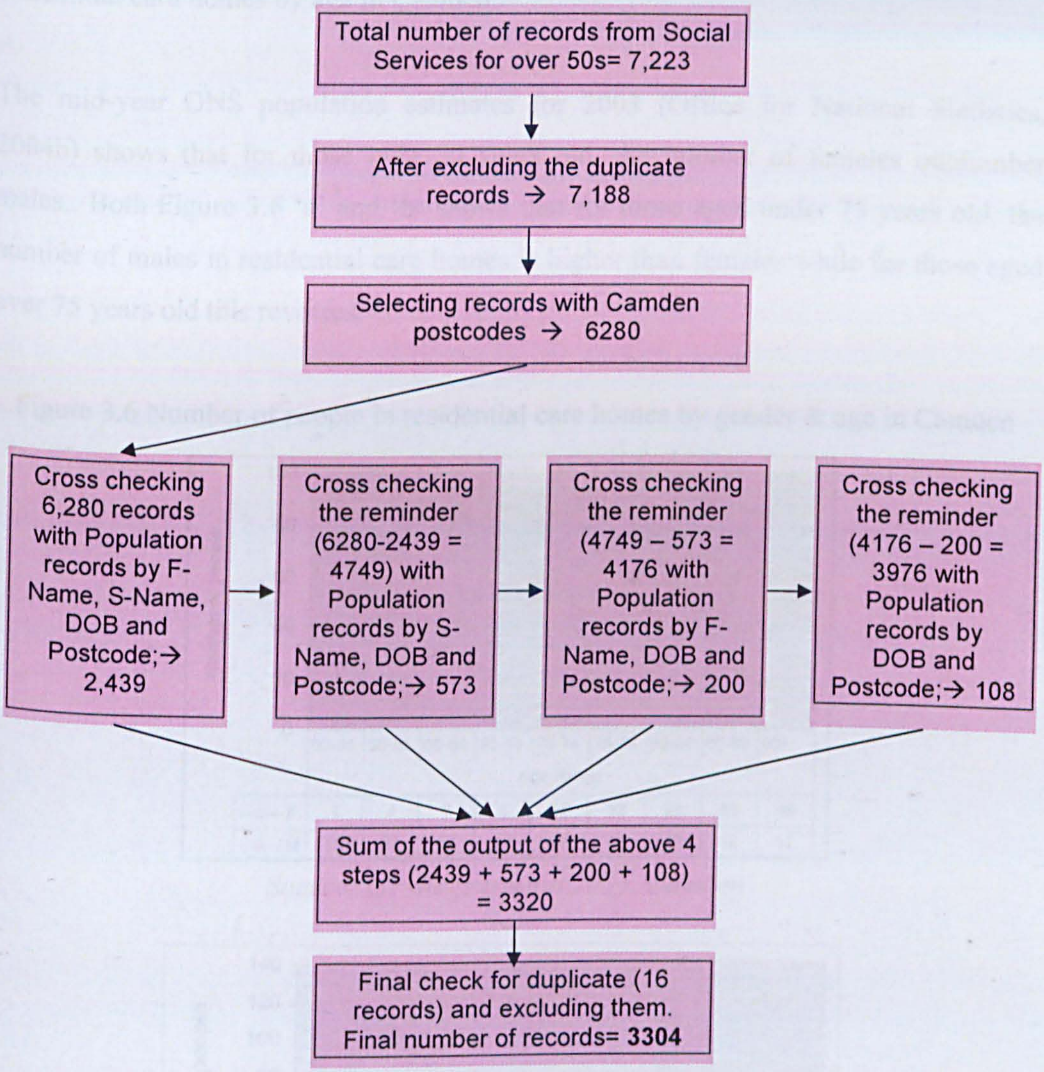
c) The 573 records extracted in '3*b*' were excluded from the above 4,749 records and the reminder of the social services records (4,176 records) were again cross checked with the population records by three factors: First Name, DOB and Postcode. The output of this stage was 200 more matched records.

d) The above procedures were repeated once again with 3,976 records of social services and the population by using two factors: DOB and Postcode only. This operation returned 108 matches.

iv) The output of four stages in '3*c*' were added together (providing 3,320 records). These records were checked for duplication once again and 16 duplicate records were found and removed from the list. The final list of matched records of social services and population contains 3,304 cases.

After the completion the above four stages (*i-iv*) a column representing the variable *'In contact with social services'* for the final list of the matched records (3,304 individuals) was added to the population list. A number '1' for those who had been in contact with the social services and a '0' for those who had not been in contact with social services were entered into the column. The process of Social Service data preparation is illustrated in Figure 3. 5.

Figure 3.5 process of data preparation for social services data



The decision to focus on the non-institutionalised population of Camden aged 50 years and above was as a direct consequence of the broad policy aim of the project which was to increase the chance of this group remaining independent and able to stay in their homes. Any institutionalised people in this age group were excluded from the list of the population. A more comprehensive explanation for the exclusion of institutionalized people from the analysis will now be found in Section 3.4.

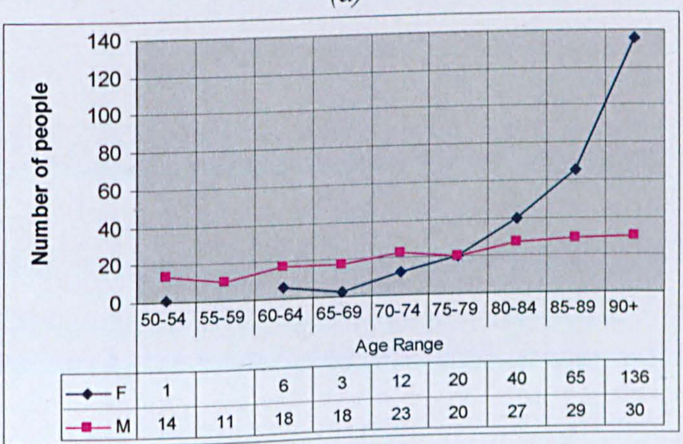
3.4 Population coverage

Excluding the institutionalised people from the entire list of the population could be a source of bias if the analyst is interested in making broad general inferences for all people aged over 5o years. Figure 3.6 shows the number of males and females in residential care homes by age in Camden.

The mid-year ONS population estimates for 2003 (Office for National Statistics, 2004b) shows that for those over 50 years old, the number of females outnumber males.. Both Figure 3.6 'a' and 'b' shows that for those aged under 75 years old, the number of males in residential care homes is higher than females while for those aged over 75 years old this reverses.

Figure 3.6 Number of people in residential care homes by gender & age in Camden

Source: GP-Registration 2003, Camden
(a)



Source: Mortality 2002-04, Camden
(b)

Grundy & Sloggett (2003) state that excluding institutionalised persons from sample may cause a bias. They add "...single persons, persons with poor health and women are more likely to be in a nursing home and thus they are more likely to be underrepresented in the sample". This bias was also assessed by Huisman, Kunst & Ackenbach (2003). Their finding shows that excluding institutionalized persons from sample will lead to underestimation of socioeconomic health differences in older ages.

However, in this analysis presented in this thesis I am attempting to identify opportunities to review social service intervention and practice in order to prevent admission to residential care and thereby enabling people to live independently in their own homes. Therefore the institutional population is not considered to be 'in scope' for the purposes of this study.

The analysis of the resulting data will now be discussed in part-II (Chapters 4, 5 and 6).

PART II

ANALYSIS,

EVALUATION & FINDINGS

4 Findings using risk ladders

4.1 Introduction

A definition of the risk ladder was provided in Chapter-2. in which it was shown that the observed risk/probability of death, p_i for a particular combination, 'i 'can be calculated by:

$$p_i = {x_i}/{n_i} \tag{4.1}$$

Where x_i and n_i are respectively the reported number of deaths and the entire number of individuals in the combination 'i', with a standard error given by:

$$SE = \sqrt{p_i(1-p_i)/n_i} \tag{4.2}$$

And the confidence interval for a 95% level of accuracy, assuming the normal approximation for binary outcome (Altman, 1999), can be calculated by:

$$CI = p_i \pm \left(1.96\sqrt{p_i(1-p_i)/n_i}\right) \tag{4.3}$$

In this chapter the analysis of data with risk ladder methodology will be illustrated. In Section 2 the observed risk of mortality using seven factors extracted from Camden's administrative data sources will be discussed first. In Section-3, the outcome of risk ladder analysis in Section-2 will be illustrated for various maps of Camden. Section-4 will follow the same procedure as Section-2 by changing the outcome variable from 'mortality' to 'being in contact with social services'. To assess the appropriateness of the targeted population for service delivery, in Section-5 the risk of mortality for different groups (provided in Section-2) will be compared with the probability/chance of being in contact with social services (provided in Section-4), for the same group. Finally in Section-6 some statistical implications in relation to the interval estimation for the binomial proportions will be discussed.

In this chapter four risk ladder analyses for each outcome variable (Mortality and Social services) will be reported. Each risk ladder for different outcome variables will includes the following factors:

- The four basic socio- demographic factors
- The four basic socio- demographic factors together with the incidence of Falls
- The four basic socio- demographic factors together with the incidence of Ischemic Heart-Disease
- The four basic socio- demographic factors together with the incidence of Stroke

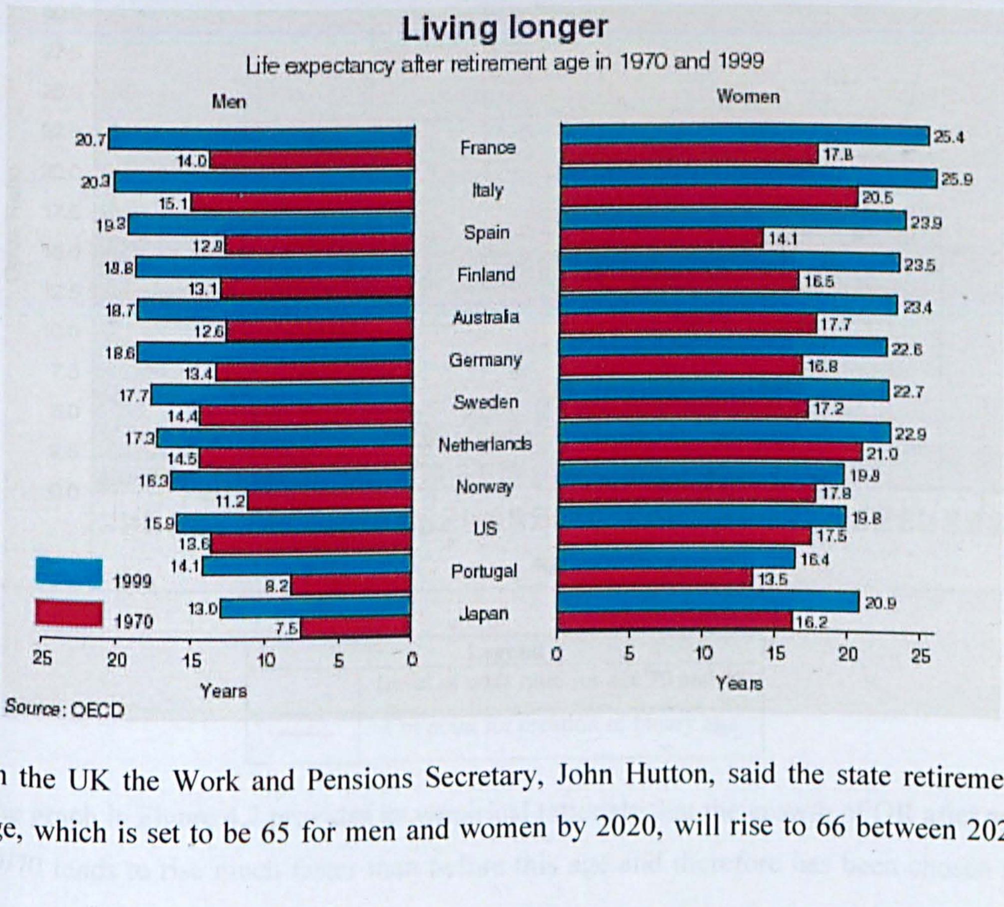
For outcome variable 'Mortality', a fifth risk ladder with seven factors will also be reported. This includes 4 basic socio- demographic factors together with the incidence of all three causes of hospital admission (falls, heart disease and strokes).

The first risk ladder is based on four socio-demographic factors; gender, age, housing tenure and council tax bands for which there are 16 (2^4) different risk factor combinations. In the subsequent risk ladder applications one of the causes of hospital admission; falls, strokes and ischemic heart disease, were added one at a time to the four basic factors, thus producing 32 combinations for each risk ladder. The number of possible combinations in the fifth risk ladder is 128 (2^7). In practice some of these combinations do not include any records, or include only a few records, and these combinations have been omitted leaving a total of 63 combinations.

For the sake of clarity the initial exposition of risk ladder analysis it was decided to use binary data for all variables. This will be relaxed in subsequent sections. Where the likelihood of the occurrence and the magnitude of the probability of an adverse event are typically lower the variable is coded '0', otherwise '1'. Thus applying the above rule, gender has been coded to female = 0 and male = 1, age to '0' if is equal or greater than 50 and less than 70 ($50\geq$ age <70) and '1' if is greater than 69 years old. Housing tenure has been coded as owner/private rented = 0 and social housing (council housing or housing association) = 1. If an address is rated as council tax band D-H (higher bands), it has been coded '0' and '1' for A-C (lower bands). Those admitted to hospital at least once for Falls, Stroke or Heart disease, are coded '1', else '0'. Those who are known to the social services are coded '1' otherwise, '0'.

Clearly when using a binary division for age there could be some arbitrariness. Typically researchers have used state retirement age as a boundary (Manhapra et al., 2004; Svendsen et al., 2004; Sjosten et al., 2007). However in recent years the changing nature of retirement coupled with increasing retirement age can be traced to increases in life expectancy. Leibfritz (2008), [Department of Economics, The Organisation for Economic Co-operation and Development (OECD)] states "In the next 50 years, low fertility rates and rising life expectancy in OECD countries will cause this old-age dependency (retirees depending on the funding of those in work for their income) rate to roughly double in size". He also continues that Life expectancy at the average effective retirement age can be as high as 18-20 years; about a third longer than it was 30 years ago. It is projected to increase further and therefore the retirement period will lengthen unless retirement itself is delayed. Effective retirement age also automatically adjusts with rising life expectancy. Figure 4.1 compares the life expectancy after retirement age in 1970 and 1999 for both men and women.

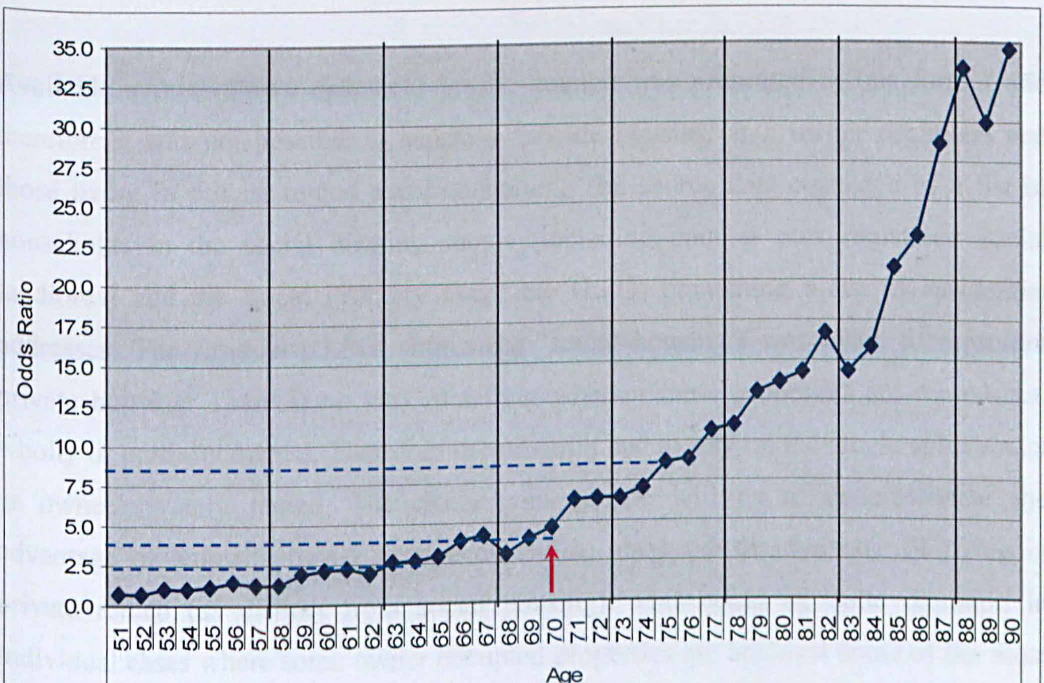
Figure 4.1 Life expectancy after retirement age in 1970 and 1999 for both men and women for selected OECD countries'



In the UK the Work and Pensions Secretary, John Hutton, said the state retirement age, which is set to be 65 for men and women by 2020, will rise to 66 between 2024 and 2026, to 67 between 2034 and 2036 and to 68 between 2044 and 2046 (Hillary Osborne and agencies, 2006)

A logistic regression analysis for mortality with four socio-economic predictors was applied as a further empirical check on how to dichotomise age. In this model the age was defined as fixed one year intervals for those aged between 50 to 90 years old in LB of Camden in 2002-2004. The odds ratios for each year of age were plotted to explore to what extent there might be 'natural step-change' in the OR and to see which age boundary would be the most appropriate cut point for creation of a binary variable for age. The output of the model is presented in Figure 4.2. All cases with age above 90 years old are excluded from the illustration as there were not enough cases for the analysis.

Figure 4.2 Graph representation of odds ratios of age from 50 to 90 years old extracted from logistic regression modelling with 4 socio-economic factors and mortality outcome



Legend	
	Level of odds ratio for age 70 and 75
→	Cut point for creation of binary age

The graph in Figure 4.2 provides an empirical rationale that the growth of OR after age 69/70 tends to rise much faster than before this age and therefore has been chosen as the cut point for binary age. The arrow in the above graph shows the point at which it was judged to make a binary division at age 70 years. Additional dotted lines show how the OR changes at 65 years and 75 years. Broadly, the OR begins a steeper ascent after 70 years.

The construction of the variable for housing tenure consists of two categories: 'owner/private rented' and 'social housing (council housing or housing association)'. The collapsing of owner occupier and private rented contradicts the accepted practice of separating owner occupiers from other categories of tenure. Owner-occupiers are on aggregate, the tenure group with the greatest financial resources, in terms of both wealth and income (Dorling et al., 2001), they have significantly greater monthly household income adjusted for family size (Macintyre et al., 2001) and are more likely to have better health related quality of life than those in rented homes (Breeze et al., 2004). White et al. (2006) also show that the risk of death for men in private rented or social housing is higher than the men in owner occupied tenure.

Available administrative data held by the council was presented in this format and therefore it was not possible to separate 'private housing' into owner occupiers and those living in private rented accommodation. The source data consisted of a list of households in the social housing sector (including council plus registered social landlords) and the Local Property Gazetteer (LPG) containing a list of residential addresses. The residual ('LPG' minus the 'social housing') was taken to represent private housing. There is no way of telling whether these properties are rented out, wholly or partially owned. Therefore the research had to rely on the single split: social or owner/privately rented. The likely consequence will be to underestimate the advantage of living in owner occupancy and to mask the disadvantage of living in private rented (as distinct from social housing). This could be quite dramatic in individual cases where some owner occupied properties are amongst some of the most expensive in the world (e.g. Hampstead) and in certain area in the privately rented sector (e.g. Somer's Town). Further discussion on this issue and investigation about the possible consequences of collapsing these categories is also taken up again in Section 8.3 (Discussion) Pages 182-184.

Shading in the risk ladders represent the different age groups (for example age 50-69 years light grey, and equal or greater than 70, 'dark grey'). All risk ladders are sorted in ascending order by the level of risk (or probability). The reason for shading the risk ladders is to assist in identifying each factor with sequences of cells with the same binary value (0 or 1). Where the sequences are located at the bottom of each column, it shows higher impact of that factor in terms of risk or probability. Where the confidence interval includes values greater than or equal to one or less than or equal to zero (indication of the sample size being small), the entire row in the risk ladder is highlighted by dark grey.

Before creating the risk ladders, a cross tabulation of the mortality records with causes of hospitalization was undertaken. Table 4.1 shows the number of deaths and the population in different combination of the three selected causes of hospital admission. The observed probability or risk of mortality for each group (category) is also shown. The letters 'F, I & S' represent 'Falls, Ischemic heart disease and Stroke' respectively. Under the column titled 'Cause', those individuals who do not have any record of hospital admission as a result of any causes of hospital admissions (FIS) are entered as 'NULL'.

Table 4.1 percentage of recorded deaths according to three different causes of hospital admissions (FIS) for 2002-04

Cause	Frequency of Death	Population	Probability (Risk) of death	Conf. Interval (95%)	
NULL	2630	41249	6.40%	6.16%	6.64%
F	186	775	24.00%	20.99%	27.01%
I	142	805	17.60%	14.97%	20.23%
S	164	471	34.80%	30.50%	39.10%
FI	11	39	28.20%	14.08%	42.32%
FS	30	67	44.80%	32.89%	56.71%
IS	24	62	38.70%	26.58%	50.82%
FIS	1	4	25.00%	-17.44%	67.44%
Total	**3188**	**43472**	**7.30%**	**7.06%**	**7.54%**

Coding Scheme	
F=Falls	I = Ischemic Heart Disease
S=Storke	Conf.Interval=Confidence Interval

The table shows that the risk of death for those who had no record of hospital admission 6.4% (no 'F', 'I' or 'S') to 44.8% for those with at least one hospital admission for Falls and one for Stroke. Between these extremes, Ischemic heart disease with 17.6% is associated with the lowest level, and Stroke with 34.8% with the highest. The number of people at risk of mortality where an admission for three diagnoses were recorded was extremely small and therefore the result is not considered reliable[9]. I will now go on to consider explicit findings for risk ladders.

9 Lower bound of confidence interval is negative clearly suggesting that conventional assumptions about the distribution are not valid.

4.2 Risk ladder analysis of Camden's Mortality data (2002-04)

4.2.1 Risk ladder-1.1 with four basic factors

Table 4.2 illustrates a risk ladder with four basic risk factors. The observed risks have been sorted in ascending order and vary from 1.4% to 21.3% for different combinations of the 4 factors. For example, the sequence '0000' (columns 2 to 5) is read as 'age less than 70 years old, female, living in private housing and in high tax bands (D-H). For the highest level of risk in the row number 16, the code is'1111' which will be interpreted as 'aged over 69 years old, male, living in social housing with a low tax band (A-C).

Table 4.2 Risk ladder-1.1; risk of mortality with four basic socio-demographic factors

Seq	Age	Gender	Tenure	Tax Band	Number of Death	Population	Observed Risk	Conf.Interval	
1	0	0	0	0	87	6074	1.4%	1.1%	1.7%
2	0	0	0	1	9	498	1.8%	0.6%	3.0%
3	0	1	0	0	135	7081	1.9%	1.6%	2.2%
4	0	0	1	0	89	4268	2.1%	1.7%	2.5%
5	0	0	1	1	93	2864	3.2%	2.6%	3.9%
6	0	1	1	0	151	3949	3.8%	3.2%	4.4%
7	0	1	0	1	29	730	4.0%	2.6%	5.4%
8	0	1	1	1	185	3423	5.4%	4.6%	6.2%
9	1	0	1	0	325	2354	13.8%	12.4%	15.2%
10	1	1	0	0	374	2571	14.5%	13.2%	15.9%
11	1	0	0	0	458	3113	14.7%	13.5%	16.0%
12	1	0	1	1	438	2486	17.6%	16.1%	19.1%
13	1	0	0	1	56	311	18.0%	13.7%	22.3%
14	1	1	1	0	319	1668	19.1%	17.2%	21.0%
15	1	1	0	1	49	248	19.8%	14.8%	24.7%
16	1	1	1	1	391	1834	21.3%	19.4%	23.2%
	14585	21504	22846	12394	3188	43472	7.3%	7.1%	7.6%

By sorting the risk in ascending order in risk ladder-1.1, the 8 groups with higher level of risk are all in the older age group, coded '1', and 8 groups with lower risk are in the younger age group, coded '0'. The table thus confirms well known phenomenon that age plays an important role in predicting mortality.

Another factor which appears to be associated with mortality is a person's council tax band. People living in the lower tax bands (coded '1'), in both age groups are at the bottom of the ladder (with the higher levels of risk). Four out of five combinations of people in older age group (70 years old and above) and three out of 4 combinations in younger age group (50 – 69) with highest level of risk are those living in lower tax bands (A-C).

Turning to gender, the risk of mortality for male is typically (but not always) higher than female. The three combinations with highest risk of mortality all comprise males.

Social housing is also associated with a higher risk of mortality. In age group 50-69, four out of five combinations with highest risk live in social housing. For those aged 70 years and above, there is no obvious tendency for social housing to confer greater risk. This could be because advancing years decreases the differences in socioeconomic mortality (Hoffmann, 2005; Liang et al., 2002; Marmot & Shipley, 1996), but it could also be a result of reduction in the importance of socioeconomic differences in old age.

4.2.2 Risk ladder-1.2 with four basic factors and the incidence of an admission for a 'Fall'

By adding a hospital admission factor (Fall) to the previous four socio-economic factors, the number of factors now increase to five and the number of risk ladder combinations to 32 (see Table 4.3).

As we can see in the risk ladder in Table 4.3, 'falls' are directly related to age. After sorting risk into ascending order, nine combinations includes the older age group (those aged 70 years or more), are located at the bottom of the table. The table shows 1.3% for the lowest level of risk (best case) and 43.2% for the worst case combinations.

After age the next factor showing a clear impact on the risk of mortality is experiencing a 'fall'. Eight combinations with highest risk (at the bottom of the table) include those who were admitted to hospital at least once during 2002-04 as a consequence of a fall.

Turning to gender, four out of five combinations with the highest risk include men. The table does not show a strong relationship with the tax band but for tenure it shows that three out of four groups with highest risk live in private housing.

Table 4.3 Risk ladder-1.2; risk of mortality with four basic socio-demographic factors and the incidence of an admission for a fall

Seq	Age	Gender	Tenure	Tax Band	Fall	No. of Death	Population	Observed Risk	Conf.Interval	
1	0	0	0	1	1	0	2	0.0%	0.0%	0.0%
2	0	0	0	0	0	81	6044	1.3%	1.1%	1.6%
3	0	0	0	1	0	9	496	1.8%	0.6%	3.0%
4	0	1	0	0	0	129	7045	1.8%	1.5%	2.1%
5	0	0	1	0	0	87	4241	2.1%	1.6%	2.5%
6	0	0	1	1	0	89	2840	3.1%	2.5%	3.8%
7	0	1	1	0	0	148	3917	3.8%	3.2%	4.4%
8	0	1	0	1	0	28	725	3.9%	2.5%	5.3%
9	0	1	1	1	0	177	3376	5.2%	4.5%	6.0%
10	0	0	1	0	1	2	27	7.4%	-2.5%	17.3%
11	0	1	1	0	1	3	32	9.4%	-0.7%	19.5%
12	1	0	1	0	0	287	2212	13.0%	11.6%	14.4%
13	1	1	0	0	0	350	2501	14.0%	12.6%	15.4%
14	1	0	0	0	0	413	2937	14.1%	12.8%	15.3%
15	1	0	0	1	0	47	289	16.3%	12.0%	20.5%
16	0	0	1	1	1	4	24	16.7%	1.8%	31.6%
17	0	1	0	0	1	6	36	16.7%	4.5%	28.8%
18	0	1	1	1	1	8	47	17.0%	6.3%	27.8%
19	1	0	1	1	0	402	2345	17.1%	15.6%	18.7%
20	1	1	1	0	0	300	1624	18.5%	16.6%	20.4%
21	1	1	0	1	0	45	238	18.9%	13.9%	23.9%
22	0	0	0	0	1	6	30	20.0%	5.7%	34.3%
23	0	1	0	1	1	1	5	20.0%	-15.1%	55.1%
24	1	1	1	1	0	368	1757	20.9%	19.0%	22.8%
25	1	0	1	1	1	36	141	25.5%	18.3%	32.7%
26	1	0	0	0	1	45	176	25.6%	19.1%	32.0%
27	1	0	1	0	1	38	142	26.8%	19.5%	34.0%
28	1	1	1	1	1	23	77	29.9%	19.6%	40.1%
29	1	1	0	0	1	24	70	34.3%	23.2%	45.4%
30	1	1	0	1	1	4	10	40.0%	9.6%	70.4%
31	1	0	0	1	1	9	22	40.9%	20.4%	61.5%
32	1	1	1	0	1	19	44	43.2%	28.5%	57.8%
	14585	21504	22846	12394	885	3188	43472	7.3%	7.1%	7.6%

4.2.3 Risk ladder-1.3 with four basic factors and the incidence of an admission for 'Ischemic Heart-Disease'

In the next table (Ischemic heart disease) replaces falls. Table 4.4 illustrates a risk ladder for these combinations which include the previous socio-economic factors. In this case the level of risk between best and worst group varies from 1.4% to 57.1%.

As in previous cases, the impact of age is very strong as might be expected. After sorting the risk in ascending order, all combinations of factors for the older age group (16 combinations) are located at the bottom of the table.

Table 4.4 Risk ladder-1.3; risk of mortality with four basic socio-demographic factors and Ischemic heart disease

Seq	Age	Gender	Tenure	Tax Band	Heart Disease	No. of Death	Population	Observed Risk	Conf.Interval	
1	0	0	0	1	1		3	0.0%	0.0%	0.0%
2	0	0	0	0	0	85	6050	1.4%	1.1%	1.7%
3	0	0	0	1	0	9	495	1.8%	0.6%	3.0%
4	0	1	0	0	0	128	7026	1.8%	1.5%	2.1%
5	0	0	1	0	0	87	4223	2.1%	1.6%	2.5%
6	0	0	1	1	0	90	2830	3.2%	2.5%	3.8%
7	0	1	1	0	0	144	3841	3.7%	3.1%	4.3%
8	0	1	0	1	0	28	721	3.9%	2.5%	5.3%
9	0	0	1	0	1	2	45	4.4%	-1.6%	10.5%
10	0	1	1	1	0	176	3331	5.3%	4.5%	6.0%
11	0	1	1	0	1	7	108	6.5%	1.8%	11.1%
12	0	0	0	0	1	2	24	8.3%	-2.7%	19.4%
13	0	0	1	1	1	3	34	8.8%	-0.7%	18.4%
14	0	1	1	1	1	9	92	9.8%	3.7%	15.9%
15	0	1	0	1	1	1	9	11.1%	-9.4%	31.6%
16	0	1	0	0	1	7	55	12.7%	3.9%	21.5%
17	1	0	1	0	0	305	2267	13.5%	12.0%	14.9%
18	1	1	0	0	0	350	2481	14.1%	12.7%	15.5%
19	1	0	0	0	0	437	3042	14.4%	13.1%	15.6%
20	1	0	1	1	0	407	2383	17.1%	15.6%	18.6%
21	1	0	0	1	0	52	304	17.1%	12.9%	21.3%
22	1	1	1	0	0	293	1576	18.6%	16.7%	20.5%
23	1	1	0	1	0	46	240	19.2%	14.2%	24.1%
24	1	1	1	1	0	373	1752	21.3%	19.4%	23.2%
25	1	1	1	1	1	18	82	22.0%	13.0%	30.9%
26	1	0	1	0	1	20	87	23.0%	14.1%	31.8%
27	1	1	0	0	1	24	90	26.7%	17.5%	35.8%
28	1	1	1	0	1	26	92	28.3%	19.1%	37.5%
29	1	0	0	0	1	21	71	29.6%	19.0%	40.2%
30	1	0	1	1	1	31	103	30.1%	21.2%	39.0%
31	1	1	0	1	1	3	8	37.5%	4.0%	71.0%
32	1	0	0	1	1	4	7	57.1%	20.5%	93.8%
	14585	21504	22846	12394	910	3188	43472	7.3%	7.1%	7.6%

In the 'Ischemic heart-disease' factor in this risk ladder, for older age, eight groups with highest risk (out of 16 different combinations), were admitted to hospital at least once in the period 2002-04. For the younger age group, six out of 16 combinations included at least one admission for Ischemic heart-disease.

For the variable 'council tax band', three combinations of people with highest risk of mortality are living in lower tax bands. Drawing conclusions based on gender and tenure in this risk ladder is difficult as there is no obvious separation of effect.

4.2.4 Risk ladder-1.4 with four basic factors and the incidence of an admission for 'Stroke'

By adding stroke to the first four socio-economic variables another risk ladder with 32 combinations was produced (see Table 4.5).

Table 4.5 Risk ladder-1.4; risk of mortality with four basic socio-demographic factors and Stroke

Seq	Age	Gender	Tenure	Tax Band	Strokes	No. of Death	Population	Observed Risk	Conf.Interval	
1	0	0	0	1	1	0	2	0.0%	0.0%	0.0%
2	1	1	0	1	1		1	0.0%	0.0%	0.0%
3	0	0	0	0	0	82	6058	1.4%	1.1%	1.6%
4	0	0	0	1	0	9	496	1.8%	0.6%	3.0%
5	0	1	0	0	0	130	7056	1.8%	1.5%	2.2%
6	0	0	1	0	0	85	4241	2.0%	1.6%	2.4%
7	0	0	1	1	0	88	2843	3.1%	2.5%	3.7%
8	0	1	1	0	0	133	3898	3.4%	2.8%	4.0%
9	0	1	0	1	0	27	726	3.7%	2.3%	5.1%
10	0	1	1	1	0	174	3376	5.2%	4.4%	5.9%
11	1	0	1	0	0	297	2296	12.9%	11.6%	14.3%
12	1	0	0	0	0	423	3028	14.0%	12.7%	15.2%
13	1	1	0	0	0	351	2510	14.0%	12.6%	15.3%
14	0	0	1	0	1	4	27	14.8%	1.4%	28.2%
15	1	0	1	1	0	403	2416	16.7%	15.2%	18.2%
16	1	0	0	1	0	52	306	17.0%	12.8%	21.2%
17	1	1	1	0	0	303	1604	18.9%	17.0%	20.8%
18	1	1	0	1	0	49	247	19.8%	14.9%	24.8%
19	0	1	0	0	1	5	25	20.0%	4.3%	35.7%
20	1	1	1	1	0	363	1767	20.5%	18.7%	22.4%
21	0	1	1	1	1	11	47	23.4%	11.3%	35.5%
22	0	0	1	1	1	5	21	23.8%	5.6%	42.0%
23	1	1	1	0	1	16	64	25.0%	14.4%	35.6%
24	0	0	0	0	1	5	16	31.3%	8.5%	54.0%
25	0	1	1	0	1	18	51	35.3%	22.2%	48.4%
26	1	1	0	0	1	23	61	37.7%	25.5%	49.9%
27	1	0	0	0	1	35	85	41.2%	30.7%	51.6%
28	1	1	1	1	1	28	67	41.8%	30.0%	53.6%
29	1	0	1	0	1	28	58	48.3%	35.4%	61.1%
30	1	0	1	1	1	35	70	50.0%	38.3%	61.7%
31	0	1	0	1	1	2	4	50.0%	1.0%	99.0%
32	1	0	0	1	1	4	5	80.0%	44.9%	115.1%
	14585	21504	22846	12394	604	3188	43472	7.3%	7.1%	7.6%

For two combinations in this risk ladder there is no report of any deaths and the number of deaths for one of the combinations is relatively small. While the observed risk for cases at lowest risk (a combination of people under 70 years old; females; living in private housing; those that live in higher council tax band properties; and those who were not admitted to the hospital for stroke) is 1.4%, whereas for the highest risk cases (which is a combination of: people under 70 years old; males; those living in private housing; those living in lower council tax bands; and those admitted to the hospital at least once for stroke) it increases to 50%.

Age is not as influential as it was for the two previous risk ladders. Six out of seven combinations with highest risk fall in the older age group. Gender does not appear to be an important influence in this risk ladder. With regards to the council tax band, four out of five groups with highest level of risk are those living in lower tax bands (A-C) and three groups at the bottom of the table are those living in social housing. The most powerful factor in this risk ladder is 'stroke'. Twelve groups (combinations) with highest level of risk of mortality are those who had at least one stroke during 2002 – 04.

4.2.5 Risk ladder-1.5 with seven factors (4 socio-demographic factors and 3 causes of hospital admissions)

The final risk ladder includes seven risk factors including the four socio-economic factors plus all three causes of hospital admissions (FIS). The total number of combinations for this risk ladder is now 128 (2^7). By distributing 3188 deaths across these combinations, in 69 groups the number of deaths or population is zero or very small (with the confidence interval including values greater than or equal to one or less than or equal to zero). These cases have been omitted reducing the numbers of combinations from 128 to 59 (to less than half). Note that the omitted groups only include 50 deaths out of the total number of 3188 deaths (1.6%), which is not expected to have a large effect on the conclusion. The results are shown in Table 4.6.

The number of rows (combinations) in Table 4.6 is relatively high but as we are more interested in the combinations at the bottom of table some rows (falling between 6 and 44) are omitted and marked by a break. The risk of mortality for different combinations in risk ladder-1.5 now varies from 1.2% to 57%[10].

Age remains one of the most influential factors, as might be expected, given previous risk ladders. Most of the combinations located at the bottom of the table fall in the older age group.

Table 4.6 Risk ladder-1.5; risk of mortality with four basic socio-demographic factors and the incidence of up to 3 causes of hospital admissions (FIS)

Seq	Age	Gender	Tenure	Tax Band	Fall	Heart Disease	Stroke	No. of Death	Population	Observed Risk	Conf.Interval	
1	0	0	0	0	0	0	0	75	6005	1.2%	1.0%	1.5%
2	0	1	0	0	0	0	0	119	6968	1.7%	1.4%	2.0%
3	0	0	0	1	0	0	0	9	491	1.8%	0.6%	3.0%
4	0	0	1	0	0	0	0	81	4171	1.9%	1.5%	2.4%
5	0	0	1	1	0	0	0	81	2788	2.9%	2.3%	3.5%
45	1	0	0	0	0	0	1	23	60	38.3%	26.0%	50.6%
46	1	1	1	0	1	0	0	15	39	38.5%	23.2%	53.7%
47	1	0	0	0	1	0	1	7	18	38.9%	16.4%	61.4%
48	0	1	1	0	0	0	1	16	41	39.0%	24.1%	54.0%
49	1	1	1	1	0	0	1	19	48	39.6%	25.7%	53.4%
50	1	1	0	0	1	0	1	3	7	42.9%	6.2%	79.5%
51	1	1	0	1	0	1	0	3	7	42.9%	6.2%	79.5%
52	1	1	0	1	1	0	0	4	9	44.4%	12.0%	76.9%
53	1	1	1	1	0	1	1	5	11	45.5%	16.0%	74.9%
54	0	1	0	1	0	0	1	2	4	50.0%	1.0%	99.0%
55	1	1	0	0	0	1	1	4	8	50.0%	15.4%	84.6%
56	1	0	1	1	0	0	1	26	51	51.0%	37.3%	64.7%
57	1	0	1	0	0	0	1	23	45	51.1%	36.5%	65.7%
58	1	0	1	1	1	0	1	7	13	53.8%	26.7%	80.9%
59	1	1	1	1	1	0	1	4	7	57.1%	20.5%	93.8%
	14585	21504	22846	12394	885	910	604	3188	43472	7.33%	7.1%	7.6%

It is difficult to provide firm evidence of differential risk between male and female, at the bottom of the risk ladder (e.g. the last ten combinations). However, if we look at all cases from row 45 onwards we can see that men tend to be at higher risk than women. For tenure it is also difficult to draw comparisons between those living in social housing and private housing. Making any conclusion about differences in tax bands from risk ladder-1.5 also is not straightforward. The majority of groups who are

[10] Note: there is a break in the table of sequences (for convenient).

located at the bottom of the list (with highest risk of mortality) are those groups who were admitted to the hospital at least twice, each time for different cause (e.g. for fall and stroke or ischemic heart disease and stroke etc). Between the three causes of hospital admissions, stroke is the most influential factor and compared to all seven factors in this risk ladder, it is as powerful as age.

4.2.6 A risk ladder with different age dichotomy (current retirement age of 65 and 66+)

The reason for choosing the age dichotomy of 50- 69 and age 70 was discussed earlier in Section 4.1 However for the purpose of further investigation, a risk ladder with a different age dichotomy (50-65 years old and 66+) will be included here as a 'quick' or 'crude' sensitivity check on the findings. Table 4.7 below shows a risk ladder with four socio-economic factors and outcome mortality with binary age 50-65 years and 66+ (identical to risk ladder 1.1 in Table 4.2).

Table 4.7 Risk ladder-1.6 A risk ladder with 4 socio-economic factors and outcome mortality with binary age 50-65 years = 0 and 66+ years =1 (equivalent to risk ladder 1.1 in Table 4.2)

Mortality Risk with 4 basic factors										
Seq	Age	Gender	Tenure	Tax Band	Combination 'AGTB'	Number of Death	Population	Observed Risk	Conf.Interval L. Bound	Conf.Interval U. Bound
1	0	0	0	0	0000	63	5223	1.2%	0.9%	1.5%
2	0	0	0	1	0001	6	428	1.4%	0.3%	2.5%
3	0	1	0	0	0100	93	6168	1.5%	1.2%	1.8%
4	0	0	1	0	0010	67	3626	1.8%	1.4%	2.3%
5	0	0	1	1	0011	60	2318	2.6%	1.9%	3.2%
6	0	1	1	0	0110	101	3305	3.1%	2.5%	3.6%
7	0	1	0	1	0101	20	634	3.2%	1.8%	4.5%
8	0	1	1	1	0111	137	2908	4.7%	3.9%	5.5%
9	1	0	1	0	1010	347	2996	11.6%	10.4%	12.7%
10	1	1	0	0	1100	416	3484	11.9%	10.9%	13.0%
11	1	0	0	0	1000	482	3964	12.2%	11.1%	13.2%
12	1	0	1	1	1011	59	381	15.5%	11.9%	19.1%
13	1	0	0	1	1001	471	3032	15.5%	14.2%	16.8%
14	1	1	1	0	1110	369	2312	16.0%	14.5%	17.5%
15	1	1	0	1	1101	58	344	16.9%	12.9%	20.8%
16	1	1	1	1	1111	439	2349	18.7%	17.1%	20.3%
	18862	21504	22846	12394		3188	43472	7.3%	7.1%	7.6%

Dichotomising age at 69 years compared to 65 has a very modest impact on the estimates of risk: the level of risk of mortality for the group of people with the lowest risk decreases 0.2% (from 1.4% to 1.2%) and for those with the highest risk of mortality increases by 2.3% (from 18.7% to 21.3%). This alteration shows a steady change in level of risk and there is no evidence of a sudden jump in risk of mortality for the both best and worth cases.

Next section includes risk ladder analysis using data from Camden's social services (2002-04).

4.3 Risk ladder analysis including data from Camden's social services (2002-04)

Risk ladders in this section are similar to those discussed in the previous section (Section 2) apart from the outcome variable which is changed from 'Mortality' to 'Being known to the social services' or 'Being in contact with social services'. I seek to identify whether there is a relationship between risk of mortality in the one hand and being in contact with social services in the other. A close relationship may indicate for example whether services are being well targeted. There are a range of services provided by the social services such as; day services, direct payments for purchasing care, equipment allocation, home based services, meals, nursing care, professional support, residential care, respite for carers, transport etc and a person may be in contact with the social services for one or more of these reasons. The aim of this analysis is not to go through a detailed explanation of different type of service. It is to assess the overall activities of the organization, its direction and the preferred service delivery for different groups of people based upon different combinations of risk factors.

In the following sub-sections, the four risk ladders introduced earlier in this chapter will be analyzed. Hereafter the outcome variable 'Being known by or Being in contact with social services' as a column title in the tables will be replaced by 'social services' or 'SS'. In the following tables, the column title of 'Known to SS' represents the number of people from a specific combination in contact with 'SS' and letter probability is used as the label as previously for probability.

4.3.1 Risk ladder-2.1 with four basic factors

Risk ladder 2.1 illustrated in table 4.7, is identical to the risk ladder 1.1 (table 4.2) discussed in Section 2 except the outcome variable is changed from 'Mortality' in 1.1 to 'social services' in 2.1. Table 4.7 is sorted by the value of the column '*P* of known to SS' in ascending order. It shows that all combinations of people in second age group (greater than or equal to 70) are located in the bottom of the table, similar to risk ladder 1.1 for the mortality outcome.

For the variable 'Gender', the three combinations of people with the highest probability of receiving social services are females.

Housing tenure for both age groups (under 70 and 70 years old or more) plays an important role. Four out of five groups with the highest probability of using services are living in social housing and are included in the combination of older age group (older than 69 years). It is also notable that for the younger age groups (50-69 years old); the four groups with the highest chance of being entitled to services are living in social housing.

Table 4.8 Risk ladder-2.1 including four basic socio-demographic factors and 'social services' as the outcome variable

Seq	Age	Gender	Tenure	Tax Band	Known to 'SS'	Population	*P* of Known to 'SS'	Conf.Interval	
1	0	1	0	0	60	7081	0.8%	0.6%	1.1%
2	0	0	0	0	79	6074	1.3%	1.0%	1.6%
3	0	0	0	1	11	498	2.2%	0.9%	3.5%
4	0	1	0	1	19	730	2.6%	1.4%	3.8%
5	0	1	1	0	147	3949	3.7%	3.1%	4.3%
6	0	0	1	0	168	4268	3.9%	3.4%	4.5%
7	0	0	1	1	147	2864	5.1%	4.3%	5.9%
8	0	1	1	1	188	3423	5.5%	4.7%	6.3%
9	1	1	0	0	225	2571	8.8%	7.7%	9.8%
10	1	1	0	1	26	248	10.5%	6.7%	14.3%
11	1	0	0	0	440	3113	14.1%	12.9%	15.4%
12	1	1	1	0	249	1668	14.9%	13.2%	16.6%
13	1	1	1	1	351	1834	19.1%	17.3%	20.9%
14	1	0	0	1	65	311	20.9%	16.4%	25.4%
15	1	0	1	0	492	2354	20.9%	19.3%	22.5%
16	1	0	1	1	637	2486	25.6%	23.9%	27.3%
	14585	21504	22846	12394	3304	43472	7.6%	7.4%	7.8%

The impact of the variable 'Council tax band' in this risk ladder is also noticeable. For older age groups, three out of four and for younger age two groups with the highest probability of receiving social services are living in lower tax band properties.

4.3.2 Risk ladder-2.2 with four basic factors and the incidence of an admission for a 'Fall'

In Risk ladder-2.2 variable 'Fall' is added to the previous four socio-economic factors and the probability of different combinations (32 combinations) of five factors with the outcome variable of 'social services ' is calculated. Risk ladder 2.2 is shown in table 4.8 which is identical to risk ladder 1.2 of table 4.3.

It is apparent that the variable 'Age' yet again has an important role in determination of the social services resources, as is demonstrated in the column 'Age' in which the eight combinations with the highest probability of using social services resources are from the older age group.

Table 4.9 Risk ladder-2.2 including four basic socio-demographic factors and the incidence of an admission for a fall with 'social services' as the outcome variable

Seq	Age	Gender	Tenure	Tax Band	Fall	Known to 'SS'	Population	*P* of Known to 'SS'	Conf.Interval	
1	0	1	0	0	0	56	7045	0.8%	0.6%	1.0%
2	0	0	0	0	0	78	6044	1.3%	1.0%	1.6%
3	0	0	0	1	0	10	496	2.0%	0.8%	3.3%
4	0	1	0	1	0	18	725	2.5%	1.4%	3.6%
5	0	0	0	0	1	1	30	3.3%	-3.1%	9.8%
6	0	1	1	0	0	143	3917	3.7%	3.1%	4.2%
7	0	0	1	0	0	163	4241	3.8%	3.3%	4.4%
8	0	0	1	1	0	142	2840	5.0%	4.2%	5.8%
9	0	1	1	1	0	178	3376	5.3%	4.5%	6.0%
10	1	1	0	0	0	196	2501	7.8%	6.8%	8.9%
11	1	1	0	1	0	24	238	10.1%	6.3%	13.9%
12	0	1	0	0	1	4	36	11.1%	0.8%	21.4%
13	0	1	1	0	1	4	32	12.5%	1.0%	24.0%
14	1	0	0	0	0	372	2937	12.7%	11.5%	13.9%
15	1	1	1	0	0	234	1624	14.4%	12.7%	16.1%
16	1	1	1	1	0	311	1757	17.7%	15.9%	19.5%
17	1	0	0	1	0	53	289	18.3%	13.9%	22.8%
18	0	0	1	0	1	5	27	18.5%	3.9%	33.2%
19	1	0	1	0	0	426	2212	19.3%	17.6%	20.9%
20	0	1	0	1	1	1	5	20.0%	-15.1%	55.1%
21	1	1	0	1	1	2	10	20.0%	-4.8%	44.8%
22	0	0	1	1	1	5	24	20.8%	4.6%	37.1%
23	0	1	1	1	1	10	47	21.3%	9.6%	33.0%
24	1	0	1	1	0	563	2345	24.0%	22.3%	25.7%
25	1	1	1	0	1	15	44	34.1%	20.1%	48.1%
26	1	0	0	0	1	68	176	38.6%	31.4%	45.8%
27	1	1	0	0	1	29	70	41.4%	29.9%	53.0%
28	1	0	1	0	1	66	142	46.5%	38.3%	54.7%
29	0	0	0	1	1	1	2	50.0%	-19.3%	119.3%
30	1	1	1	1	1	40	77	51.9%	40.8%	63.1%
31	1	0	1	1	1	74	141	52.5%	44.2%	60.7%
32	1	0	0	1	1	12	22	54.5%	33.7%	75.4%
	14585	21504	22846	12394	885	3304	43472	7.6%	7.4%	7.8%

For 'Gender' like the previous risk ladder the probability of females 'being in contact with social services' is higher than for males.

Although the highest probability of being in contact with social services is associated with private housing (with a probability of 54.5%), the majority of the combinations in the bottom half of the table are those groups of people living in social housing.

Council tax band has a noticeable impact on the probability level, indicating that living in lower tax band properties increases the chance of being known to the social services.

As highlighted in table 4.8, the strongest factor in determining the use of social services' resources is 'Fall'. If the rows highlighted with dark grey (with a confidence intervals including values greater than or equal to one or less than or equal to zero) are excluded, nine out of ten combinations with highest probability of being in contact with social services are those who had at least one admission to the hospital because of 'Fall'.

4.3.3 Risk ladder-2.3 with four basic factors and the incidence of an admission for 'Ischemic Heart-Disease'

Risk ladder 2.3 illustrated in table 4.9, is identical to risk ladder 1.3 in table 4.4. In different combinations of the four socio-economic factors and 'heart disease', the variable 'Age' again has the most influence on the probability of someone being in contact with social services. Twelve out of thirteen groups of people with the highest probability of being in contact with social services are from combinations including the older age group.

For the variable 'Gender' the probability level for combinations with female are higher than for those with male. From eight groups with the highest probability of being in contact with social services, six groups are from combinations including female.

In this case, the variable 'Tenure' is an important variable in determination of the probability of being in contact with social services.
Living in a lower council tax band is also a reasonably dominant factor in increasing the possibility of being in contact with social services.

The incidence of an admission to hospital as a result of heart disease strongly increases the probability of allocation of social services' resources to a person, although not to the same extent as the variable 'Fall' discussed earlier.

Table 4.10 Risk ladder-2.3 including four basic socio-demographic factors and the incidence of an admission for heart disease with 'social svices' as the outcome variable

Seq	Age	Gender	Tenure	Tax Band	Heart Disease	Known to 'SS'	Population	*P* of Known to 'SS'	Conf.Interval	
1	0	0	0	1	1		3	0.0%	0.0%	0.0%
2	0	1	0	0	0	57	7026	0.8%	0.6%	1.0%
3	0	0	0	0	0	76	6050	1.3%	1.0%	1.5%
4	0	1	0	1	0	16	721	2.2%	1.1%	3.3%
5	0	0	0	1	0	11	495	2.2%	0.9%	3.5%
6	0	1	1	0	0	135	3841	3.5%	2.9%	4.1%
7	0	0	1	0	0	165	4223	3.9%	3.3%	4.5%
8	0	0	1	1	0	144	2830	5.1%	4.3%	5.9%
9	0	1	1	1	0	179	3331	5.4%	4.6%	6.1%
10	0	1	0	0	1	3	55	5.5%	-0.5%	11.5%
11	0	0	1	0	1	3	45	6.7%	-0.6%	14.0%
12	1	1	0	0	0	212	2481	8.5%	7.4%	9.6%
13	0	0	1	1	1	3	34	8.8%	-0.7%	18.4%
14	0	1	1	1	1	9	92	9.8%	3.7%	15.9%
15	1	1	0	1	0	24	240	10.0%	6.2%	13.8%
16	0	1	1	0	1	12	108	11.1%	5.2%	17.0%
17	0	0	0	0	1	3	24	12.5%	-0.7%	25.7%
18	1	0	0	0	0	422	3042	13.9%	12.6%	15.1%
19	1	1	0	0	1	13	90	14.4%	7.2%	21.7%
20	1	1	1	0	0	232	1576	14.7%	13.0%	16.5%
21	1	1	1	0	1	17	92	18.5%	10.5%	26.4%
22	1	1	1	1	0	332	1752	18.9%	17.1%	20.8%
23	1	0	1	0	0	464	2267	20.5%	18.8%	22.1%
24	1	0	0	1	0	63	304	20.7%	16.2%	25.3%
25	1	1	1	1	1	19	82	23.2%	14.0%	32.3%
26	1	1	0	1	1	2	8	25.0%	-5.0%	55.0%
27	1	0	1	1	0	600	2383	25.2%	23.4%	26.9%
28	1	0	0	0	1	18	71	25.4%	15.2%	35.5%
29	1	0	0	1	1	2	7	28.6%	-4.9%	62.0%
30	1	0	1	0	1	28	87	32.2%	22.4%	42.0%
31	0	1	0	1	1	3	9	33.3%	2.5%	64.1%
32	1	0	1	1	1	37	103	35.9%	26.7%	45.2%
	14585	21504	22846	12394	885	3304	43472	7.6%	7.4%	7.8%

4.3.4 Risk ladder-2.4 with four basic factors and the incidence of an admission for 'Stroke'

This risk ladder in Table 4.10 (Risk ladder 2.4) includes four basic socio-economic factors and the incidence of at least one admission to hospital for stroke.

In the first column of the risk ladder 2.4 which represents the variable 'Age', five combinations with the highest level of probability of 'being in contact with social services' are the combinations including older age groups. While the variable age in these combinations influences the outcome it is not as strong as the previous two risk ladders (Risk ladders 2.2 and 2.3) which include 'fall' and 'heart disease'. Also, the variable 'Gender' in this risk ladder does not seem to affect the outcome.

Table 4.11 Risk ladder 2.4 including four basic socio-demographic factors and the incidence of an admission for stroke with 'social services' as the outcome variable

Seq	Age	Gender	Tenure	Tax Band	Stroke	Known to 'SS'	Population	*P* of Known to 'SS'	Conf.Interval	
1	0	0	0	0	0	78	6058	1.3%	1.0%	1.6%
2	0	1	0	0	0	56	7056	0.8%	0.6%	1.0%
3	0	0	0	1	0	10	496	2.0%	0.8%	3.3%
4	0	1	0	1	0	18	726	2.5%	1.3%	3.6%
5	0	1	1	0	0	136	3898	3.5%	2.9%	4.1%
6	0	0	1	0	0	161	4241	3.8%	3.2%	4.4%
7	0	0	1	1	0	145	2843	5.1%	4.3%	5.9%
8	0	1	1	1	0	176	3376	5.2%	4.5%	6.0%
9	0	0	0	0	1	1	16	6.3%	-5.6%	18.1%
10	1	1	0	0	0	205	2510	8.2%	7.1%	9.2%
11	0	0	1	1	1	2	21	9.5%	-3.0%	22.1%
12	1	1	0	1	0	25	247	10.1%	6.4%	13.9%
13	1	0	0	0	0	412	3028	13.6%	12.4%	14.8%
14	1	1	1	0	0	235	1604	14.7%	12.9%	16.4%
15	0	1	0	0	1	4	25	16.0%	1.6%	30.4%
16	1	1	1	1	0	327	1767	18.5%	16.7%	20.3%
17	1	0	0	1	1	1	5	20.0%	-15.1%	55.1%
18	1	0	1	0	0	471	2296	20.5%	18.9%	22.2%
19	1	0	0	1	0	64	306	20.9%	16.4%	25.5%
20	0	1	1	0	1	11	51	21.6%	10.3%	32.9%
21	1	1	1	0	1	14	64	21.9%	11.7%	32.0%
22	0	1	0	1	1	1	4	25.0%	-17.4%	67.4%
23	1	0	1	1	0	609	2416	25.2%	23.5%	26.9%
24	0	1	1	1	1	12	47	25.5%	13.1%	38.0%
25	0	0	1	0	1	7	27	25.9%	9.4%	42.5%
26	1	1	0	0	1	20	61	32.8%	21.0%	44.6%
27	1	0	0	0	1	28	85	32.9%	22.9%	42.9%
28	1	1	1	1	1	24	67	35.8%	24.3%	47.3%
29	1	0	1	0	1	21	58	36.2%	23.8%	48.6%
30	1	0	1	1	1	28	70	40.0%	28.5%	51.5%
31	0	0	0	1	1	1	2	50.0%	-19.3%	119.3%
32	1	1	0	1	1	1	1	100.0%	100.0%	100.0%
	14585	21504	22846	12394	885	3304	43472	7.6%	7.4%	7.8%

Housing tenure in risk ladder 2.4 is a strong factor in determining the outcome variable 'being in contact with social services'. Ten out of thirteen groups with the highest likelihood of benefiting from the social services' resources contain people living in social housing.

The variable 'council tax band' does not appear to have a particular impact on the outcome variable, whereas the variable 'stroke' has a very strong influence. Nine out of ten groups with the highest level of observed probability of being in contact with social services are those who had at least one admission to the hospital for stroke.

So far the influence of different factors on outcome variables 'Mortality' and 'being in contact with social services' in different combinations has been discussed in detail. The outcome of each risk ladder alone for the both outcome variables ('Mortality' and 'being in contact with social services') has been shown helpful in partitioning risk. In the risk ladders with the outcome 'Mortality', the degree to which different factors determine the health inequalities was quite clear. In the case of the outcome variable 'being in contact with social services', the affect of different factors on allocation of resources was seen to be sensitive to the variable 'hospital admission' and to the cause for that admission. For the purpose of further investigation and in order to assess the relationship between health inequalities and the allocation of social services' resources, all risk ladders with different outcomes will be compared graphically in the following section.

4.4 Comparing the risk of mortality and the probability of being in contact with social services

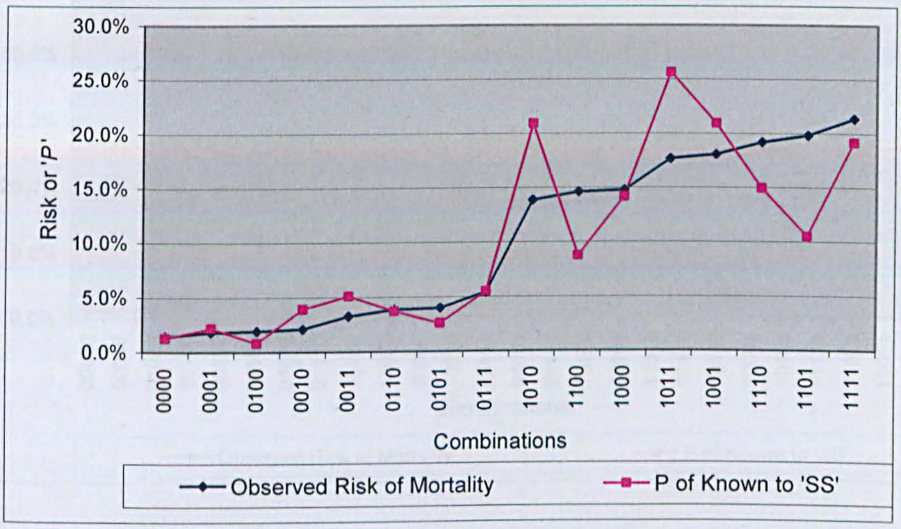
In this section each of the risk ladders with outcome variables 'Mortality' and 'being in contact with social services' (i.e. Risk ladder 1.1 with 2.1, Risk ladder 1.2 with 2.2 and so on) will be contrasted. The outcomes of the comparisons are illustrated in the following figures (Figures 4.2-4.5). Each figure includes combinations of different factors with their relevant 'risk of mortality' and 'probability of being in contact with social services'. Appendix-D includes the tables containing the combined risk ladders (risk ladders with similar factors but different outcomes) and their detailed information which are the foundation of the following figures. In Appendix-D where the confidence interval includes values greater than or equal to one or less than or equal to zero, the entire row for both outcomes (with the same combination) is omitted.

4.4.1 Comparison with four socio-economic factors

A comparison of the two outcome columns (risk of mortality and the probability of being known to the social services) for sixteen different combinations of four factors of both Risk ladder 1.1 (Table 4.2) and Risk ladder 2.1 (Table 4.7) are illustrated in Figure 4.2. In general there is a reasonable correspondence between the outcome variables for each combination. However, the curve representing the probability of being in contact with social services does not follow the curve of 'observed risk of mortality' for some of the combinations. Five points on the social services curve with codes; '0010', '0011', '1010', '1011' and '1001' located above the mortality curve include females and four points on social services curve with codes; '1100', '1110', '1101' and '1111' located under the mortality curve include males. In another words, while the comparison of the 'risk of mortality' and 'the probability of being in contact with social services, shows for the same level of risk of mortality, females are more in contact with social services than males. The reason for this inequality could be that females are more likely to outlive males and live alone with no informal carer in the household which will be discussed further in Chapter 7 on policy implications.

For the variable 'tenure', it is evident that four points out of five, located above the mortality curve with the highest level of probability of being in contact with social services (coded; '0010', '0011', '1010' and '1011') include people living in social housing. All four points include females. Also, four groups of people with lowest level of probability of being in contact with social services are males living in private housing (with codes; '0100', '0101', '1100' and '1101').

Figure 4.3 comparing the risk of mortality and the probability of being in contact with social services for different combinations of four factors

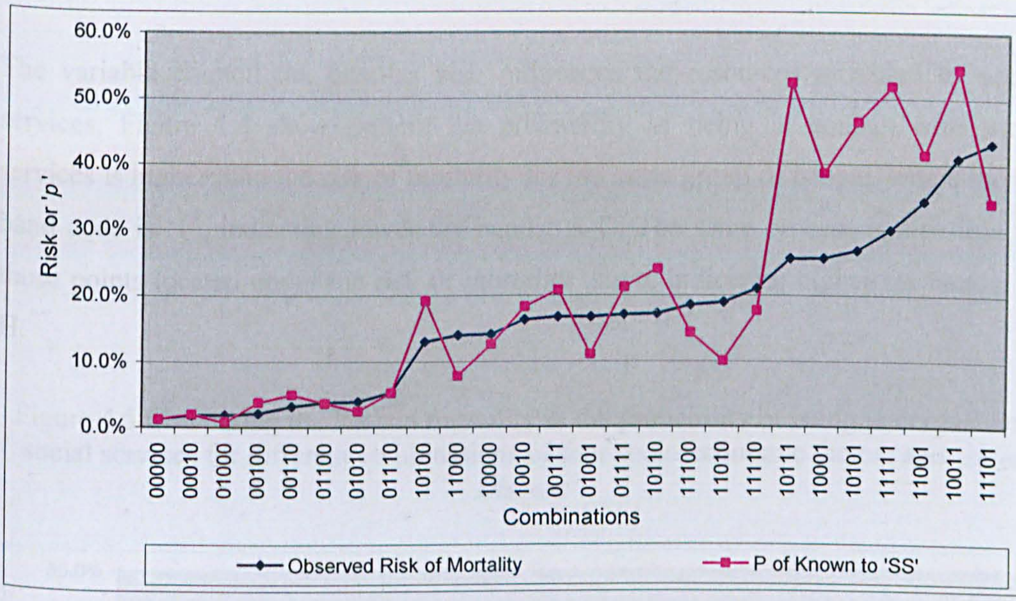


4.4.2 Comparison with four socio-economic factors and incidence of an admission for a 'Fall'

The outcomes of Risk ladder-1.2, Table 4.3 (risk of mortality) and Risk ladder-2.2, Table 4.8 (the probability of being known to the social services) which include the combinations of four socio-economic factors and 'falls' are contrasted in Figure 4.3 below. The outcomes of six combinations with confidence interval with the values greater than or equal to one or less than or equal to zero are omitted.

In Figure 4.3 the social services curve for some of the combinations located to the right of the chart, including; '10111', '10001', '10101', '11111', '11001' and '10011', is positioned above the mortality curve. The last digit of the above six combinations is '1' which represents the incidence of at least one hospital admission for falls. Therefore we deduce that 'Falls' play an important role in the allocation of social services' resources, especially in combination with age, as the above six combinations all fit into older age groups. Figure 4.3 also confirms that in general females are more in contact with the social services than males.

Figure 4.4 comparing the risk of mortality and the probability of being in contact with social services for different combinations of four socio-economic factors and *Falls*



For variable tenure, again most of the points on 'social services' curve located above the mortality line, are those combinations that include social housing. The majority of the points positioned under the mortality line tend to include private housing.

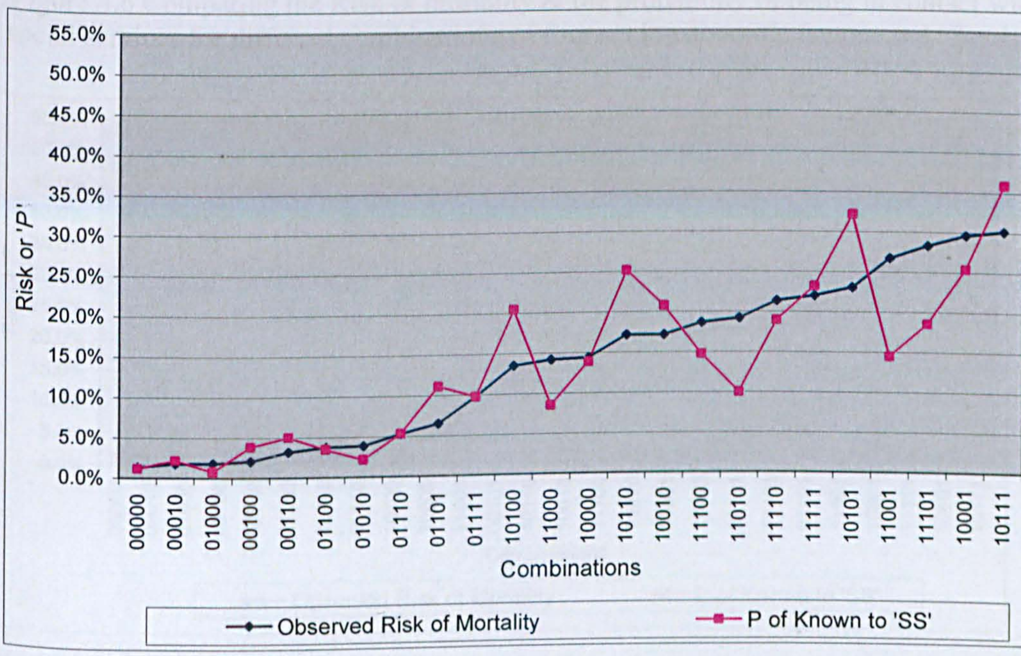
4.4.3 Comparison with four socio-economic factors and incidence of an admission for a 'Heart disease'

Figure 4.4 illustrates the comparison of the 'observed risk' of mortality (from Risk ladder 1.3, Table 4.4) and the 'probability of being in contact with social services' (from Risk ladder 2.3, Table 4.9). Eight combinations with confidence interval with the values greater than or equal to one or less than or equal to zero are excluded from the chart.

The analysis of the following chart make it clear that from combinations with code '10100' onward, all points on the social services curve located above the risk of mortality curve are a combination of female gender with other factors. The same tendency is also true for those points on the social services curve, located under the mortality curve, but this time all points are a combination of male gender and four other factors. The comparison of the two risk ladders again confirms that females are more in contact with the social services than males.

The variable council tax band is also influences the resources provided by social services. Figure 4.4 shows where the probability of being in contact with social services is higher than the risk of mortality for the same group of people, where the tax band code is '1', indicating lower tax bands: A-C. The same process is also true for those points located under the risk of mortality curve, indicating higher tax bands: D-H.

Figure 4.5 Comparing the Risk of mortality & the probability of being in contact with social services for different combinations of four socio-economic factors and '*Heart Disease*'



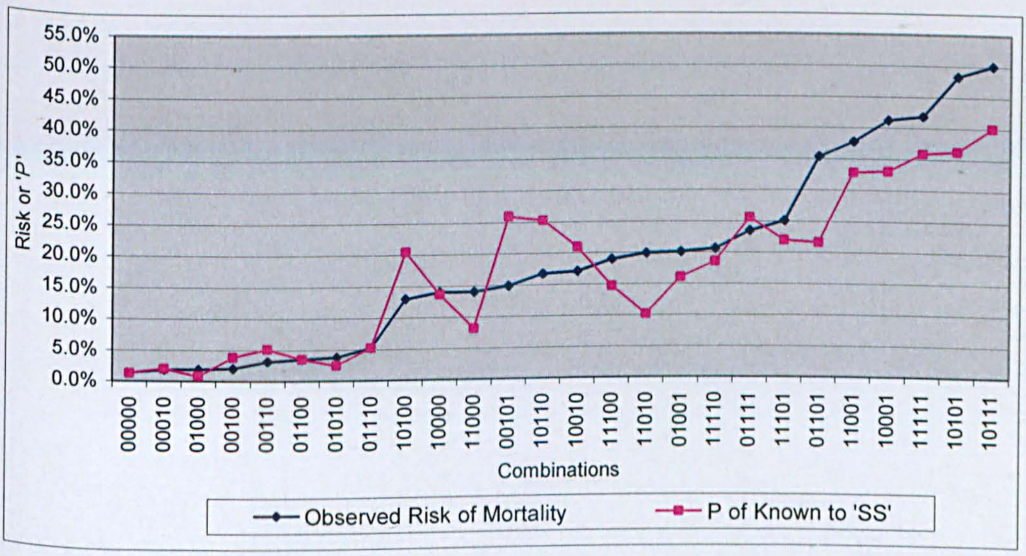
Once again the variable tenure is an important factor on allocation of social services' resources. Most of the combinations with social housing tenure are located above the 'risk of mortality' line; whereas most of points under the mortality line contain private housing tenure.

4.4.4 Comparison with four socio-economic factors and incidence of an admission for a 'Stroke'

Figure 4.5 contains the observed risk of mortality and the probability of being in contact with social services for 26 combinations of four socio-economic factors and 'Stroke'. Six combinations with the confidence interval including values greater than or equal to one or less than or equal to zero are excluded from the chart.

In twelve combinations on the right hand side of the Figure 4.5, only one point is located above the mortality curve and the reminder of combinations (eleven combinations) are under it. Within the eleven combinations, eight combinations include male as gender and the other three combinations include female. On the other hand, those points positioned above the risk of mortality curve, are a combination of different factors include females living in social housing.

Figure 4.6 Comparing the Risk of mortality & the probability of being in contact with social services for different combinations of four socio-economic factors and '*Strokes*'



It is noteworthy that Figure 4.5 shows a relatively (by comparing the levels of risk of mortality with probability of being in contact with social services) smaller percentage of people (with incidence of admission to hospital for stroke) in contact with the social services than the two previous figures (Figures 4.3 and 4.4 based on falls and heart disease). A possible reason for why these groups of people are less in contact with social services could be its relationship with the factors 'age' and 'gender'. This issue will be discussed further in the policy chapter (Chapter 7).

Another factor that influences the outcome in this case, is tax banding. Approximately two thirds of the combinations with low percentage use of social services live in higher tax bands (i.e. they are wealthier).

As mentioned earlier there are some statistical implications in relation to the interval estimation for the binomial proportions, which will be discussed in next section.

4.5 The use and interpretation of confidence interval estimates

The number of cases in any combination of the factors in the risk ladders can range from a small number of people to many thousands. This means that the statistical confidence we have in any given level of risk will vary depending on the sample size and the level of risk/probability. Confidence intervals (or p-values) can be used whenever there is a need to describe the uncertainty in a point estimate wherever the estimate is derived from a sample. The data used in this study refers to all deaths, uses of social services and admission to the hospital for the population of Camden's older citizens (aged over 50 years) living in non-institutionalised residences for a three year period (2002-04). Our rationale for using confidence intervals rests upon the assumption that these data represent a single sample of the older population Camden for a specific time interval. Our population of inference would be considered to be a broader universe 'in time' or a super population (Cassel et al., 1977).

The overall probabilities of 'mortality' and 'contact with social services' for various combinations of risk factors have been presented in terms of the rank order of the resulting point estimate in each risk ladder. Caution must be exercised whenever examining these tables as the range or width of the confidence intervals will sometimes overlap. For example in Table 4.2 the confidence intervals suggest that there is no real statistical difference between sequences 1-7 or in Table 4.3, the implication that sequences 16-18 should be below sequences 19-24. Appendix-E includes a graphical illustration of the observed probabilities for each risk ladder presented in this chapter together with their estimated 95% confidence intervals (or 'high-low' bars). The graphs display the point estimates of risk are likely to have overlapping interval estimates and therefore can not be distinguished.

4.6 Methodological considerations about Confidence Interval estimation for binomial proportions

Some combinations in risk ladders 1.2 to 1.5 (tables 4.3-4.6) have no reported deaths (e.g. sequence number 1 in risk ladder-1.2 & 1.3 and sequence numbers 1 and 2 in risk ladder-1.4). Their resulting probability value and confidence interval can not be calculated and so are set to '0'. There are also some other combinations in the risk ladders for which there are only small number of reported deaths (e.g. sequence numbers 10, 11, 23 in risk ladder-1.2 and sequence numbers 9, 12, 13 and 15 in risk ladder-1.3). The level of observed risk for these cases is very low which may result in their lower bound confidence interval being negative. A third group for which the number of reported deaths compared with the total population are very close to each other (e.g. sequence number 32 in risk ladder-1.4 and sequence numbers 60-63 in risk ladder-5), the upper confidence interval may be greater than 100%. Similar cases can also be found in risk ladders 2.2 – 2.4 with the outcome variable 'social services'.

The reason for including the above three groups from the risk ladders, is for completeness. The reason why confidence intervals can fall outside the conventional range of 0 -1, is because the Wald formula is an approximation and breaks down for very small and very large values of p.

In relation to the above mentioned cases, Agresti & Coull (1998) recommend a method they term the modified Wald method. These authors argue that the confidence intervals may include numbers greater than 1 or negative numbers, but proportions obviously cannot exceed 1 or be negative. Agresti & Coull (1998) and similarly Altman (1999; 1999) also suggest where the lower limit calculated using modified Wald method is less than zero, set the lower limit to 0.0 and similarly where the calculated upper limit is greater than 1.0, set the upper limit to 1.0.

The probability assigned to negative numbers and numbers greater than 1 is usually small when the sample size is large and the proportion being estimated is not too close to 0 or 1". Lawrence et al (2001) in highlighting the problem with standard confidence interval claim that: "We begin by showing that the chaotic coverage properties of the Wald interval are far more persistent than is appreciated. Furthermore, common textbook prescriptions regarding its safety are misleading and defective in several respects and can not be trusted".

Lawrence et al (2001) also explain the non-negligible oscillation of actual coverage probability of confidence interval as a result of variation of both p and n (which an example of it from this work was mentioned earlier), add: "There exist some 'Lucky' pairs *(p, n)* such that the actual coverage probability *C(p, n)* is very close to or larger than the nominal level. On the other hand, there also exist 'unlucky' pairs *(p, n)* such that the corresponding *C (p, n)* is much smaller than the nominal level...Furthermore, drastic changes in coverage occur in nearby p for fixed n and in nearby n for fixed p".

An alternative to the Wald standard interval is the Wilson interval is recommended by Lawrence et al (2001) for a small '*n*'. The difference between the Wald and Wilson intervals is that the Wilson interval only adds two successes and two failures to the observed counts of the adjusted Wald interval, changing the probability or risk probability in 4.1 to:

$$\tilde{p}_i = {x_i + 2}\big/{n_i + 4} \tag{4.4}$$

And the confidence interval for a 95% level of accuracy, can be calculated by

$$CI = \tilde{p}_i \pm \left(1.96\sqrt{\tilde{p}_i(1-\tilde{p}_i)/\tilde{n}_i}\right) \tag{4.5}$$

Where $\tilde{n}_i = n_i + 4$.

For the purpose of comparison risk ladders 1.2-1.4, reproduced with both Wald and Wilson confidence interval, are presented in Appendix-F.

Summary: In this chapter, first the risk ladder methodology was introduced. In section-2 the observed risk of mortality with combinations of different factors (four socio-economic factors and three causes of hospital admissions) with risk ladders were calculated. In Section 4 the probabilities of 'Being in contact with social services' for different combinations were calculated. The output of Section 2 and Section 4 were compared in Section 5 in order to assess whether or not the services provided by the social services department, are allocated to the most vulnerable people. In total, 994 individuals out of 3188 (31%) who died between 2002 and 2004 were in contact with the social services.

Key findings:

The key findings from using the risk ladder methodology in this chapter are:

i) The variables 'age' and causes of hospital admissions (FIS) are strong factors in the determination of both outcomes ('risk of mortality' and 'allocation of social services' resources).

ii) For the variable 'gender', while men are relatively at higher risk of mortality, females have higher chance of being in contact with social services.

iii) Housing tenure and Council tax banding also have a fairly high impact on both outcomes.

In next chapter the relative importance of different variables/risk factors will be assessed using logistic regression.

5 Assessing the relative importance of risk factors

5.1 Introduction

In Chapter 4 risk ladder methodology was introduced and applied to the data. Risk ladder methodology can be thought of as a method of clustering for grouping the population with similar characteristics in order to assess the probability of an adverse event (observed risk) for each group. The similar characteristics for a group of people in this case will be those from the same gender, age group, housing tenure, reason for hospital admission and so on.

In this chapter the relative importance of the risk factors previously used to define groups in risk ladder analysis will be assessed by logistic regression (Altman, 1999; Hosmer & Lemeshow, 2000; Tabachnick & Fidel, 2006). Whereas risk ladders show the clustering of risk for different combinations of risk factors, logistic regression quantifies the relative importance of each risk factor (predictors) and its contribution to overall risk. Further logistic regression enables one to discard weak risk factors and also to predict risk in cases where there are not enough observations.

A forward stepwise logistic regression method (Tabachnick & Fidel, 2006) with four basic socio-demographic variables in the initial model was used to achieve the best possible predictive model from the available data. The improvement of each model is presented by the value of the LogLikelihood (the probability that the observed values of the dependent variable may be predicted from the observed values of the independent variables) and x^2 (the goodness of fit). The Pseudo R^2 (Ender, 2006) value as the percentage of variance in the dependent variable explained by the independent variables, are also provided for each model.

It is also worth noting that for logistic regression models, by applying the prediction equation of the logistic regression (2.5), risk/probability for all covariate patterns (different combination of the factors) can be estimated. An example of the estimated risk of mortality based on logistic regression coefficient for sixteen combinations of four binary socio-economic predictors; age, gender, housing tenure and council tax bands, equivalent to the Risk Ladder -1 (Table 4.2), will be illustrated and compared later in this Chapter (in Section 5.5).

5.1.1 Coding scheme and analytical strategy

Table5.1 includes the coding scheme applied for the variables in the analysis. As all of the variables are categorical, the reference category column identifies the baseline for comparing parameter estimates. In order to use as much of the available information, those variables that have the potential to be divided into more categories such as age, housing tenure and council tax banding, additional categories will be considered further.

Table 5.1 Coding scheme for variables used in logistic regression

Variables Name	Reference-Category	Value
Gender binary (female/male)	Female	0
Age binary (50-69 & 70 years old or more)	50-69	0
Age 3 groups (50-64=1/ 65-79=2/ 80 years old or more = 3)	50-64	1
Age 4 groups (50-64=1/ 65-74=2/ 75-84=3/ 85 years old or more = 4)	50-64	1
Age 5 groups (50-59=1/ 60-69=2/ 70-79=3/80-89=4/90 years old or more = 5)	50-59	1
Tenure binary (private housing/social housing)	Private housing	0
Tenure 3 categories (private housing=1/ housing association = 2 & council housing = 3)	Private housing	1
Tax bands binary (D-H/A-C)	D-H	0
Tax bands 3 categories (A-C=3/ D-E=2/ F-H=1)	F-H	1
Hospital admission for falls - binary (no/yes)	No	0
Hospital admission for heart disease (no/ yes)	No	0
Hospital admission for stroke (no/ yes)	No	0
Social services	In contact (known)	1

The following rules also apply in the proceeding tables (representation of different models):

i) The reference categories are not displayed in the list of variables in the model.

ii) Column (2) contains the estimated 'odds' ratio which compares the odds of an outcome in each category with the reference category.

iii) Column (3) provides a probability value for this estimate.

iv) Column (4) contains a 95% confidence interval, in particular if the lower boundary of the interval falls below 1 there is little statistical evidence for any effect.

All models in this section are created by Stata (StataCorp, 2005). Stata unlike SPSS provides the Log of the Likelihood (LogL) value to summarize the 'goodness of fit' of any model as opposed to -2LogL which is common in other packages like SPSS. Tabachnick and Fidel (2006) define Log-likelihood as sum of the probabilities associated with the predicted and actual outcomes and for each model:

$$\text{Log-likelihood} = \sum_{i=1}^{N}[Y_i \ln(\hat{Y}_i) + (1 - Y_i)\ln(1 - \hat{Y}_i)] \tag{7.1}$$

Where Y_i is the actual outcome for case i and $\hat{Y}$ is the estimated probability that the ith case (I = 1... n) is in one of the (binary) categories. We expect LogL to increase in magnitude (incline towards '0') as more terms are included, similarly, R^2 would increase towards '1' as more terms are added.

The R^2 reported by Stata is McFadden's Pseudo R^2 (Ender, 2006). It compares the likelihood for the intercept only model ($LogL_0$) to the likelihood for the model with all of the predictors for the current model ($LogL_{Full}$) which can be calculated by:

$$Pseudo\, R^2 = 1 - \frac{LogL_{full}}{LogL_0} \tag{7.2}$$

Chi-square (χ^2) is used to assess the relative contribution that different terms make to the model. The conditional χ^2 can only be applied to any model with fewer terms (the smaller model) than the larger model; models are provided 'nested' or 'hierarchical'. i.e. the degrees of freedom (d.f) represent the difference in the number of parameters fitted. The χ^2 also can be calculated as:

$$\chi^2 = 2[(\text{LogL for bigger model}) - (\text{LogL for smaller model})] \tag{7.3}$$

5.1.2 Modelling strategy

In the previous chapter (Chapter 4) with use of risk ladder methodology the observed risk or probability of an outcome for different groups of people (with similar characteristics) were discussed. In this chapter the relative importance of each predictor (risk/protective factor) will be investigated using logistic regression modelling. Whilst risk ladder obviously could identify the spatial differences (inequalities) between different groups of people with the same combination of factors, its limitation is that it can only deal with a certain number of factors (or levels for categorical variables) depending on sample size. If too many factors are included, some estimates may be unreliable due to the small number of observations for some factor combinations. Therefore the information on impact of different factors on those groups of people with small number of observation will be lost. This drop will rise by increasing the number of factors which will result in increasing the number of combinations (consequently, the number of ladders in the table).

In this chapter first the process of model construction will be investigated. The aim of the model construction is to find the model with the most precise prediction of the outcome. In the following sections the logistic regression modelling will be extended step by step and the amount of improvement at each stage will be assessed by the value of Pseudo R^2 and the level of significance of all predictors included in the model. Firstly the relative importance of the four binary (socio-economic) predictors which evolves out of the risk ladder analysis in the previous chapter will be assessed. This basic model is followed by a model with seven binary predictors (including four socio-economic predictors and three causes of hospital admission). Subsequently, more logistic regression models will be introduced by increasing the number of levels for the categorical predictors specifically age, housing tenure and council tax band.

When introducing risk ladder methodology it was more convenient from the perspective of providing a simple illustration of the approach to treat all variables as binary. In this chapter I will begin to discuss the application of logistic regression with a binary value for age in order to assist the exposition of risk ladder methodology. As a general rule to treat age as anything other than continuous would represent a 'loss of information'. In the analysis that follows age is handled first as binary, then as three and five categories. Finally the model with age coded as five categories is compared to a model with continuous age.

For further potential model improvement, two model refinement techniques will be examined: firstly; using age as a covariate predictor; and secondly: interaction terms between variables. Model refinement is also followed by a discussion on advantages and disadvantages of using age as either a continuous or as a categorical variable.

Finally a subset of the models with 'death' as an outcome variable will be used to assess the relative importance of these factors on probability of someone 'being in contact with social services'. The odds ratio of each factor in a model with outcome 'death' will also be compared with the odds ratio of the same factor in its identical model with outcome variable 'being in contact with social services'. The rationale behind this comparison is to examine the possible relationship between the relative importance of each predictor on risk of mortality and allocation of the resources by social services.

The next section will systematically examine different models to test the impact of each factor on mortality in order to find the best model.

5.2 Examining different models in order to find the best model by testing the impact of each factor on mortality

In order to find the best logistic regression model, seven models will be examined in this section. In the first stage, two models with four and seven binary factors are constructed. The first two binary models are followed by five more models by increasing the number of levels for the categorical predictors namely age, housing tenure and council tax band. The outcome we want to predict is mortality.

5.2.1 Findings from logistic regression, all variables binary

Firstly, two models with four socio-economic factors (including gender, age, housing tenure and council tax bands) and seven binary factors (including the above four socio-economic factors plus falls, heart disease and strokes) will be studied.

5.2.1.1 Logistic regression with four socio-economic factors (Model-1)

The first model is based on four socio-economic factors; gender, age, housing tenure and council tax bands. The output in Table 5.2 shows that the effects of all four variables on deaths are highly significant. As is shown in Table 5.2, men are 1.3 times more likely to die compared to women. Older people (70 years old or more) have a 7-fold increase in the odds of mortality as compared with those under 70. Living in social housing increases the probability of death 1.24 times as also does living in lower tax band by 1.33 times.

Table 5.2 Odds ratios based on logistic regression modelling with four basic socio-economic factors with confidence intervals (Model-1)

Death	Odds Ratio	Sig.	[95% Conf. Interval]	
Gender	1.31	0.000	1.21	1.41
Age	7.12	0.000	6.55	7.75
Tenure	1.24	0.000	1.14	1.36
Tax band	1.33	0.000	1.22	1.45

Number of obs = 43472 χ^2 (4 d.f) = 2727.8 , P = 0.0000

Pseudo R^2 = 0.120 Log likelihood = -10033.6

5.2.1.2 Logistic regression with four socio-economic factors and three causes of hospital admission (Model-2)

In the second model, three causes of hospital admission including; falls, ischemic heart disease and strokes have been added to model-1. The output for model-2 is shown in Table 5.3 and shows that all seven variables are highly significant, in particular, 'strokes'. The second model shows an improvement of 1.6% in the Pseudo R^2 as compared with model-1 (the proportion of variance explained by predictors), changing from 12% to 13.6%.

Table 5.3 Odds ratios based on logistic regression modelling with four socio-economic factors and three causes of hospital admissions (Model-2)

Death	Odds Ratio	Sig.	[95% Conf. Interval]	
Gender	1.32	0.000	1.22	1.42
Age	6.50	0.000	5.97	7.08
Tenure	1.21	0.000	1.11	1.32
Tax band	1.34	0.000	1.23	1.46
Falls	2.27	0.000	1.92	2.68
Heart Disease	1.82	0.000	1.51	2.18
Strokes	4.27	0.000	3.55	5.14

Number of obs = 43472 χ^2 (7d.f) = 3100.2, P = 0.0000

Pseudo R^2 = 0.136 Log likelihood = -9847.4

5.2.2 Extending the logistic regression analysis by increasing the number of levels for categorical predictors including age, tax bands and tenure (final model)

In models 1 and 2 discussed earlier, all variables are binary. In order to utilise the amount of information available and increase the value of Pseudo R^2 to maximise the prediction of death for different variables (predictors), it was decided to increase the number of levels of the predictors wherever possible. Some variables such as age, tenure and tax bands can be broken down into more than two categories as follows:

i) Increasing the number of levels for 'age' to 3 categories (Model-3): All variables in this model are the same as the previous model (model-2) except age has been recoded from having two levels to three categories (50-64, 65-79 and 80 years old or more).

The reference category (by default in Stata) is the first category which is 50-64. The new model shows an improvement of 1.9% in Pseudo R^2 compared to model-2 in table 5.3.

ii) Increasing the number of levels for 'age' to 4 categories (Model-4): The age variable in this model has been recoded to four categories including; 50-64 (reference category), 65-74, 75-84 and 85 years old or more. By changing the age from 3 categories in model-3 to 4 categories in model-4, there is an increase of 1.1% in Pseudo R^2, again the effect of all variables on mortality is highly significant.

iii) Increasing the number of levels for 'age' to 5 categories (Model-5): Finally age was divided to 5 categories of 10 year bands from 50 to 89 and 90 years old or more, with age 50-59 the reference category. The model output shows all variables have highly significant impact on deaths outcome but only a small increase (0.2%) in the Pseudo R^2 value.

By including more information about age, we observe (as illustrated in Appendix-G) a slight but steady improvement in 'model fit', and stronger evidence of an age gradient in the estimated odds ratios.

iv) Changing Tax band from binary to three categories (Model-6): The initial aim of using variable tax band as a predictor is the role tax band plays as a 'marker' or proxy of wealth or material circumstances. There are a considerable difference between the price of properties with tax band 'D' or 'E' with the lower or higher tax bands. Therefore in model-6 eight tax bands are divided into three categories; A-C, D-E and F-H (reference category) and all of the other variables remains coded as in model-5.

The change of Pseudo R^2 in this model is not very evident (0.1% improvement). Although all variables still have highly significant effect on predicting death. Appendix-G includes more detailed explanations of model-3 to model-6.

v) Changing tenure from binary to three categories (Model-7, final model): In this model, the number of categories for variable tenure have been increased from two to three including; private housing (reference category), social housing and housing association. The reason for splitting the social housing into two categories in this model is to test whether living in council housing or housing association property have a differential impact upon mortality.

The results show that the change in Pseudo R^2 value is very small (0.08%) in this case. An interesting result in this model is the difference between the risk of living in council housing or a housing association property. While the impact of council housing as a predictor of deaths is no longer significant (p=0.213) suggesting little to distinguish the impact of living in council housing to private housing, the impact of living in a housing association property becomes highly significant, as is shown in Table 5.4. Further discussion of this issue is provided in Chapter 7 (policy implications).

Table 5.4 Odds ratios based on logistic regression modelling by increasing the number of levels for predictors; age, housing tenure and council tax band (Model-7, final model)

Death	Odds Ratio	Sig.	[95% Conf. Interval]	
Gender	1.49	0.000	1.38	1.61
Age-2	2.47	0.000	2.13	2.86
Age-3	6.56	0.000	5.73	7.52
Age-4	14.68	0.000	12.79	16.84
Age-5	33.41	0.000	28.16	39.64
Housing Association	1.43	0.000	1.24	1.66
Council Housing	1.07	0.213	0.96	1.18
Tax band_2 (D-E))	1.35	0.000	1.20	1.51
Tax band_3 (A-C)	1.65	0.000	1.46	1.86
Falls	1.75	0.000	1.47	2.08
Heart Disease	1.76	0.000	1.46	2.11
Strokes	4.03	0.000	3.34	4.86

Number of obs = 43472 χ^2 (12 d.f) = 3864.3, P = 0.0000

Pseudo R^2 = 0.170 Log likelihood = -9465.3

We will now discuss the impact of age as a continuous variable or covariate (as an alternative to the categorical age) and the interaction terms on further attempts at improving our models will be discussed.

5.3 Further Model Refinement

This section will firstly examine the effect of the variable 'age' as a continuous variable on model improvement. Thus, the models discussed in Section 5.2 (model-1 with four socio-economic binary factors, model-2 with seven binary factors and model-7 the final model) will be compared using identical models but simply changing the age from 'dichotomous' to 'continuous' variable. Then the impact of the different variables on each other and consequently on the dependent variable (in this study, death) with use of interaction effects between different variables will be assessed.

The rationale behind choosing the three models (model-1, -2 and -7) is due to the fact that these three models include distinguishable components from one another. Model-1 starts with four binary socio-economic variables and in Model-2 three more binary variables (three causes of hospital admission) are added. Finally those categorical variables which potentially could be expanded to more levels (as long as the contribution of the predictor variables remains significant), were extended and model-7 was created. The differences between the other four models (model-3, -4, -5 and -6) are negligible; and so they will be excluded from any further investigation.

5.3.1 Creating logistic regression models with a continuous variable 'age'; comparing these models with previous models (with dichotomous 'age')

i) Comparing the result of the first logistic model with four variables

Table 5.5 illustrates the output of two logistic regressions with four socio-economic factors (age, gender, housing tenure and council tax bands); Table 5.5a shows the logistic model with three binary factors and age as a continuous variable, Table 5.5b is a copy of model-1 (considered in Section 5.2) with four binary factors.

By comparing the two models in Table 5.5, the odds ratio (OR) in Table 5.5-a with the continuous variable age shows a higher probability of death for male than the Table

5.5-b with binary age. The differences between the OR of tenure and tax bands for two models is not very large.

Table 5.5 Comparison of two logistic regression models with four factors; *a)* with 3 binary factors and the continuous variable age *b)* all 4 factors are binary.

Death	Odds Ratio	Sig.
Gender	1.53	0.000
Age	1.10	0.000
Tenure	1.32	0.000
Tax band	1.30	0.000

$R^2 = 0.163$

(a)

Death	Odds Ratio	Sig.
Gender	1.31	0.000
Age	7.12	0.000
Tenure	1.24	0.000
Tax band	1.33	0.000

$R^2 = 0.120$

(b)

The Pseudo R^2 for the model with the continuous variable age shows a 4.3% increase compared to the model based on use of binary age and thus provides a more accurate explanation of the dependent variable (death) by the predictors and better adjustment of variables. The OR for Gender changes from 1.31 in the model with binary age to 1.53 in the model with age as a continuous variable which indicates a higher risk of mortality for males. In the model with the continuous variable age, the OR for tenure also increases but the OR for tax band decreases from 1.33 to 1.3.

ii) Comparing the result of the second logistic models with seven factors

Table 5.6 illustrates the output of two logistic regressions with seven factors including the four socio-economic factors and three causes of hospital admission (FIS). In Table 5.6-a the factor age is a continuous variable and in Table 5.6-b all factors are binary (discussed in Section 5.2).

There is a noticeable improvement in Pseudo R^2 in the model using age as a continuous variable of about 4%. Again in the model with the continuous variable age the OR for 'Gender' increased from 1.32 to 1.52, showing higher risk of death for males. In this model, the OR for tenure also shows an increase but for all other factors there is a decline in ORs. Among the factors which had a decrease in terms of their OR value, the changes for falls and strokes were considerable. The OR for falls reduces from 2.27 in the model with binary age to 1.67 in the model with the age as a continuous variable. For stroke it also changes from 4.27 to 3.87. The changes in ORs for other variables are not very large.

Table 5.6 Comparison of two logistic regression models with seven factors; *a)* with age as a continuous variable *b)* all 7 factors are binary.

Death	Odds Ratio	Sig.
Gender	1.52	0.000
Age	1.10	0.000
Tenure	1.28	0.000
Tax band	1.30	0.000
Falls	1.67	0.000
Heart Diseas	1.71	0.000
Strokes	3.87	0.000

$R^2 = 0.175$

(a)

Death	Odds Ratio	Sig.
Gender	1.32	0.000
Age	6.50	0.000
Tenure	1.21	0.000
Tax band	1.34	0.000
Falls	2.27	0.000
Heart Diseas	1.82	0.000
Strokes	4.27	0.000

$R^2 = 0.136$

(b)

iii) Comparing the result of the two logistic models with increasing the levels of categories for tenure, tax bands and age to more than two categories

Table 5.7 illustrates the output of two logistic regressions with all seven factors. In Table 5.7-a the number of levels for tenure and tax bands is increased to three categories but the age is a continuous variable and in Table 5.7-b the number of levels for age is also increased to five categories (equal to model-7 in Section 5.2).

By comparing the above two models it is clear that the value of OR for some factors such as tax band are exactly the same and for some other factors the terms are very close (e.g. for tenure).

The difference between the Pseudo R^2 value of the above two models is only 0.7%. However the advantage of the model with dichotomous age in five categories (Table 5.7-b) is that the OR for different age groups clearly explains the variation of the effect of age on mortality for these groups and therefore clearer from a presentational stand point.

Table 5.7 Comparison of two logistic regression models *a)* a model similar to the final model in Section 5-2 except the variable age in this model is a continuous variable *b)* a model identical to the final model in Section 5-2.

Death	Odds Ratio	Sig.
Gender	1.53	0.000
Age	1.10	0.000
Housing Association	1.42	0.000
Council Housing	1.08	0.166
Tax band_2 (D-E)	1.35	0.000
Tax band_3 (A-C)	1.65	0.000
Falls	1.67	0.000
Heart Diseas	1.72	0.000
Strokes	3.85	0.000

$R^2 = 0.177$

(a)

Death	Odds Ratio	Sig.
Gender	1.49	0.000
Age-2	2.47	0.000
Age-3	6.56	0.000
Age-4	14.68	0.000
Age-5	33.41	0.000
Housing Association	1.43	0.000
Council Housing	1.07	0.213
Tax band_2 (D-E)	1.35	0.000
Tax band_3 (A-C)	1.65	0.000
Falls	1.75	0.000
Heart Diseas	1.76	0.000
Strokes	4.03	0.000

$R^2 = 0.170$

(b)

In further investigation, the model illustrated in Table 5.7-a, was used independently for each of the five different age categories of the final model (illustrated in Table 5.7-b). The output is shown in Table 5.8 below. In Table 5.7-a, the OR age as a continuous variable is 1.10 while in Table 5.8 it varies for the five different age groups from 1.07 to 1.10. Table 5.8 clearly illustrates the better adjustments of OR for all variables when the logistic regression model is used for different age groups independently. In Table 5.8 the cells containing the OR of those variables that are not significant at the 0.05 level of probability, are highlighted. One reason for these ORs not being significant could be that the data set used for this study does not include enough cases for all groups (e.g. those aged over 90 years). However for a larger data set, where there are enough cases of all different groups for analysis, the above method could be ideal.

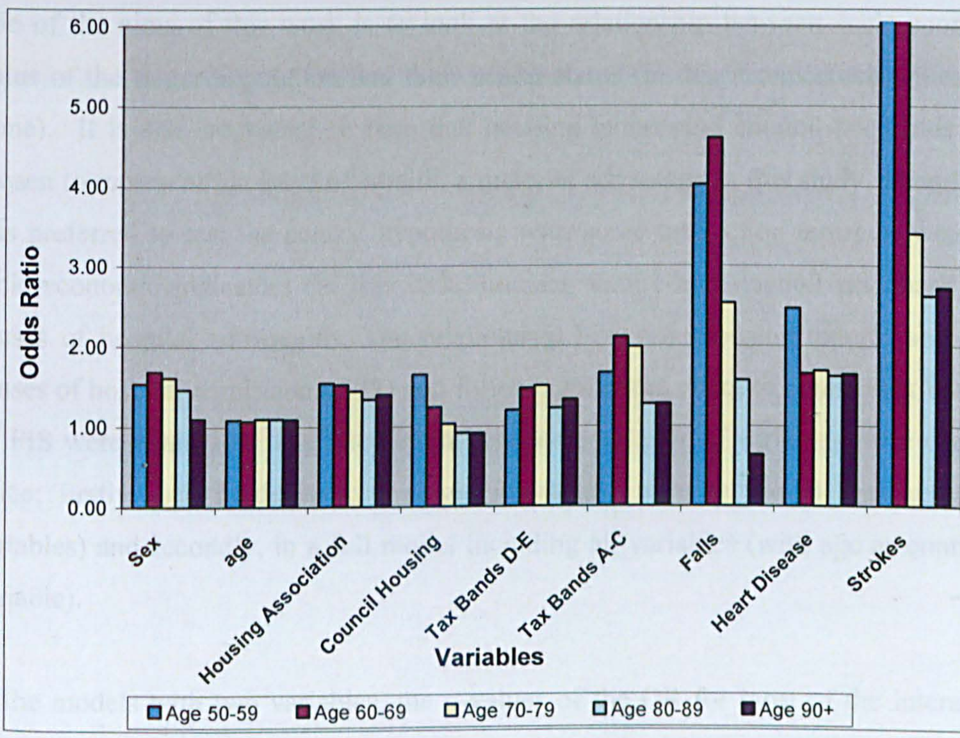
Table 5.8 The comparison of the Odds Ratio and p-value of all factors for 5 different Age groups

Age Group	50-59		60-69		70-79		80-89		90+	
Outcome-Death	OR	Sig.	OR	Sig.	OR	Sig.	OR	Sig.	OR	Sig.
Gender	1.54	0.001	1.69	0.000	1.61	0.000	1.46	0.000	1.10	0.479
age	1.08	0.000	1.07	0.000	1.10	0.000	1.10	0.000	1.09	0.000
Housing Association	1.54	0.066	1.52	0.016	1.44	0.007	1.34	0.039	1.40	0.155
Council Housing	1.65	0.002	1.24	0.082	1.03	0.728	0.92	0.396	0.91	0.593
Tax Bands D-E	1.20	0.308	1.47	0.006	1.50	0.000	1.23	0.052	1.33	0.097
Tax Bands A-C	1.67	0.007	2.11	0.000	1.99	0.000	1.27	0.039	1.28	0.219
Falls	4.02	0.001	4.60	0.000	2.54	0.000	1.57	0.000	0.64	0.029
Heart Disease	2.47	0.011	1.66	0.039	1.69	0.001	1.62	0.003	1.65	0.072
Strokes	14.91	0.000	6.11	0.000	3.41	0.000	2.63	0.000	2.73	0.001

Whilst the estimated coefficients (exponential (B_j's)) for age appear constant within each age group, the relative importance of age in terms of additional years within a category will be different. Equally, it is not appropriate to simply compare the absolute difference between estimated ORs between categories as the OR is multiplicative.

The output of Table 5.8 is also illustrated in Figure-5.1. In Figure-5.1 the OR of each variable for five different age groups is demonstrated. OR with value of '1' is the reference for all cases.

Figure-5.1 Illustration of Table 5.8



Comparing the ORs of four different age groups (by excluding those aged 90 years or more) in Table 5.8 and Figure-5.1 shows the gap, in this case inequalities, within younger age groups (under 70s) is higher than the inequalities within older age groups (70 years old or more). This inequality, particularly for those aged 80-89, is much smaller. The differences in ORs for variables gender, housing association, tax bands A-C and three causes of hospital admissions are more evident. The reason for exclusion of age group 90 years old or more is that there are not enough cases in the data for this age group and as a result the p-value for most of variables in relation to this age group is not significant. .

5.3.2 Examining the interaction effects

In a statistical model an interaction is a term in which the effect of two or more variables is assumed to be multiplicative in which effect of one variable on the outcome depends on the value of another variable (Box, 1990; Hosmer & Lemeshow, 2000).

One of the aims of this work is to look at the relationship between socio-economic status of the target population and their health status (in this instance using mortality alone). It is also important to note that housing tenure and council tax bands were chosen to represent the level of wealth, a material advantage in this study. Therefore it was preferred to test the central hypothesis with some interaction terms between the socio-economic indicators (in this case, housing tenure and council tax bands) and causes of hospital admissions. The relationship between 'housing tenure' and three causes of hospital admissions (FIS) and following this the effect of council tax banding on FIS were examined. The interaction between each pair of variables were checked twice; firstly as a model with a single interaction term (a model with only two variables) and secondly, in a full model including all variables (with age as continues variable).

In the models with two variables, the p-values of the OR for most of the interaction terms are significant or highly significant but the Pseudo R^2 values for the six different models vary between 1% and 3% which is not very high. However, when all the net impact on interaction terms are included in a full model (similar to the final model in Section 5.2 but with the continuous variable age) the ORs for most of the interaction terms are no longer significant. The Pseudo R^2 value for all models is broadly the same as the model in Table 5.7a (17.7%). Appendix-H includes a detailed explanation of all models tested with interaction terms.

For the purpose of further exploration, a model with categorical age (model-7 in Section 5.2) was examined by adding an interaction term between tenure (with 3 categories) and age (with 5 categories), the number of variables increases to 20. The new model includes 8 more variables than the model-7. The Pseudo R^2 value increases very little (from 16.95% to 17.14%) and the impact of five of the interaction terms on predicting mortality are not significant, with the p-value of greater than 0.05. However the joint effect of council housing and being in an older age group (70 years and above), is highly significant and would indeed lead to a reduction in the relative odds for these groups. In Table 5.9 the non-significant variables are highlighted.

As we can see in Table 5.9, all interactions between tenure-2 (council housing) and all 4 age groups (2 to 5) including the interaction between tenure-3 (housing association) and age group-2 are not significant. Thus, when a logistic regression model with more variables is used, the OR for each variable will be more adjusted and consequently the joint effect of the different variables on each other (interaction) decreases.

Table 5.9 Odds ratios based on logistic regression modelling equal to the final model (in Section-2) including an interaction between tenure (with 3 categories) and age (with 5 categories)

Death	Odds Ratio	Sig.	[95% Conf.	Interval]
Gender	1.49	0.000	1.38	1.61
Age-2	2.64	0.000	2.05	3.41
Age-3	8.07	0.000	6.38	10.21
Age-4	22.28	0.000	17.69	28.06
Age-5	48.19	0.000	36.85	63.01
Housing Association (Ten_2)	1.58	0.043	1.02	2.45
Council Housing (Ten_3)	1.66	0.000	1.28	2.15
Tax band_2 (D-E)	1.35	0.000	1.21	1.52
Tax band_3 (A-C)	1.67	0.000	1.48	1.89
Falls	1.74	0.000	1.47	2.07
Heart Disease	1.75	0.000	1.45	2.10
Strokes	3.97	0.000	3.29	4.79
Ten2 X Age2	1.12	0.675	0.65	1.93
Ten2 X Age3	0.97	0.912	0.59	1.61
Ten2 X Age4	0.75	0.274	0.45	1.25
Ten2 X Age5	0.79	0.444	0.43	1.45
Ten3 X Age2	0.86	0.371	0.63	1.19
Ten3 X Age3	0.68	0.010	0.50	0.91
Ten3 X Age4	0.47	0.000	0.35	0.64
Ten3 X Age5	0.50	0.000	0.35	0.71

Number of obs = 43472 x^2 (20) = 3906.5 R^2 = 0.171

Log likelihood = -9444.3 Prob > x^2 = 0.0000

In order to examine the association between other variables, one more interaction between tax band and gender were added to the previous model. The number of variable in the model increased to 22 and the p-values for seven variables were not significant. The Pseudo R^2 value increased by 0.07% from 17.14% to 17.21%.

In terms of the improvement of Pseudo R^2, introducing more complexity into the model does not have a significant effect on our outcome variable. As Tabachnick & Fidel (2006) state; "interactions may complicate a model without reliably improving the prediction". Thus far, it is therefore evident that model-7 is the best fitting model.

By examining additional models with interactions, improvements in Pseudo R^2 were obtained but some terms were not significant any more ($p > 5\%$) as highlighted in Table 5.9.

In next section the effect of different factors on a person being in contact with social services will be examined using the most appropriate models discussed in the previous two sections (Sections-5.2 and 5.3).

5.4 Examining the relative impact of each factor on whether or not someone is in contact with social services

Different logistic regression models with variables extracted from the available data set for the outcome variable mortality have already been discussed in Section-2. In Section-3 the possibility of further model refinement with use of a continuous variable for age and various interaction terms were discussed. In this section models 1, 2 and 7 of Section-2 are reconstructed by changing the outcome variable from 'mortality' to 'being in contact with social services'.

The aim is to examine the relationship between the relative importance of risk factors and the allocation of resources by 'social services'. The hypothesis to be tested is that those groups of the population with higher risk of death have a tendency to be more in contact with social services than those at lower risk of death.

In order to test the above hypothesis and consequently to assess the possible association between risk of death and allocation of the social services' resources, the results (odds ratio) of the three models for each outcome (mortality and social services) are compared in the following three Tables 5.10, 5.11 and 5.12.

5.4.1 Comparing the result of the two logistic regression models with four binary variables and different outcome variables ('mortality' and 'social services')

Table 5.10 illustrates the output of two logistic regressions with four binary socio-economic factors (age, gender, housing tenure and council tax bands); in Table 5.10a the outcome variable is 'social services' and in Table 5.10-b is 'mortality' (a copy of model-1 in Section-2).

By comparing the result of the two logistic regressions for each variable in Table 5.10-a and 5.10-b, I show that females tend to be much more in contact with the social services than males. Furthermore, while the risk of mortality for males is 1.31 times higher than females, the probability of females being in contact with social services is

(1/0.71 = 1.41) times higher than it is for males. The product of the above two probabilities indicates that assuming the equal chance of being in contact with social services for both genders based on risk of mortality, the females are 1.85 times more than males in contact with the social services.

The estimated probability of being in contact with social services according to age (6.5 times more for those aged 70 years old or more than those between 50-69 years old) is close to the estimated probability of mortality for the same age groups (with an OR of 7.12).

Based on tenure, those living in social housing are 1.6 times more in benefit of the social services than those in private housing (including owner occupied and private renting), comparing with the risk of mortality for these two groups.

The estimated probability of the variables council tax banding for both outcomes (mortality and social services) is very close to each other, but this can not be said for tenure.

Table 5.10 Odds ratios based on logistic regression modelling with four binary variables; a) the outcome variable, 'social services' b) the outcome variable, 'mortality.

Social Services	Odds Ratio
Gender	0.71
Age	6.50
Tenure	1.98
Tax band	1.38

(a)

Death	Odds Ratio
Gender	1.31
Age	7.12
Tenure	1.24
Tax band	1.33

(b)

5.4.2 Comparing the result of the two logistic regression models with seven binary variables and different outcome variables ('mortality' and 'social services')

Table 5.11 illustrates the output of two logistic regressions with seven binary factors including the four socio-economic factors and three causes of hospital admissions (FIS). In Table 5.11a the outcome variable is 'social services' and in Table 5.11-b is 'mortality'.(a copy of model-2 in Section-2).

The outputs from the following two models for four socio-economic factors are very similar to the previous two models in Table 5.10. By comparing the OR of the three causes of hospital admissions in two models it shows a 1.87 times increase in social services usage for falls' patients, 0.63 times increase (or 1.58 times decrease) for stroke patients and for heart disease it shows about the same level as the risk of mortality.

Table 5.11 Odds ratios based on logistic regression modelling with seven binary variables; a) the outcome variable, 'social services' b) the outcome variable, 'mortality.

Social Services	Odds Ratio
Gender	0.72
Age	5.82
Tenure	1.98
Tax band	1.38
Falls	4.24
Heart Diseas	1.71
Strokes	2.70

(a)

Death	Odds Ratio
Gender	1.32
Age	6.50
Tenure	1.21
Tax band	1.34
Falls	2.27
Heart Diseas	1.82
Strokes	4.27

(b)

5.4.3 Comparing the result of the two logistic regression models (final models) with different outcome variables ('mortality' and 'social services')

Table 5.12 illustrates the output of two logistic regressions with all factors included. The number of levels for age, tenure and tax bands is increased to more than two categories (equal to model-7 in Section-2). In Table 5.12-a the outcome variable is 'social services' and in Table 5.12-b is 'mortality' (a copy of model-7 in Section-2).

Here again by comparing the ORs of two identical factors in Table 5.12-a (with outcome 'social services') and Table 5.12-b (with outcomes 'mortality'), like previous models in Table 5.11, the disparity between risk of death and probability of being in contact with social services for some factors including: housing tenure, falls and stroke is evident. The estimated probability of being in contact with social services for all four age categories, are close to the estimated probability of mortality for the same age groups which is reasonable. In general the ORs for the same factor with different outcome variables ('mortality' and 'social services') in Table 5.12 is similar to Table 5.11. However the ORs of different factors in models introduced in Table 5.12 are adjusted more precisely (the extent to which each predictor is adjusted for the impact of the other predictors, leading to an improvement in 'model fit') as discussed earlier in Section 5.2.

Table 5.12 Odds ratios based on logistic regression modelling with age, tenure and tax band more than two categories; a) the outcome variable, 'social services' b) the outcome variable, 'mortality.

Social Services	Odds Ratio
Gender	0.80
Age-2	2.17
Age-3	5.05
Age-4	12.64
Age-5	27.41
Housing Association	2.12
Council Housing	1.71
Tax band_2 (D-E)	1.54
Tax band_3 (A-C)	1.90
Falls	3.38
Heart Diseas	1.66
Strokes	2.53

(a)

Death	Odds Ratio
Gender	1.49
Age-2	2.47
Age-3	6.56
Age-4	14.68
Age-5	33.41
Housing Association	1.43
Council Housing	1.07
Tax band_2 (D-E)	1.35
Tax band_3 (A-C)	1.65
Falls	1.75
Heart Diseas	1.76
Strokes	4.03

(b)

5.4.4 Comparing the result of the final model including a continuous variable for age

Here once again two final models with different outcomes and age as a continuous variable are contrasted. Table 5.13 shows these two models; Table 5.13-a is the model with outcome variable 'social services' and Figure Table 5.13-b is the model with 'Mortality' as outcome variable.

Table 5.13 Odds ratios based on logistic regression modelling with the continuous variable age a) the outcome variable, 'social services' b) the outcome variable, 'mortality.

Social Services	Odds Ratio
Gender	0.82
Age	1.09
Housing Association	2.10
Council Housing	1.71
Tax band_2 (D-E)	1.55
Tax band_3 (A-C)	1.91
Falls	3.28
Heart Diseas	1.60
Strokes	2.39

$R^2 = 0.199$

(a)

Death	Odds Ratio
Gender	1.53
Age	1.10
Housing Association	1.42
Council Housing	1.08
Tax band_2 (D-E)	1.35
Tax band_3 (A-C)	1.65
Falls	1.67
Heart Diseas	1.72
Strokes	3.85

$R^2 = 0.177$

(b)

Comparing the differences between the ORs for each variable in the above two models (in Table 5.13) with the differences between the ORs for the same variables in the models illustrated in Table 5.12 (with categorical age), a high level of similarity is evident.

5.5 Risk/probability estimation of covariate patterns

In the Introduction to this chapter it was noted that for logistic regression models, by applying the prediction equation of the logistic regression (2.5), risk/probability for all covariate patterns (different combination of the factors) can be estimated. Risk of mortality for sixteen combinations of four binary socio-economic predictors; age, gender, housing tenure and council tax bands (Model-1, Table 5.2) which is equivalent to the Risk Ladder -1 (Table 4.2), are estimated and are illustrated in Table 5.14 bellow. Table 5.14 also includes the observed probability of Risk ladder-1 for the purpose of comparison.

Table 5.14 Comparing observed probability of mortality for sixteen different combinations (groups) of four basic socio-economic factors (produced in risk ladder-1, Table 4.2) with estimated probability by logistic regression for the same combinations of the same variables for people aged 50 years old or more in London Borough of Camden by observed probability in ascending order.

Seq	Age	Gender	Tenure	Tax Band	Observed Probability	Estimated Probability	Observed Probability %	Estimated Probability %
1	0	0	0	0	0.0143233	0.019048	1.4%	1.9%
2	0	0	0	1	0.0180723	0.025263	1.8%	2.5%
3	0	1	0	0	0.0190651	0.024804	1.9%	2.5%
4	0	0	1	0	0.0208529	0.023596	2.1%	2.4%
5	0	0	1	1	0.0324721	0.031247	3.2%	3.1%
6	0	1	1	0	0.0382375	0.030683	3.8%	3.1%
7	0	1	0	1	0.039726	0.032834	4.0%	3.3%
8	0	1	1	1	0.0540462	0.040536	5.4%	4.1%
9	1	0	1	0	0.1380629	0.146814	13.8%	14.7%
10	1	1	0	0	0.1454687	0.153341	14.5%	15.3%
11	1	0	0	0	0.147125	0.121473	14.7%	12.1%
12	1	0	1	1	0.1761866	0.186777	17.6%	18.7%
13	1	0	0	1	0.1800643	0.155797	18.0%	15.6%
14	1	1	1	0	0.191247	0.183939	19.1%	18.4%
15	1	1	0	1	0.1975806	0.194675	19.8%	19.5%
16	1	1	1	1	0.2131952	0.231268	21.3%	23.1%

The observed and estimated probability/risk for most of the sixteen combinations in Table 5.14 are very close to each other indicating the consistency of the model.

Summary

In this chapter, seven models were used in a 'forward selection logistic regression' to assess the relative importance of risk factors and their contribution to overall risk of mortality, before selecting the best models. Then, two statistical enhancement tools including the use of the continuous variable instead of the categorical variable (for variable age) and interaction terms were examined. It was concluded that in this study, these two refinements did not have a significant impact on model improvement. The relationship between each variable (risk factors) and mortality has also been discussed through each model.

Subsequently the relationship between each variable and the 'social services' were studied using four models developed in earlier stage (with mortality outcome). These four models include model-1, -2, -7 (final) and one similar to the model-7 but with the continuous variable age. At the same time, each of these four models was contrasted with its identical model for the 'Mortality' outcome. The reason for this comparison was explained in the introduction to Section 4.

Key findings

The outcome of the final model (model-7) with mortality as an outcome illustrated in table 5.4 shows all factors and their relevant categories used in this study except for 'council housing' are highly significant in increasing the risk of mortality. It has also been confirmed that all variables (factors) and their relevant categories with 'social service' outcome are also highly significant. However, there are some disparities found by comparing the two final models with different outcomes ('mortality' and 'social services'). These disparities in the role played by the following predictors; gender, housing tenure, falls and strokes are quite noticeable. Their implications will be discussed in the chapter related to policy implication (Chapter-7).

The examination of the model refinement by using age as a continuous variable and interaction effects also shows no significant impact on model improvement.

In the next Chapter the Receiver Operating Characteristic Curve (ROC) will be introduced as an evaluation tool and will subsequently be used to evaluate the findings of Chapters 4 and 5.

6 Evaluation of the models and their outcomes

6.1 Introduction

The aim of this chapter is to investigate how well the models in previous chapter are able to discriminate between those subjects who experience the outcome of interest versus those who do not (Hosmer & Lemeshow, 2000). The first part of this chapter will consist of a discussion of Receiver Operating Characteristic (ROC) curves and how they will be used as a tool to evaluate the outcome of the models analysed in the previous chapters. Following this exploration, I will then use the 'Gini coefficient'; a measure of inequality which usually used to measure income inequality, as an alternative evaluation tool and explore its relationship with ROC curve.

ROC curves are used as a tool to evaluate the discrimination ability of various statistical methods that combine a variety of evidence, test results, etc. for predictive purposes (Hanley & McNeil, 1982). The history of ROC goes back to the World War II and based on Mason & Graham (2002) "...it was first employed in the study of discriminator systems for the detection of radio signals in the presence of noise in the 1940s, following the attack on Pearl Harbor. The initial research was motivated by the desire to determine how the US radar (receiver operators) had missed the Japanese aircraft".

The construction of the ROC curves depends on the relationship between the sensitivity and specificity of an outcome. Where sensitivity indicates how likely the outcome of a test will be positive for actual positive cases and specificity indicates how likely the outcome of a test will be negative for actual negative cases (Peat & Barton, 2005). Peat & Barton (2005) state that sensitivity and specificity are used to estimate the utility of a test in predicting the presence of a condition or a disease. They also continue: "...If the outcome...is binary, a likelihood ratio (LR) can be calculated directly. If the test result is on a continuous scale, a ROC curve is used to determine the point that maximizes the LR". The ROC curve by Pepe (2003) has been defined as a graphical plot of the sensitivity vs. (1 - specificity) for a binary classifier system as its discrimination threshold is varied.

ROC curves can also be defined as a graphical representation of the trade-off between sensitivity and specificity. "It plots the probability of detecting true signal (sensitivity) and false signal (1-specificity) for an entire range of possible cut-points" (Hosmer & Lemeshow, 2000).

ROC curves have been used in psychophysics, to assess the detection of weak signals in humans (and occasionally animals) since 1960s. They are also used for the evaluation of machine learning such as internet search engines, epidemiology and medical research (Pepe, 2003).

The aim of this chapter is to evaluate the result of the estimated probabilities of different models produced in Chapter 5, by ROC curve. In addition, through use of ROC curves, the discrimination ability of one of the risk ladders (with four socio-economic factors, Table 4.2) will be compared with the discrimination ability of the identical model derived from logistic regression modeling. However, before going into a detailed explanation of the ROC, it is necessary to explain and define the concepts of sensitivity and specificity which derive from epidemiology and the use of screening tests. In addition, the calculation of sensitivity and specificity and the steps need to be taken in the construction of a ROC curve also will be discussed in this chapter.

In the next section (Section 6.2) the relevant terminologies related to the application of the ROC curves will be introduced.

6.2 Terminologies and definition

The basis of the ROC curve is the classification table. Peat and Barton (2005) state that for diagnostic statistics, it is best to code the variable indicating 'disease status' (as present or absent) and 'test result' (as positive or negative). In our case instead of variables 'disease status' and 'test result'; we can use two binary variables 'mortality' (as deceased or living) and 'housing tenure' (as private housing or social housing). This coding can be presented as a classification table, similar to the table 6.1 below. In table 6.1: TP, FP, FN and TN represent True Positive, False Positive, False Negative and True Negative respectively.

Table 6.1 an example of a classification table

		Housing Tenure		
		Social Housing	Private Housing	
Mortality	Deceased +	a (TP)	b (FP)	a + b
	living -	c (FN)	d (TN)	c + d
		a + c	b + d	a + b + c + d

We can assume that there are four possible groups (combinations) of peoples, as indicated a, b, c and d in Table 6.1. From the above table, we determine the sensitivity and specificity as follows:

Sensitivity refers to the proportion of TP cases (e.g. people living in social housing who died in the period of 2002-04) and can be calculated as:

$$\text{Sensitivity} = a / (a+c). \qquad (6.1)$$

Specificity refers to the proportion of TN cases (e.g. residents of private housing who were living in the period of 2002-04) and can be calculated as:

$$\text{Specificity} = d / (b+d). \qquad (6.2)$$

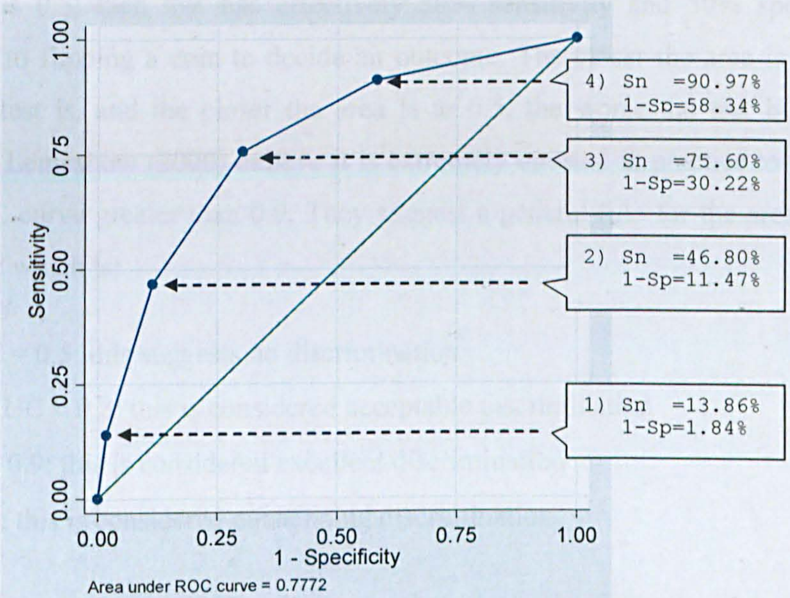
The ratio of the probability of being deceased and living in social housing with the corresponding probability if they were living in private housing, is called the 'likelihood ratio' (Altman, 1999) and therefore defined as follows:

$$\text{Likelihood ratio} = \frac{Sensitivity}{(1-Specificity)} = \frac{a/(a+c)}{b/(b+d)} = \frac{a(b+d)}{b(a+c)} \qquad (6.3)$$

6.2.1 ROC Curves

A ROC curve plots the false positive rate (1-Specificity) on the X axis against Sensitivity (the true positive rate) on the Y axis for various combinations of explanatory factors. Figure 6.1 illustrates a ROC curve.

Figure 6.1 an illustration of a ROC curve



A detailed explanation of a ROC curve construction can be find in Appendix-I.

The evaluation and the key criteria of a ROC curve will be discussed in next section.

6.3 Evaluating a ROC curve

In general when a ROC curve climbs rapidly towards the upper left hand corner of the graph, the test result is good. This means that the sensitivity (1- FN) is high and the false positive rate (1-Specificity) is low. When the ROC curve follows a diagonal path it means that the null hypothesis is true or in other words, the possibility of a positive test result is the same amongst those with the disorder as those without the disorder. In relation to the utility of ROC curve as a tool, Hosmer & Lemeshow (2000) state: "the area under the ROC curve which ranges from zero to one, provides a measure of the model's ability to discriminate between those subjects who experience the outcome of interest versus those who do not".

The larger the area under the curve (AUC), the better the test result. If the AUC is equal to 1 it is an ideal test (it achieves both 100% sensitivity and 100% specificity). If the AUC is 0.5, then test has effectively 50% sensitivity and 50% specificity or equivalent to flipping a coin to decide an outcome. The closer the area is to 1.0, the better the test is, and the closer the area is to 0.5, the worse the test is. However, Hosmer & Lemeshow (2000) believe it is extremely unusual in practice to get an area under ROC curve greater than 0.9. They suggest a general rule for the area under the ROC curve which is:

If the AUC = 0.5: this suggests no discrimination

If 0.7 ≤ the AUC < 0.8: this is considered acceptable discrimination

If 0.8 ≤ the AUC < 0.9: this is considered excellent discrimination

If the AUC ≥ 0.9: this is considered outstanding discrimination.

Hosmer & Lemeshow (2000) also note that a poorly fitting model (i.e. poorly calibrated as assessed by goodness-of-fit measures) may still have good discrimination. Therefore they suggest that model performance should be assessed by considering both calibration and discrimination.

We now go on to consider the production of standard errors for the ROC curve and the key criteria used in its evaluation.

Hanley & McNeil (1982) provided the methods of calculating the standard error for the area under a ROC curve. They calculate standard error (SE) as:

$$SE(A) = \sqrt{\frac{A(1-A)+(n_a-1)(Q1-A^2)+(n_n-1)(Q2-A^2)}{n_a n_n}} \tag{6.4}$$

'A' is the area under the curve, 'n_n' is the number of normal cases (those subjects who experience the outcome of interest, in Table 6.1, those living in private housing which include the sum of $b+d$) and 'n_a' is the number of abnormal cases (those subjects who do not experience the outcome of interest, those living in social housing, the sum of $a+c$).

Q1 and Q2 are estimated by:

$$Q1 = A / (2\text{-}A) \tag{6.5}$$

$$Q2 = 2A^2 / (1+A) \tag{6.6}$$

Q1 and *Q2* are also defined by (Hanley & McNeil, 1982) as:

Q1 = the probability of two randomly chosen abnormal cases (i.e. in Table 6.1 those living in social housing) will both be ranked with greater suspicion than a randomly chosen normal case (i.e. in Table 6.1 those living in private housing).

Q2 = the probability of one randomly chosen abnormal case will be ranked with greater suspicion than two randomly chosen normal cases.

The formula to calculate the confidence interval for Sensitivity is written:

$$Sn \pm Z_{1-a/2}\sqrt{\frac{Sn(1-Sn)}{n_a}} \tag{6.7}$$

And for Specificity is written:

$$Sp \pm Z_{1-a/2}\sqrt{\frac{Sp(1-Sp)}{n_n}} \qquad (6.8)$$

In the next Section first the construction of ROC curve in logistic regression will be discussed. Then the ROC curve will then be used as a tool to evaluate the different logistic models produced earlier in Chapter 5.

6.4 Logistic regression and ROC

In Chapter 5 the analysis of the Camden mortality data using logistic regression was presented in detail. Seven models were introduced which conditioned on basic socio-demographic information and information about three popular causes of hospital admissions (FIS) either sequentially or in a 'block'. Models were also adjusted to include different ways of handling categorical items. In this sectionIuse the ROC curve as an additional tool to evaluate different logistic models produced earlier in Chapter 5. Firstly,Ineed to define how a ROC curve is produced in the context of a logistic regression model.

6.4.1 Classification Table

The use of a classification table is a customary way of summarizing a fitted logistic model. Hosmer & Lemeshow (2000) define it as: "...the result of cross-classifying the outcome variable, y, with a dichotomous variable whose values are derived from the estimated logistic probabilities". In order to obtain the derived dichotomous variable, a cut-point, *c*, needs to be defined. By defining *c* (say, 0.5, the default value in SPSS and Stata), each estimated probability derived from logistic regression for every individual would be compared with *c*. By setting up the *c* value = 0.5 means that any individual with an estimated probability of mortality value over 0.5 is assigned to be a case i.e. the derived value will be equal to 1, otherwise an individual is not a case or has a derived value equal to 0.

However, Hosmer & Lemeshow (2000) with detailed explanation prove that the classification always favours the larger group between the two component groups (misclassification) "...a fact that is also independent of the fit of the model". The main point here is while the overall rate of correct classification is high or reasonable, it happens that the rate of the negative cases (0) classified by model is high but the rate of the positive cases (1) is low. The aim is to maximise the true positive rate (sensitivity) and to minimize the false negative rate. If the true positive rate (sensitivity) and the false positive rate (1-specificity) can be plotted, then we can decide how well the data fits the model.

These authors have provided 'rules of thumb' to assist the analyst in deciding on the discriminatory value of a particular model. In order to assign a case to a binary category (death or not) based on the model a 'cut-off' probability is determined by equating 'sensitivity' and 'specificity'. This is equivalent to assuming that the cost of a false negative is the same as the cost of a false positive and therefore, the rule is 'context dependent'. In different applications of ROC analysis ranging from their original use in interpreting radar signals Mason & Graham (2002) and subsequent use in screening and clinical diagnosis (Pepe, 2003) the costs attached to false negatives and false positives may be different e.g. the costs of treating someone who is not a case may be greater or less than the costs of failing to treat someone who is a genuine case. For this reason the evaluation of the discriminatory power of the models has to be taken as 'indicative' despite the fact the application of Hosmer & Lemeshow's evaluation is now routinely provided by SPSS and Stata in ROC analysis.

6.4.2 Plot of Sensitivity and Specificity

By setting the cut-point to a different value, the values of sensitivity and specificity will change. To choose an optimal cut-point for the purpose of classification, we need to select a cut-point that maximises the sum of sensitivity and specificity (Altman, 1999) or as Lemeshow (2007) suggests, "...choosing a cutoff that makes both Sensitivity' and Specificity relatively high". As Hosmer & Lemeshow (2000) emphasize, this choice will be facilitated by a 'plot of sensitivity and specificity' where an optimal choice for a cut-point might be approximately anywhere the sensitivity and specificity curves cross.

To illustrate this idea, will be helpful to consider a specific example of the construction of a ROC curve produced for four basic binary factors; age, gender, housing tenure and tax banding. This example analysed by both SPSS and Stata. Whereas SPSS provides a table of 'coordinates of the curve' which includes all possible probability cut-points, sensitivity and 1-specificity, Stata simply plots sensitivity and specificity versus all possible probability cut-points.Iwill use both outputs to develop an understanding of varying the cut-points.

The 'coordinates of the curve', produced by SPSS has been modified by adding the sequence, combination, estimated probability and 'specificity' columns to it, is illustrated in Table 6.2. The second column's name (positive if ≥ a) is a default name given by SPSS where 'a' represents the value of the estimated probabilities. The sensitivity and specificity in each row are calculated based on the value of 'a' as a cut-point and any cases with a value greater or equal to 'a' are assigned to being positive. As we can see in Table 6.2, the number of probability cut-points for four variables is 16 (2^4), the number of different combinations for four binary variables.

In Stata instead of a 'coordinates of the curve' table, we can obtain a 'plot of sensitivity and specificity versus all possible cut-points. Figure 6.2-a is an illustration of this plot and Figure 6.2-b is the plot of sensitivity and 1- specificity (ROC curve) for the above ex ample.

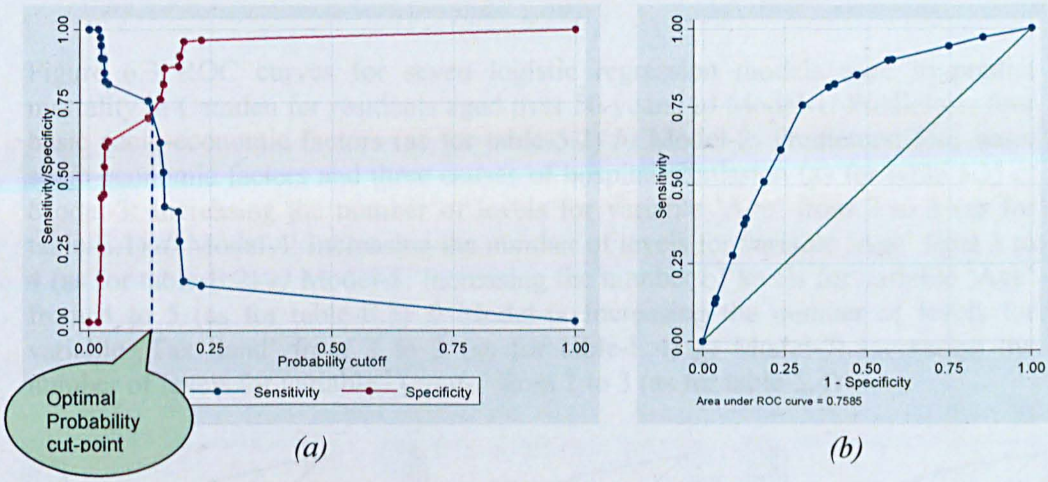
Table 6.2 Coordinates of the Curve from SPSS - modified

Seq	Combination 'ABCD'	Estimated Probability	Cut-Point (positive if ≥a)	Sensitivity	Specificity	1 - Specificity
			0	1	0	1
1	0000	0.019048	**0.021322**	0.973	0.149	0.851
2	0010	0.023596	**0.024200**	0.945	0.252	0.748
3	0100	0.024804	**0.025033**	0.902	0.425	0.575
4	0001	0.025263	**0.027973**	0.9	0.437	0.563
5	0110	0.030683	**0.030965**	0.852	0.531	0.469
6	0011	0.031247	**0.032040**	0.823	0.6	0.4
7	0101	0.032834	**0.036685**	0.814	0.617	0.383
8	0111	0.040536	**0.081005**	0.756	0.698	[illegible]
9	1000	0.121473	**0.134143**	0.612	0.764	0.236
10	1010	0.146814	**0.150078**	0.51	0.814	0.186
11	1100	0.153341	**0.154569**	0.393	0.869	0.131
12	1001	0.155797	**0.169868**	0.375	0.875	0.125
13	1110	0.183939	**0.185358**	0.275	0.908	0.092
14	1011	0.186777	**0.190726**	0.138	0.959	0.041
15	1101	0.194675	**0.212971**	0.123	0.964	0.036
16	1111	0.231268	**1**	0	1	0

Optimal Probability cut-point

The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group.
'a' The smallest cutoff value is the minimum observed test value minus 1, and the largest cutoff value is the maximum observed test value plus 1. All the other cutoff values are the averages of two consecutive ordered observed test values.

Figure 6.2 *a)* plot of sensitivity and specificity versus all possible probability cut-points, generated by a logistic regression for four binary variables; gender, age, housing tenure and tax bands *b)* ROC Curve or plot of sensitivity and 1-specificity (Stata output)

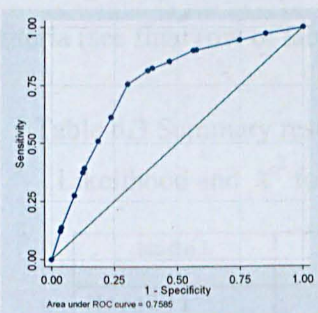


We will now continue to apply this evaluation to the previous seven models as presented in Chapter 5.

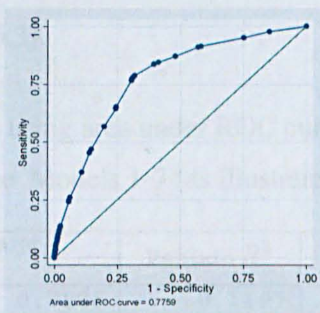
6.4.3 ROC curves for seven logistic models in Section 5.2

Figure 6.3 (*a-g*) includes a summary of the output from logistic regression and a graph showing the Area Under ROC Curve (AUC).

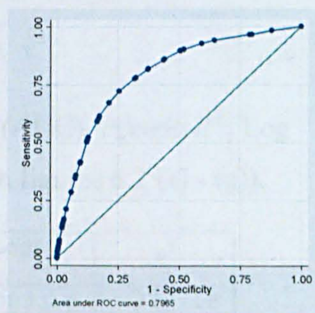
Figure 6.3 ROC curves for seven logistic regression models used to predict mortality in Camden for residents aged over 50 years; *a)* Model-1: Predictors; four basic socio-economic factors (as for table-5.2) *b)* Model-2: Predictors; four basic socio-economic factors and three causes of hospital admission (as for table 5.3) *c)* Model-3: Increasing the number of levels for variable 'Age' from 2 to 3 (as for table-E.1) *d)* Model-4: Increasing the number of levels for variable 'Age' from 3 to 4 (as for table-E.2) *e)* Model-5: Increasing the number of levels for variable 'Age' from 4 to 5 (as for table-E.3) *f)* Model-6: Increasing the number of levels for variable 'Tax band' from 2 to 3 (as for table-E.4) *g)* Model-7: Increasing the number of levels for variable 'Tenure' from 2 to 3 (as for table-5.4)



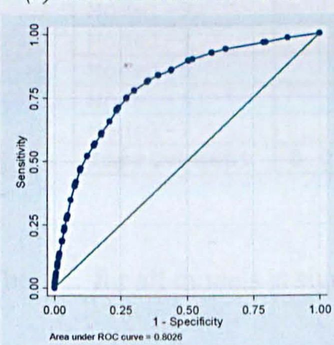
(a) Model-1 AUC=0.7585



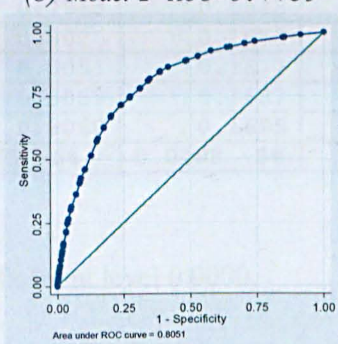
(b) Model-2 AUC=0.7759



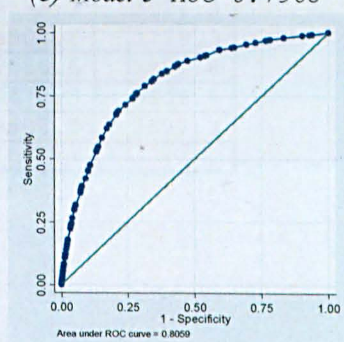
(c) Model-3 AUC=0.7965



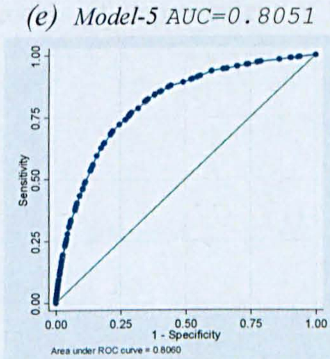
(d) Model-4 AUC=0.8026



(e) Model-5 AUC=0.8051



(f) Model-6 AUC=0.8059



(g) Model-7 AUC=0.8060

By improving the logistic models either by increasing the number of terms in the model or deepening the number of levels used to define categorical variables, the ROC curve at each step gradually changes from a staggered shape to a smoother curve form and the AUC also increases slowly. In the figure 6.3 (*a-g*), in each model the AUC are in bold.

The values of the AUC and Pseudo R^2 for comparison are presented in Table 6.3 below. By comparing these two values we see a small but steady improvement in both AUC and Pseudo R^2 as models increase in complexity. As a brief resume comparing 'model-1' which simply predicts mortality using four socio-economic factors with 'model-7' which expands the categories for age, tenure and tax banding as well as including hospital admissions report, we see a modest 5% improvement for both criteria (see final row of table 6.3).

Table 6.3 Summary results using area under ROC curve (AUC), Pseudo R^2, Log Likelihood and x^2 for the Models 1-7 (as illustrated in figures 6.2 (a) - (g)).

Model	**AUC**	**Pseudo R^2**	**LogL**	x^2
Model-1	**0.7585**	**0.1197**	-10033.6	2727.8
Model-2	0.7759	0.1360	-9847.4	3100.2
Model-3	0.7965	0.1548	-9633.2	3528.5
Model-4	0.8026	0.1658	-9507.8	3779.3
Model-5	0.8051	0.1677	-9486.6	3821.8
Model-6	0.8059	0.1687	-9474.3	3846.5
Model-7	**0.8060**	**0.1695**	-9608.8	3577.3
Improvement	**0.0475 ~5%**	**0.0498 ~5%**		

The x^2 for all models is significant at level 0.0000.

Appendix-J summarises the process of improvement of the seven logistic regression models.

6.5 Examining the model refinement presented in Chapter-5 by the use of ROC curves

The aim of this section is to examine the process of model refinement in Section 5.3 once again, with use of ROC curves. To achieve this ROC curve first will be used as an evaluation tool to contrast different models with 'continuous' and 'dichotomous' age, as discussed in 5.3.1. Then, the effect of interaction terms on model refinement (discussed in 5.3.2) also will be assessed with ROC curves. Finally, the area under ROC curve of the final model (model-7 in Section 5.2, Table 5.4) will be compared with the model introduced in Section 5.3 (Table 5.6) which includes an interaction term between housing tenure and age.

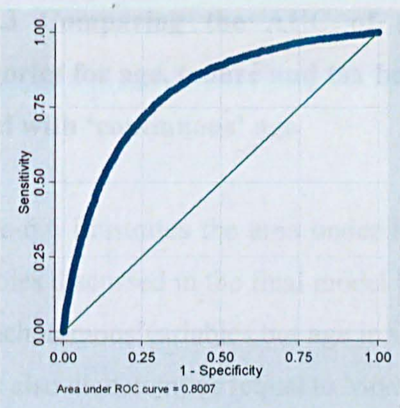
6.5.1 Comparing the AUC of the models with 'continuous' and 'dichotomous' age discussed in Section 5.3.1

In order to assess the effect of the continuous variable age on model refinement, three models including model-1, -2 and -7, with 'continuous' and 'dichotomous' age variable were contrasted in Chapter 5. Here once again the differences between each pair of models with dichotomous and continuous age will be assessed with use of ROC curve.

6.5.1.1 Comparing the AUC of the model with four binary socio-economic variables (first logistic model in Chapter 5) and the equivalent model with 'continuous' age

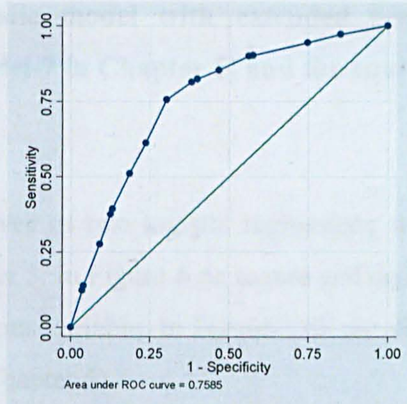
Figure-6.4 illustrates the area under ROC curves of two logistic regression models with four socio-economic factors (age, gender, housing tenure and council tax bands). Figure-6.4a is an illustration of the logistic model with three binary factors and continuous age and Figure-6.4b is a copy of model-1 in Chapter 5 with four binary factors. The area under ROC curve for the model with the continuous variable age shows a 4% increase compared to the model with binary age, which provides a better level of discrimination.

Figure-6.4 Illustration of the AUC of two logistic regression models with four factors; *a)* with age as a continuous variable *b)* all 4 factors are binary.



AUC = 0.8007

(a)



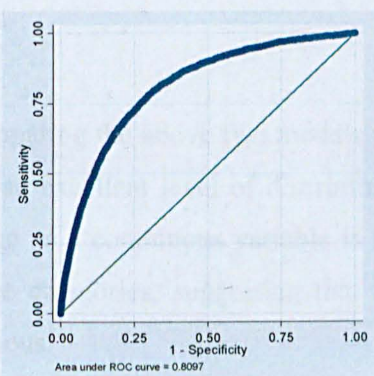
AUC = 0.7585

(b)

6.5.1.2 Comparing the AUC of the model with seven binary socio-economic variables (the second logistic model in Chapter 5) and the equivalent model with 'continuous' age

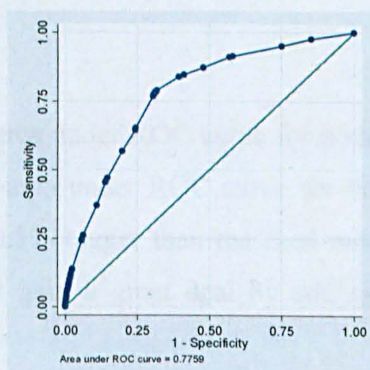
Figure-6.5 illustrates the area under the ROC curves of two logistic models with seven factors including the four socio-economic factors and three causes of hospital admissions (FIS). In Figure-6.5a the variable age is a continuous variable and in Figure-6.5b all variables are binary (a copy of model-2 in Chapter 5).

Figure-6.5 Illustration of the AUC of two logistic regression models with seven factors; *a)* with age as a continuous variable *b)* all 7 factors are binary.



AUC = 0.8097

(a)



AUC = 0.7759

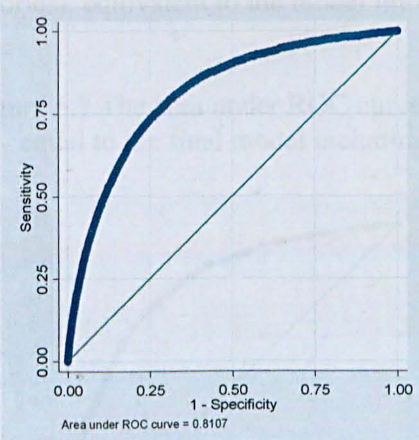
(b)

The improvement of the area under ROC curve in model with the continuous variable age is about 3.4% which is noticeable.

6.5.1.3 Comparing the AUC of the logistic model with extended levels of categories for age, tenure and tax band (model-7 in Chapter 5) and the equivalent model with 'continuous' age

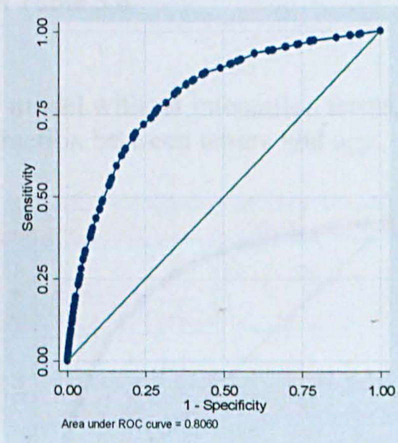
Figure-6.6 illustrates the area under ROC curves of two logistic regressions with all variables discussed in the final model in Chapter 5. In Figure-6.6a tenure and tax bands are dichotomous variables but age is a continuous variable. In Figure-6.6b the variable age is also dichotomous (equal to Model-7 in Chapter 5).

Figure-6.6 Illustration of the AUC of two logistic regression models; *a)* equivalent to the final model with continuous age *b)* Final model (Model-7).



AUC = 0.8107

(a)



AUC = 0.8060

(b)

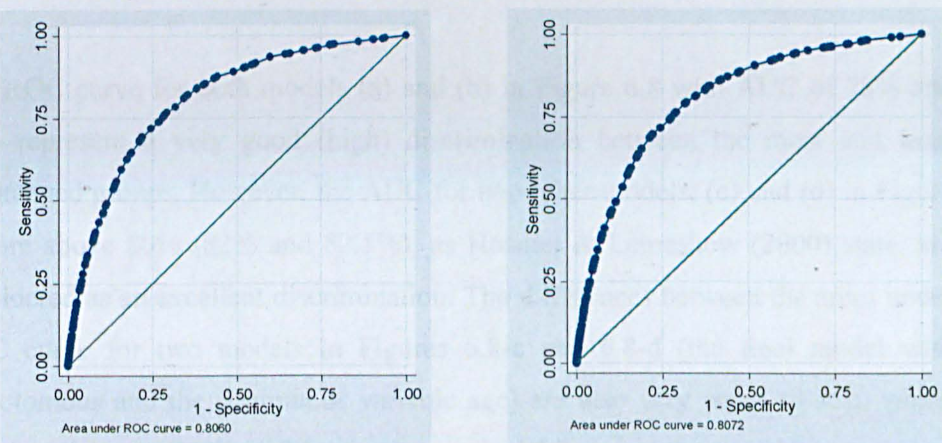
By comparing the above two models while the area under ROC curve for both models shows an excellent level of discrimination, the area under ROC curve for the model with age as a continuous variable is less than 0.5% bigger than the final model with five age categories, suggesting that we do not gain a great deal by adding age as continuous.

Here, again, the effect of the continuous variable age on three models, discussed in Section 5.3, has been assessed with ROC curves. The outcome confirms that differences exist between the first two models with binary age and age as a continuous variable. However, in the last models (illustrated in Figure 6.6) there is a little to choose between the model with 5 level ages and the one with the continuous variable age.

6.5.2 Examining the impact of interaction effects by using ROC curves

In Chapter 5 the impact of allowing for interaction in different models were discussed in detail. Again, the AUC of the final model is compared with another model which includes the same variables including some interaction terms. Figure 6.7a shows the area under ROC curve of the final model and Figure 6.7b is the illustration of the final model with interaction between tenure (with three categories) and age (with five categories), equivalent to the model introduced in Table 5.6.

Figure 6.7 The area under ROC curve; a) final model without interaction terms, b) equal to the final model including the interaction between tenure and age.



AUC = 0.8060 (a)

AUC = 0.8072 (b)

The area under ROC curves in Figure 6.7 clearly shows that the interaction effects is very minimal and can be safely discarded.

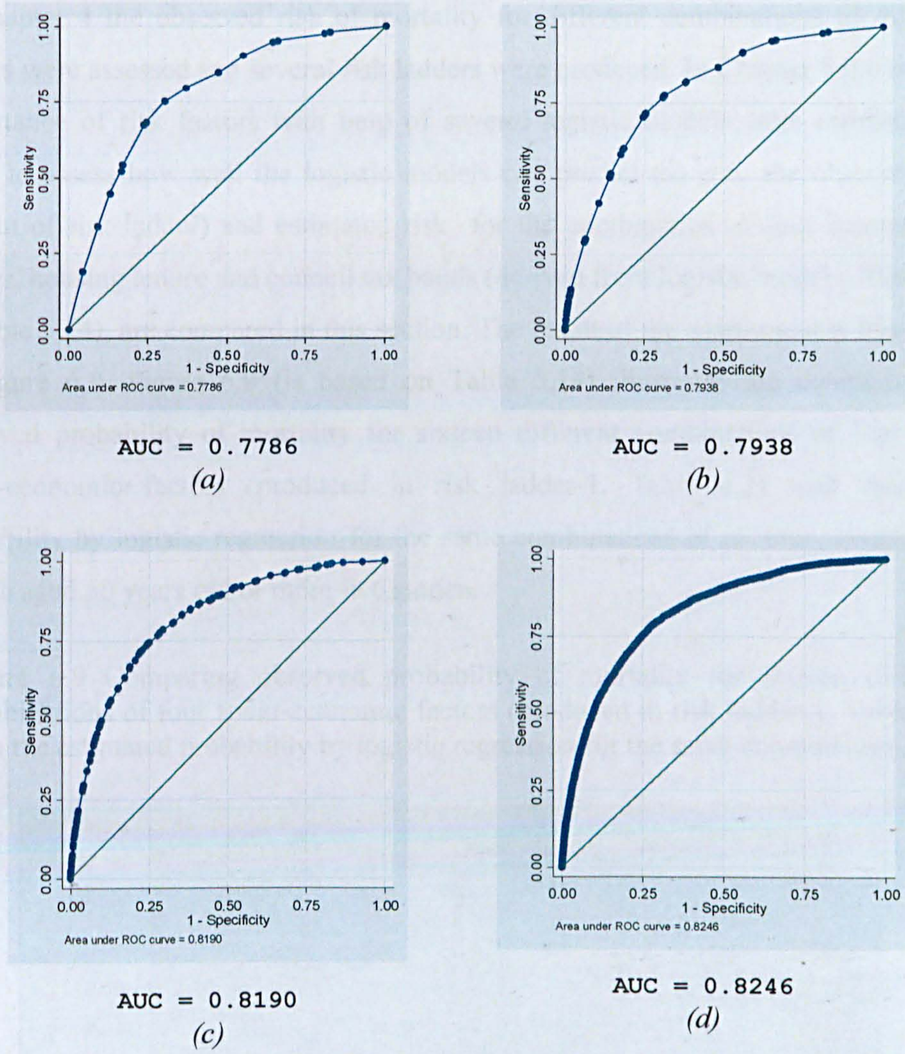
6.6 Evaluation of different models with 'social services' as outcome variable by ROC curves

In Section 5.4 four different logistic models were discussed with the outcome variable 'social services'. These four models include: the basic model with four binary variables (equal to Model-1); model with seven binary variables (equal to Model-2); final model (Model-7) and finally the model equivalent to the final model but with the continuous variable age.

In this section the level of discrimination of each of these models are evaluated with ROC curves. Figure 6.8 shows four ROC curves of four different models including; (a) the model with four binary variables (age, gender, housing tenure and council tax band), (b) the model with seven binary variables by adding three causes of hospital admissions (FIS) to the previous model in (a), (c) final model and (d) the model equivalent to the final model with age as a continuous variable.

The ROC curve for both models (a) and (b) in Figure 6.8 with AUC of 78% and 79% represent a very good (high) discrimination between the most and least advantaged groups. However, the AUC for two other models; (c) and (d) in Figure 6.8 are above 80% (82% and 82.5%), as Hosmer & Lemeshow (2000) state, are considered as an excellent discrimination. The differences between the areas under ROC curve for two models in Figures 6.8-c and 6.8-d (the final model with dichotomous and the continuous variable age) are also very small (0.5%), which indicates the contribution of the continuous variable age to the model improvement is not significant.

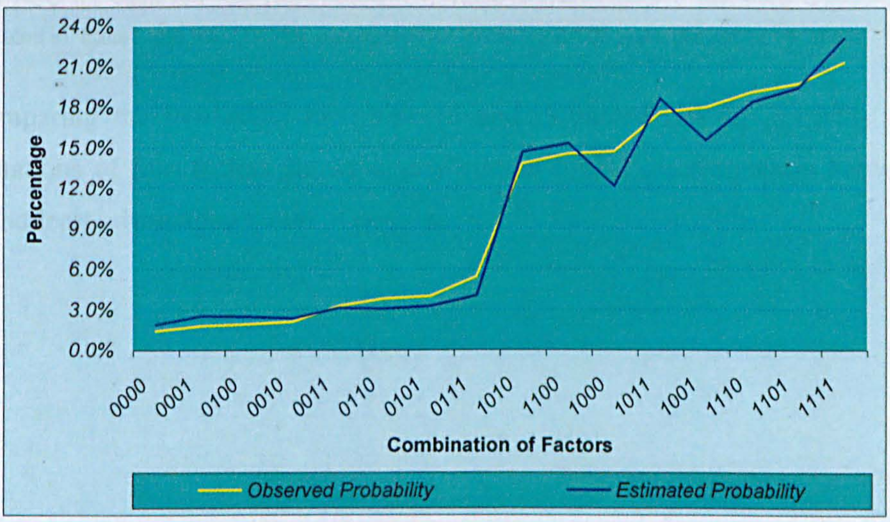
Figure 6.8 ROC curves for four different models with 'social services' outcome; a) equal to model-1, b) equal to model-2, c) equal to final model and d) equal to the final model with the continuous variable age.



6.7 Comparing the AUC of the Observed and Estimated risk

In Chapter 4 the observed risk of mortality for different combinations of different factors were assessed and several risk ladders were produced. In Chapter 5 the relative importance of risk factors with help of several logistic models were estimated. In order to assess how well the logistic models can predict the risk, the observed risk (output of risk ladder) and estimated risk for the combination of four factors; age, gender, housing tenure and council tax bands (derived from logistic models, illustrated in Table 5.14), are compared in this section. The result of the assessment is illustrated in Figure 6.9. Figure 6.9 (is based on Table 5.14) illustrates the comparison of observed probability of mortality for sixteen different combinations of four basic socio-economic factors (produced in risk ladder-1, Table 4.2) with estimated probability by logistic regression for the same combinations of all four variables for people aged 50 years old or more in Camden.

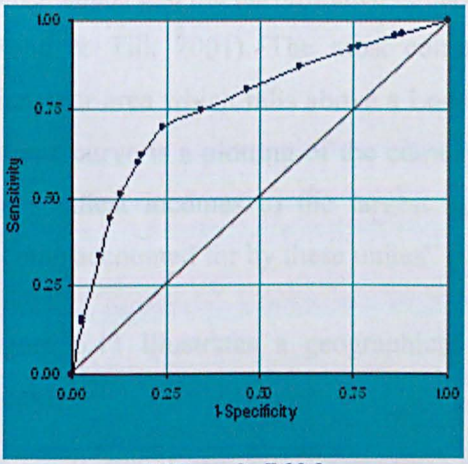
Figure 6.9 Comparing observed probability of mortality for sixteen different combinations of four socio-economic factors (produced in risk ladder-1, Table 4.2) with the estimated probability by logistic regression for the same combinations.



By comparing the outcomes from risk ladder methodology (observed probability) and the estimated probability with logistic regression, we get very similar patterns of probabilities, validating the precision of our models.

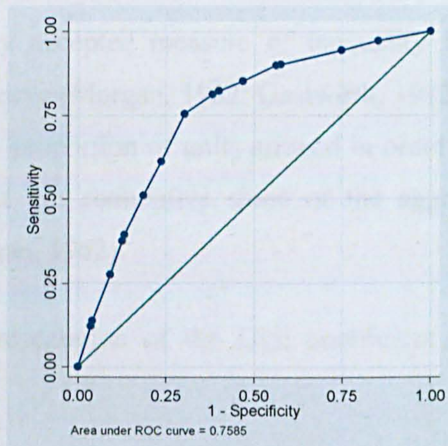
The AUC of the observed and estimated risk of four socio-economic factors for the purpose of comparison are also illustrated in Figure 6.10.

Figure 6.10 AUC of the Observed and Estimated risk for 4 socio-economic factors; a) Based on observed risk of risk ladder-1 in Table 4.2, b) Based on estimated risk extracted from logistic regression model-1.



AUC = 0.7606

(a)



AUC = 0.7585

(b)

- The AUC for observed risk (Risk ladder-1.1, Table 4.2) in Figure 6.10a were created and calculated in 'Excel' and the AUC for estimated risk in Figure 6.10b were produced by 'Stata'.

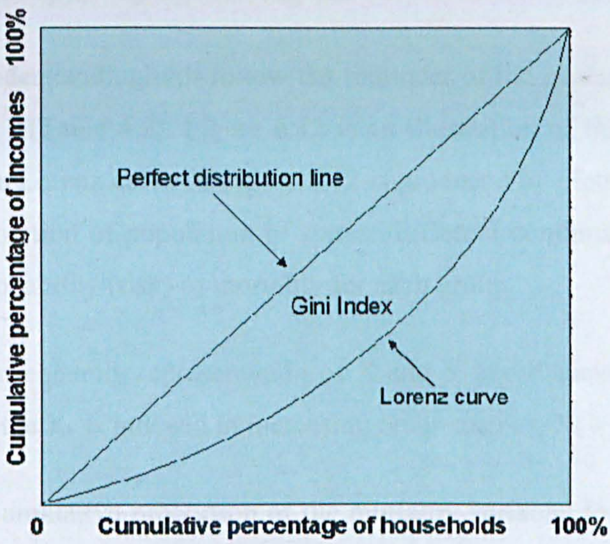
By comparing the two AUC for observed and estimated risk of mortality for the combinations of four factors shows a very similar (76%) discrimination between the most and least advantaged group of people.

6.8 Gini Coefficient and its relationship with ROC Curves

The Gini coefficient is one of the popular measures of inequality of a distribution, mostly used to measure income and wealth inequality Gastwirth (1972). It has also been applied in other disciplines to measure the health inequality (Van Doorslaer & Jones, 2003) and the performance of supervised classification rules in machine learning (Hand & Till, 2001). The most commonly accepted measure of inequality is the triangular area which falls above a Lorenz curve (Morgan, 1962; Gastwirth, 1972). "A Lorenz curve is a plotting of the cumulative proportion of units arrayed in order from the smallest incomes to the largest against the cumulative share of the aggregate income accounted for by these unites" (Morgan, 1962).

Figure 6.11 illustrates a geographical representation of the Gini coefficient (Gini index).

Figure 6.11 Geographical representation of the Gini coefficient



6.8.1 Estimation of Gini coefficient

For perfect equality the Lorenz curve would overlap the diagonal (indication of a Gini coefficient of '0') and in the case of a perfect inequality the Lorenz curve would overlap the bottom and right straight line indicates that a single household receives all of the income, will result a Gini coefficient of '1'. In other words, the closer the Lorenz curve to the diagonal the less the inequality.

The formula for approximating the Gini coefficient (Morgan, 1962) is:

$$G = \frac{Area\,between\,Lorenze\,curve\,and\,diagonal}{Area\,under\,diagonal}$$

$$= \frac{0.5 - Area\,under\,Lorenze\,curve}{Area\,under\,diagonal}$$

$$= 1 - (Area\,under\,Lorenz\,curve \times 2) \qquad (6.9)$$

Sometimes the Lorenz curve can not be defined across its whole range but the values at certain intervals is available or can be estimated. In this case the Gini coefficient can be approximated by interpolating any missing values.

For the ease of understandingIwill follow the reminder of the explanation with help of the Risk ladder-1.1 (Table 4.2). Figure 6.12 is an illustration of the Lorenz curve for Risk ladder-1. The Lorenz curve in Figure 6.12 is produced by plotting the cumulative values of the proportion of population of sixteen different combinations (groups) and the cumulative probability (risk) of mortality for each group.

If ($X_{k,}$ Y_k) are consequently representation of X and Y coordinates of each points on Lorenz curve and the X_k is indexed in increasing order such as, $X_k > X_{k-1}$ and:

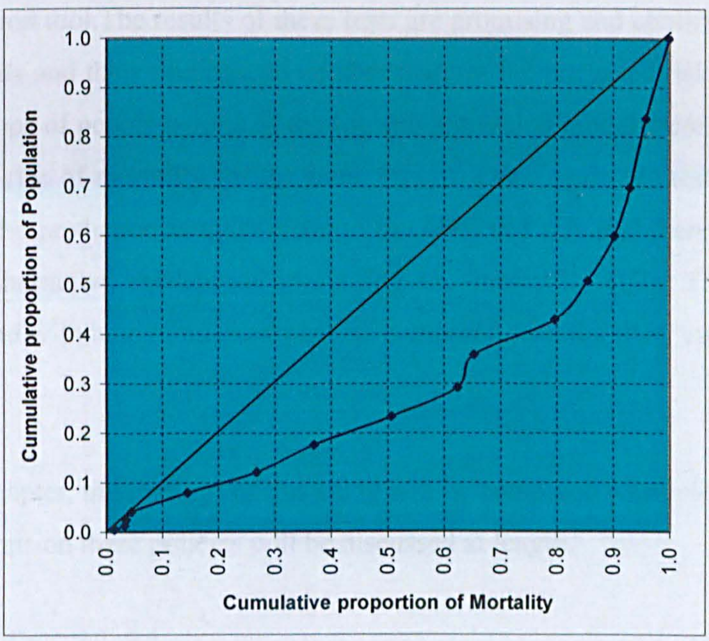
- X_k is the cumulative proportion of the mortality variable, for K= 0, ..., n, with $X_0 = 0$ and $X_n = 1$.
- Y_k is the cumulative proportion of the population variable, for K = 0, ..., n, with $Y_0 = 0$ and $Y_n = 1$.

The Gini coefficient can be approximated by using the values of X and Y of n points, so:

$$G \geq 1 - \sum_{k=1}^{n} (X_k - X_{k-1})(Y_k + Y_{k-1}) \qquad (6.10)$$

The value of Gini coefficient for Risk ladder-1.1 using equation 6.10 is 0.42. As we can see in 6.10, the real value of 'G' usually is greater than the output of the above equation. Indeed as Gastwirth (1972) states, the above approximation leads to an under-estimation of 'G' since the straight line connecting the two points on Lorenz curve lies above the convex curve. So, the output will represent the lower bound of the 'G' value. In order to produce a more accurate result, Gastwirth (1972) also developed a method to estimate the upper bound, which the further discussion on it, is not the purpose of this study. While Morgan (1962) believes: "For eight or more groups this approximation should be quite close".

Figure 6.12 Lorenz curve of the Risk ladder-1.1 in Table 4.2



6.8.2 The relationship between the area under ROC curve and Gini coefficient

Hanley & McNeil (1982) show that the area under ROC curve is equivalent to Wilcoxon test of rank. It is also closely related to the Gini coefficient which sometimes is used as alternative measure (Hand & Till, 2001). Gini coefficient is actually twice the area between the diagonal and ROC curve (Breiman et al., 1984). Hand and Till (2001) state that Gini +1 = 2 × AUC.

The area under ROC curve for Risk ladder-1.1 as stated in Section 6.7 is 0.76 and the area above the diagonal is 0.76 – 0.5 = 0.26. However the Gini coefficient estimated by 6.10, is 0.42 and by diving it to 2, it would result 0.21 which is smaller than the area between diagonal and the ROC curve in Risk ladder-1. This difference could be as a result of the under estimation by 6.10 which leading to lower bound as stated by Gastwirth (1972). One way of reducing this difference could be increasing the number of points on Lorenz curve. More precisely, the equation 6.10 would work better for the studies that dealing with more groups.

Summary In this chapter the performance of different models and their capability in justification of the data in the previous chapter are examined with use of ROC curves as an evaluation tool.The results of these tests are promising and confirm the reliability of the methods and their findings. In another attempt the predicted risk of mortality of different groups of people by one of the logistic regression model were contrasted with the observed risk of mortality for the same groups. Once again the test shows that the outcome of the prediction is quite close to the observed risk and therefore consistent. Finally an alternative method of measuring the inequality (Gini Coefficient) was introduced and with help of an example was compared with the ROC curve.

In the next chapter, the findings of this study will be compared with relevant policy and the implications on these policies will be discussed at length.

PART III

POLICY IMPLICATIONS,

DISCUSSION & CONCLUSIONS

7 Policy implications

"The primary determinants of disease are mainly economic and social, therefore its remedies must also be economic and social." (Rose, 1992)

7.1 Introduction

In the introductory chapter the importance of health promotion and prevention of ill health for older people was discussed, and in particular how these areas have became key aspects in social and national policy across health and social care in the new millennium. In addition, it was also acknowledged that the promotion of health is part of a wider strategy for reducing social inequalities in health with a particular focus on those in poor health (Department of Health, 1999; Godfrey, 2001). Considering these factors, it is important that as we try to address these socio-economic differences in health we also try to quantify and measure any reduction. This issue is critical as the way the gaps are quantified and measured can affect the results (Low & Low, 2006).

In Chapter-1 the policy context at national and local government level and the existing strategies and delivery plans were reviewed. The National Service Framework (NSF) for Older People; the central guidance to this study and particularly Standard-8 of NSF for older people ('to extend the healthy life expectancy of older people' by 'modification of risk factors for disease') were discussed in more detail. In Chapter 3 a number of socio-economic and health related risk factors were identified by combining the health and local authority administrative data sources in Camden. The risk factors were assessed empirically by a combination of methods and analytical tools in Chapters 4, 5 and 6. In this chapterIwill discuss how the empirical findings of this study could assist and improve existing policies.

In Section 7.2 a set of implications for intervention will be described based on the risk ladder analysis for combinations of different factors. In Section 7.3 these findings will be converted into recommendations using the results of logistic regression to explore the relative importance of the various factors.

It is also worth noting that in the following two sections there might be some issues regarding disparity in the findings for the two principal methods of analysis; risk ladder and logistic regression. As was previously discussed in the introductory section of Chapter 5, it was explained that the risk ladder analysis shows the level of risk for different combination of factors (or different groups of people with similar socio-economic and/or health related characteristics) taken altogether in each combination. Whereas in the case of logistic regression it quantifies the contribution of each risk factor on outcome for the entire target population. Any combination of factors present (used to define analogous combinations identified in risk ladder) can be then used to estimate an overall estimated probability (an example of it is presented in Figure 6.9).

7.2 Policy implications based on observed risk/probability from risk ladders

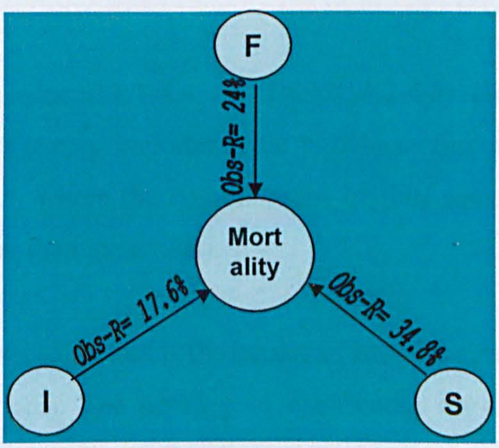
In this section policy implications based on the findings of risk ladder analysis (observed risk/probability) in Chapter 4 will be discussed. The first part of this section is based on observed risk of mortality and the second part is rooted in observed probability of someone being in contact with social services.

7.2.1 Policy implications based on observed risk of mortality

i) Policy implications based on observed risk of mortality as a result of three causes of hospital admissions (falls, ischemic heart disease and strokes)

Figure 7.1 below illustrates the relationship between each of the three causes of hospital admissions and mortality. The percentage shows the observed risk of mortality for those people who had at least one incident of admission to the hospital as a result of one of three causes (F, I or S). Figure 7.1 shows that between these three causes, Ischemic heart disease has the lowest risk (17.6%) and Stroke has the highest effect on mortality.

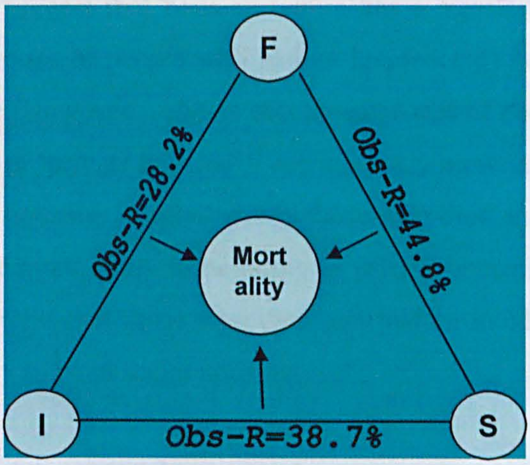
Figure 7.1 Illustration of the observed risk (Obs-R) of mortality as a result of each of the three causes of hospital admissions in the period 2002-04



Note: based upon Table 4.1

Figure 7.2 below, also shows the mutual effect of two causes of hospital admissions on mortality. The risk of mortality as a result of the joint effect of stroke and fall is the highest (44.8%). The risk resulting from the combined effect of heart disease and falls is the lowest (28.2%), though this is still relatively high compared with 6.4% for those group of people who did not have any record of an incidence of hospital admission as a result of the three causes (falls, ischemic heart disease and strokes).

Figure 7.2 Illustration of the observed risk (Obs-R) of mortality as a result of the joint effect of two causes of hospital admissions in the period 2002-04



Note: based upon Table 4.1

ii) Policy implications from observed risk in Risk ladders 1.1 – 1.4 (combinations of four socio-economic factors and three causes of hospital admissions, illustrated in Tables 4.2-4.5)

The findings of risk ladders 1.1-1.4 (in Tables 4.2-4.5) clearly show that in all combinations age plays a very important role in determining mortality (except in the case of Risk ladder 1.4, where the combinations of four socio-economic factors and 'Stroke', is a little lower than the others).

The second important set of factors is the causes of hospital admissions (falls, ischemic heart disease and Strokes). The potency of each cause on mortality was discussed earlier and was illustrated in Figure 7.1.

The third factor is council tax banding which, in general, for the combinations of four socio-economic factors (Risk ladder 1.1), the combination of four socio-economic factors plus ischemic heart disease (Risk ladder 1.2) and stroke (Risk ladder 1.4) is influential but not for the combination of four socio-economic factors and falls (Risk ladder 1.2). The variable gender also alters the risk of mortality for the combination of four socio-economic factors and the combination of four socio-economic factors plus falls but not for the combinations of four socio-economic factors with heart disease and strokes.

The findings do not suggest that housing tenure has a significant impact on risk of mortality for those groups of people who had an incidence of hospital admission as a result of heart disease. However it shows that those groups of people who experienced a fall and were living in 'private housing'[11] are relatively more at risk of mortality than those living in social housing. In other words those who died and had an incidence of fall before death were more likely to be living in private housing than social housing. The findings also suggest that those who died and had an incidence of stroke before death were more likely to be in social housing.

So far the outcome of the risk ladder analysis suggests (once age is excluded as the most dominant factor for hospital admission) those who were admitted as a result of stroke were the most vulnerable group(Centre for Disease Control and Prevention, 1999; Philp, 2004). Standard-5 of the National Service Framework for older people (Department of Health, 2001) also emphasises that: "Stroke is the single biggest cause of severe disability and the third most common cause of death in the UK and other developed countries". The relationship between stroke and other factors (falls and heart disease) is also strong. Therefore there is an argument that the priority of resource allocation should be concentrated on those groups of people who had an incidence of stroke.

The outcome from the risk ladder analysis shows that the risk of mortality is quite high for patients with heart disease, which is one of the top three most likely causes of death (American Heart Association's Heart Disease and Stroke Statistics, 2004; Philp, 2004).

[11] The issue of housing tenure will be discussed in the final chapter.

There is also a strong relationship between those who had at least one incidence of hospital admission for each of the two causes, falls and strokes, which leads to the conclusion that prevention of strokes will reduce incidence of falls (by considering the age factor). The studies on relationship between strokes and falls support the above assertion. The study conducted by Olsson et al. (2004) shows the positive correlation between patients in stroke rehabilitation and risk of fall. Nyberg & Gustafson (1995) also consider falls as a significant problem in stroke rehabilitation and Poole et al. (2002) state: "hip fracture after stroke is an increasingly recognized problem".

Falls in combination with age factor were also found to have stronger relationship with private housing compared with social housing (a combination of council housing and housing association properties). This could indicate that in general more affluent people have a higher chance of living longer and consequently, the risk of having an incidence of admission to the hospital as a result of a fall is higher for them. Further research could expand on these analyses by investigating the sequence of occurrence of each cause of hospital admission to identify which incident comes first.

7.2.2 Policy implications based on observed probability of someone being known to social services

In Section 4.5 the risk of mortality and the probability of being in contact with social services were discussed in detail. A resume of the findings suggests the following:

i) Comparing the risk of mortality and the probability of being in contact with social services shows that for the same level of risk of mortality, females are more likely to be in contact with social services than males (except in the case of Risk ladder 2.4, the combination of four socio-demographic factors and stroke).

ii) Those living in social housing are more likely to be in contact with social services than those living in private housing. This could indicate that being in contact with social services is protective for falls.

iii) The factor 'fall' plays an important role in allocation of the social services' resources particularly when taken in combination with age.

iv) Those living in lower council tax band properties (i.e. poorer residents of Camden) are more likely to be known to social services than those living in higher tax bands (wealthier).

7.2.3 Further analysis of social services with risk ladders

Once again in order to assess the relationship between social services and the risk of mortality, or in other words, to examine if the services provided by social services are allocated to the people most at risk or not, the following risk ladders were created. The risk ladder illustrated in Table 7.1 includes four socio-economic factors and contact with social services as predictors with outcome variable, mortality. After sorting the risk of death for all combinations in ascending order, the effect of each variable on the outcome is as follows:

The most powerful factor in this risk ladder, as highlighted in its related column, is contact with social services. Twelve combinations at the bottom of the social services column, with the highest level of risk being those combinations which are in contact with social services, indicates that the services are already provided to those at most risk of death.

The second strongest factor is age. Eight out of nine groups with the highest level of risk of mortality and in contact with social services are from older age groups.

Generally females are more in contact with social services than males (as established earlier, when 'social services' was included as outcome variable). However, when social services are included in the model as a predictor and the risk of mortality is the outcome variable, males are relatively more in contact with social services. This means males are more likely to be in contact with social services when they are in critical health condition. This would seem to suggest that men should be specifically targeted as a group for receipt of early intervention services, particularly given that "Loneliness resulting from the death of a spouse, poor social support and physical illness or disability can lead to self-harm and suicide in old age – particularly amongst older men" (Kelly & Bramwell, 2006).

For the variables 'housing tenure' and 'tax banding' the difference between the residents of social and private housing, and similarly for those living in low or high council tax bands properties, is not very large. However, for both variables the differences are considerable. Table 7.1 shows that the probability (p) of death for those living in private housing, higher tax bands and in contact with social services is higher than the reverse group (for the variable housing tenure, risk ladder 1.2 in Table 4.3 also shows similar output).

Table 7.1 A risk ladder with four socio-economic factors and social services with outcome variable 'Mortality'

Seq	Age	Gender	Tenure	Tax Band	SS	No. of Death	Population	p of Death	Conf.Interval	
1	0	0	0	0	0	74	5995	1.2%	1.0%	1.5%
2	0	0	0	1	0	8	487	1.6%	0.5%	2.8%
3	0	1	0	0	0	116	7021	1.7%	1.4%	2.0%
4	0	0	1	0	0	69	4100	1.7%	1.3%	2.1%
5	0	0	1	1	0	75	2717	2.8%	2.1%	3.4%
6	0	1	1	0	0	124	3802	3.3%	2.7%	3.8%
7	0	1	0	1	0	25	711	3.5%	2.2%	4.9%
8	0	1	1	1	0	145	3235	4.5%	3.8%	5.2%
9	0	0	0	1	1	1	11	9.1%	-7.9%	26.1%
10	1	0	1	0	0	193	1862	10.4%	9.0%	11.7%
11	1	0	0	0	0	307	2673	11.5%	10.3%	12.7%
12	0	0	1	0	1	20	168	11.9%	7.0%	16.8%
13	0	0	1	1	1	18	147	12.2%	6.9%	17.5%
14	1	1	0	0	0	293	2346	12.5%	11.2%	13.8%
15	1	0	1	1	0	257	1849	13.9%	12.3%	15.5%
16	1	0	0	1	0	36	246	14.6%	10.2%	19.1%
17	1	1	1	0	0	222	1419	15.6%	13.8%	17.5%
18	1	1	0	1	0	35	222	15.8%	11.0%	20.6%
19	0	0	0	0	1	13	79	16.5%	8.3%	24.6%
20	1	1	1	1	0	265	1483	17.9%	15.9%	19.8%
21	0	1	1	0	1	27	147	18.4%	12.1%	24.6%
22	0	1	0	1	1	4	19	21.1%	2.7%	39.4%
23	0	1	1	1	1	40	188	21.3%	15.4%	27.1%
24	1	0	1	0	1	132	492	26.8%	22.9%	30.7%
25	1	0	1	1	1	181	637	28.4%	24.9%	31.9%
26	1	0	0	1	1	20	65	30.8%	19.5%	42.0%
27	0	1	0	0	1	19	60	31.7%	19.9%	43.4%
28	1	0	0	0	1	151	440	34.3%	29.9%	38.8%
29	1	1	1	1	1	126	351	35.9%	30.9%	40.9%
30	1	1	0	0	1	81	225	36.0%	29.7%	42.3%
31	1	1	1	0	1	97	249	39.0%	32.9%	45.0%
32	1	1	0	1	1	14	26	53.8%	34.7%	73.0%

An alternative interpretation is that social services tend to allocate their services more to those who are living in private housing and higher tax bands (wealthier). One reason for this could be that social services concentrate more on those who have experienced a fall (as discussed earlier on Risk ladder 2.2, Table 4.8), which usually includes older age groups (over 80 years old) and they are older because they are typically wealthy. This relationship could be simplified as:

The wealthier → the longer life expectancy → the higher risk of fall → the higher probability of being in contact with social services

However, this relationship cannot be generalized to the entire population as it only explains the level of risk *(p)* for some of the combinations or group of people without considering the size of each group (as discussed in Section 7.1).

In relation to age factors, and in support of the above assertion, in 'A practical guide for older people' published by 'Age concern' and the 'Royal Society for the Prevention of Accidents' (2004) it is stated : " an 85 years old is five times more likely to have a fall than a 65 year old". In relation to wealth, in Chapter 1 the relationship between wealth and ethnicity also was discussed. In the UK 16% of white people are aged 65 or older while 9% of Black Caribbean and only 2% of Black African or Mixed race are aged 65 or older (Age Concern, 2005).

Another possibility could be that richer residents have higher levels of educational attainment and consequently, better understanding of entitlement to the available services. It suggests that for those who are not aware of their entitlement to the available services; (i.e. one in four older people in Camden, based on health inequality report by Camden and Islington Health Authority (2001)), more sources of help and advice need to be provided.

Table 7.2 below also shows a risk ladder with combinations of four factors; gender, falls, stroke and social services. In this risk ladder the outcome variable is mortality and 'stroke' plays an important role in increasing the level of risk of mortality. However contact with social services is an interesting factor when the combinations of the other factors for different groups are held constant. Two groups with the highest probability of death are males and females who had experienced at least one fall and one stroke and are *not* in contact with the social services. By contrast, those groups of people with the same combination of factors but who are also in contact with social services had a much lower probability of death. This differences in probability of death for females changes from 60% to 28% (comparing the combinations with sequence numbers 16 ‘0110’ with 8 ‘0111) and for males it changes from 50% to 40% (comparing the combinations with sequence numbers 15 ‘1110’ with 13 ‘1111’). Clearly the effect of contact with social services is much greater for females than it is for males.

Table 7.2 A risk ladder with four factors; Gender, Falls, Stroke, social services with outcome variable ‘Mortality’

Seq	Gender	Fall	Stroke	SS	Death	Population	*p* of Death	Conf.Interval	
1	0	0	0	0	866	19422	4.5%	4.2%	4.7%
2	1	0	0	0	1098	19806	5.5%	5.2%	5.9%
3	0	1	0	1	48	207	23.2%	17.4%	28.9%
4	0	1	0	0	73	312	23.4%	18.7%	28.1%
5	1	1	0	0	50	200	25.0%	19.0%	31.0%
6	0	0	0	1	452	1743	25.9%	23.9%	28.0%
7	1	1	0	1	26	95	27.4%	18.4%	36.3%
8	0	1	1	1	7	25	28.0%	10.4%	45.6%
9	1	0	1	1	22	77	28.6%	18.5%	38.7%
10	1	0	1	0	69	217	31.8%	25.6%	38.0%
11	1	0	0	1	356	1083	32.9%	30.1%	35.7%
12	0	0	1	0	68	175	38.9%	31.6%	46.1%
13	1	1	1	1	4	10	40.0%	9.6%	70.4%
14	0	0	1	1	29	64	45.3%	33.1%	57.5%
15	1	1	1	0	8	16	50.0%	25.5%	74.5%
16	0	1	1	0	12	20	60.0%	38.5%	81.5%

The outcome of the second risk ladder in Table 7.2 suggests that the preference of service allocation must be based on the following order:

i) Those groups of people who were admitted to hospital at least once for fall and for stroke.

ii) Those who had at least one stroke.

iii) Those who had at least one fall.

However the above order is a broad guideline and more factors should be taken into consideration when making decisions concerning particular cases.

To expand our exploration of the effect of each factor on mortality and the uptake of provision of social services, it will be more appropriate to examine the relative importance of risk factors using logistic regression models.

In this section a set of implications for intervention based on risk ladder analysis for combinations of different factors were discussed. In order to explore the relative importance of the various factors, in the following section these findings will be converted into recommendations based on the results of logistic regression.

7.3 Policy implications based on estimated risk/probability from logistic regression modelling

In Section 7.2 the observed risk of mortality and the probability of being in contact with social services, based on several risk ladder analyses presented in Chapter 4 were discussed and policy implications were extracted. In this section some suggestions on policy, with the help of the findings from logistic regression modelling and the relative importance of each factor, will be discussed.

7.3.1 Policy implications based on estimated impact of the socio-economic factors on Falls (F), Ischemic heart disease (I), and Strokes (S)

In order to assess how much each of the three causes of hospital admission (falls, ischemic heart disease and stroke) are influenced by the available socio-economic factors, three logistic regression models were fitted based on the final model presented in Chapter 5, for each of the three causes. Table 7.3 shows the Odds Ratios (OR) of all variables for the three models. The non-significant ORs are highlighted.

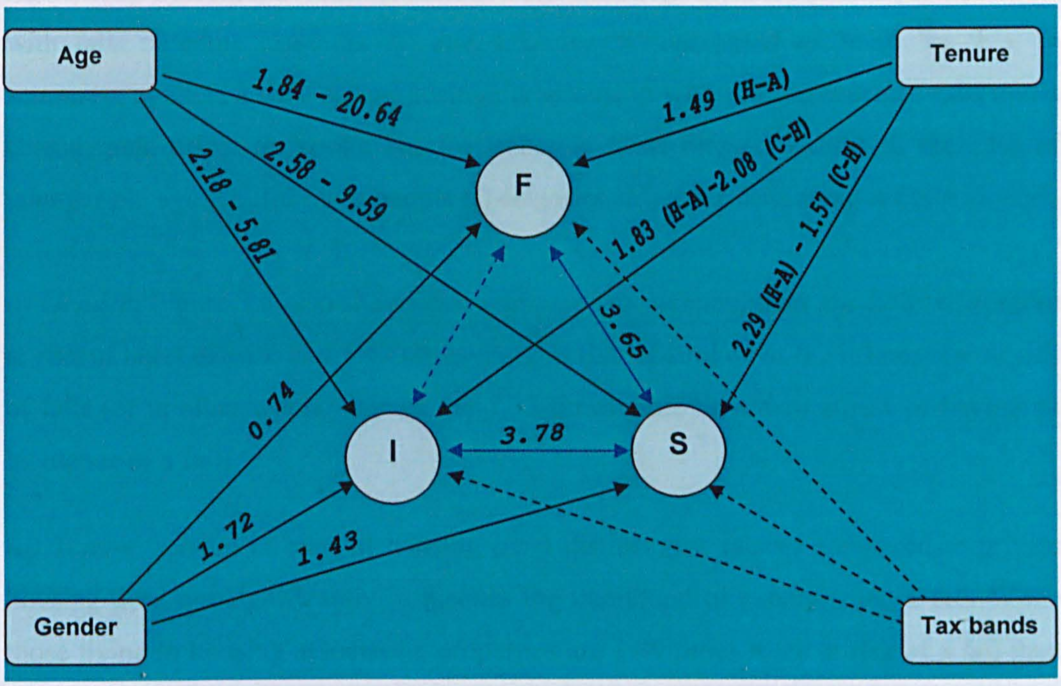
Table 7.3 Odds ratios of the three logistic regression models with outcome falls, heart disease and strokes.

	Outcome Variable					
	Falls		Heart-D		Stroke	
Predictor	OR	Sig.	OR	Sig.	OR	Sig.
Gender	0.74	0.000	1.72	0.000	1.43	0.000
Age-2	1.84	0.000	2.18	0.000	2.58	0.000
Age-3	4.48	0.000	3.80	0.000	5.13	0.000
Age-4	12.31	0.000	4.36	0.000	7.11	0.000
Age-5	20.64	0.000	5.81	0.000	9.59	0.000
Tenure_2	1.49	0.002	1.83	0.000	2.29	0.000
Tenure_3	1.09	0.359	2.08	0.000	1.57	0.000
Tax band_2	0.99	0.902	1.04	0.717	0.99	0.904
Tax band_3	1.17	0.155	1.02	0.883	0.97	0.794
Falls			1.38	0.053	3.68	0.000
Heart Diseas	1.37	0.056			3.81	0.000
Strokes	3.62	0.000	3.75	0.000		

Figure 7.3 below is also an illustration of Table 7.3. The numbers above the arrows in Figure 7.3 are the ORs shown in Table 7.3. In Figure 7.3 the variable tenure 'H-A' is short for Housing Association and 'C-H' represents Council Housing. For the variable

age in Figure 7.3, only the OR of Age-2 and Age-5 (minimum and maximum value of OR in Table 7.3) are demonstrated. The relationship between each of the four socio-demographic variables with three causes of hospital admissions shows:

Figure 7.3 The relationship between 4 socio-economic factors and three causes of hospital admissions (F, I and S) with odds ratios based on logistic regression modelling in Table 7.3



Legend	
Direction of the effect of the predictors	⟶
The effect of two predictors on each other (mutual effect)[12]	⟷
Non significant predictors	- - - - - - -> <- - - - - - ->
Numerical information is OR, 95% confidence interval	

i) Age: age has the strongest direct relationship with the increase incidence of falls. The risk of fall for those aged 60-69 years old (Age-2) compared with the reference category which includes those 50-59 years old (Age-1), is 1.84 times higher. However,

[12] Mutual effect is an average of the different OR for two predictors (causes of hospital admission) where one OR represents the first variable (e.g. Falls) as predictor (independent) and the second variable (e.g. Stroke) as outcome. The second OR represents the second variable (Stroke) as predictor and Falls as outcome. This is because we do not have the information on precedence of occurrence of the two factors (causes).

the risk of falls for those aged over 90 years old (Age-5) compared to the reference category is 20.64 times higher.

Figure 7.3 also shows the relationship between age and the two other outcome variables including heart disease and stroke. The ORs for the younger age group (Age-2) for both outcomes ('I' and 'S', respectively 2.18 and 2.59) is higher than the OR for Age-2 with outcome 'F' but the ORs for the oldest age is much smaller than the one with falls outcome (5.81 for 'I' and 9.59 for 'S' compared to 20.64 for 'F'). In summary, age has a very strong positive relationship with the incidence of falls and a considerable effect on stroke but for ischemic heart disease, except in the case of younger old age (i.e. for those people 60-69 years old) this relationship is quite modest.

ii) Gender: Figure 7.3 also shows that compared to women, men are 1.72 times more at risk of heart disease and 1.43 times more at risk of stroke but 0.74 times less at risk of falls (or in other words, women are 1.35 times more than men at risk of having an incidence of a fall).

iii) Tenure: Living in council housing (as a distinct risk factor) compared to private housing does not significantly influences the likelihood of experiencing a fall. While those living in housing association properties are 1.49 times more at risk of a fall than those living in private housing. The relationship between tenure and heart disease is quite strong. For those living in council housing, the likelihood of suffering from heart disease is greater than for those living in housing association property (respectively 2.08 and 1.83 times higher than for those living in private housing). The influence of tenure on stroke is significant for both council housing and housing association property but contrary to heart disease, the risk of stroke is higher for those living in housing association property. So, the high risk of falls could be the consequence of the high risk of stroke for those living in housing association property (this was discussed earlier in Section 7.2).

iv) Council tax bands: the effect of council tax bands is not significant on any of the causes of hospital admissions (shown by dotted lines in Figure 7.3).

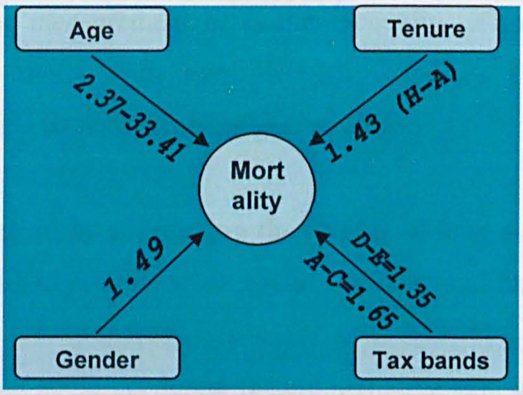
v) The effect of the three causes of hospital admissions on each other: apart from the consequences of four socio-demographic factors on three causes of hospital admissions, there are also some mutual influences of the three causes on each other which are illustrated with a double-sided arrow in Figure 7.3. The numbers next to each arrow are the average of two ORs of two variables in two sides of the arrow (i.e. the average of the ORs of the variable F with outcome S and the variable S with outcome F). The reason for calculating the average is that each of the two causes (on either side of the arrow) acts as both a predictor and an outcome variable in each of the models. For example in Table 7.3 when F is an outcome variable, the OR of S is 3.62 and when S is an outcome variable, the OR of F is 3.68 and the average of these two is 3.65 (shown in Figure 7.3). Thus the mutual effect of F and S on each other is 3.65 and the mutual effect of I and S on each other is 3.78, though this relationship between F and I is not significant. However, more accurate analysis of the above impact could be carried out if information about the sequence of occurrence of each cause were available. More precisely, if the date of each incident was available, then it would be possible to assess the relative impact of each cause on the other.

7.3.2 Assessing the impact of each factor on mortality

The impact of each predictor on each of the three causes of hospital admissions was discussed earlier. In this section, I am going to discuss the impact of each of the predictors together with socio-demographic factors and causes of hospital admissions on mortality. In Figure 7.4 the OR of four socio-demographic factors and their categories are shown. For the variable age, the risk of mortality for different age groups compared to the reference category (Age-1, 50-59 years old), varies from 2.37 times for Age-2 (60-69 years old) to 33.41 times for Age-5 (90 years old or more). For gender, the risk of mortality for men is approximately 1.5 times higher than women.

The risk of mortality for those living in council housing compared to those living in private housing is not significant but those living in housing association (H-A) properties, are 1.43 times more at risk of mortality than those who live in private housing. Those living in properties with tax bands D or E are 1.35 times more at risk and those living in properties with tax bands A, B or C are 1.65 times more at risk of mortality than those living in the highest tax bands properties (with band F, G or H).

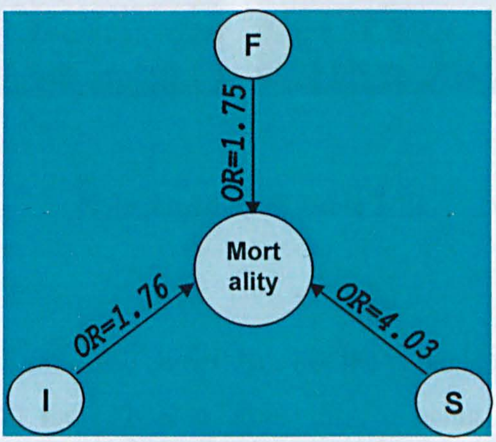
Figure 7.4 Illustration of the relative impact of four socio-demographic factors on mortality using odds ratios



Note: based upon Table 5.4

In Figure 7.5 the impact of the three causes of hospital admissions (F, I and S) on mortality with their odds ratios is illustrated. The odds ratios for falls and ischemic heart disease are 1.75 and 1.76 respectively which means those who had an incidence of fall and heart disease, in that order, are 1.75 and 1.76 times more at risk of mortality than those who did not have. This increase of risk for stroke is much higher. Those who had an incidence of stroke are more than four times at risk of mortality than those who did not have an incidence. It suggests that the victims of stroke should be at the top priority of the social services for allocation of their resources.

Figure 7.5 Illustration of the impact of three causes of hospital admissions (F, I and S) on mortality with their odds ratios



Note: based upon Table 5.4

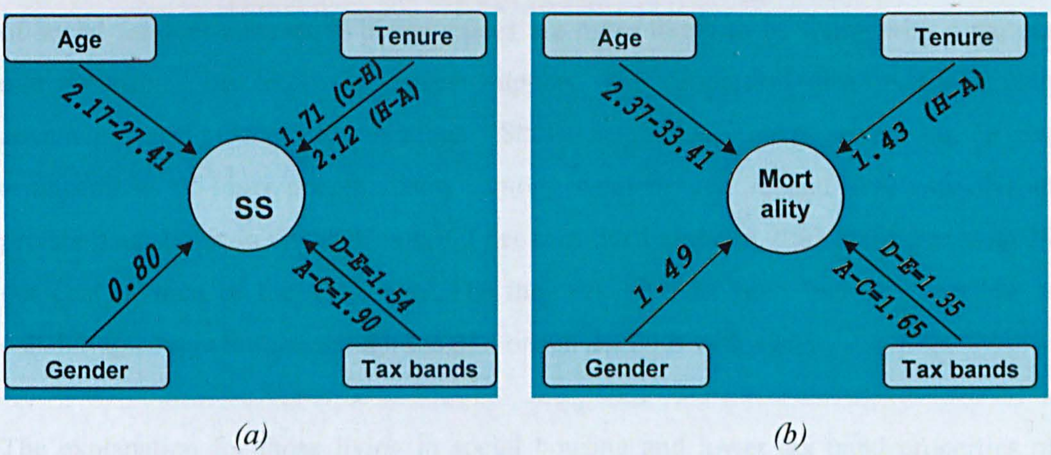
7.3.3 Policy implications by assessing the impact of each factor on probability of being in contact with social services

Logistic regression modelling with social services outcome provides evidence for the relative importance of each factor. Earlier in this study I discussed that mortality is a good indicator for the measurement of health inequality. Therefore, comparing the values of the same factor with two different outcomes ('social services' and 'mortality') may provide us with some insight for policy.

7.3.3.1 Policy implication by comparing the impact of four socio-economic factors on probability of being in contact with 'social services' and 'mortality' outcome

The odds ratios for four socio-economic factors including age, gender, housing tenure and council tax banding for both outcome variables 'social services' and 'mortality' are shown in Figure 7.6. Figure 7.6-a represents the model with 'social services' as the outcome and Figure 7.6-b is a copy of Figure 7.4.

Figure7.6 Odds ratios for age, gender, tenure and tax band a) the outcome variable 'social services' b) the outcome variable 'mortality'



Note: based upon Table 5.12

Comparing Figure 7.6-a and 7.6-b shows that: For the variable age (for all 4 categories exposed in Table 5.12) the estimated probability of being in contact with social services, is close to the estimated probability of mortality for the same age groups which is reasonable. It indicates that regarding the variable age, people with 'equal risk' of mortality have 'equal access' to being in contact with social services.

For gender, the model suggests men are 0.8 times less likely than women (or women are 1.25 times more than men) to be in contact with social services, while men are 1.49 times more at risk of mortality. Assuming equal possibility of using resources of social services based on risk of mortality, women are 1.86 times more in contact with social services than men.

Working with the same assumption, those living in housing association properties are 1.48 times more likely to be in contact with social services and those living in council housing are 1.6 times more likely compared to those living in private housing.

For both lower tax band categories (including 'D-E' and 'A-C') the model shows a slightly higher level of services by social services than level of their relative impact on mortality, which is reasonable.

The reason for women being more in contact with the social services than men could be that females tend to out-live males and could therefore be 'living alone' and in need of social services support. Whereas males are more likely to be living with a partner and therefore, less likely to request support. The proportion also increases with advancing age. Office for National Statistics (2004a) in a report on 'living arrangements of older people states: "among women aged 75 and over who live in private households in Great Britain, 60 per cent lived alone in 2002 compared with 29 per cent of men of the same age...The majority of older men live in a married or cohabiting couple family, though the proportion declines with age".

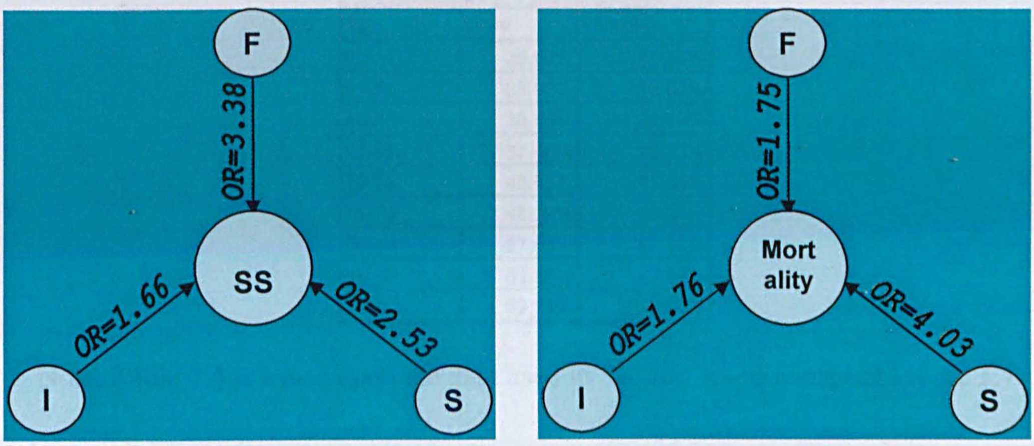
The explanation for those living in social housing and lower tax band properties of being more in contact with social services than those people living in private housing and higher tax band could also be related to the level of wealth, e.g. where wealthier people are in a position to pay for private carers.

7.3.3.2 Comparing the impact of three causes of hospital admissions on probability of being in contact with 'social services' and 'mortality' outcome: Policy implications

The odds ratios of three causes of hospital admissions including falls, ischemic heart disease and strokes for both outcome variables 'social services' and 'mortality' are illustrated in Figure 7.7. Figure 7.7-a represents the model with 'social services' outcome and Figure 7.7-b is a copy of the Figure 7.5.

Again, comparing the odds ratios of each cause of hospital admission in Figures 7.7-a and 7.7-b shows that those who had at least one incidence of hospital admission as a result of 'fall' are 1.93 times more likely to be in contact with social services than those who did not have. For those who had a stroke the odds are reversed. Those who had a stroke are 0.63 times less (negative) likely to be under the care of social services. For ischemic heart disease with ORs of 1.66 with social services outcome and 1.76 with mortality outcome, it shows an appropriate allocation of social services resources.

Figure7.7 Odds ratios for falls, heart disease and strokes; a) the outcome variable, 'social services' b) the outcome variable, 'mortality'



(a) *(b)*

Note: based upon Table 5.12

One explanation for this disparity in allocation of the social services resources for those who had at least one incidence of hospital admission as a result of falls or strokes is that, age is an important factor in determination of the likelihood of falls (as discussed in Section 7.2, illustrated in Figure 7.3). Therefore one possibility could be that older people with an incidence of fall tend to be less able to participate in physical activities than those of a younger age with an incidence of stroke. However this assertion requires further investigation. A second possibility also mentioned earlier (the relationship between gender and the incidence of falls in Figure 7.3) is that females are more at risk of having a fall than males. It has also been suggested that women in older age are more likely to live alone than men. Thus, the factors of being in an older age group, female and experiencing falls helps to explain the gender disparity of being in contact with social services.

It is noteworthy that the limiting long term illness (LLTI) data from the 2001 census also supports the above claim. Table 7.4 below is produced from the 2001 census data. Table 7.4 shows the percentage of males and females aged 50 years old or more in Camden with LLTI for every five years age group. As can be seen in Table 7.4, the proportion of females with LLTI aged 70 years old or more is higher than for males.

Table 7.4 Percentage of male and female aged 50 years old or more in the Borough of Camden with LLTI for every five years age group based on census 2001

Percentage of people with LLTI		
Age	Male	Female
50-54	25.92%	25.32%
55-59	28.96%	30.07%
60-64	36.90%	32.17%
65-69	37.46%	35.83%
70-74	43.63%	43.90%
75-79	48.72%	50.34%
80-84	57.88%	59.10%
85-89	61.60%	67.67%
90+	69.71%	73.96%

Note: Table 7.4 is based upon the data used in Section 1.4 to construct Figure 1.7

In summary, the findings from risk ladder and logistic regression modelling suggest that, in general, males, living in social housing and lower council tax bands properties (indicative of lower socio-economic status) are most at risk of mortality confirming evidence spanning a broad swathe of history (Chadwick, 1842; Dorling et al., 2001; Macintyre et al., 2001; Washington State Department of Health, 2002b; White et al., 2006). It also confirms that those people, who had at least one incidence of hospital admission as a result of stroke, fall and heart disease respectively, are at higher risk of mortality. The findings of logistic regression also clearly indicate that poor health is related to the lower socio-economic position of individuals (Auerbach & Krimgold, 2001; Bowling, 2004; Buchanan, 2003; Mackenbach et al., 1997).

The findings also suggest that the allocation of resources by social services, at least in terms of being aware of a person's needs, in general, targets the most in need and the most vulnerable people. However there are also some disparities which would benefit from further investigation.

8 Discussion and Conclusions

8.1 Introduction

In the previous chapter (Chapter-7) the interpretation of the substantive findings from the study were discussed and some constructive policy suggestions were provided. This final chapter brings together a resume of aims and objectives, a discussion on the limitations of the study itself and potential future work. The final section will also provide a succinct conclusion based on the research in previous sections.

8.2 Resume of aims and objectives

The principle aim of this project was to enhance Camden Primary Care Trust's strategy to specifically characterize the needs of the older people in response to NSF Standard 8. More specifically the precise aims of this research were:

i) To combine several administrative data sources in the London Borough of Camden for 2002-04 in order to draw relevant explanatory factors (potential risk factors) of the population of study for further analysis.

ii) To examine the relationship between mortality rates and the extracted risk factors including; age, gender, housing tenure, council tax banding and three popular causes of hospital admission (falls, ischemic-heart disease and strokes).

iii) To examine the relationship between those who have been in contact with 'social services' and the above potential risk factors in order to assess the appropriateness of the targeted population for service delivery.

iv) To enhance Camden Primary Care Trust's strategies by identifying older people's needs in ill-health prevention and health promotion in specific geographical locations in response to NSF Standard 8.

The study met all the above objectives. The details about the data preparation procedure were discussed in Chapter 3. In Chapter 4 and 5 the relationship between different risk factors with both outcomes including risk of mortality and the probability of being in contact with social services were examined. Subsequently the appropriateness of the degree of involvement of social services was assessed. In order to prevent the older people's ill-health and to promote their health, by assessing the main health related risk factors, some constructive policy implications were provided in Chapter 7.

The approach employed within this environment could be applied to conduct research in many other areas such as bioinformatics, medicine, social science, and economics including health economics. It could also be modified or improved upon based on the available data source and the requirements.

8.3 Discussion

i) Ethical issues surrounding the merging individual records: In any research that uses administrative data about people there is an inevitable 'trade-off' between protecting the individual's right to anonymity and privacy, and the use of the data for the 'public good'. There is an overall agreement by all experts that countless lives have been saved or improved as a result of using health information in medical research. The research and development of Department of Health and National Health Services (1998) states: "The aim of the Department of Health is to improve the health and wellbeing of the population and to secure high quality care for those who need it. Research is a powerful means of achieving these objectives".

A report 'Personal Data for public good: using health information in medical research' by the Academy of Medical Science (2006) says that over-strict interpretation of data protection rules is stifling health research and may be causing tens of thousands of unnecessary deaths and injuries each year. There is, however, a need to find a balance between facilitating important research and protecting the confidentiality of patients (Strobl et al., 2000).

During this research Camden PCT and Camden Council agreed an Overarching Information Sharing Protocol. The primary objective of the protocol is "...to improve the speed and efficiency of health and social care within Camden, without compromising the confidentiality and integrity of personal information" (London Borough of Camden, 2007b). The protocol sets out the agreement for the sharing of information between Camden Primary Care Trust (PCT) and Camden Council taking into account the effect of relevant legislation, guidance, plus common law, upon the way information is shared and used. The PCT has also agreed a service specific information sharing agreement – 'Integrating Services for children and young people (ISA)' with Camden Council. The overall aim of ISA is "...to improve services for children and young people through better multi agency working and information sharing" (London Borough of Camden, 2007c).

ii) Housing tenure: The information on housing tenure used in this study is based on Camden's Local Area Shared Information Resources (LASIR) data base. The basis of the housing tenure in LASIR is the list of Camden's Registered Social Landlords (RSL) and Camden housing stock. Based on information from Camden Council, more than 20% of the council properties are owned by tenants following the 'right to buy' legislation in 1990. So, the occupants of these properties are no longer counted as council tenants but as leaseholders. In the list of RSL those properties that are leasehold are not flagged separately. In 'LASIR' the housing tenure is divided into three categories; council housing, housing association and private housing. Therefore anyone who does not live in council housing or housing association properties, (e.g. those living in 'private rented properties' or 'living rent free', a term used in census 2001) are listed in private housing. Based on information from census 2001, around 13% of people aged 50 years old or more in Camden are living in 'private rented' and 2.4% are living in 'rent free' accommodations (Office for National Statistics, 2006a). The analysis of extracted data from Census 2001 related to the housing tenure in Scotland shows that a large number of people (around 65%) of those identified as 'living rent free' were actually social housing tenants (Boag, 2003). Similar information for England, London or Camden could not be found.

The socio-economic status of those living in private rented accommodations is also a matter of concern. The question remains: can we place those living in private rented homes in the same category as those people who own their home? Easterlow & Smith (2004) argue that private rented properties not only are in worse repair than owner-occupied properties but also are in worse repair than the local authority housing stock. In the 2004 English House Condition Survey (EHCS) some 27% of the owner-occupied stock is defined as non-decent, compared with 31% of social housing and 43% of private rented accommodation (Department for Communities and Local Government, 2006). Table 8.1 adapted from the annual report of EHCS by Department for Communities and Local Government (2006) shows the percentage of non-decent homes by tenure.

Table 8.1 Change over time, 1996–2004- Non-decent homes by tenure (Department for Communities and Local Government, 2006)

	owner occupied	private rented	all private	local authority	RSL	all social	all dwellings
% within tenure							
1996	39.7	62.4	42.6	53.9	47.6	52.6	44.7
2001	29.2	50.7	31.9	41.8	33.2	38.9	33.3
2003	27.7	47.5	30.2	39.6	28.8	35.3	31.2
2004	26.6	42.6	28.7	34.9	26.2	31.3	29.2

Base: all dwellings

The information on housing tenure for older people in Camden (those aged 50 years old and more) from census 2001 shows 42.2% own their home, 42.4% live in Council homes or housing association rented homes and 15.4% are living in private rented or 'living rent free' accommodations (Office for National Statistics, 2006a). Assuming 2.4% of those living in rent free accommodations are having their rent paid by the Local Authorities (such as housing benefit) and 13% of the reminder are living in private rented properties, the total percentage of private housing and social housing accordingly is 55% and 45%. However based on information from LASIR, these figures for private and social housing in the same order are 47.5% and 52.5% (it shows 7.5% increases on social housing and decrease on private housing comparing with the information from census 2001).

These differences between the census figures and LASIR could have several explanations. For example one of them could be the response rates for the 2001 Census. As a whole the response rate for all age groups in London Borough of Camden in census 2001 was 77%, one of the lowest rates in the UK, and compared with 88% for inner London. However in the 2001census in general the response rate for older people was higher than the overall rate. In Camden the response rate for those aged 50 years old or more varies between 80% - 95% (for those aged 50-80 it was less than 90% and for those over 80 years old which doesn't include too many people it was above 90%).

This highlights that still more than 10% of population are missing (Office for National Statistics, 2006b) which could be one of the sources of the uncertainty in extraction of the proportion of private and social housing. The different interpretation of the terms used in Census 2001 such as 'private rented' and 'living rent free' as discussed earlier could also be other sources of difference. Inclusion of those properties bought by former council tenants following the 'right to buy' legislation is also another potential source of difference.

Despite these limitations, it seems the data from LASIR used in this study is more reliable as it is based on the actual existing data in the London Borough of Camden. The only weakness of the LASIR housing tenure data might be the inclusion of the council leaseholders in the social housing list. However, including the council's leaseholders in social housing seems to be more appropriate than including them in private housing because the socio-economic background of the majority of this group of people is closer to those living in social housing than to those in private housing. Any future research using housing tenure as a factor needs careful consideration.

iii) Sequence of occurrence of causes of hospital admissions: The data used in this study shows that a large proportion of older people were admitted to hospital more than once and each time for different causes. One of the best and most efficient uses of the hospital admissions data could be the search for the sequence of occurrence of multiple causes of hospital admission for each individual. Keeping the record of the sequences of each admission via provided information on date of each incidence will help to draw more confident conclusions on prevention of ill-health. For example by knowing for the majority of cases of multiple hospital admissions, strokes comes before falls will help to divert the available resources first on prevention of strokes. Furthermore by having the knowledge of someone having already had an incidence of stroke, prevention of an incidence of falls for the same person with consideration of some other factors such as 'age' etc, would be more feasible.

iv) Further potential expansion of the study: The methodological approach adopted in this thesis could be extended developed by combining and enhancing the data from different sources at both individual and area level. Additional individual level data from local authorities such as information on marital status (single, widowed/divorced, married etc) might be included. The area level information on deprivation (e.g. from the publication of Office of the Deputy Prime Minister (ODPM)), data provided by

Office for National Statistics and aggregated data extracted from Census like levels of Limiting Long Term Illness (LLTI) would also be useful. In addition further investigation on any other causes of hospital admissions, apart from those discussed in this study could also be extracted and analysed. Similar approaches include (Raymer et al., 2007; Agerbo et al., 2007; Johnston et al., 2006).

v) Use of Artificial Intelligence's prediction techniques for health: The evidence points to predictive models employed having impressive predictive ability. However, within the category of predictive modelling there are a large variety of techniques, some of which are more developed than others. Literature on the subject is extensive, yet it is clear that there is no single consensus as to which technique is best. The most developed approach uses regression models but there is emerging interest in using artificial intelligence (King's Fund, 2005). In recent years, new models for predicting risk have been developed based on artificial intelligence. These models can utilise neural networks, regression, decision trees, fuzzy logic etc (Axelrod & Vogel, 2003). Some studies suggests that models that use artificial intelligent techniques provide a higher predictive power than typical regression models; Axelrod & Vogel (2003) claim that the accuracy of the R^2 statistic of the artificial intelligence models is more than twice that of the traditional regression model. Therefore artificial intelligence techniques could potentially be used to develop more advanced models for health risk predictions and management.

vi) Potential of Geographical Information System (GIS) for further application: In GIS applications maps represent a graphical means of visualizing the extent to which there is a geographical patterning in the various risk factors. The GIS application can be used to present the output of risk ladder analysis as geographical maps and to identify high pockets of risk across the Borough. All individuals under study can be geo-referenced (assigned an x, y co-ordinate) so that the level of risk for each individual in each risk ladder could be mapped (Mayhew, 2004).

Estimated risks/probabilities from logistic regression analyses could be presented as a mapping exercise. In principle, by allocating the estimated risk/probability of each combination of risk factors to the relevant individuals, a risk map using the GIS techniques can be produced.

The thesis was originally conceived to include GIS to aid the interpretation of risk ladder methodology and logistic regression analysis. However, in application it was decided that the resolution of the maps available in MapInfo did not provide sufficient detail to add anymore to the interpretation of the findings.

8.4 Conclusions

The use of routinely and daily collected administrative data is going to be one of the main sources of many research projects in public and private sectors in the immediate and long-term future both nationally and internationally (Bruhn, 2001; Jones & Elias, 2006; Redfern, 2004). The process of data preparation; including data collection, cleaning, linkage, integration and variable creation is a time consuming process. However, by improving the tools, techniques and methods involved in different stages of the work, it can be done much faster. The ideal system of data collection and processing for any organization with any dimension is a Relational Database Management System (RDBMS). A reliable RDBMS will make a huge reduction in the cost and time, along with a big boost in the quality of most of the research projects based on administrative data. A RDBMS particularly could significantly enhance research related to the public health including epidemiology, primary care, health policy, health economy and many other areas of medical research, which deals with a large amount of the applicable administrative data. For the purpose of this research the data preparation process was successfully completed and discussed in Chapter 3.

In addition, the findings of risk ladder approach in Chapter-4 explain the observed risk of mortality and the probability of someone being in contact with social services for many different group of people with similar socio-economic and health related characteristics. The findings of the risk ladder also suggest that the variables 'age' and the three causes of hospital admissions (FIS) are the most influential factors in determination of both 'risk of mortality' and allocation of social services' resources. It also shows that for variable 'gender' while men are at relatively higher risk of mortality, females have more chance of being in contact with social services. The impact of housing tenure and council tax bands based on result of the risk ladders is also high on both outcomes.

The relative impact of each factor on outcome variables 'mortality' and 'social services' with help of several models and rigorous tests derived from logistic regression modelling were examined in Chapter 5. The findings of logistic regression modelling confirm that all socio-economic factors, including health related factors and their relevant categories are highly significant in determination of the risk of mortality and the allocation of the social services' resources. The results also confirm the findings of risk ladders relating to the higher effect of variable 'age' and three causes of hospital admissions, especially 'strokes'.

In Chapter 7 through a further exploration on policy implication, the relative impacts of the socio-economic factors on three causes of hospital admission were examined with clear outcomes. They clearly highlighted that ill-health is typically rooted in the lower socio-economic status of individuals. Having found that the low socio-economic status of a person determines their ill-health and also being aware of the negative correlation between ill-health and mortality, the impact of low socio-economic position on pre-mature death is clear, and indeed an obvious link.

In this work it has also been identified that all variables and their relevant categories with outcome 'social services' are highly significant. By comparing the findings from mortality and social services outcomes some visible disparities for variables; gender, housing tenure, falls and strokes were exposed. Examination of the models with continuous age and interaction effects did not show a significant impact on model improvement.

The findings of different models and their capability in justification of the data were studied with use of ROC curves. The results of the tests are promising and confirm the reliability of the methods both in justification of the available data and findings. The examination also corroborates that the outcome of the predicted risks are quite close to the observed risks and therefore consistent.

Some policy implications were also drawn from the analysis. Generally the allocation of resources by social services targets the most vulnerable people. However the study shows some disparities between the level of the risk of mortality and the allocation of resources made by social services, which could be the subject of further investigations.

The study commenced using the records at individual level which were aggregated for the purpose of the analyses. The aggregated records included different groups (clusters) of people with similar characteristics. Yet, for the purpose of investigation or further research, whenever it is required to break down each group to a smaller unit such as postcode level, household level and even at individual level, it would be possible to track them back and to identify them.

It is a key point to emphasize that so far many studies nationally and internationally have been conducted to measure the socio-economic and health inequalities and their relevant factors. It is also apparent that based on those studies, strategies and delivery plans/policies have been implemented and executed by the relevant authorities. Some of the most recent key policy strands that were aimed to reduce the socio-economic and subsequently health inequalities in UK are discussed in Chapter 1 (Section 1.1). However the central issue in inequalities, the 'spatial inequality', not only exists but the evidence shows (as discussed in Section 1.2), it is also getting wider.

Ultimately, the challenge for the future remains the necessity to ensure that resources are allocated to those most in need and as this study has shown this is often a very complex task. Mkandawire (2005) states: "For much of its history, social policy has involved choices about whether the core principle behind social provisioning will be universalism or selectivity through targeting". Besley and Kanbur (1990) also pointedly observe; "...improved targeting means that more poverty alleviation can be achieved with less expenditure".

This study provides a clear approach to identifying health inequalities and measuring the relevant factors at the individual level. As a result through adapting the right 'social policy' based on identifying and targeting those most in need the spatial inequality could be tackled. While logistic regression methodology provides us with a broad and clear measure of the relative importance of each factor on the outcome in general (on the whole population under study), risk ladder approach seems to be a useful tool to ensure we are targeting those people most in need as opposed to a universalism approach.

LIST OF APPENDICES

Appendices

Appendix-A A summary of eight standards of NSF for older people

Standard One: Rooting out age discrimination

Standard: NHS services will be provided, regardless of age, on the basis of clinical need alone. Social care services will not use age in their eligibility criteria or policies, to restrict access to available services.

In some health and social care services, older people and their carers have experienced age-based discrimination in access to and availability of services. Older people from black and minority ethnic groups can be particularly disadvantaged and are likely to suffer more discrimination in accessing services.

This standard has been set up to ensure that older people are never unfairly discriminated against in accessing NHS or social care services as a result of their age.

Standard Two: Person-centred care

Standard: NHS and social care services treat older people as individuals and enable them to make choices about their own care. This is achieved through the single assessment process, integrated commissioning arrangements and integrated provision of services, including community equipment and continence services.

Proper assessment of the range and complexity of older people's needs and prompt provision of care (including community equipment and continence services) can improve their ability to function independently, reduce the need for emergency hospital admission and decrease the need for premature admission to a residential care setting. Person-centred care needs to be supported by services that are organised to meet needs. This includes the introduction of a single assessment process in health and social care to ensure that older people's needs are assessed and evaluated fully; improved access to community equipment; and the establishment of integrated continence services.

Standard Three: Intermediate care

Standard: Older people will have access to a new range of intermediate care services at home or in designated care settings, to promote their independence by providing enhanced services from the NHS and councils to prevent unnecessary hospital admission and effective rehabilitation services to enable early discharge from hospital and to prevent premature or unnecessary admission to long-term residential care.

Standard three requests a new range of acute and rehabilitation services to bridge the gap between acute hospital and primary and community care. For example the National Beds Inquiry (NBI) found that significant numbers of older people stay in acute hospitals longer than is necessary or desirable.

Standard Four: General hospital care

Standard: Older people's care in hospital is delivered through appropriate specialist care and by hospital staff who have the right set of skills to meet their needs.

At any one time, older people occupy around two-thirds of hospital beds. Too often the older person's experience of hospital care has been of outdated and unclean wards which have undermined their need for privacy and damaged their confidence in other aspects of care.

Action is needed to improve the clinical care of older people in general hospitals, through ensuring; early access to the specialist team in a general acute hospital, appropriate attention to the health status of the older person while in hospital, privacy and overall quality of care – and through new investment to convert many old 'Nightingale' wards to older people-friendly environments; single sex accommodation, more privacy, and more space for rehabilitation equipment.

Standard Five: Stroke

Standard: The NHS will take action to prevent strokes, working in partnership with other agencies where appropriate.

People who are thought to have had a stroke have access to diagnostic services, are treated appropriately by a specialist stroke service, and subsequently, with their carers, participate in a multidisciplinary programme of secondary prevention and rehabilitation.

Stroke is the single biggest cause of severe disability and the third most common cause of death in the UK and other developed countries. Some population groups are at higher risk of stroke than others. The risk is higher for men from African-Caribbean and South Asian communities and in those in lower socioeconomic groups.

This standard sets out four main components for the development of integrated stroke services: prevention, immediate care, early and continuing rehabilitation and long-term support for stroke patient and their carers.

Standard Six: Falls

Standard: The NHS, working in partnership with councils, takes action to prevent falls and reduce resultant fractures or other injuries in their populations of older people.

Older people who have fallen receive effective treatment and, with their carers, receive advice on prevention through a specialised falls service.

Falls are a major cause of disability and the leading cause of mortality due to injury in older people aged over 75 in the UK. Every year, over 400,000 older people in England attend A&E Departments following an accident and up to 14,000 people a year dies in the UK as a result of an osteoporotic hip fracture.

A fall can precipitate admission to long-term care. Fear of falling can provide a significant limitation on daily activities.

The aim of Standard six is to reduce the number of falls which result in serious injury and ensure effective treatment and rehabilitation for those who have fallen. Action will also be taken on prevention; to reduce the incidence of falls and treatment of osteoporosis.

Standard Seven: Mental health in older people

Standard: Older people who have mental health problems have access to integrated mental health services, provided by the NHS and councils to ensure effective diagnosis, treatment and support, for them and for their carers.

The aim of this standard is to promote good mental health in older people and to treat and support those older people with dementia and depression.

Mental health services for older people should be able to respond effectively to individual needs, and take account of the social and cultural factors affecting recovery and support. Improving prevention, care and treatment of mental health problems in old age depends on: promoting good mental health, early recognition and management of mental health problems and access to specialist care. Mental health services for older people should be community-orientated and provide seamless packages of care and support for older people and their carers.

Standard Eight: The promotion of health and active life in older age

Standard: The health and well-being of older people is promoted through a co-ordinated programme of action led by the NHS with support from councils.

There is a growing body of evidence to suggest that the modification of risk factors for disease even late in life can have health benefits for the individual; longer life, increased or maintained levels of functional ability, disease prevention and an improved sense of wellbeing. Integrated strategies for older people aimed at promoting good health and quality of life, and to prevent or delay frailty and disability can have significant benefits for the individual and society. Therefore, the NHS and local partners should re-focus on helping and supporting older people to continue to live healthy and fulfilling lives by:

- Access to mainstream health promotion and disease prevention programmes.
- plan for increasing physical activity, improved diet and nutrition, immunisation and management programmes for influenza
- Wider initiatives involving a multi-sectoral approach to promoting health, independence and well-being in old age: exercise services, healthy eating, *keep Warm, Keep Well* campaign, Home Energy Efficiency Scheme.

Appendix-B Steps were taken in finding/allocation of an ethnicity to each record in the mortality list

1. Hospital admission data includes a field for the ethnicity and has been used as the main source of the ethnicity information. There are approximately 175,000 hospital records for the period of 1996-2004 and 66,000 for the period of 2002-04 for Camden's citizens over 50 years old. 93,000 out of 175,000 of the hospital admissions for the period of 1996-2004 and 56,000 out of 66,000 of hospital admissions for the period of 2002-04 are without address and therefore are not possible to allocate a Unique Property Reference Number (UPRN) for them. Without having a UPRN for each record in Hospital Admission list, it is difficult to link the two tables together. Therefore; instead of UPRN, the NHS number which is in the both tables has been used as a primary key to link them together;

a) Approximately for 1,400 records in mortality list an ethnicity value were extracted from hospital admission records.

b) By cross-checking the mortality list with 'Social Service' the ethnicity values were found for around 250 records.

c) By cross-checking the mortality list with 'School pupil-roll' for 162 records the ethnicity values were found. For the total number of 412 records from both 'Social Service' and 'School Pupil Roll', 212 records had already been allocated an ethnicity value from the hospital admission, therefore for 200 more records the ethnicity was extracted from Social Service and School Pupil Roll.

d) For around 1200 records from the combination of the 'Place of birth' (e.g. Bangladesh, Ireland or India) + First name and Surname, the ethnicity was extracted.

e) The most difficult task was to allocate an ethnicity to around 400 records in which place of birth was stated 'England or a specific city or borough within England), to solve this problem the following steps were taken:

i) To sort by surname and to look at other similar surname and Ethnicity

ii) To sort by First name and to find other similar First name and Ethnicity

iii) In case of similar First name and Surname with different ethnicity; the list was sorted by postcode first, then by looking at the ethnicity of the majority of people living in the same postcode, an ethnicity was assigned to that person (e.g. if majority of people in the same postcode are Irish, then the ethnicity 'Irish' was assigned to that individual).

Appendix-C the basis of Tax Bands

(Valuation Office Agency, 2006)

The Tax bands for England are as follows:

The starting point

The basis of valuation for a dwelling which is not used for any business purpose is the amount which, subject to certain assumptions, it would have sold for on the 'open market' by a 'willing vendor' on 1 April 1991.

- 'open market' means a market where the property is offered openly with adequate publicity being given to the sale. Please note that if your property was purchased under a discount scheme (such as 'Right to Buy') this does not fall within the definition of 'open market' and therefore will not apply.
- 'willing vendor' means someone who sells the property as a free agent and not someone who is forced to do so.

Band	Value
A	up to £40,000
B	£40,001 to £52,000
C	£52,001 to £68,000
D	£68,001 to £88,000
E	£88,001 to £120,000
F	£120,001 to £160,000
G	£160,001 to £320,000
H	£320,001 and above

Why 1 April 1991?

Council Tax came into effect on 1 April 1993. However, the process of valuing every domestic property in England and Wales for banding purposes started some time before this. Therefore, we had to adopt a valuation date prior to 1 April 1993 so that all properties would be valued on a common footing. Even if your property was built after 1 April 1993, we must band the property according to what we think that its value would have been on 1 April 1991. This means that recent sale prices are not necessarily a good guide to the correct band for a property.

Appendix-D Tables including the combined risk ladders with similar factors and different outcomes

Appendix-D includes four tables, each one a combination of two risk ladders with similar factors but different outcomes ('risk of mortality' and 'probability of being in contact with social services') discussed in Sections 4.2 and 4.4. in the following tables if the confidence interval in a row (combination) for one outcome variable includes values greater than or equal to one or less than or equal to zero, the entire row for both outcomes (with the same combination) is omitted.

D.1 Combination of two risk ladders with four socio-economic factors (Risk ladders 1.1 and 2.1)

Sixteen different combinations of four factors of both Risk ladder 1.1 (Table 4.2) and Risk ladder 2.1 (Table 4.7) are collectively presented in Table D.1 below.

Table D.1 Combination of two risk ladders with four socio-economic factors (Risk ladders 1.1 and 2.2)

	Combination	Mortality					Social Services				
Seq	'AGTB'	Number of Death	Population	Observed Risk	Conf. Interval		Known to 'SS'	Population	*P* of Known to 'SS'	Conf. Interval	
1	**0000**	87	6074	**1.4%**	1.1%	1.7%	79	6074	**1.3%**	1.0%	1.6%
2	**0001**	9	498	**1.8%**	0.6%	3.0%	11	498	**2.2%**	0.9%	3.5%
3	**0100**	135	7081	**1.9%**	1.6%	2.2%	60	7081	**0.8%**	0.6%	1.1%
4	**0010**	89	4268	**2.1%**	1.7%	2.5%	168	4268	**3.9%**	3.4%	4.5%
5	**0011**	93	2864	**3.2%**	2.6%	3.9%	147	2864	**5.1%**	4.3%	5.9%
6	**0110**	151	3949	**3.8%**	3.2%	4.4%	147	3949	**3.7%**	3.1%	4.3%
7	**0101**	29	730	**4.0%**	2.6%	5.4%	19	730	**2.6%**	1.4%	3.8%
8	**0111**	185	3423	**5.4%**	4.6%	6.2%	188	3423	**5.5%**	4.7%	6.3%
9	**1010**	325	2354	**13.8%**	12.4%	15.2%	492	2354	**20.9%**	19.3%	22.5%
10	**1100**	374	2571	**14.5%**	13.2%	15.9%	225	2571	**8.8%**	7.7%	9.8%
11	**1000**	458	3113	**14.7%**	13.5%	16.0%	440	3113	**14.1%**	12.9%	15.4%
12	**1011**	438	2486	**17.6%**	16.1%	19.1%	637	2486	**25.6%**	23.9%	27.3%
13	**1001**	56	311	**18.0%**	13.7%	22.3%	65	311	**20.9%**	16.4%	25.4%
14	**1110**	319	1668	**19.1%**	17.2%	21.0%	249	1668	**14.9%**	13.2%	16.6%
15	**1101**	49	248	**19.8%**	14.8%	24.7%	26	248	**10.5%**	6.7%	14.3%
16	**1111**	391	1834	**21.3%**	19.4%	23.2%	351	1834	**19.1%**	17.3%	20.9%
		3188	**43472**	**7.3%**	**7.1%**	**7.6%**	**3304**	**43472**	**7.6%**	**7.4%**	**7.8%**

Legend		
A = Age	G = Gender	T = Tenure
B =Tax Band	SS = Social Services	

D.2 Combination of two risk ladders with four socio-economic factors and incidence of an admission for a 'Fall' (risk ladders 1.2 and 2.2)

The second risk ladders for both outcome variables (Risk ladder 1.2 in Table 4.3 and Risk ladder 2.2 in Table 4.8) include the combinations of four socio-economic factors and 'Fall'. Table D.2 contains the outcomes of both risk ladders except for six combinations with confidence intervals either negative or greater than '1' which are omitted from the table.

Table D.2 Combination of two risk ladders with four socio-economic factors and incidence of an admission for a 'Fall' (risk ladders 1.2 and 2.2)

	Combination	Mortality					Social Services				
Seq	'AGTBF'	Number of Death	Population	Observed Risk	Conf.Interval		Known to 'SS'	Population	P of Known to 'SS'	Conf. Interval	
1	00000	81	6044	**1.3%**	1.1%	1.6%	78	6044	**1.3%**	1.0%	1.6%
2	00010	9	496	**1.8%**	0.6%	3.0%	10	496	**2.0%**	0.8%	3.3%
3	01000	129	7045	**1.8%**	1.5%	2.1%	56	7045	**0.8%**	0.6%	1.0%
4	00100	87	4241	**2.1%**	1.6%	2.5%	163	4241	**3.8%**	3.3%	4.4%
5	00110	89	2840	**3.1%**	2.5%	3.8%	142	2840	**5.0%**	4.2%	5.8%
6	01100	148	3917	**3.8%**	3.2%	4.4%	143	3917	**3.7%**	3.1%	4.2%
7	01010	28	725	**3.9%**	2.5%	5.3%	18	725	**2.5%**	1.4%	3.6%
8	01110	177	3376	**5.2%**	4.5%	6.0%	178	3376	**5.3%**	4.5%	6.0%
9	10100	287	2212	**13.0%**	11.6%	14.4%	426	2212	**19.3%**	17.6%	20.9%
10	11000	350	2501	**14.0%**	12.6%	15.4%	196	2501	**7.8%**	6.8%	8.9%
11	10000	413	2937	**14.1%**	12.8%	15.3%	372	2937	**12.7%**	11.5%	13.9%
12	10010	47	289	**16.3%**	12.0%	20.5%	53	289	**18.3%**	13.9%	22.8%
13	00111	4	24	**16.7%**	1.8%	31.6%	5	24	**20.8%**	4.6%	37.1%
14	01001	6	36	**16.7%**	4.5%	28.8%	4	36	**11.1%**	0.8%	21.4%
15	01111	8	47	**17.0%**	6.3%	27.8%	10	47	**21.3%**	9.6%	33.0%
16	10110	402	2345	**17.1%**	15.6%	18.7%	563	2345	**24.0%**	22.3%	25.7%
17	11100	300	1624	**18.5%**	16.6%	20.4%	234	1624	**14.4%**	12.7%	16.1%
18	11010	45	238	**18.9%**	13.9%	23.9%	24	238	**10.1%**	6.3%	13.9%
19	11110	368	1757	**20.9%**	19.0%	22.8%	311	1757	**17.7%**	15.9%	19.5%
20	10111	36	141	**25.5%**	18.3%	32.7%	74	141	**52.5%**	44.2%	60.7%
21	10001	45	176	**25.6%**	19.1%	32.0%	68	176	**38.6%**	31.4%	45.8%
22	10101	38	142	**26.8%**	19.5%	34.0%	66	142	**46.5%**	38.3%	54.7%
23	11111	23	77	**29.9%**	19.6%	40.1%	40	77	**51.9%**	40.8%	63.1%
24	11001	24	70	**34.3%**	23.2%	45.4%	29	70	**41.4%**	29.9%	53.0%
25	10011	9	22	**40.9%**	20.4%	61.5%	12	22	**54.5%**	33.7%	75.4%
26	11101	19	44	**43.2%**	28.5%	57.8%	15	44	**34.1%**	20.1%	48.1%
		3188	43472	7.3%	7.1%	7.6%	3304	43472	7.6%	7.4%	7.8%

Legend		
A = Age	G = Gender	T = Tenure
B =Tax Band	F = Falls	SS = Social Services

D.3 Combination of two risk ladders with four socio-economic factors and incidence of an admission for a 'Heart Disease' (risk ladders 1.3 and 2.3)

The comparison of the two risk ladders 1.3 (Table 4.4) and 2.3 (Table 4.9), the combinations of four socio-economic factors and Ischemic heart disease are illustrated in Table D.3 below. Eight combinations with confidence interval including values greater than or equal to one or less than or equal to zero are excluded from the table.

Table D.3 Comparing the Risk of mortality & the probability of being in contact with social services for different combinations of four socio-economic factors and 'Heart Disease'

	Combination	Mortality					Social Services				
Seq	"AGTBI"	Number of Death	Population	Observed Risk	Conf. Interval		Known to 'SS'	Population	*P* of Known to 'SS'	Conf. Interval	
1	**00000**	85	6050	**1.4%**	1.1%	1.7%	76	6050	**1.3%**	1.0%	1.5%
2	**00010**	9	495	**1.8%**	0.6%	3.0%	11	495	**2.2%**	0.9%	3.5%
3	**01000**	128	7026	**1.8%**	1.5%	2.1%	57	7026	**0.8%**	0.6%	1.0%
4	**00100**	87	4223	**2.1%**	1.6%	2.5%	165	4223	**3.9%**	3.3%	4.5%
5	**00110**	90	2830	**3.2%**	2.5%	3.8%	144	2830	**5.1%**	4.3%	5.9%
6	**01100**	144	3841	**3.7%**	3.1%	4.3%	135	3841	**3.5%**	2.9%	4.1%
7	**01010**	28	721	**3.9%**	2.5%	5.3%	16	721	**2.2%**	1.1%	3.3%
8	**01110**	176	3331	**5.3%**	4.5%	6.0%	179	3331	**5.4%**	4.6%	6.1%
9	**01101**	7	108	**6.5%**	1.8%	11.1%	12	108	**11.1%**	5.2%	17.0%
10	**01111**	9	92	**9.8%**	3.7%	15.9%	9	92	**9.8%**	3.7%	15.9%
11	**10100**	305	2267	**13.5%**	12.0%	14.9%	464	2267	**20.5%**	18.8%	22.1%
12	**11000**	350	2481	**14.1%**	12.7%	15.5%	212	2481	**8.5%**	7.4%	9.6%
13	**10000**	437	3042	**14.4%**	13.1%	15.6%	422	3042	**13.9%**	12.6%	15.1%
14	**10110**	407	2383	**17.1%**	15.6%	18.6%	600	2383	**25.2%**	23.4%	26.9%
15	**10010**	52	304	**17.1%**	12.9%	21.3%	63	304	**20.7%**	16.2%	25.3%
16	**11100**	293	1576	**18.6%**	16.7%	20.5%	232	1576	**14.7%**	13.0%	16.5%
17	**11010**	46	240	**19.2%**	14.2%	24.1%	24	240	**10.0%**	6.2%	13.8%
18	**11110**	373	1752	**21.3%**	19.4%	23.2%	332	1752	**18.9%**	17.1%	20.8%
19	**11111**	18	82	**22.0%**	13.0%	30.9%	19	82	**23.2%**	14.0%	32.3%
20	**10101**	20	87	**23.0%**	14.1%	31.8%	28	87	**32.2%**	22.4%	42.0%
21	**11001**	24	90	**26.7%**	17.5%	35.8%	13	90	**14.4%**	7.2%	21.7%
22	**11101**	26	92	**28.3%**	19.1%	37.5%	17	92	**18.5%**	10.5%	26.4%
23	**10001**	21	71	**29.6%**	19.0%	40.2%	18	71	**25.4%**	15.2%	35.5%
24	**10111**	31	103	**30.1%**	21.2%	39.0%	37	103	**35.9%**	26.7%	45.2%
		3188	**43472**	**7.3%**	0.000	**7.1%**	**3304**	**43472**	**7.6%**	**7.4%**	**7.8%**

Legend		
A = Age	G = Gender	T = Tenure
B =Tax Band	I = Ischemic Heart Disease	SS = Social Services

D.4 Combination of two risk ladders with four socio-economic factors and incidence of an admission for a 'Stroke' (risk ladders 1.4 and 2.4)

Table D.4 is a combination of Risk ladder 1.4 (Table 4.5) and Risk ladder 2.4 (Table 4.10). Table D.4 contains the observed risk of mortality and the probability of being in contact with 'SS' (for 26 combinations of four socio-economic factors and 'Stroke'. Six combinations with the confidence interval including values greater than or equal to one or less than or equal to zero are excluded from the table.

Table D.4 Combination of two risk ladders with four socio-economic factors and incidence of an admission for a 'Stroke' (risk ladders 1.4 and 2.4)

	Combination	Mortality					Social Services				
Seq	'AGTBS'	Number of Death	Population	Observed Risk	Conf. Interval		Known to 'SS'	Population	*P* of Known to 'SS'	Conf. Interval	
1	**00000**	82	6058	**1.4%**	1.1%	1.6%	78	6058	**1.3%**	1.0%	1.6%
2	**00010**	9	496	**1.8%**	0.6%	3.0%	10	496	**2.0%**	0.8%	3.3%
3	**01000**	130	7056	**1.8%**	1.5%	2.2%	56	7056	**0.8%**	0.6%	1.0%
4	**00100**	85	4241	**2.0%**	1.6%	2.4%	161	4241	**3.8%**	3.2%	4.4%
5	**00110**	88	2843	**3.1%**	2.5%	3.7%	145	2843	**5.1%**	4.3%	5.9%
6	**01100**	133	3898	**3.4%**	2.8%	4.0%	136	3898	**3.5%**	2.9%	4.1%
7	**01010**	27	726	**3.7%**	2.3%	5.1%	18	726	**2.5%**	1.3%	3.6%
8	**01110**	174	3376	**5.2%**	4.4%	5.9%	176	3376	**5.2%**	4.5%	6.0%
9	**10100**	297	2296	**12.9%**	11.6%	14.3%	471	2296	**20.5%**	18.9%	22.2%
10	**10000**	423	3028	**14.0%**	12.7%	15.2%	412	3028	**13.6%**	12.4%	14.8%
11	**11000**	351	2510	**14.0%**	12.6%	15.3%	205	2510	**8.2%**	7.1%	9.2%
12	**00101**	4	27	**14.8%**	1.4%	28.2%	7	27	**25.9%**	9.4%	42.5%
13	**10110**	403	2416	**16.7%**	15.2%	18.2%	609	2416	**25.2%**	23.5%	26.9%
14	**10010**	52	306	**17.0%**	12.8%	21.2%	64	306	**20.9%**	16.4%	25.5%
15	**11100**	303	1604	**18.9%**	17.0%	20.8%	235	1604	**14.7%**	12.9%	16.4%
16	**11010**	49	247	**19.8%**	14.9%	24.8%	25	247	**10.1%**	6.4%	13.9%
17	**01001**	5	25	**20.0%**	4.3%	35.7%	4	25	**16.0%**	1.6%	30.4%
18	**11110**	363	1767	**20.5%**	18.7%	22.4%	327	1767	**18.5%**	16.7%	20.3%
19	**01111**	11	47	**23.4%**	11.3%	35.5%	12	47	**25.5%**	13.1%	38.0%
20	**11101**	16	64	**25.0%**	14.4%	35.6%	14	64	**21.9%**	11.7%	32.0%
21	**01101**	18	51	**35.3%**	22.2%	48.4%	11	51	**21.6%**	10.3%	32.9%
22	**11001**	23	61	**37.7%**	25.5%	49.9%	20	61	**32.8%**	21.0%	44.6%
23	**10001**	35	85	**41.2%**	30.7%	51.6%	28	85	**32.9%**	22.9%	42.9%
24	**11111**	28	67	**41.8%**	30.0%	53.6%	24	67	**35.8%**	24.3%	47.3%
25	**10101**	28	58	**48.3%**	35.4%	61.1%	21	58	**36.2%**	23.8%	48.6%
26	**10111**	35	70	**50.0%**	38.3%	61.7%	28	70	**40.0%**	28.5%	51.5%
		3188	**43472**	**7.3%**	**7.1%**	**7.6%**	**3304**	**43472**	**7.6%**	**7.4%**	**7.8%**

Legend		
A = Age	G = Gender	T = Tenure
B =Tax Band	S = Strokes	SS = Social Services

Appendix–E Graphical illustration of observed risk for Risk ladders 1.1 – 1.4, 1.6 and 2.1-2.4 with 'high-low' 95% Confidence Interval (CI) bars

Note: In this appendix a copy of each risk ladder is located above the relevant 'high-low' 95% confidence interval bars. 'High' corresponds to the upper limit and 'low' the lower limit of each 95% confidence interval. The sequence number defined in each table represents a specific combination of risk factors is reproduced along the horizontal axis of each graph.

Table 4.2 Risk ladder-1.1; risk of mortality with four basic socio-demographic factors

Seq	Age	Gender	Tenure	Tax Band	Number of Death	Population	Observed Risk	Conf.Interval	
1	0	0	0	0	87	6074	1.4%	1.1%	1.7%
2	0	0	0	1	9	498	1.8%	0.6%	3.0%
3	0	1	0	0	135	7081	1.9%	1.6%	2.2%
4	0	0	1	0	89	4268	2.1%	1.7%	2.5%
5	0	0	1	1	93	2864	3.2%	2.6%	3.9%
6	0	1	1	0	151	3949	3.8%	3.2%	4.4%
7	0	1	0	1	29	730	4.0%	2.6%	5.4%
8	0	1	1	1	185	3423	5.4%	4.6%	6.2%
9	1	0	1	0	325	2354	13.8%	12.4%	15.2%
10	1	1	0	0	374	2571	14.5%	13.2%	15.9%
11	1	0	0	0	458	3113	14.7%	13.5%	16.0%
12	1	0	1	1	438	2486	17.6%	16.1%	19.1%
13	1	0	0	1	56	311	18.0%	13.7%	22.3%
14	1	1	1	0	319	1668	19.1%	17.2%	21.0%
15	1	1	0	1	49	248	19.8%	14.8%	24.7%
16	1	1	1	1	391	1834	21.3%	19.4%	23.2%
	14585	21504	22846	12394	3188	43472	7.3%	7.1%	7.6%

FigureE1 Illustration of observed risks and 'high-low' 95% confidence interval bars for Risk ladder 1.1 in Table 4.2

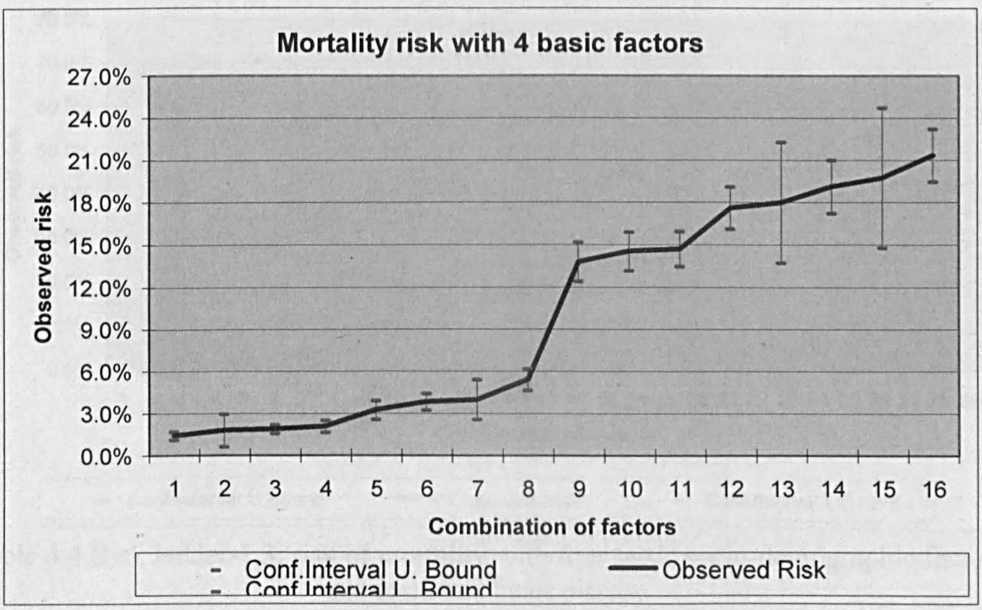


Table 4.3 Risk ladder-1.2; risk of mortality with four basic socio-demographic factors and the incidence of an admission for a fall

Seq	Age	Gender	Tenure	Tax Band	Fall	No. of Death	Population	Observed Risk	Conf.Interval	
1	0	0	0	1	1	0	2	0.0%	0.0%	0.0%
2	0	0	0	0	0	81	6044	1.3%	1.1%	1.6%
3	0	0	0	1	0	9	496	1.8%	0.6%	3.0%
4	0	1	0	0	0	129	7045	1.8%	1.5%	2.1%
5	0	0	1	0	0	87	4241	2.1%	1.6%	2.5%
6	0	0	1	1	0	89	2840	3.1%	2.5%	3.8%
7	0	1	1	0	0	148	3917	3.8%	3.2%	4.4%
8	0	1	0	1	0	28	725	3.9%	2.5%	5.3%
9	0	1	1	1	0	177	3376	5.2%	4.5%	6.0%
10	1	0	1	0	0	287	2212	13.0%	11.6%	14.4%
11	1	1	0	0	0	350	2501	14.0%	12.6%	15.4%
12	1	0	0	0	0	413	2937	14.1%	12.8%	15.3%
13	1	0	0	1	0	47	289	16.3%	12.0%	20.5%
14	0	0	1	1	1	4	24	16.7%	1.8%	31.6%
15	0	1	0	0	1	6	36	16.7%	4.5%	28.8%
16	0	1	1	1	1	8	47	17.0%	6.3%	27.8%
17	1	0	1	1	0	402	2345	17.1%	15.6%	18.7%
18	1	1	1	0	0	300	1624	18.5%	16.6%	20.4%
19	1	1	0	1	0	45	238	18.9%	13.9%	23.9%
20	0	0	0	0	1	6	30	20.0%	5.7%	34.3%
21	1	1	1	1	0	368	1757	20.9%	19.0%	22.8%
22	1	0	1	1	1	36	141	25.5%	18.3%	32.7%
23	1	0	0	0	1	45	176	25.6%	19.1%	32.0%
24	1	0	1	0	1	38	142	26.8%	19.5%	34.0%
25	1	1	1	1	1	23	77	29.9%	19.6%	40.1%
26	1	1	0	0	1	24	70	34.3%	23.2%	45.4%
27	1	1	0	1	1	4	10	40.0%	9.6%	70.4%
28	1	0	0	1	1	9	22	40.9%	20.4%	61.5%
29	1	1	1	0	1	19	44	43.2%	28.5%	57.8%

Figure E2 Illustration of observed risk and 'high-low' 95% confidence interval bars for Risk ladder 1.2 in Table 4.3

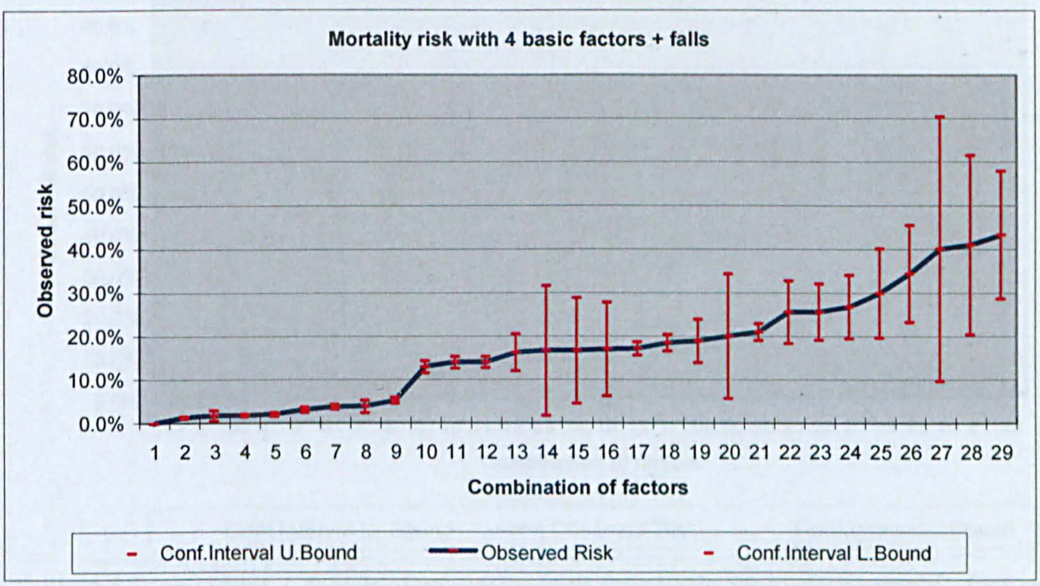


Table 4.4 Risk ladder-1.3; risk of mortality with four basic socio-demographic factors and Ischemic heart disease

Seq	Age	Gender	Tenure	Tax Band	Heart Disease	No. of Death	Population	Observed Risk	Conf.Interval	
1	0	0	0	1	1		3	0.0%	0.0%	0.0%
2	0	0	0	0	0	85	6050	1.4%	1.1%	1.7%
3	0	0	0	1	0	9	495	1.8%	0.6%	3.0%
4	0	1	0	0	0	128	7026	1.8%	1.5%	2.1%
5	0	0	1	0	0	87	4223	2.1%	1.6%	2.5%
6	0	0	1	1	0	90	2830	3.2%	2.5%	3.8%
7	0	1	1	0	0	144	3841	3.7%	3.1%	4.3%
8	0	1	0	1	0	28	721	3.9%	2.5%	5.3%
9	0	1	1	1	0	176	3331	5.3%	4.5%	6.0%
10	0	1	1	0	1	7	108	6.5%	1.8%	11.1%
11	0	1	1	1	1	9	92	9.8%	3.7%	15.9%
12	0	1	0	0	1	7	55	12.7%	3.9%	21.5%
13	1	0	1	0	0	305	2267	13.5%	12.0%	14.9%
14	1	1	0	0	0	350	2481	14.1%	12.7%	15.5%
15	1	0	0	0	0	437	3042	14.4%	13.1%	15.6%
16	1	0	1	1	0	407	2383	17.1%	15.6%	18.6%
17	1	0	0	1	0	52	304	17.1%	12.9%	21.3%
18	1	1	1	0	0	293	1576	18.6%	16.7%	20.5%
19	1	1	0	1	0	46	240	19.2%	14.2%	24.1%
20	1	1	1	1	0	373	1752	21.3%	19.4%	23.2%
21	1	1	1	1	1	18	82	22.0%	13.0%	30.9%
22	1	0	1	0	1	20	87	23.0%	14.1%	31.8%
23	1	1	0	0	1	24	90	26.7%	17.5%	35.8%
24	1	1	1	0	1	26	92	28.3%	19.1%	37.5%
25	1	0	0	0	1	21	71	29.6%	19.0%	40.2%
26	1	0	1	1	1	31	103	30.1%	21.2%	39.0%
27	1	1	0	1	1	3	8	37.5%	4.0%	71.0%
28	1	0	0	1	1	4	7	57.1%	20.5%	93.8%

Figure E3 Illustration of observed risk and 'high-low' 95% confidence interval bars for Risk Ladder 1.3 in Table 4.4

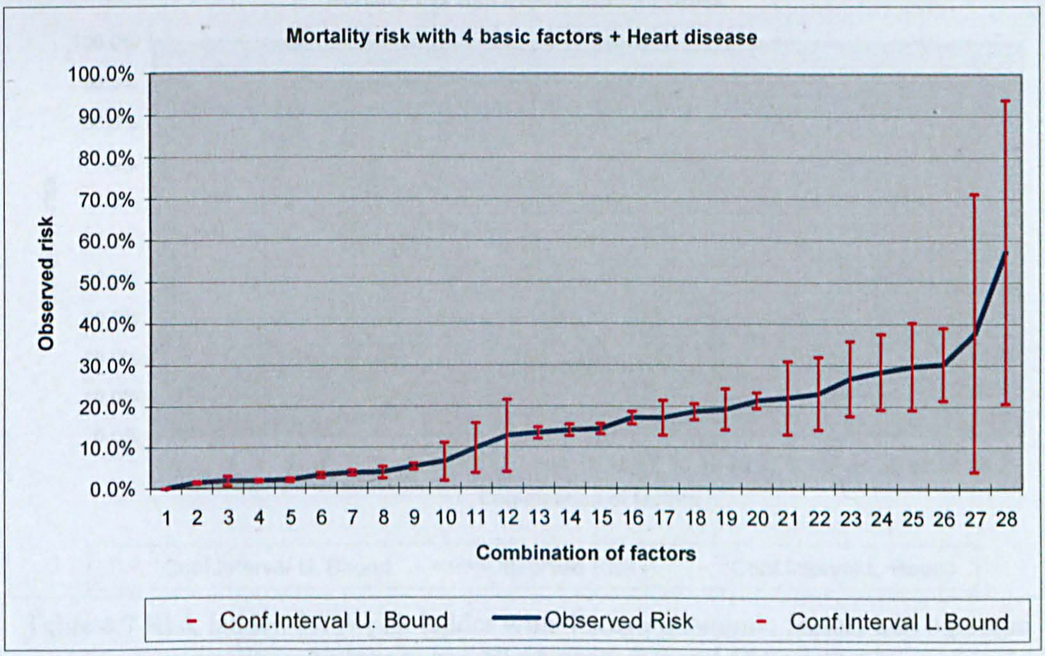


Table 4.5 Risk ladder-1.4; risk of mortality with four basic socio-demographic factors and Stroke

Seq	Age	Gender	Tenure	Tax Band	Strokes	No. of Death	Population	Observed Risk	Conf.Interval	
1	0	0	0	0	0	82	6058	1.4%	1.1%	1.6%
2	0	0	0	1	0	9	496	1.8%	0.6%	3.0%
3	0	1	0	0	0	130	7056	1.8%	1.5%	2.2%
4	0	0	1	0	0	85	4241	2.0%	1.6%	2.4%
5	0	0	1	1	0	88	2843	3.1%	2.5%	3.7%
6	0	1	1	0	0	133	3898	3.4%	2.8%	4.0%
7	0	1	0	1	0	27	726	3.7%	2.3%	5.1%
8	0	1	1	1	0	174	3376	5.2%	4.4%	5.9%
9	1	0	1	0	0	297	2296	12.9%	11.6%	14.3%
10	1	0	0	0	0	423	3028	14.0%	12.7%	15.2%
11	1	1	0	0	0	351	2510	14.0%	12.6%	15.3%
12	0	0	1	0	1	4	27	14.8%	1.4%	28.2%
13	1	0	1	1	0	403	2416	16.7%	15.2%	18.2%
14	1	0	0	1	0	52	306	17.0%	12.8%	21.2%
15	1	1	1	0	0	303	1604	18.9%	17.0%	20.8%
16	1	1	0	1	0	49	247	19.8%	14.9%	24.8%
17	0	1	0	0	1	5	25	20.0%	4.3%	35.7%
18	1	1	1	1	0	363	1767	20.5%	18.7%	22.4%
19	0	1	1	1	1	11	47	23.4%	11.3%	35.5%
20	0	0	1	1	1	5	21	23.8%	5.6%	42.0%
21	1	1	1	0	1	16	64	25.0%	14.4%	35.6%
22	0	0	0	0	1	5	16	31.3%	8.5%	54.0%
23	0	1	1	0	1	18	51	35.3%	22.2%	48.4%
24	1	1	0	0	1	23	61	37.7%	25.5%	49.9%
25	1	0	0	0	1	35	85	41.2%	30.7%	51.6%
26	1	1	1	1	1	28	67	41.8%	30.0%	53.6%
27	1	0	1	0	1	28	58	48.3%	35.4%	61.1%
28	1	0	1	1	1	35	70	50.0%	38.3%	61.7%
29	0	1	0	1	1	2	4	50.0%	1.0%	99.0%

Figure E4 Illustration of observed risk and 'high-low' 95% confidence interval bars for Risk Ladder 1.4 in Table 4.5

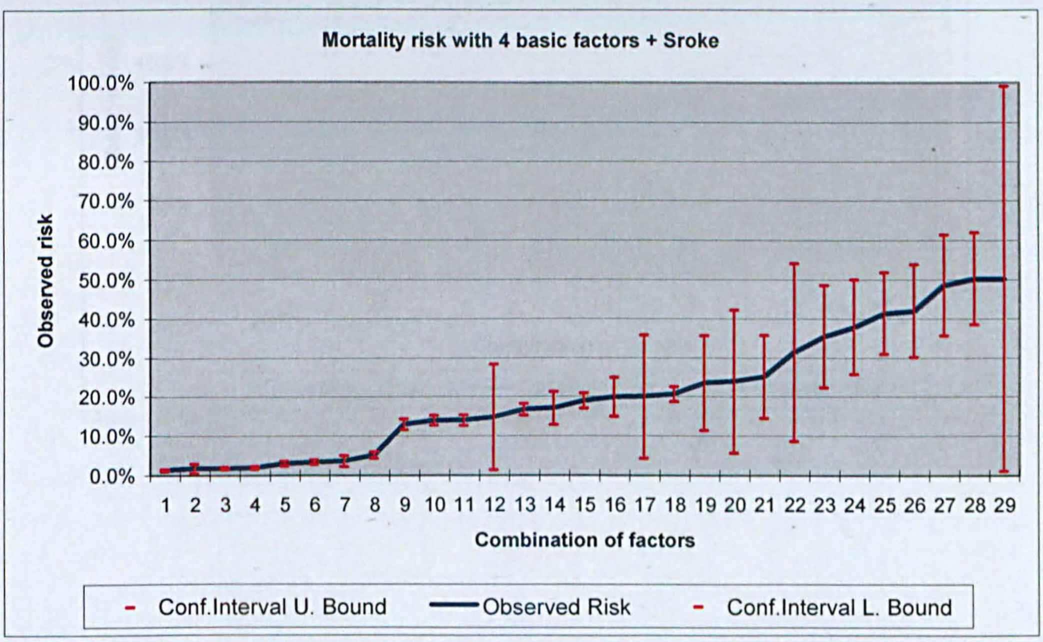


Table 4.7 Risk ladder-1.6 A risk ladder with 4 socio-economic factors and outcome mortality with binary age 50-65 years = 0 and 66+ years =1 (equivalent to risk ladder 1.1 in Table 4.2)

Seq	Age	Gender	Tenure	Tax Band	Combination 'AGTB'	Number of Death	Population	Observed Risk	Conf.Interval L. Bound	Conf.Interval U. Bound
1	0	0	0	0	0000	63	5223	1.2%	0.9%	1.5%
2	0	0	0	1	0001	6	428	1.4%	0.3%	2.5%
3	0	1	0	0	0100	93	6168	1.5%	1.2%	1.8%
4	0	0	1	0	0010	67	3626	1.8%	1.4%	2.3%
5	0	0	1	1	0011	60	2318	2.6%	1.9%	3.2%
6	0	1	1	0	0110	101	3305	3.1%	2.5%	3.6%
7	0	1	0	1	0101	20	634	3.2%	1.8%	4.5%
8	0	1	1	1	0111	137	2908	4.7%	3.9%	5.5%
9	1	0	1	0	1010	347	2996	11.6%	10.4%	12.7%
10	1	1	0	0	1100	416	3484	11.9%	10.9%	13.0%
11	1	0	0	0	1000	482	3964	12.2%	11.1%	13.2%
12	1	0	1	1	1011	59	381	15.5%	11.9%	19.1%
13	1	0	0	1	1001	471	3032	15.5%	14.2%	16.8%
14	1	1	1	0	1110	369	2312	16.0%	14.5%	17.5%
15	1	1	0	1	1101	58	344	16.9%	12.9%	20.8%
16	1	1	1	1	1111	439	2349	18.7%	17.1%	20.3%

Figure E5 A graph illustration of observed risk and 'high-low' 95% confidence interval bars for the Risk ladder-1.6 in Table 4.7

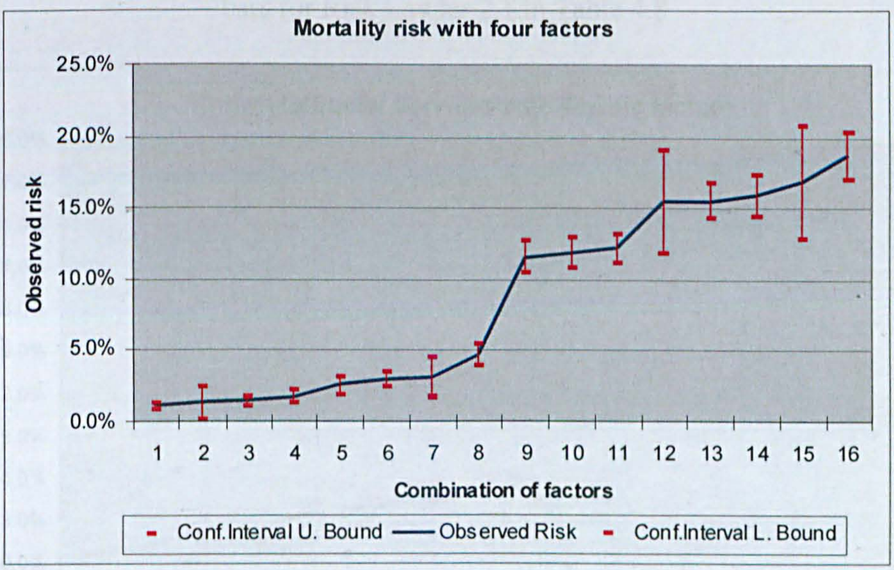


Table 4.8 Risk ladder-2.1 including four basic socio-demographic factors and 'social services' as the outcome variable

Seq	Age	Gender	Tenure	Tax Band	Known to 'SS'	Population	*P* of Known to 'SS'	Conf.Interval	
1	0	1	0	0	60	7081	0.8%	0.6%	1.1%
2	0	0	0	0	79	6074	1.3%	1.0%	1.6%
3	0	0	0	1	11	498	2.2%	0.9%	3.5%
4	0	1	0	1	19	730	2.6%	1.4%	3.8%
5	0	1	1	0	147	3949	3.7%	3.1%	4.3%
6	0	0	1	0	168	4268	3.9%	3.4%	4.5%
7	0	0	1	1	147	2864	5.1%	4.3%	5.9%
8	0	1	1	1	188	3423	5.5%	4.7%	6.3%
9	1	1	0	0	225	2571	8.8%	7.7%	9.8%
10	1	1	0	1	26	248	10.5%	6.7%	14.3%
11	1	0	0	0	440	3113	14.1%	12.9%	15.4%
12	1	1	1	0	249	1668	14.9%	13.2%	16.6%
13	1	1	1	1	351	1834	19.1%	17.3%	20.9%
14	1	0	0	1	65	311	20.9%	16.4%	25.4%
15	1	0	1	0	492	2354	20.9%	19.3%	22.5%
16	1	0	1	1	637	2486	25.6%	23.9%	27.3%
	14585	**21504**	**22846**	**12394**	**3304**	**43472**	7.6%	**7.4%**	**7.8%**

Figure E6 Illustration of observed probability and 'high-low' 95% confidence interval bars for Risk Ladder 2.1 in Table 4.8

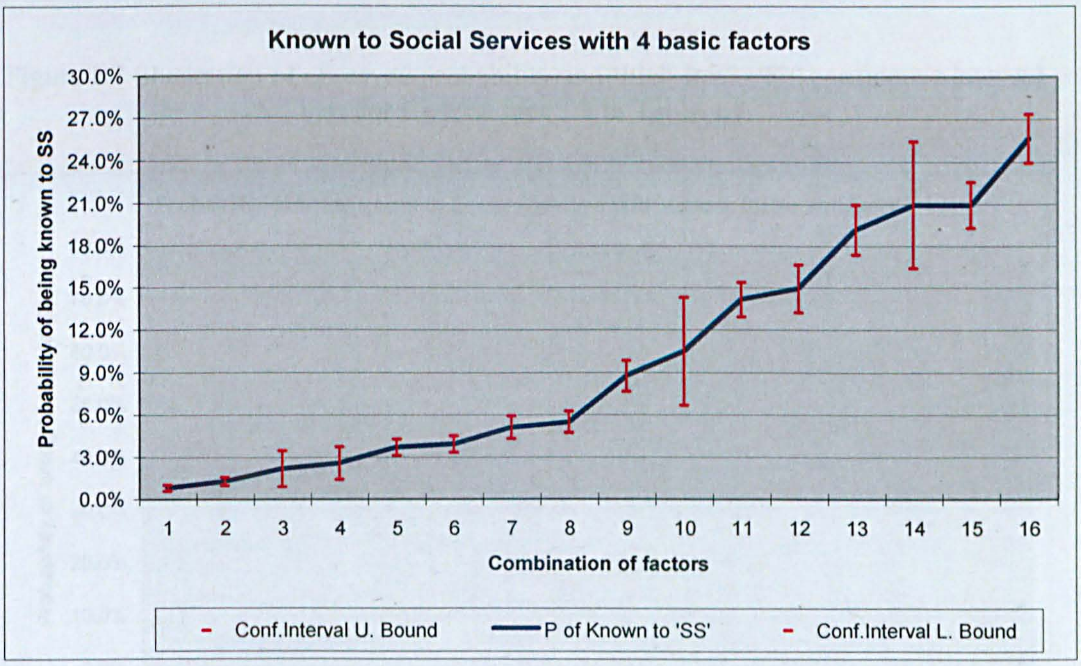


Table 4.9 Risk ladder-2.2 including four basic socio-demographic factors and the incidence of an admission for a fall with 'social services' as the outcome variable

Seq	Age	Gender	Tenure	Tax Band	Fall	Known to 'SS'	Population	*P* of Known to 'SS'	Conf.Interval	
1	0	1	0	0	0	56	7045	0.8%	0.6%	1.0%
2	0	0	0	0	0	78	6044	1.3%	1.0%	1.6%
3	0	0	0	1	0	10	496	2.0%	0.8%	3.3%
4	0	1	0	1	0	18	725	2.5%	1.4%	3.6%
5	0	1	1	0	0	143	3917	3.7%	3.1%	4.2%
6	0	0	1	0	0	163	4241	3.8%	3.3%	4.4%
7	0	0	1	1	0	142	2840	5.0%	4.2%	5.8%
8	0	1	1	1	0	178	3376	5.3%	4.5%	6.0%
9	1	1	0	0	0	196	2501	7.8%	6.8%	8.9%
10	1	1	0	1	0	24	238	10.1%	6.3%	13.9%
11	0	1	0	0	1	4	36	11.1%	0.8%	21.4%
12	0	1	1	0	1	4	32	12.5%	1.0%	24.0%
13	1	0	0	0	0	372	2937	12.7%	11.5%	13.9%
14	1	1	1	0	0	234	1624	14.4%	12.7%	16.1%
15	1	1	1	1	0	311	1757	17.7%	15.9%	19.5%
16	1	0	0	1	0	53	289	18.3%	13.9%	22.8%
17	0	0	1	0	1	5	27	18.5%	3.9%	33.2%
18	1	0	1	0	0	426	2212	19.3%	17.6%	20.9%
19	0	0	1	1	1	5	24	20.8%	4.6%	37.1%
20	0	1	1	1	1	10	47	21.3%	9.6%	33.0%
21	1	0	1	1	0	563	2345	24.0%	22.3%	25.7%
22	1	1	1	0	1	15	44	34.1%	20.1%	48.1%
23	1	0	0	0	1	68	176	38.6%	31.4%	45.8%
24	1	1	0	0	1	29	70	41.4%	29.9%	53.0%
25	1	0	1	0	1	66	142	46.5%	38.3%	54.7%
26	1	1	1	1	1	40	77	51.9%	40.8%	63.1%
27	1	0	1	1	1	74	141	52.5%	44.2%	60.7%
28	1	0	0	1	1	12	22	54.5%	33.7%	75.4%

Figure E7 Illustration of observed probability and 'high-low' 95% confidence interval bars for Risk Ladder 2.2 in Table 4.9

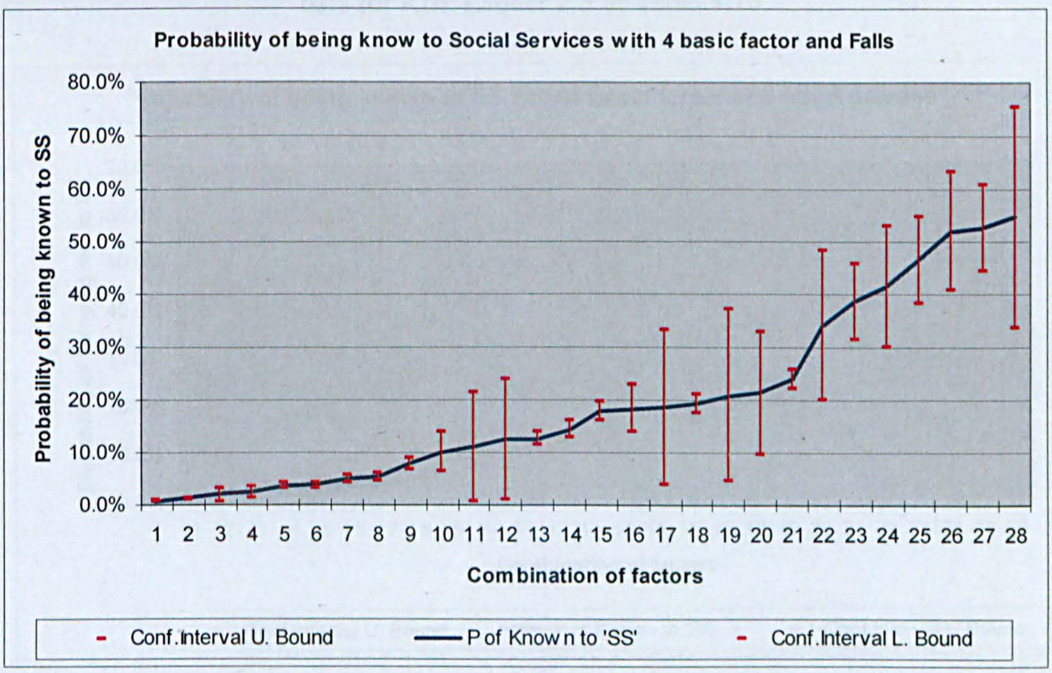


Table 4.10 Risk ladder-2.3 including four basic socio-demographic factors and the incidence of an admission for heart disease with 'social services' as the outcome variable

Seq	Age	Gender	Tenure	Tax Band	Heart Disease	Known to 'SS'	Population	*P* of Known to 'SS'	Conf.Interval	
1	0	0	0	1	1		3	0.0%	0.0%	0.0%
2	0	1	0	0	0	57	7026	0.8%	0.6%	1.0%
3	0	0	0	0	0	76	6050	1.3%	1.0%	1.5%
4	0	1	0	1	0	16	721	2.2%	1.1%	3.3%
5	0	0	0	1	0	11	495	2.2%	0.9%	3.5%
6	0	1	1	0	0	135	3841	3.5%	2.9%	4.1%
7	0	0	1	0	0	165	4223	3.9%	3.3%	4.5%
8	0	0	1	1	0	144	2830	5.1%	4.3%	5.9%
9	0	1	1	1	0	179	3331	5.4%	4.6%	6.1%
10	1	1	0	0	0	212	2481	8.5%	7.4%	9.6%
11	0	1	1	1	1	9	92	9.8%	3.7%	15.9%
12	1	1	0	1	0	24	240	10.0%	6.2%	13.8%
13	0	1	1	0	1	12	108	11.1%	5.2%	17.0%
14	1	0	0	0	0	422	3042	13.9%	12.6%	15.1%
15	1	1	0	0	1	13	90	14.4%	7.2%	21.7%
16	1	1	1	0	0	232	1576	14.7%	13.0%	16.5%
17	1	1	1	0	1	17	92	18.5%	10.5%	26.4%
18	1	1	1	1	0	332	1752	18.9%	17.1%	20.8%
19	1	0	1	0	0	464	2267	20.5%	18.8%	22.1%
20	1	0	0	1	0	63	304	20.7%	16.2%	25.3%
21	1	1	1	1	1	19	82	23.2%	14.0%	32.3%
22	1	0	1	1	0	600	2383	25.2%	23.4%	26.9%
23	1	0	0	0	1	18	71	25.4%	15.2%	35.5%
24	1	0	1	0	1	28	87	32.2%	22.4%	42.0%
25	0	1	0	1	1	3	9	33.3%	2.5%	64.1%
26	1	0	1	1	1	37	103	35.9%	26.7%	45.2%

Figure E8 Illustration of observed probability and 'high-low' 95% confidence interval bars for Risk Ladder 2.3 in Table 4.10

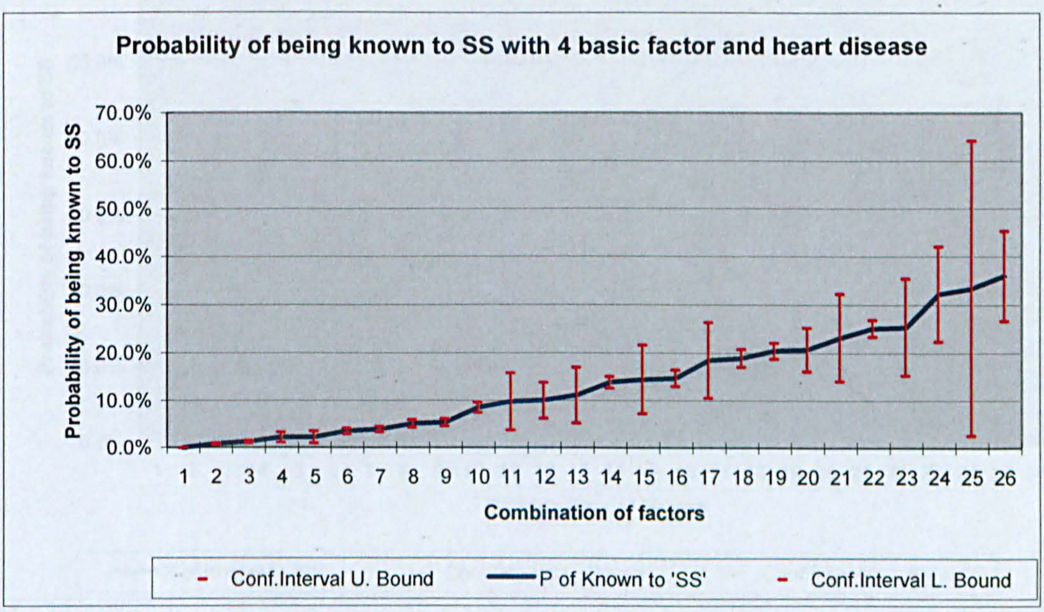
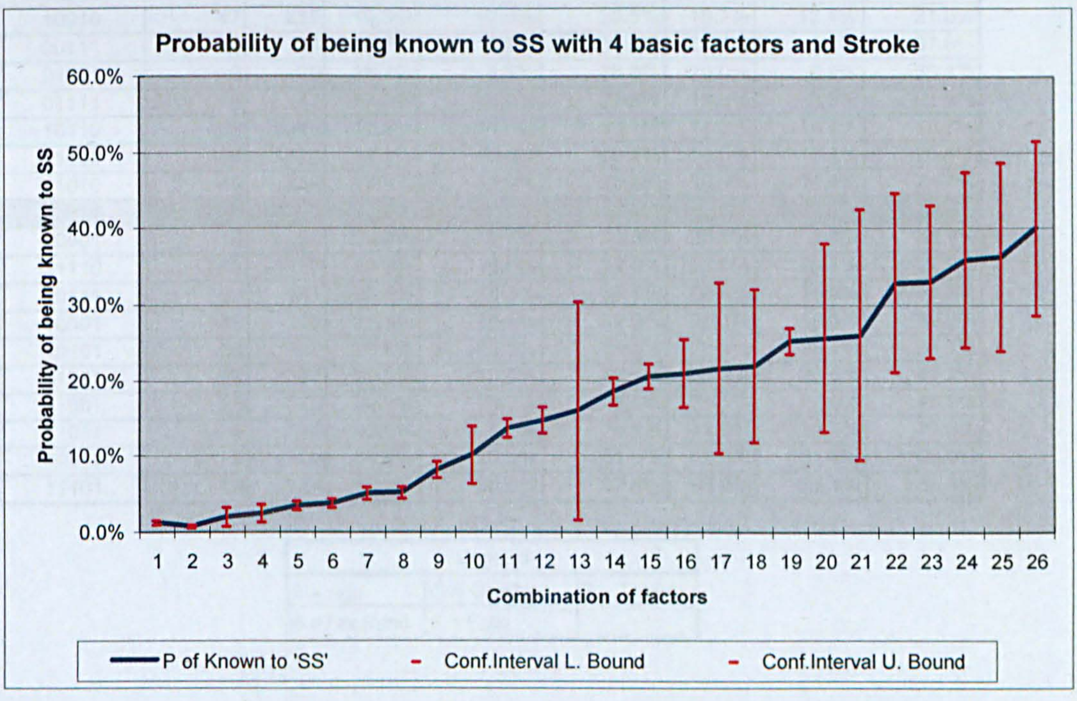


Table 4.11 Risk ladder 2.4 including four basic socio-demographic factors and the incidence of an admission for stroke with 'social services' as the outcome variable

Seq	Age	Gender	Tenure	Tax Band	Stroke	Known to 'SS'	Population	*P* of Known to 'SS'	Conf.Interval	
1	0	0	0	0	0	78	6058	1.3%	1.0%	1.6%
2	0	1	0	0	0	56	7056	0.8%	0.6%	1.0%
3	0	0	0	1	0	10	496	2.0%	0.8%	3.3%
4	0	1	0	1	0	18	726	2.5%	1.3%	3.6%
5	0	1	1	0	0	136	3898	3.5%	2.9%	4.1%
6	0	0	1	0	0	161	4241	3.8%	3.2%	4.4%
7	0	0	1	1	0	145	2843	5.1%	4.3%	5.9%
8	0	1	1	1	0	176	3376	5.2%	4.5%	6.0%
9	1	1	0	0	0	205	2510	8.2%	7.1%	9.2%
10	1	1	0	1	0	25	247	10.1%	6.4%	13.9%
11	1	0	0	0	0	412	3028	13.6%	12.4%	14.8%
12	1	1	1	0	0	235	1604	14.7%	12.9%	16.4%
13	0	1	0	0	1	4	25	16.0%	1.6%	30.4%
14	1	1	1	1	0	327	1767	18.5%	16.7%	20.3%
15	1	0	1	0	0	471	2296	20.5%	18.9%	22.2%
16	1	0	0	1	0	64	306	20.9%	16.4%	25.5%
17	0	1	1	0	1	11	51	21.6%	10.3%	32.9%
18	1	1	1	0	1	14	64	21.9%	11.7%	32.0%
19	1	0	1	1	0	609	2416	25.2%	23.5%	26.9%
20	0	1	1	1	1	12	47	25.5%	13.1%	38.0%
21	0	0	1	0	1	7	27	25.9%	9.4%	42.5%
22	1	1	0	0	1	20	61	32.8%	21.0%	44.6%
23	1	0	0	0	1	28	85	32.9%	22.9%	42.9%
24	1	1	1	1	1	24	67	35.8%	24.3%	47.3%
25	1	0	1	0	1	21	58	36.2%	23.8%	48.6%
26	1	0	1	1	1	28	70	40.0%	28.5%	51.5%

Figure E9 Illustration of observed probability and 'high-low' 95% confidence interval bars for Risk Ladder 2.4 in Table 4.11



Appendix-F Reproduction of risk ladders1.2-1.4 with both Wald (Standards) and Wilson confidence intervals

In the following tables where the confidence intervals for a combination with the Wald method is not significant, the relevant row is partially coloured and where both the Wald and Wilson methods are not significant, the entire row is coloured.

Table F.1 Reproduction of the Risk ladder-1.2 (table 4.3), with Wald and Wilson confidence intervals

Combination 'AGTBF'	Number of Death	Population	Observed Risk	Standard Conf.Interval L.Bound	Standard Conf.Interval U.Bound	$\widetilde{p}$	Wilson Conf.Interval L.Bound	Wilson Conf.Interval U.Bound
00011	0	2	0.0%	0.0%	0.0%	33.3%	-32.0%	98.7%
00000	81	6044	1.3%	1.1%	1.6%	1.4%	1.1%	1.7%
00010	9	496	1.8%	0.6%	3.0%	2.2%	0.9%	3.5%
01000	129	7045	1.8%	1.5%	2.1%	1.9%	1.5%	2.2%
00100	87	4241	2.1%	1.6%	2.5%	2.1%	1.7%	2.5%
00110	89	2840	3.1%	2.5%	3.8%	3.2%	2.6%	3.8%
01100	148	3917	3.8%	3.2%	4.4%	3.8%	3.2%	4.4%
01010	28	725	3.9%	2.5%	5.3%	4.1%	2.7%	5.6%
01110	177	3376	5.2%	4.5%	6.0%	5.3%	4.5%	6.1%
00101	2	27	7.4%	-2.5%	17.3%	12.9%	0.3%	25.5%
01101	3	32	9.4%	-0.7%	19.5%	13.9%	1.9%	25.9%
10100	287	2212	13.0%	11.6%	14.4%	13.0%	11.6%	14.4%
11000	350	2501	14.0%	12.6%	15.4%	14.1%	12.7%	15.4%
10000	413	2937	14.1%	12.8%	15.3%	14.1%	12.9%	15.4%
10010	47	289	16.3%	12.0%	20.5%	16.7%	12.4%	21.0%
00111	4	24	16.7%	1.8%	31.6%	21.4%	5.0%	37.8%
01001	6	36	16.7%	4.5%	28.8%	20.0%	6.9%	33.1%
01111	8	47	17.0%	6.3%	27.8%	19.6%	8.3%	31.0%
10110	402	2345	17.1%	15.6%	18.7%	17.2%	15.7%	18.7%
11100	300	1624	18.5%	16.6%	20.4%	18.6%	16.7%	20.4%
11010	45	238	18.9%	13.9%	23.9%	19.4%	14.4%	24.4%
01011	1	5	20.0%	-15.1%	55.1%	33.3%	-8.0%	74.7%
00001	6	30	20.0%	5.7%	34.3%	23.5%	8.4%	38.7%
11110	368	1757	20.9%	19.0%	22.8%	21.0%	19.1%	22.9%
10111	36	141	25.5%	18.3%	32.7%	26.2%	18.9%	33.5%
10001	45	176	25.6%	19.1%	32.0%	26.1%	19.6%	32.6%
10101	38	142	26.8%	19.5%	34.0%	27.4%	20.1%	34.7%
11111	23	77	29.9%	19.6%	40.1%	30.9%	20.5%	41.2%
11001	24	70	34.3%	23.2%	45.4%	35.1%	24.0%	46.3%
11011	4	10	40.0%	9.6%	70.4%	42.9%	12.2%	73.5%
10011	9	22	40.9%	20.4%	61.5%	42.3%	21.7%	63.0%
11101	19	44	43.2%	28.5%	57.8%	43.8%	29.1%	58.4%

Legend		
A = Age	G = Gender	T = Tenure
B =Tax Band	F = Falls	

Table F.2 Reproduction of the Risk ladder-1.3 (table 4.4), with Wald and Wilson confidence intervals

Combination 'AGTBI'	Number of Death	Population	Observed Risk	Standard Conf.Interval L.Bound	Standard Conf.Interval U. Bound	$\widetilde{p}$	Wilson Conf.Interval L.Bound	Wilson Conf.Interval U. Bound
00011	0	3	0.0%	0.0%	0.0%	28.6%	-22.5%	79.7%
00000	85	6050	1.4%	1.1%	1.7%	1.4%	1.1%	1.7%
00010	9	495	1.8%	0.6%	3.0%	2.2%	0.9%	3.5%
01000	128	7026	1.8%	1.5%	2.1%	1.8%	1.5%	2.2%
00100	87	4223	2.1%	1.6%	2.5%	2.1%	1.7%	2.5%
00110	90	2830	3.2%	2.5%	3.8%	3.2%	2.6%	3.9%
01100	144	3841	3.7%	3.1%	4.3%	3.8%	3.2%	4.4%
01010	28	721	3.9%	2.5%	5.3%	4.1%	2.7%	5.6%
00101	2	45	4.4%	-1.6%	10.5%	8.2%	0.2%	16.2%
01110	176	3331	5.3%	4.5%	6.0%	5.3%	4.6%	6.1%
01101	7	108	6.5%	1.8%	11.1%	8.0%	2.9%	13.2%
00001	2	24	8.3%	-2.7%	19.4%	14.3%	0.3%	28.3%
00111	3	34	8.8%	-0.7%	18.4%	13.2%	1.8%	24.5%
01111	9	92	9.8%	3.7%	15.9%	11.5%	4.9%	18.0%
01011	1	9	11.1%	-9.4%	31.6%	23.1%	-4.4%	50.6%
01001	7	55	12.7%	3.9%	21.5%	15.3%	5.8%	24.8%
10100	305	2267	13.5%	12.0%	14.9%	13.5%	12.1%	14.9%
11000	350	2481	14.1%	12.7%	15.5%	14.2%	12.8%	15.5%
10000	437	3042	14.4%	13.1%	15.6%	14.4%	13.2%	15.7%
10110	407	2383	17.1%	15.6%	18.6%	17.1%	15.6%	18.6%
10010	52	304	17.1%	12.9%	21.3%	17.5%	13.3%	21.8%
11100	293	1576	18.6%	16.7%	20.5%	18.7%	16.7%	20.6%
11010	46	240	19.2%	14.2%	24.1%	19.7%	14.6%	24.7%
11110	373	1752	21.3%	19.4%	23.2%	21.4%	19.4%	23.3%
11111	18	82	22.0%	13.0%	30.9%	23.3%	14.1%	32.4%
10101	20	87	23.0%	14.1%	31.8%	24.2%	15.2%	33.2%
11001	24	90	26.7%	17.5%	35.8%	27.7%	18.4%	36.9%
11101	26	92	28.3%	19.1%	37.5%	29.2%	19.9%	38.5%
10001	21	71	29.6%	19.0%	40.2%	30.7%	19.9%	41.4%
10111	31	103	30.1%	21.2%	39.0%	30.8%	21.9%	39.8%
11011	3	8	37.5%	4.0%	71.0%	41.7%	7.5%	75.8%
10011	4	7	57.1%	20.5%	93.8%	54.5%	17.7%	91.4%

Legend		
A = Age	G = Gender	T = Tenure
B =Tax Band	I = Ischemic Heart Disease	

Table F.3 Reproduction of the Risk ladder-1.4 (table 4.5), with Wald and Wilson confidence intervals

Combination 'AGTBS'	Number of Death	Population	Observed Risk	Standard Conf.Interval L. Bound	Standard Conf.Interval U. Bound	$\widetilde{p}$	Wilson Conf.Interval L. Bound	Wilson Conf.Interval U. Bound
00011	0	2	0.0%	0.0%	0.0%	33.33%	-32.0%	98.7%
11011	0	1	0.0%	0.0%	0.0%	40.00%	-56.0%	136.0%
00000	82	6058	1.4%	1.1%	1.6%	1.39%	1.1%	1.7%
00010	9	496	1.8%	0.6%	3.0%	2.20%	0.9%	3.5%
01000	130	7056	1.8%	1.5%	2.2%	1.87%	1.6%	2.2%
00100	85	4241	2.0%	1.6%	2.4%	2.05%	1.6%	2.5%
00110	88	2843	3.1%	2.5%	3.7%	3.16%	2.5%	3.8%
01100	133	3898	3.4%	2.8%	4.0%	3.46%	2.9%	4.0%
01010	27	726	3.7%	2.3%	5.1%	3.97%	2.6%	5.4%
01110	174	3376	5.2%	4.4%	5.9%	5.21%	4.5%	6.0%
10100	297	2296	12.9%	11.6%	14.3%	13.00%	11.6%	14.4%
10000	423	3028	14.0%	12.7%	15.2%	14.02%	12.8%	15.3%
11000	351	2510	14.0%	12.6%	15.3%	14.04%	12.7%	15.4%
00101	4	27	14.8%	1.4%	28.2%	19.35%	4.5%	34.3%
10110	403	2416	16.7%	15.2%	18.2%	16.74%	15.2%	18.2%
10010	52	306	17.0%	12.8%	21.2%	17.42%	13.2%	21.7%
11100	303	1604	18.9%	17.0%	20.8%	18.97%	17.0%	20.9%
11010	49	247	19.8%	14.9%	24.8%	20.32%	15.3%	25.3%
01001	5	25	20.0%	4.3%	35.7%	24.14%	7.4%	40.9%
11110	363	1767	20.5%	18.7%	22.4%	20.61%	18.7%	22.5%
01111	11	47	23.4%	11.3%	35.5%	25.49%	13.0%	37.9%
00111	5	21	23.8%	5.6%	42.0%	28.00%	8.8%	47.2%
11101	16	64	25.0%	14.4%	35.6%	26.47%	15.7%	37.3%
00001	5	16	31.3%	8.5%	54.0%	35.00%	11.6%	58.4%
01101	18	51	35.3%	22.2%	48.4%	36.36%	23.2%	49.6%
11001	23	61	37.7%	25.5%	49.9%	38.46%	26.3%	50.7%
10001	35	85	41.2%	30.7%	51.6%	41.57%	31.1%	52.1%
11111	28	67	41.8%	30.0%	53.6%	42.25%	30.4%	54.1%
10101	28	58	48.3%	35.4%	61.1%	48.39%	35.5%	61.2%
01011	2	4	50.0%	1.0%	99.0%	50.00%	1.0%	99.0%
10111	35	70	50.0%	38.3%	61.7%	50.00%	38.3%	61.7%
10011	4	5	80.0%	44.9%	115.1%	66.67%	25.3%	108.0%

Legend		
A = Age	G = Gender	T = Tenure
B =Tax Band	S = Strokes	

Appendix-G detailed explanation of the logistic regression models-3, -4, -5 and -6

All models in this appendix condition on model-2 in Section 5.2 and contain detailed explanation of model-3 to -6 discussed in Section 5.2. In this appendix, Subsection G.1 focuses on 'age' and G.2 considers 'tax band'.

G.1 increasing the number of levels for 'age' variable

G.1.1 increasing the number of levels for 'age' to 3 categories (Model-3)

The effect of all variables on mortality in this model like the previous one (model-2) is highly significant. By comparing the new model with the previous one, we can see a noticeable change in the value of OR for all variables especially for falls, stroke and gender. While for gender and tenure there is an increase in the value of OR, for other variables the OR values have decreased. Increasing the number of levels for age means that the OR of all variables will be more adjusted. So, for some variables, the gap between the reference category and other categories increases (e.g. gender and tenure, indication of higher discrimination on risk of mortality between different groups for the same variable), and for other variables the gap decreases. However there is evidence of a strong gradient for age. Table G.1 includes the output for the model-3.

Table G.1 Model-3: Odds ratios based on logistic regression modelling by increasing the number of levels for predictor age to three categories

Death	Odds Ratio	Sig.	[95% Conf. Interval]	
Gender	1.44	0.000	1.33	1.55
Age-2	4.04	0.000	3.62	4.50
Age-3	13.75	0.000	12.32	15.35
Tenure	1.24	0.000	1.14	1.35
Tax band	1.32	0.000	1.21	1.44
Falls	1.88	0.000	1.59	2.23
Heart Disease	1.80	0.000	1.50	2.16
Strokes	4.12	0.000	3.42	4.96

Number of obs = 43472 χ^2 (8 d.f) = 3528.5, P = 0.0000
Pseudo = 0.155 Log likelihood = -9633.2

G.1.2 increasing the number of levels for 'age' to 4 categories (Model-4)

Like previous models, there is an increase in the OR of gender and tenure and a reduction in the OR of other variables as well as a clear gradient in OR for age. The output for the model-4 is shown in Table G.2

Table G.2 Model-4: Odds ratios based on logistic regression modelling by increasing the number of levels for predictor age to 4 categories.

Death	Odds Ratio	Sig.	[95% Conf. Interval]	
Gender	1.48	0.000	1.37	1.60
Age-2	3.14	0.000	2.78	3.53
Age-3	7.83	0.000	7.00	8.76
Age-4	19.73	0.000	17.45	22.30
Tenure	1.26	0.000	1.15	1.38
Tax band	1.31	0.000	1.20	1.44
Falls	1.79	0.000	1.51	2.13
Heart Disease	1.74	0.000	1.44	2.09
Strokes	3.90	0.000	3.23	4.70

Number of obs = 43472 χ^2 (9 d.f) = 3779.3, P = 0.0000

Pseudo R^2 = 0.166 Log likelihood = -9507.8

G.1.3 increasing the number of levels for 'age' to 5 categories (Model-5)

Unlike the previous models with dichotomous age variable, in this model, there is an increase in the OR for tax band, heart disease and strokes. Full information for model-5 is shown in Table G.3.

Table G.3 Model-5: Odds ratios based on logistic regression modelling by increasing the number of levels for predictor age to 5 categories.

Death	Odds Ratio	Sig.	[95% Conf. Interval]	
Gender	1.48	0.000	1.37	1.60
Age-2	2.46	0.000	2.12	2.85
Age-3	6.53	0.000	5.69	7.48
Age-4	14.61	0.000	12.73	16.77
Age-5	33.47	0.000	28.22	39.70
Tenure	1.27	0.000	1.17	1.39
Tax band	1.31	0.000	1.20	1.43
Falls	1.75	0.000	1.48	2.08
Heart Disease	1.74	0.000	1.45	2.09
Strokes	4.05	0.000	3.35	4.88

Number of obs = 43472 χ^2 (10 d.f) = 3821.8, P = 0.0000

Pseudo R^2 = 0.168 Log likelihood = -9486.6

G.2 Changing Tax band from binary to three categories (Model-6)

By increasing the number of levels for the tax band from two to three categories, the risk of someone living in tax band A-C increases from 1.3 times in the previous models to 1.6 times in model-6, comparing with tax bands F-H. The risk of someone living in Tax bands D-E (which in the previous models was not included in the model as it was combined with F-H) comparing with tax bands F-H, is 1.3 times higher.

The most evident changes in OR are in tenure, reducing from 1.27 to 1.12 and the level of significance from less than 0.001 decrease to 0.02. It can also be interpreted that tenure appears to be the weakest predictor of death outcome between all seven variables as shown in Table G.4. This is interesting in the sense that the effect of tenure tends to be ameliorated once the tax banding becomes more sensitive.

Table G.4 Model-6: Odds ratios based on logistic regression modelling by increasing the number of levels for predictor tax band to three categories.

Death	Odds Ratio	Sig.	[95% Conf. Interval]	
Gender	1.49	0.000	1.38	1.61
Age-2	2.47	0.000	2.13	2.86
Age-3	6.55	0.000	5.71	7.50
Age-4	14.63	0.000	12.75	16.79
Age-5	33.49	0.000	28.23	39.73
Tenure	1.12	0.022	1.02	1.24
Tax band_2 (D-E)	1.33	0.000	1.19	1.49
Tax band_3 (A-C)	1.62	0.000	1.43	1.84
Falls	1.76	0.000	1.48	2.09
Heart Disease	1.75	0.000	1.45	2.10
Strokes	4.06	0.000	3.37	4.90

Number of obs = 43472 χ^2 (11 d.f) = 3846.5, P = 0.0000

Pseudo R^2 = 0.169 Log likelihood = -9474.3

Appendix-H Models with Interaction effects

H.1 Models with only the pair of variables in the interaction

In the first part of this appendix the interaction between 'housing tenure' and three causes of hospital admissions (FIS) will be considered and in the second part I will look at the interaction between 'Council tax banding' and FIS.

H.1.1 Interaction between Housing tenure and FIS

In this part the effect of housing tenure on each causes of hospital admission are illustrated in a separate table (model) extracted from Stata.

G.1.1.1 Interaction between Tenure and Falls:

The result from interaction between both housing association properties and council housing with the hospital admission as a result of fall, which is illustrated in Table-H.1, is highly significant.

Table-H.1 Interaction between Tenure and Falls

Death	Odds Ratio	Sig.	[95% Conf.	Interval]
Housing Association	1.99	0.000	1.74	2.26
Council Housing	1.52	0.000	1.40	1.64
Falls	6.46	0.000	5.06	8.23
Housing-Ass X Falls	.35	0.000	.20	.61
Council-Hou X Falls	.62	0.005	.45	.86
Number of obs = 43472		χ^2 (5 d.f) = 448.83, P = 0.0000		
Pseudo R^2 = 0.02		Log likelihood = -11173.08		

H.1.1.2 Interaction between Tenure and Heart Disease

Table-H.2 shows the interaction between tenure and heart disease. In this model the effect of interaction between housing association and heart disease is significant but between council housing and heart disease is highly significant.

Table-H.2 Interaction between Tenure and Heart Disease

Death	Odds Ratio	Sig.	[95% Conf.	Interval]
Housing Association	1.94	0.000	1.70	2.21
Council Housing	1.50	0.000	1.39	1.63
Heart Disease	5.12	0.000	3.83	6.85
Housing-Ass X HD	.47	0.014	.25	.85
Council-Hou X HD	.47	0.000	.32	.68
Number of obs = 43472 Pseudo R^2 = 0.013		χ^2 (5 d.f)=300.10, P = 0.0000 Log likelihood = -11247.45		

H.1.1.3 Interaction between Tenure and Stroke

In this model the interaction between housing association and stroke is not significant. The level of significance for interaction between council housing and stroke is also weaker than the earlier two models; for falls and heart disease. The result of interaction between tenure and stroke is illustrated in Table-H.3.

Table-H.3 Interaction between Tenure and Stroke

Death	Odds Ratio	Sig.	[95% Conf.	Interval]
Housing Association	1.86	0.000	1.63	2.12
Council Housing	1.49	0.000	1.37	1.61
Stroke	10.18	0.000	7.59	13.65
Housing-Ass X Stroke	.67	0.146	.38	1.15
Council-Hou X Stroke	.59	0.006	.41	.86
Number of obs = 43472 Pseudo R^2 = 0.025		χ^2(5 d.f)= 560.64, P = 0.0000 Log likelihood = -11117.18		

In the above three models, the interaction between both housing association and council housing with FIS, except for one case (housing association and stroke) are significant.

H.1.2 Interaction between Council tax banding and FIS

In this part the interaction between council tax banding and each of the three causes of hospital admissions will be examined.

H.1.2.1 Interaction between Tax banding and Falls

The result for this test shows that the interaction between tax banding D-E and falls is not significant but between tax banding A-C and falls is highly significant. The result of the interaction between housing tenure and falls are illustrated in Table-H.4.

Table-H.4 Interaction between Tax banding and Falls

Death	Odds Ratio	Sig.	[95% Conf.	Interval]
Tax Bands D-E	1.45	0.000	1.31	1.60
Tax Bands A-C	2.10	0.000	1.91	2.31
Falls	6.34	0.000	4.72	8.52
Tax Bands D-E X Falls	.78	0.219	.53	1.16
Tax Bands A-C X Falls	.52	0.001	.35	.76
Number of obs = 43472 Pseudo R^2 = 0.023		χ^2(5 d.f) = 527.46, P = 0.0000 Log likelihood = -11133.77		

H.1.2.2 Interaction between Tax band and Heart Disease

In this model the level of significance for interaction between tax banding D-E and heart disease just over 0.05 which means is not significant but for tax banding A-C and heart disease is highly significant, as it is shown in Table-H.5.

Table-H.5 Interaction between Tax banding and Heart disease

Death	Odds Ratio	Sig.	[95% Conf.	Interval]
Tax Bands D-E	1.44	0.000	1.30	1.58
Tax Bands A-C	2.08	0.000	1.89	2.29
Heart Disease	4.71	0.000	3.34	6.62
Tax Bands D-E X HD	.65	0.053	.42	1.00
Tax Bands A-C X HD	.50	0.002	.32	.78
Number of obs = 43472 Pseudo R^2 = 0.017		χ^2(5 d.f)=381.62, P = 0.0000 Log likelihood = -11206.688		

H.1.2.3 Interaction between Tax band and Stroke

Table-H.6 bellow shows the interaction between tax banding and stroke. In this model for strokes again the interaction effect is not significant for bands D-E but for bands A-C is significant.

Table-H.6 Interaction between Tax banding and Stroke

Death	Odds Ratio	Sig.	[95% Conf.	Interval]
Tax Bands D-E	1.44	0.000	1.30	1.58
Tax Bands A-C	2.07	0.000	1.89	2.28
Stroke	10.05	0.000	7.20	14.04
Tax Bands D-E X Stroke	.73	0.161	.47	1.13
Tax Bands A-C X Stroke	.61	0.024	.39	.93
Number of obs = 43472		χ^2(5 d.f) = 652.38, P = 0.0000		
Pseudo R^2 = 0.029		Log likelihood = -11071.309		

The output of the above three models confirms that the interaction between tax banding A-C and FIS are significant but it is not significant for tax band D-E.

H.2 The full model with all variables (age as a continuous variable)

In this section the interaction between tenure and tax banding with three causes of hospital admissions (all six models in Section H.1, above) once again have been examined in a full model (including all variables).

H.2.1 Interaction between Housing tenure and FIS in full model

In this part interaction between tenure and FIS will be assessed.

H.2.1.1 Full model including the interaction between Tenure and Falls

Table-H.7 above illustrates the full model including the interaction between tenure and falls. As the model suggests, by including other factors (such as sex, age etc), neither the variable council housing itself nor the interaction between council housing and falls are significant.

Table-H.7 Full model including the interaction between Tenure and Falls

Death	Odds Ratio	Sig.	[95% Conf.	Interval]
Sex	1.53	0.000	1.41	1.66
Age	1.10	0.000	1.09	1.10
Housing Association	1.50	0.000	1.29	1.74
Council Housing	1.09	0.117	.98	1.21
Tax Bands D-E	1.34	0.000	1.20	1.51
Tax Bands A-C	1.64	0.000	1.45	1.86
Falls	1.96	0.000	1.50	2.57
Heart Disease	1.72	0.000	1.43	2.07
Stroke	3.83	0.000	3.17	4.63
Housing-Ass X Falls	.43	0.007	.23	.79
Council Hou X Falls	.86	0.425	.60	1.24
Number of obs = 43472 Pseudo R^2 = 0.177		χ^2(11 d.f) = 4032.94, P = 0.0000 Log likelihood = -9381.029		

H.2.1.2 Full model with interaction between Tenure and Heart disease:

Unlike the previous model (in Table-H.7) where the interaction between council housing and falls was not significant, this model, as illustrated in Table-H.8, shows that the interaction between council housing and heart disease is significant and also that the interaction between housing association and heart disease is not significant.

Table-H.8 Full model including the interaction between Tenure and Heart disease.

Death	Odds Ratio	Sig.	[95% Conf.	Interval]
Sex	1.53	0.000	1.41	1.65
Age	1.10	0.000	1.09	1.10
Housing Association	1.44	0.000	1.24	1.67
Council Housing	1.10	0.080	.99	1.22
Tax Bands D-E	1.35	0.000	1.21	1.51
Tax Bands A-C	1.65	0.000	1.46	1.87
Falls	1.67	0.000	1.40	1.98
Heart Disease	2.35	0.000	1.70	3.24
Stroke	3.85	0.000	3.19	4.65
Housing-Ass X HD	.74	0.388	.38	1.46
Council-Hou X HD	.63	0.023	.42	.94
Number of obs =43472 Pseudo R^2 = 0.177		χ^2(11 d.f) = 4030.30, P = 0.0000 Log likelihood = -9382.349		

H.2.1.3 Full model with interaction between Tenure and Stroke:

The interaction between strokes and either of the social housing tenure is significant as the following model in Table-H.9 implies.

Table-H.9 Full model including the interaction between Tenure and Stroke

Death	Odds Ratio	Sig.	[95% Conf.	Interval]
Sex	1.53	0.000	1.41	1.66
Age	1.10	0.000	1.09	1.10
Housing Association	1.42	0.000	1.22	1.66
Council Housing	1.08	0.149	.97	1.20
Tax Bands D-E	1.35	0.000	1.21	1.52
Tax Bands A-C	1.65	0.000	1.46	1.87
Falls	1.67	0.000	1.40	1.98
Heart Disease	1.72	0.000	1.43	2.07
Stroke	4.08	0.000	2.95	5.64
Housing-Ass X Stroke	.97	0.913	.52	1.79
Council-Hou X Stroke	.91	0.639	.60	1.37
Number of obs = 43472		χ^2(11 d.f) = 4025.41, P = 0.0000		
Pseudo R^2 = 0.177		Log likelihood = -9384.793		

H.2.2 Interaction between Council tax banding and FIS in full model

In the following three models, the interaction between tax banding and three causes of hospital admissions (FIS) will be examined in a full model. As has been highlighted, in the all three following models (illustrated in Tables-H.10, H.11 and H.12), the interaction between two groups of lower tax banding (categories D-E and A-C) with all three causes of hospital admissions (FIS), are not significant.

H.2.2.1 Full model including the interaction between Tax banding and Falls:

Table-H.10 Full model including the interaction between Tax banding and Falls

Death	Odds Ratio	Sig.	[95% Conf.	Interval]
Sex	1.53	0.000	1.41	1.66
Age	1.10	0.000	1.09	1.10
Housing Association	1.42	0.000	1.23	1.65
Council Housing	1.08	0.168	.97	1.19
Tax Bands D-E	1.35	0.000	1.20	1.52
Tax Bands A-C	1.68	0.000	1.48	1.91
Falls	1.89	0.000	1.37	2.62
Heart Disease	1.73	0.000	1.44	2.08
Stroke	3.85	0.000	3.18	4.65
Tax Bands D-E X Falls	1.02	0.928	.66	1.57
Tax Bands A-C X Falls	.71	0.111	.46	1.08
Number of obs = 43472		χ^2(11 d.f)= 4029.22, P = 0.0000		
Pseudo R^2 = 0.177		Log likelihood = -9382.89		

H.2.2.2 Full model including the interaction between Tax banding and Heart Disease

Table-H.11 Full model including the interaction between Tax bands and Heart disease

Death	Odds Ratio	Sig.	[95% Conf.	Interval]
Sex	1.53	0.000	1.41	1.65
Age	1.10	0.000	1.09	1.10
Housing Association	1.42	0.000	1.22	1.64
Council Housing	1.07	0.175	.97	1.19
Tax Bands D-E	1.37	0.000	1.22	1.53
Tax Bands A-C	1.68	0.000	1.48	1.91
Falls	1.66	0.000	1.40	1.98
Heart Disease	2.16	0.000	1.49	3.13
Stroke	3.85	0.000	3.19	4.66
Tax Bands D-E X HD	.81	0.388	.50	1.30
Tax Bands A-C X HD	.69	0.120	.43	1.10
Number of obs = 43472 Pseudo R^2 = 0.177		χ^2(11 d.f)= 4027.58, P = 0.0000 Log likelihood = -9383.71		

H.2.2.3 Full model including the interaction between Tax banding & Stroke:

Table-H.12 Full model including the interaction between Tax banding and Stroke

Death	Odds Ratio	Sig.	[95% Conf.	Interval]
Sex	1.53	0.000	1.41	1.65
Age	1.10	0.000	1.09	1.10
Housing Association	1.42	0.000	1.23	1.65
Council Housing	1.08	0.167	.97	1.19
Tax Bands D-E	1.36	0.000	1.21	1.53
Tax Bands A-C	1.66	0.000	1.46	1.88
Falls	1.67	0.000	1.40	1.98
Heart Disease	1.73	0.000	1.44	2.08
Stroke	4.26	0.000	2.96	6.15
Tax Bands D-E X Stroke	.88	0.584	.54	1.41
Tax Bands A-C X Stroke	.87	0.559	.54	1.40
Number of obs = 43472 Pseudo R^2 = 0.177		χ^2(11 d.f)= 4025.59, P = 0.0000 Log likelihood = -9384.71		

Appendix-I ROC curve construction

In a ROC curve, sensitivity is calculated using every value of a variable (factor) in the data set as a cut-point and is plotted against the corresponding value of (1-specificity) at that point. Thus the curve is the true positives plotted against the false positives calculated using each value of the test as a cut-point (Peat & Barton, 2005). An illustration of a ROC curve is presented in Figure I.2.Ifirst need to show how the cut-points are determined.

I.1 Calculation of cut-points and construction of ROC curves

The process of the calculation of the sensitivity and specificity in detail can be explained with help of the following two examples using Stata.

Example-1 includes the process of the calculation of sensitivity and specificity for two variables (Mortality and Tenure). Table I.1 is a classification table of the binary variables deaths and tenure (for the data analysed in earlier sections), by Stata.

Table I.1 classification table for variables 'death' and 'tenure'.

```
           |        ten_bi
  death_bi |         0          1 |     Total
-----------+----------------------+----------
         0 |    19,429     20,855 |    40,284
         1 |     1,197      1,991 |     3,188
-----------+----------------------+----------
     Total |    20,626     22,846 |    43,472
```

In first step Stata creates a classification table by allocating the value to each cell based on the convention that '0' signifies absence and '1' signifies presence ('0/–' and '1/+' title of rows and columns).

For example the cell represented by 'a' in the Table I.1, has the title of '11' or '++' (which is TP) and so on. The above table from Stata (table I.1), using conventional labels a, b, c and d to match table I.1, can be presented in the form of Table I.2.

Table H.2 Reproduction of Table H.1 with each cell in order

		tenure_bi	
		1 (+)	0(-)
death_bi	1 (+)	1,991 **(a)**	20,855 **(b)**
	0 (-)	1,197 **(c)**	19,429 **(d)**
		3,188 **a + c**	40,284 **b + d**

For the above example, the sensitivity and specificity can be calculated as:

Sensitivity = a/(a+c) = 1,991/3,188 = 62.45% (the same value as Stata output, in Table 8.4)

Specificity = d/(b+d) = 19,429/40,284 = 48.23% (the same value as Stata output, in Table 8.5)

1-Specificity = 100%-48.23% = 51.77%

The X and Y values in the graph are respectively 51.77% and 62.45%.

Table I.3 shows the detailed reports of Sensitivity and Specificity from Stata.

Table I.3 Detailed reports of Sensitivity and Specificity from Stata

Cutpoint	Sensitivity	Specificity	Correctly Classified	LR+	LR-
(>= 0)	100.00%	0.00%	7.33%	1.0000	
(>= 1)	62.45%	48.23%	49.27%	1.2064	0.7785
(> 1)	0.00%	100.00%	92.67%		1.0000

The ROC curve for two binary variables (death and tenure) produced by Stata is illustrated in Figure I.1 and Table I.4 contains detailed information.

Fig I.1 ROC curve for two binary variables; death and tenure from Stata

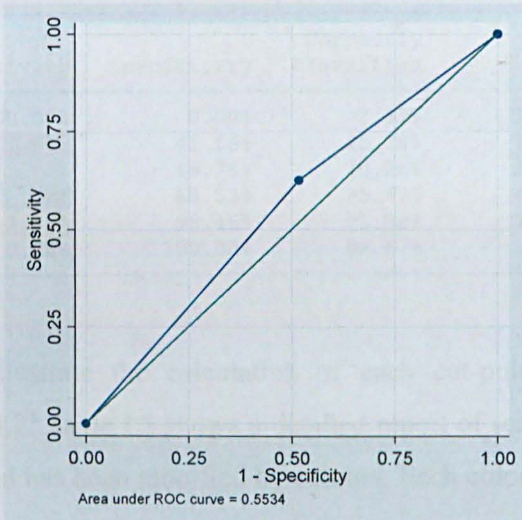


Table I.4 Detailed information of area under ROC Curve (AUC) from Stata

Obs	ROC Area	Std. Err.	-Asymptotic Normal-- [95% Conf.	Interval]
43472	0.5534	0.0045	0.54466	0.56217

In the above graph (Figure I.1) the area between the diagonal and the curve shows the size (magnitude) of differences between mortality in social housing and private housing. The area under the curve suggests there is little to distinguish in their impact on death when considered in isolation. In general, when the ROC curve climbs rapidly towards the upper left hand corner of the graph, the test result is good.

Clearly, with categorical variables which are dichotomous there is little discriminating potential in the number of points in order to construct a curve.Iwill now extend the number of cut-offs by considering age as a categorical variable with five levels (50-59, 60-69, 70-79, 80-89 and 90 years old or more).

Example-2 An example of ROC curve with the process of calculation of sensitivity and specificity with variables 'death' (binary) and 'age' (with 5 categories by increasing age from category 1 to 5):

Table I.5 Detailed reports of Sensitivity and Specificity by Stata (modified)

Cutpoint	Sensitivity	Specificity	Correctly Classified	LR+	LR-
(>= 1)	100.00%	0.00%	7.33%	1.0000	
(>= 2)	90.97%	41.66%	45.28%	1.5594	0.2168
(>= 3)	75.60%	69.78%	70.20%	2.5013	0.3497
(>= 4)	46.80%	88.53%	85.47%	4.0790	0.6009
(>= 5)	13.86%	98.16%	91.98%	7.5476	0.8775
(> 5)	0.00%	100.00%	92.67%		1.0000

The steps bellow illustrate the calculation of each cut-point in the ROC curve illustrated in Figure I.2. Table I.5 shows a detailed report of sensitivity and specificity produced by Stata and has been modified by colours. Each colour represents one of the steps below.

The process of identifying each cell with an appropriate value (labelled as a, b, c and d in balloons) for computation of the sensitivity and specificity in Table I.5 (which is the basis of ROC curve in Figure I.2) are demonstrated in the following four steps. Sensitivity at each step can be computed by dividing the value of 'a' to the sum of 'a' and 'c'. The value of sensitivity for each step can be find in Table H.5 highlighted with the same colour used in each step below. The same mechanism is also appropriate for the computation of specificity (by dividing $d/_{b+d}$). The four cut-points are determined in the following four steps:

Step-1 (cut-point 1):

death_bi	diage2 1	2	3	4	5	Total
0	16,784	11,325	7,553	3,882	740	40,284
1	288	490	918	1,050	442	3,188
Total	17,072	11,815	8,471	4,932	1,182	43,472

Step -2 (cut-point 2):

death_bi	diage2 1	2	3	4	5	Total
0	16,784	11,325	7,553	3,882	740	40,284
1	288	490	918	1,050	442	3,188
Total	17,072	11,815	8,471	4,932	1,182	43,472

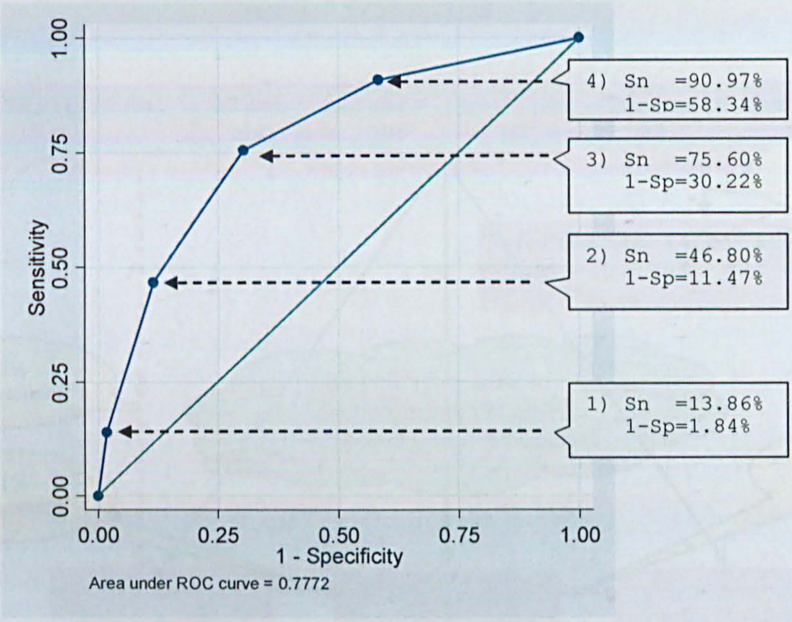
Step-3 (cut-point 3):

death_bi	diage2 1	2	3	4	5	Total
0	16,784	11,325	7,553	3,882	740	40,284
1	288	490	918	1,050	442	3,188
Total	17,072	11,815	8,471	4,932	1,182	43,472

Step-4 (cut-point 4):

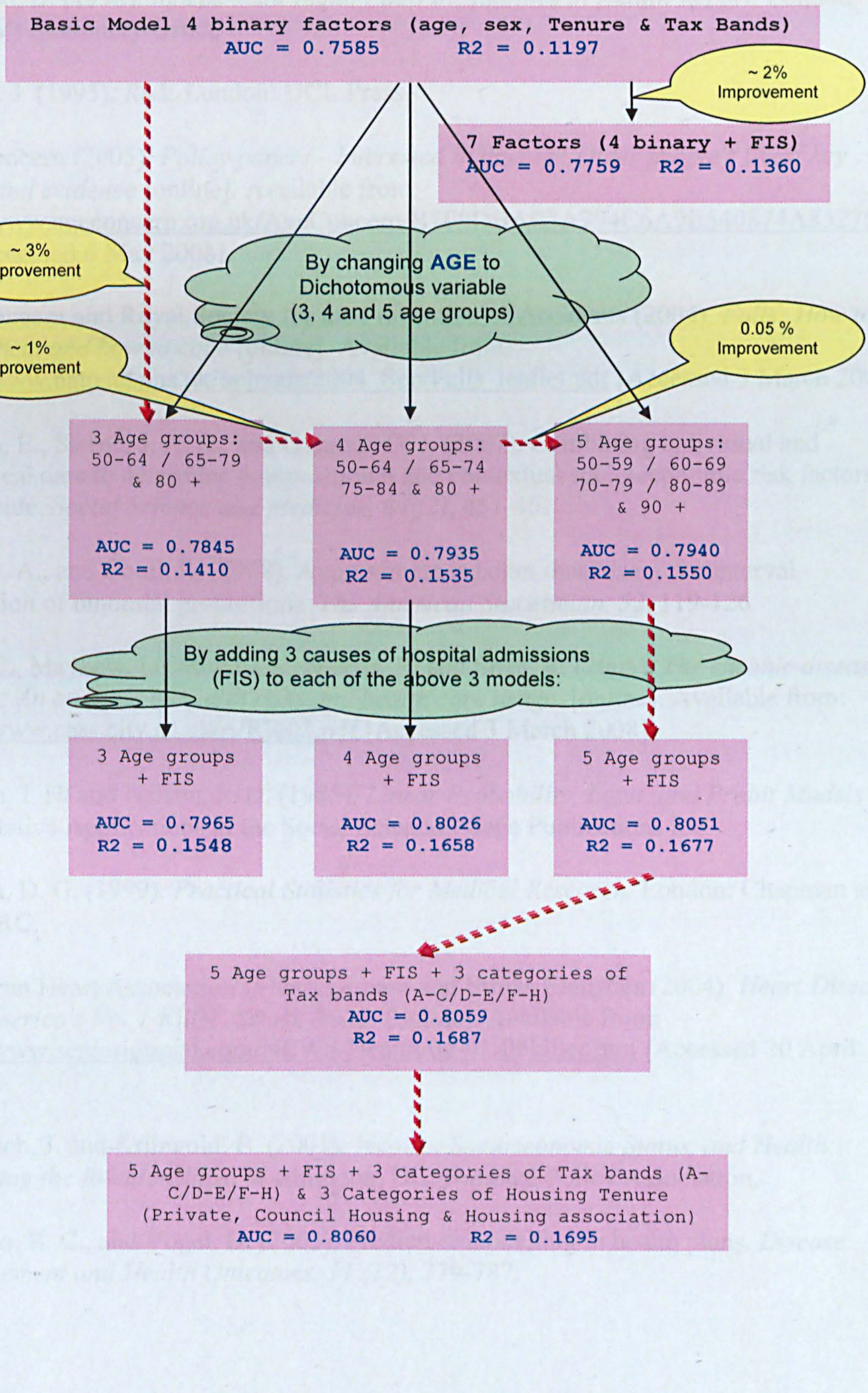
death_bi	diage2 1	2	3	4	5	Total
0	16,784	11,325	7,553	3,882	740	40,284
1	288	490	918	1,050	442	3,188
Total	17,072	11,815	8,471	4,932	1,182	43,472

Figure I.2 A ROC Curve shows the cut-point with the values for Sensitivity and 1-Specificity for each point.



The AUC in the above figure (0.78) shows a stronger result comparing with the previous one in figure I.1 with AUC = 0.55.

Appendix-J A summary of the evaluation of each stage of logistic regression modelling



Bibliography

Acheson, D. (1998). *Independent Inquiry into Inequalities in Health Report.* London: Her Majesty's Stationery Office

Adams. J (1995). *Risk.* London: UCL Press.

Age Concern (2005). *Policy papers - Interested in ageing? Older people's lives: key issues and evidence* [online]. Available from: http://www.ageconcern.org.uk/AgeConcern/B7F9DBAE3A794C6A9B540874A8527673.asp [Accessed 6 May 2008].

Age Concern and Royal Society for the Prevention of Accidents (2004). *Falls: How to avoid them and how to cope* [online]. Available from: http://www.natpact.nhs.uk/uploads/2004_Sep/Falls_leaflet.pdf [Accessed 3 March 2008].

Agerbo, E., Sterne, J. A. C., and Gunnell, D. J. (2007). Combining individual and ecological data to determine compositional and contextual socio-economic risk factors for suicide. *Social Science and Medicine, 64 (2)*, 451-461.

Agresti, A., and Coull, B. (1998). Approximate is better than 'exact' for interval estimation of binomial proportions. *The American Statistician, 52*, 119-126.

Alder, J., Mayhew, L., Moody, S., Morris, R. and Shah, R. (2005). *The chronic disease burden: An analysis of health risks and health care usage.* [online]. Available from: http://www.cass.city.ac.uk/ri/RI003.pdf [Accessed 3 March 2008].

Aldrich, J. H. and Nelson, F. D. (1985). *Linear Probability, Logit, and Probit Models* Quantitative Applications in the Social Sciences. Sage Publications Inc.

Altman, D. G. (1999). *Practical Statistics for Medical Research.* London: Chapman and Hall/CRC.

American Heart Association's Heart Disease and Stroke Statistics (2004). *Heart Disease Still America's No. 1 Killer, Stroke No. 3.* [online]. Available from: http://www.seniorjournal.com/NEWS/Health/04-01-05killer.htm [Accessed 20 April 2008].

Auerbach, J. and Krimgold, B. (2001). *Income, Socioeconomic Status, and Health: Exploring the Relationships.* Washington, DC: National Policy Association.

Axelrod, R. C., and Vogel, D. (2003). Predictive modelling in health plans. *Disease Management and Health Outcomes, 11 (12)*, 779-787.

Barros, A., and Hirakata, V. (2003). Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio. *BioMed Central, Medical Research Methodology, 3*, 21.

Besley, T. and Kanbur, R. (1990). *The Principles of Targeting. In Development Economics* [online]. Available from: http://www-wds.worldbank.org/servlet/WDSContentServer/WDSP/IB/1990/03/01/000009265_3960928225215/Rendered/PDF/multi0page.pdf [Accessed 4 June 2008].

Bhopal, R. (1997). Is research into ethnicity and health racist, unsound, or important science? *British Medical Journal, 314 (7096)*, 1751-1756.

Boag, J. (2003). Living Rent Free. Unpublished work conducted by Falkirk Council made available by personal communication.

Bowling, A. (2004). Socioeconomic differentials in mortality among older people. *Journal of Epidemiology and Community Health, 58 (6)*,438-440.

Box, G. E. P. (1990). Do interactions matter? *Quality Engineering, 2*, 365-369.

Breeze, E., Jones, D. A., Wilkinson, P., Latif, A. M., Bulpitt, C. J., and Fletcher, A. E. (2004). Association of quality of life in old age in Britain with socioeconomic position: baseline data from a randomised controlled trial. *Journal of Epidemiology and Community Health, 58 (8)*, 667-673.

Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984). *Classification and Regression Trees*. Chapman and Hall/CRC.

Bruhn, Å. (2001). *The 2005 population and housing census in Sweden will be totally register-based* [online]. Available from: http://unstats.un.org/unsd/demographic/meetings/egm/Symposium2001/docs/symposium_13.htm [Accessed 6 March 2008].

Buchanan, D. (2003). Social Epidemiology. *Health Education Research, 18 (3)*, 404-407.

Camden and Islington Health Authority (2001). *Health Inequalities in Camden and Islington*. London: Camden and Islington Health Authority.

Camden Primary Care Trust (2004). *Health in Camden: the Facts*. [online]. Available from: http://www.camdenpct.nhs.uk/your_health/your_health.htm [Accessed 20 April 2006].

Camden Primary Care Trust (2005). *Improving Health in Camden: Annual Public Health Report 2003/04*. London: Fraser Hamilton Associates.

Camden Primary Care Trust and Camden Council (2003). *Draft long-term care and support strategy for older people in Camden.* London: Camden Primary Care Trust; Camden Council.

Carter-Pokras, O., Baquet, C. R., and Poppell, C. F. (2004). Re: Four-Year Review of the Use of Race and Ethnicity in Epidemiologic and Public Health Research. *American Journal of Epidemiology, 160 (4)*, 403-404.

Cassel, C. M., Sarndal, C. E. and Wretman, J. H. (1977). *Foundations of inference in survey sampling.* New York: John Wiley and Sons.

Centre for Disease Control and Prevention (1999). *Achievements in Public Health: Decline in Deaths from Heart Disease and Stroke - United States, 1900-1999* [online]. Available from: http://www.cdc.gov/mmwr/preview/mmwrhtml/mm4830a1.htm [Accessed 6 May 2008].

Chadwick, E. (1842). *Genral Report on the Sanitary Conditions of the Labouring Population of Great Britain.* London: W. Clowes and Son.

Codd, E. F. (1979). Extending the Database Relational Model to Capture More Meaning. *ACM Transactions on Database System, 4 (4)*, 397-434.

Codd, E. F. (1990). *The Relational Model for Database Management.* Boston: Addison Wesley Publishing Company.

Connolly, T. and Begg, C. (2001). *Database Systems: A Practical Approach to Design, Implementation, and Management.* Addison Wesley.

Cousins, M. S., Shickle, L. M., and Bander, J. A. (2002). An Introduction to Predictive Modeling for Disease Management Risk Stratification. *Disease Management, 5 (3)*, 157-167.

Cowan, J. (2003). Risk management and the NSF for older people. *Clinical Governance: An International Journal, 8 (1)*, 92-95.

Curtis, S. and Jones, I. R. (1998). Is there a place for geography in the analysis of health inequality? In: M. Bartley, D. Blane and G. Davey Smith, eds. *The Sociology of Health Inequalities.* Oxford: Blackwell Publishers, 1998, pp. 85-112.

Danesh, J., Gault, S., Semmence, J., Appleby, P., Peto, R., Ben-Shlomo, Y., and Smith, G. D. (1999). Postcodes as useful markers of social class: population based study in British households Commentary: Socioeconomic position should be measured accurately. *British Medical Journal, 318 (7187)*, 843-845.

Date, C. J. (2000). *An Introduction to Database Systems.* Boston, MA: Addison-Wesley.

Department for Communities and Local Government (2006). *English House Condition Survey 2004-Annual Report: decent homes and decent places*. London: Her Majesty's Stationery Office

Department of Health (2001). *National Service Framework for Older People*. London: Her Majesty's Stationery Office.

Department of Health (2006). *Hospital and community health services gross current expenditure by age, 2002-03*. [online]. Available from: http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/AnnualReports/Browsable/DH_4947193 [Accessed 6 May 2008].

Department of Health (1989). *Caring for People*. London: Her Majesty's Stationery Office.

Department of Health (1997). *Better Services for Vulnerable People*. London: Her Majesty's Stationery Office.

Department of Health (1998). *Modernising Health and Social Services, National Priorities Guidance*. London: Her Majesty's Stationery Office.

Department of Health (1999). *Saving Lives: Our Healthier Nation*. London: Her Majesty's Stationery Office.

Dorling, D. (1997a). Death in Britain: How local mortality rates have changed: 1950s-1990s. York: Joseph Rowntree Foundation.

Dorling,D. (1997b). *Changing mortality ratios in local areas of Britain 1950s-1990s*. Bristol: Joseph Rowntree Foundation [online]. Available from: http://www.jrf.org.uk/knowledge/findings/socialpolicy/sp126.asp [Accessed 3 June 2008].

Dorling, D., Shaw, M., and Brimblecombe, N. (2001). Housing wealth and community health: exploring the role of migration. In: H. Graham, ed. *Understanding Health Inequalities*. Buckingham: Open University Press, 2001, pp. 186-199.

Dressler, W. W., Oths, K. S., and Gravlee, C. C. (2005). Race and ethnicity in public health research: Models to explain health disparities. *Annual Review of Anthropology, 34*,231-252.

Drever, F., and Whitehead, M. (1997). *Health inequalities, Decennial Supplement*. London: Office for National Statistics, Her Majesty's Stationery Office: London

Easterlow, D., and Smith, S. J. (2004). Housing for health: can the market care? *Environment and Planning A, 36 (6)*, 999-117.

Encyclopedia of Public Health (2002). *Risk Assessment* [online]. Available from: http://www.answers.com/topic/risk-assessment [Accessed 10 April 2008].

Ender, P. (2006). *Applied Categorical and Nonnormal Data Analysis* [online]. Available from: http://www.gseis.ucla.edu/courses/ed231c/notes3/fit.html [Accessed 6 May 2008].

Filakti , H., and Fox, J. (1995). Differences in mortality by housing tenure and by car access from the OPCS Longitudinal Study. *Population Trends, 81,* 27-30.

Gardner, J., and Oswald, A. (2004). How is mortality affected by money, marriage, and stress? *Journal of Health Economics, 23 (6)*, 1181-1207.

Gastwirth, J. L. (1972). The Estimation of the Lorenz Curve and Gini Index. *The Review of Economics and Statistics, 54 (3)*, 306-316.

Gennick,J. (1999). *Oracle SQL * Plus, The Definitive Guide*. O'Reilly and Associates.

Godfrey, M. (2001). Prevention: developing a framework for conceptualizing and evaluating outcomes of preventive services for older people. *Health and Social Care in the Community, 9 (2)*, 89-99.

Gould, M. I., and Jones, K. (1996). Analyzing perceived limiting long-term illness using U.K. census microdata. *Social Science and Medicine, 42 (6)*, 857-869.

Gould, W. (2000). A publication to promote communication among Stata users: Interpreting logistic regression in all its forms. *Stata Technical Bulletin, 53,* 19-29.

Graham, H. (2001). The challenge of health inequalities. In: H. Graham, ed. *Understanding Health Inequalities*. Buckingham: Open University Press, 2001, pp. 3-21.

Grundy, E., and Sloggett, A. (2003). Health inequalities in the older population: the role of personal capital, social resources and socio-economic circumstances. *Social Science and Medicine, 56 (5)*, 935-947.

Hand, D. J., and Till, R. J. (2001). A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. *Machine Learning, 45 (2)*, 171-186.

Hanley, J. A., and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143 (1)*, 29-36.

Harding, S., Bethune, A., Maxwell, R., and Brown, J. (1997). Mortality trends using the Longitudinal Study. In: F. Drever and M. Whitehead, eds. *Health Inequalities: decennial supplement*. London: Her Majesty's Stationery Office, pp. 143-155.

Hattersley, L. (1999). Trends in life expectancy by social class: an update. *Health Statistics Quarterly, 2 (1999),* 16-24.

Hillary Osborne and agencies (2006). Retirement age will rise to 68. Guardian Unlimited [25/05/2006].

Hoffmann, R. (2005). Do socioeconomic mortality differences decrease with rising age? *Demographic Research, 13 (2)*, 35-62.

Holton, G. A. (2004). Defining Risk. *Financial Analysis Journal, 60 (6)*, 19-25.

Hosmer, D. W., and Lemeshow, S. (2000). *Applied Logistic Regression-Second Edition.* New York: John Wiley and Sons. Inc.

Huisman, M., Kunst, A. E., Andersen, O., Bopp, M., Borgan, J. K., Borrell, C., Costa, G., Deboosere, P., Desplanques, G., Donkin, A., Gadeyne, S., Minder, C., Regidor, E., Spadea, T., Valkonen, T., and Mackenbach, J. P. (2004). Socioeconomic inequalities in mortality among elderly people in 11 European populations. *Journal of Epidemiology and Community Health, 58 (6)*, 468-475.

Huisman, M., Kunst, A. E., and Mackenbach, J. P. (2003). Socioeconomic inequalities in morbidity among the elderly; a European overview. *Social Science and Medicine, 57 (5)*, 861-873.

Iversen, L., Andersen, O., Andersen, P. K., Christoffersen, K., and Keiding, N. (1987). Unemployment and mortality in Denmark, 1970-80. *British Medical Journal,* 295 (6603), 879-884.

Jacobson, B., and Aspinall, P. (2006). *How to Analyse Ethnic Differences in Health, Health Care and the Workforce: A Toolkit for the NHS.* London: London Health Observatory.

James, W. P., Nelson, M., Ralph, A., and Leather, S. (1997). Socioeconomic determinants of health: The contribution of nutrition to inequalities in health. *British Medical Journal, 314 (7093)*, 1545–1549.

Jewell, N. P. (2004). *Statistics for Epidemiology.* Chapman and Hall / CRC.

Johnston, A., Cooper, J., Webb, R., and Kapur, N. (2006). Individual- and area- level predictors of self-harm repetition. *The British Journal of Psychiatry, 189 (5)*, 416-421.

Jones, P., and Elias, P. (2006). *Administrative data as research resources: a selected audit.* Warwick: The National Data Strategy (NDS); ESRC and Warwick Institute for Employment Research.

Kawachi, I., Kennedy, B. P., Lochner, K., and Prothrow-Stith, D. (1997). Social capital, income inequality, and mortality. *American Journal of Public Health, 87 (9)*, 1491-1498.

Kelly, R., and Bramwell, S. (2006). *Report to Cabinet: Older People Strategic Framework.* Gateshead Council [online]. Available from: http://online.gateshead.gov.uk/docushare/dsweb/Get/Document-14954/Item+14+-+Older+People+Strategy.doc [Accessed 3 June 2008].

King's Fund (2005). *Predictive risk project* [online]. Available from: http://www.kingsfund.org.uk/current_projects/predictive_risk/index.html [Accessed 6 May 2008].

Krieger, N. (2001). Theories for social epidemiology in the 21st century: an ecosocial perspective. *International Journal of Epidemiology, 30 (4)*, 668-677.

Krieger, N. (2000). Epidemiology and social sciences: towards a critical reengagement in the 21st century. *Epidemiol Rev, 22 (1)*, 155-163.

Kuper, L. (1975). *Race, science and society.* London: Allen and Unwin.

Laslett, P. (1996). *A Fresh Map of Life: Emergence of the Third Age.* London: Palgrave Macmillan.London.

Lauderdalei, D. S., and Kestenbaum, B. (2000). Asian American ethnic identification by surname. *Population Research and Policy Review, 19 (3),* 283-300.

Lawrence, D., Brown, T., Tony, C., and Anirban, D. (2001). Interval Estimation for a Binomial Proportion. *Statistical Science, 16 (2)*, 101-133.

Leibfritz, W. (2008). *Retiring later makes sense.* [online]. Available from: http://www.oecdobserver.org/news/fullstory.php/aid/824/Retiring_later_makes_sense.html [Accessed 6 May 2008].

Lemeshow, S. (2007). A point on Creating ROC curves. In Applied Logistic Regression; Hosmer,D and Lemeshow, S (eds.). New York: John Wiley & Sons, Inc.

Lewis, G., Bebbington, P., Brugha, T., Farrell, M., Gill, B., Jenkins, R., and Meltzer, H. (2003). Socio-economic status, standard of living, and neurotic disorder. *International Review of Psychiatry, 15 (1-2)*, 91-96.

Liang, J., Bennett, J., Krause, N., Kobayashi, E., Kim, H., Brown, J. W., Akiyama, H., Sugisawa, H., and Jain, A. (2002). Old Age Mortality in Japan: Does the Socioeconomic Gradient Interact With Gender and Age? *Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 57B (5)*, S294-S307.

London Borough of Camden (2007c). *Service-Specific Information Sharing Agreement and Protocol (ISA) between Camden ISA team and Camden PCT.* London: London Borough of Camden.

London Borough of Camden (2002). *The Quality of Life Strategy for Camden's Older Citizens*. London.

London Borough of Camden (2007a). *Camden Carers Liaison Group Meeting, 20 Feb 2007* [online]. Available from: http://www3.camden.gov.uk/templates/committees/showHTML.cfm?file=24015.htm [Accessed 6 May 2008].

London Borough of Camden (2007b). *Overarching Information Sharing Protocol Between the LB of Camdne and Camden PCT*. LB of Camden and Camden PCT.

Low, A., and Low, A. (2006). Health economics: Important of relative measures in policy on health inequalities. *British Medical Journal, 332 (7547)*, 969-971.

Macintyre, S., Hiscock, R., Kearns, A., and Ellaway, A. (2001). Housing tenure and health inequalities: a three-dimensional perspective on people, homes and neighbourhoods. In: H. Graham, ed. *Understanding Health Inequalities*. Buckingham: Open University Press, 2001, pp. 129-142.

Mackenbach, J. P., Kunst, A. E., Cavelaars,A.E., Groenhof,F., and Geurts,J.J. (1997). Socioeconomic inequalities in morbidity and mortality in western Europe. *The Lancet, 349 (9066)*, 1655-1659.

Manhapra, A., Canto, J. G., Vaccarino, V., Parsons, L., Kiefe, C. I., Barron, H. V., Rogers, W. J., Weaver, W. D., and Borzak, S. (2004). Relation of age and race with hospital death after acute myocardial infarction. *American Heart Journal, 148 (1)*, 92-98.

Marmot, M. (1998). The magnitude of social inequalities in coronary heart disease: possible explanations. In Sharp, I. (ed.), *Social inequalities in coronary heart disease*. London: Her Majesty's Stationery Office.

Marmot, M. G., Bosma, H., Hemingway, H., Brunner, E., and Stansfeld, S. (1997). Contribution of job control and other risk factors to social variations in coronary heart disease incidence. *The Lancet, 350 (9073)*, 235-239.

Marmot, M. G., and Shipley, M. J. (1996). Do socioeconomic differences in mortality persist after retirement? 25 Year follow up of civil servants from the first Whitehall study. *British Medical Journal, 313 (7066)*, 1177-1180.

Martikainen, P. T., and Valkonen, T. (1996). Excess mortality of unemployed men and women during a period of rapidly increasing unemployment. *The Lancet, 348 (9032)*, 909-912.

Mason, S. J., and Graham, N. E. (2002). Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. *Quarterly Journal of the Royal Meteorological Society, 128*, 2145-2166.

Mateos, P. (2007). A Review of Name-based Ethnicity Classification Methods and their Potential in Population Studies. *Population, Space and Place, 13 (4),* 243-263.

Mayhew Associates Ltd and Camden PCT (2005). *Youth offending in Camden: A risk analysis*. London: Camden Primary Care Trust.

Mayhew, L., and Harper, L. (2006). *Brent partnership for older people project (POPP): Baseline evaluation.* London: Mayhew Associates Ltd.

Mayhew, L. (2002). *A systematic approach to the examination of health and social risks at neighbourhood level.* London: City University. ISBN 1 901615 66 9.

McDonough, P., Duncan, G. J., Williams, D., and House, J. (1997). Income dynamics and adult mortality in the United States, 1972 through 1989. *American Journal of Public Health, 87*,1476-1483.

McLoone, P., and Ellaway, A. (1999). Postcodes don't indicate individuals' social class. *British Medical Journal, 319 (7215)*, 1003.

Microsoft Corporation (2003a). Microsoft Office Access. Microsoft.

Microsoft Corporation (2003b). Microsoft Office Excel. Microsoft.

Microsoft Corporation (2003c). Microsoft SQL Enterprise Manager. Microsoft.

Mkandawire, T. (2005). *Targeting and Universalism in Poverty Reduction.* United Nations Research Institute for Social Development [online]. Available from: http://www.unrisd.org/unrisd/website/document.nsf/d2a23ad2d50cb2a280256eb300385855/955fb8a594eea0b0c12570ff00493eaa/$FILE/mkandatarget.pdf [Accessed 6 May 2008].

Modood, T., Berthoud, R., Lakey, J., Nazroo, J., Smith, P., Virdee, S., and Beishon, S. (1997). *Ethnic Minorities in Britain: Diversity and Disadvantage - Fourth National Survey of Ethnic Minorities*. London: Policy Studies Institute.

Moreau, P., and Jordan, L. T. (2005). A framework for the animal health risk analysis of biotechnology-derived animals: a Canadian perspective. *Revue Scientifique et technique (Office International des Epizooties), 24 (1)*, 51-60.

Morgan, J. (1962). The Anatomy of Income Distribution. *The Review of Economics and Statistics, 44 (3)*, 270-283.

Morris, J. K., Cook, D. G., and Shaper, A. G. (1994). Loss of employment and mortality. *British Medical Journal, 308 (6937)*, 1135-1139.

Moser, K. A., Fox, A. J., and Jones, D. R. (1984). Unemployment and mortality in the OPCS longitudinal study. *The Lancet, 8;2 (8415)*, 1324-1329.

Nyberg, L., and Gustafson, Y. (1995). Patient Falls in Stroke Rehabilitation : A Challenge to Rehabilitation Strategies. *Stroke, 26 (5)*, 838-842.

Office for National Statistics (2006a). *2001 Census: Standard Tables for London Borough of Camden, Table TT007*. [online]. Available from: http://neighbourhood.statistics.gov.uk/dissemination/LeadDatasetList.do?a=3andb=276749andc=camdenandd=13andg=329603andi=1001x1003andm=0andenc=1anddomainId=7 [Accessed 6 May 2008].

Office for National Statistics (2004b). *Mid-2003 Population Estimates* . [online]. Available from: http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=14357 [Accessed 6 May 2008].

Office for National Statistics (2006b). *Census 2001: Neighbourhood Statistics, Camden Population* . [online]. Available from: http://www.statistics.gov.uk/census2001/pyramids/pages/00ag.asp [Accessed 6 May 2008].

Office for National Statistics (2006c). *Census 2001: People and their homes in England and Wales* . [online]. Available from: http://www.statistics.gov.uk/census2001/profiles/commentaries/housing.asp#background [Accessed 6 May 2008].

Office for National Statistics (2004a). *Focus on Older People: Living arrangements*. [online]. Available from: http://www.statistics.gov.uk/cci/nugget.asp?id=1264 [Accessed 6 May 2008].

Office for National Statistics (2006d). *Super Output Areas (SOAs)* . [online]. Available from: http://www.statistics.gov.uk/ [Accessed 6 May 2008].

Oliver, S., and Peersman, G. (2001). *Using Research for Effective Health Promotion.* Buckingham: Open University Press.

Olsson, B., Löfgren, B., Gustafson, Y., and Nyberg, L. (2004). Validation of a fall risk index in stroke rehabilitation. *Journal of Stroke and Cerebrovascular Diseases, 14*, 23-28.

Pampel, F. C. (2000). *Logistic regression: A primer. Quantitative Applications in the Social Sciences*. Thousand Oaks, CA: Sage Publications.

Pearce, N., Foliaki, S., Sporle, A., and Cunningham, C. (2004). Genetics, race, ethnicity, and health. *British Medical Journal, 328 (7447)*, 1070-1072.

Peat, J., and Barton, B. (2005). *Medical Statistics: A Guide to Data Analysis and Critical Appraisal*. Blackwell Publishing.

Pepe, M. S. (2003). *The statistical evaluation of medical tests for classification and prediction*. Oxford University Press.

Philp, I. (2004). *Better health in old age*. London: Her Majesty's Stationery Office.

Poole, K. E. S., Reeve, J., and Warburton, E. A. (2002). Falls, Fractures, and Osteoporosis After Stroke: Time to Think About Protection? *Stroke, 33 (5)*, 1432-1436.

Power, A. (2000). Poor Areas and Social Exclusion. In: A. Power and J. Wilson, eds. *Social Exclusion and the Future of Cities*. London: London School of Economics: CASE paper 35.

Putnam, R. D. (1995). Bowling Alone: America's Declining Social Capital. *Journal of Democracy, 6 (1)*, 65-78.

Raine, R., Godden, S., and MacKee, M. (2006). Information and intelligence for healthy populations. *British Medical Journal, 332 (7552)*, 1226-1227.

Ratcliffe, P. (1998). 'Race', Housing and Social Exclusion. *Housing Studies, 13*,807-818.

Raymer, J., Abel, G., and Smith, P. W. F. (2007). Combining census and registration data to estimate detailed elderly migration flows in England and Wales. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 170 (4)*, 891-908.

Redfern, P. (2004). An alternative view of the 2001 census and future census taking. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 167 (2)*, 209-228.

Report of the Steering Committee of the Better Government for Older People Programme (2000). *All Our Futures, Better Government for Older People* [online]. Available from: http://www.bl.uk/welfarereform/issue15/care.html [accessed 4 June 2008].

Robertson, A., and Minkler, M. (1994). New health promotion movement: a critical examination. *Health Education Quarterly, 21(3)*, 295-312.

Rose, G. (1992). *The Strategy of Preventive Medicine*. Oxford University Press.

Royal Society (1983). *Risk Assessment: A Study Group Report* . London.

Senior, P. A., and Bhopal, R. (1994). Ethnicity as a variable in epidemiological research. *British Medical Journal, 309 (6950)*, 327-330.

Shaw, M., Gordon, D., Dorling, D., Mitchell, R., and Smith, G. D. (2000). Increasing mortality differentials by residential area level of poverty: Britain 1981-1997. *Social Science and Medicine, 51 (1),* 151-153.

Shaw, M., Dorling, D., and Brimblecombe, N. (1998). Changing the map: health in Britain 1951-1991. *Sociology of Health and Illness, 20 (5)*, 694-709.

Sjosten, N. M., Salonoja, M., Piirtola, M., Vahlberg, T. J., Isoaho, R., Hyttinen, H. K., Aarnio, P. T., and Kivela, S. L. (2007). A multifactorial fall prevention programme in the community-dwelling aged: predictors of adherence. *The European Journal of Public Health, 17 (5)*, 464-470.

Smith, J. P. (1999). Healthy Bodies and Thick Wallets: the Dual Relation between Health and Economic Status. *Journal of Economic Perspectives, 13 (2)*, 145-166.

SPSS (2003). SPSS 12.0 for Windows. Chicago: SPSS inc.

StataCorp (2005). Stata Statistical Software. Texas: StataCorp LP.

Statistics Commission (2006). *Statistics Commission says legislation must transform public trust in official statistics* [online]. Available from: http://www.statscom.org.uk/C_1074.aspx [Accessed 3 June 2008].

Strobl, J., Cave, E., and Walley, T. (2000). Data protection legislation: interpretation and barriers to research. *British Medical Journal, 321 (7265)*, 890-892.

Svendsen, S. W., Bonde, J. P., Mathiassen, S. E., Stengaard-Pedersen, K., and Frich, L. H. (2004). Work related shoulder disorders: quantitative exposure-response relations with reference to arm posture. *Occupational and Environmental Medicine, 61 (10)*, 844-853.

Tabachnick, B. G., and Fidel, L. S. (2001). *Using Multivariate Statistics*. 4th ed. Allyn and Bacon.

Tabachnick, B. G., and Fidel, L. S. (2006). *Using Multivariate Statistics*. 5th ed. Allyn and Bacon

The Academy of Medical Sciences (2006). *Personal data for public good: using health information in medical research*. London: The Academy of Medical Sciences.

Valuation Office Agency (2006). *Council Tax Bands* [online]. Available from: http://www.voa.gov.uk/council_tax/index.htm [Accessed 3 June 2008].

Van Doorslaer, E., and Jones, A. M. (2003). Inequalities in self-reported health: validation of a new approach to measurement. *Journal of Health Economics, 22 (1)*, 61-87.

Ward, L. S. (2003). Race as a variable in cross-cultural research. *Nursing Outlook, 51 (4)*, 120-125.

Washington State Department of Health (2002a). *Major Risk and Protective Factors* [online]. Available from: http://www.doh.wa.gov/HWS/doc/RPF/RPF.doc [Accessed 3 June 2008].

Washington State Department of Health (2002b). *Social Determinants of Health* [online]. Available from: http://www.doh.wa.gov/HWS/doc/RPF/RPF_Soc.doc [Accessed 12 March 2008].

Washington State Department of Health (2004). *Health of Washington State*. Washington: Olympia, WA.

White, C., Wiggins, R. D., Blane, D., Whitworth, A., and Glickman, M. (2006). Person, place or time? The effect of individual circumstances, area and changes over time on mortality in men, 1995-2001. *Health Statistics Quarterly, 28*, 18-26.

WHO (1998). Health Promotion Glossary. World Health Organization.

WHO (2001). Health promotion. World Health Organization.

Wilkinson, R.G. (1999) Putting the picture together: prosperity, redistribution, health and welfare. In: R.G. Wilkinson and M. Marmot, eds. *Social Determinants of Health*. Oxford: OUP, 1999, pp. 256-274.