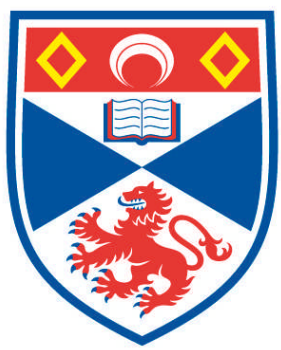
EXPLORING THE INCIDENCE OF LUNG CANCER IN SMALL AREAS ACROSS SCOTLAND

James R. Pearce

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



2003

EXPLORING THE INCIDENCE OF LUNG CANCER IN SMALL AREAS ACROSS SCOTLAND

A thesis submitted to the University of St Andrews for the Degree of Ph.D.

James R. Pearce

School of Geography and Geosciences,
University of St Andrews

August 2003

Declaration

i. I, James Pearce, hereby certify that this thesis, which is approximately 97,500 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

Date 12/8/03... Signature of candidate . 

ii I was admitted as a research student in February 2000 and as a candidate for the degree of Ph.D. in February 2001; the higher study for which this is a record was carried out in the University of St Andrews between 2000 and 2003.

Date 12/8/03... Signature of candidate . 

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

Date 12/8/03... Signature of supervisor . 

A. Unrestricted

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker.

Date 12/8/03.. Signature of candidate .. 

TABLE OF CONTENTS

ABSTRACT

Lung cancer is one of the most important causes of 'avoidable deaths' globally and is responsible for approximately 900,000 deaths per year. However, lung cancer rates tend to be higher for males than for females and the disease also varies geographically, as rates are far higher in developed countries compared to developing countries. Scotland has the highest rate of lung cancer of any country where lung cancer data is available. However, explaining the spatial distribution of this disease is difficult because lung cancer has a number of known causes that operate at a range of different spatial scales. This is further complicated by the lag time between the period of exposure to a risk factor and the date of diagnosis.

This thesis examines the causes of lung cancer across Scotland, using lung cancer registrations for the period 1988 to 1991. Exploratory methods are presented for examining the geographical distribution of the disease in small areas using methods of age-standardisation and cluster detection to identify areas with unusual rates. Estimates of the key risk factors potentially associated with lung cancer are calculated for the same small areas. These include estimates of smoking behaviour, air pollution levels in 1971 and 1991, radon gas potential, coal mining activity, quarrying activity and area deprivation. The risk factors are incorporated into a set of regression models to examine which factors are significant in explaining lung cancer incidence. Finally, the residual values derived from the optimum model of lung cancer incidence in Scotland are examined to identify areas where lung cancer incidence is particularly high and low.

This study revealed that there were marked geographical differences in lung cancer rates, with higher rates in the large urban areas, especially Glasgow, and also the more deprived areas of Scotland. Smoking was consistently significant in explaining lung cancer incidence for all cohorts, types of lung cancer and urban-rural areas. The estimated air pollution levels in 1971 were also found to be significant, but the 1991 estimates were not. Exposure to radon was only significant in explaining lung cancer in the younger age groups. However, the coal mining and quarrying variables did not independently influence the incidence of the disease. Area deprivation and measure of urbanness both had a significant effect on lung cancer incidence in Scotland that was independent of the key risk factors. The analysis of the residual values showed that, having controlled for the key risk factors, lung cancer incidence is higher than expected in rural rather than urban areas.

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

There are a number of people I would like to thank for their advice and support throughout the writing of this thesis.

I would firstly like to thank my supervisor, Paul Boyle, who has been a constant source of ideas and encouragement. Thanks also to Robin Flowerdew who helped me so much both in Lancaster and now here in St Andrews.

I would like to acknowledge the help of Louise Findlayson at the Scottish Household Suvery, Lesley Bhatti at the Information and Statistics Division of NHS Scotland and Graham Law at the Leukaemia Research Fund for their assistance with accessing various datasets.

Thanks must also go to a number of people who have offered very useful advice on different aspects of the research, particularly Liz Twigg at Portsmouth, Andy Jones at UEA, Christine Dunn at Durham and Ed Stephens here in St Andrews.

The technical help of Daniel Exeter, Matt Cox, Clive Sabel and Zhiquang Feng has been very much appreciated. Zhiquang, in particular, has shown endless patience in helping me with a range of technical issues and his skills in computer programming are to be marvelled at. Thanks also to Pig, Matt and Clive for their assistance with proof reading.

Finally, I would like to thank a very special group of friends who have helped to make my time in St Andrews so much fun and I will never quite recover from the shock of my surprise leaving party. Therefore, thank you especially to Pig, Matt, Alix, Kim, Wil, Poby, Jacks, Andrew, Emilie and Gavin - I'll miss you all.

1. INTRODUCTION

1.1 Background

Lung cancer is the most common form of cancer both in terms of incidence and mortality, accounting for one in seven new cases. In the year 2000, there were approximately 1.25 million cases of lung cancer and of these 900,000 were male and 350,000 female. It is also the most deadly cancer, responsible for 900,000 deaths globally each year (Murray and Lopez, 1997) and the rate is increasing by approximately 0.5 percent per year (Haugen, 2000).

However, rates of lung cancer incidence vary between and within different countries. For example, there are far higher rates in developed countries compared to less developed countries. Using a world standard population, it has been estimated that the more developed countries have an age-standardised rate of lung cancer of 55.6 for men and 15.6 for women. In less developed countries, the age-standardised rate is 24.8 for men and 8.4 for women (GLOBOCAN, 2000). This demonstrates that lung cancer is more prevalent in high income compared to low income countries and also that the rates of lung cancer are consistently higher for men than they are for women.

There are also differences at smaller scales such as between seemingly similar countries, regions of a country or between small areas within cities. For example, the rate of lung cancer in Sweden is twice that of Denmark for males and females (GLOBOCAN, 2000). Similarly, the lung cancer rates within the UK show distinct

geographical differences. Using the European standard population, the age-standardised lung cancer rate in the UK is 90.2 for males and 34.9 for females[1]. However, this disguises some important variations within regions of the UK. For example, the ASR for the East of England is 74.3 for males and 26.8 for females, whereas in Wales it is 112.8 for males and 49.5 for females. Furthermore, the age-standardised rates of lung cancer in Scotland demonstrate that it has a particularly high rate of lung cancer (123.5 for males and 67.5 for females) compared to most other countries (Babb *et al.*, 2001). Importantly, Scotland has the tenth highest rate of lung cancer for males but the highest for females compared to all other countries in the world (WHO, 2002).

The geographical differences in lung cancer rates provide some important clues about the causes of the disease. Although smoking has long been accepted as a cause of lung cancer (Doll, 1950), there remains considerable debate about the relative importance of other factors. Smoking is thought to account for approximately 85 percent of lung cancer cases (Schottenfeld, 1996) and the increase in lung cancer incidence at a global scale is largely a reflection of increased smoking rates. Furthermore, the geographical differences in lung cancer rates that have been highlighted are largely a reflection of variations in smoking behaviour. This has resulted in decreasing rates of lung cancer for men and increasing rates for women in the developed world (GLOBOCAN, 2000). However, a number of studies have suggested that many other risk factors are important in explaining geographical variations in lung cancer, although these are less well understood. Most noticeably, a number of studies have examined the role of air pollution (Pless-Mulloli *et al.*,

[1] The age-standardised rates for the UK cannot be compared to the rates in other counties as they use different standard populations.

1998) and radon levels (Darby *et al.*, 1999). However, there has been work on other potential causes such as passive smoking, nutrition, previous lung disease, genetic predisposition and a number of occupational hazards, especially asbestos. The evidence of a link between lung cancer and each of the potential risk factors, with the exception of smoking and asbestos, remains inconclusive. Therefore, given that lung cancer is a major cause of death globally, it is important to examine the disease in order to gain a better understanding of its aetiology.

This study focuses upon lung cancer in Scotland, a country with particularly high lung cancer rates, and attempts to examine the importance of a number of potential risk factors in small areas throughout the country. This is achieved by comparing the incidence of lung cancer in small geographical areas (1991 census output-areas) with estimates of the key risk factors for the same areas. This will enable localised variations in these risk factors to be examined which may help to explain variations in the disease and provide new insights into the causes of lung cancer. This is the first time that the incidence of lung cancer has been examined in such small geographical areas.

This is a quantitative study of the geographical variations in lung cancer incidence in Scotland. Quantitative methods allowed the lung cancer rates in small areas to be directly compared, the exposure to different risk factors to be estimated and an exploration of the extent to which the risk factors explained variations in lung cancer incidence to be undertaken. However, it is important to appreciate that there are alternative philosophical standpoints and hence methodological approaches that can be taken to research lung cancer incidence in Scotland. For example, a qualitative

approach could provide different insights into the nature of lung cancer such as the perception of cancer care and delivery, or contextualise the multiple explanations of smoking behaviour. It is therefore important to consider the philosophical standpoint taken here and to appreciate that there may be many alternative approaches to examining the geography of lung cancer.

1.2 Geographies of Health

There is a wide literature on the philosophical developments in Geography and it is important to be aware of and reflect upon the philosophical approach taken (Graham, 1997). Even within the Geography of Health subdiscipline there are multiple interpretations of, and possible explanations for, particular health issues in geographical settings (Gatrell, 2002). The following discussion does not intend to provide a full discussion of the developments of philosophical approaches but rather to consider the key theoretical frameworks that have been employed by health geographers. This is not to suggest that each piece of research should be pigeon holed into one stance or another but rather to provide a framework for reflection on the various approaches taken.

A number of authors have attempted to identify a list of the key approaches taken by health geographers (e.g. Aggleton, 1990; Litva and Eyles, 1995; Curtis and Taket, 1996; Gatrell, 2002). There is some agreement between the suggested classifications and although the terminology may differ between authors they can generally be thought of as: biomedical; social interactionist; structuralist; structurationist and post-structuralist. Although these do represent a loose chronological development,

this list should not be interpreted as one approach superseding another, leading to the rejection of a longer established approach, nor should it imply superiority of one strand over another, but rather the development and diversification of the subdiscipline. For example, Kearns (1994) states that the relatively recent changes in Medical Geography should not be interpreted as an attempt to discard the important contributions of the spatial analytical tradition. Instead, some of these approaches reflect interests that have widened beyond the more traditional empirical work to incorporate the wider social and cultural sphere. This widening agenda is to some extent a response to some of the criticisms made by some authors (e.g. Kearns, 1993; Dorns and Laws, 1994) about the previously dominant positivist paradigm.

The biomedical perspective, guided by the positivist principle, is based on scientific rationality, an emphasis on objective, numerical measurements and on physical and chemical data (Bowling, 1997). It views the body as a machine whereby disease is generated by specific aetiological agents that lead to changes in the body's structure and form. It considers nature as a piece of clockwork that is governed by exact mathematical rules and disease as a temporary or permanent impairment in the functioning of any single component, or the relationship between the components making up the individual (Polgar, 1968). The biomedical approach therefore tends to consider there to be a risk factor or malfunction that causes a disease, usually by a specific biological process (Jones and Moon, 1987).

Geographical studies traditionally adopted the philosophy of the biomedical model under the broad framework of Medical Geography. Such studies have attempted to understand the geographical variations of diseases using empirical evidence in a

largely statistical way (Learmonth, 1988). Medical Geography retains important associations with other non-geographical disciplines, particularly with research in the medical profession and more specifically with the field of epidemiology. The approach is predominantly statistical in its methods and the conclusions are usually expressed in terms of probabilities. Epidemiological research uses quantitative measures and techniques to help to explore and understand the relationship between a disease of interest and potential causal factors, and its primary objective is to identify preventable (avoidable) causes of diseases (Moon *et al.*, 1997). Geographical examples have tended to focus more on the spatial nature and distribution of the disease of interest (Litva and Eyles, 1995) and many studies have demonstrated that the spatial properties of a disease have often proven to be highly important in understanding its aetiology.

Biomedical approaches have been used by health geographers to consider a range of health issues. Geographers have studied particular diseases such as influenza (Cliff *et al.*, 1986) or AIDS (e.g. Gould and Wallace, 1994), issues such as links between deprivation and health (e.g. Dorling, 1997) or health service delivery (e.g. Haynes, 1987; Powell, 1990; Legetic *et al.*, 1996). Taking the example of cancer, different quantitative approaches have been taken in order to try and reveal new insights into the aetiology of the disease and also to help plan and manage cancer service provision. The spatial pattern of cancer incidence has been related to particular environmental factors such as high powered radio stations (e.g. Michelozzi, *et al.*, 2002), pesticides (e.g. Meinert *et al.*, 2000) or ultraviolet radiation (e.g. Tenkate, 1998). Some authors have developed methods of cluster detection to identify localised patterns of cancer around a particular location that is considered to pose a

possible threat such as a nuclear reprocessing plant (e.g. Openshaw *et al.*, 1987), or within a region of interest to help decide where to target epidemiological work (e.g. Schneider *et al.*, 1993; Rigby and Gatrell, 2000). Alternatively, the framework is used to help decision makers decide how and where to supply services, or to help them understand how patients utilise health services (Curtis and Taket, 1996). This includes work that examines where the optimal location to place a cancer unit is or evaluate the potential effects of different alternatives (e.g. Smallman-Raynor *et al.*, 1998). The example of cancer studies starts to reveal the diversity of methods that this approach encompasses, but this discussion reflects just a fraction of the quantitative methods used and research that has been undertaken by geographers which examine not only a range of diseases (e.g. cancer, heart disease, tuberculosis) but also many types of health behaviour (e.g. smoking) and health outcomes (e.g. limiting long term illness).

Critics have argued that the highly quantitative nature of this geographical research has resulted in techniques being used in circumstances for which they may not be suitable or has ignored other important elements that may be influential in understanding the disease in question (e.g. Kearns, 1993). For example, Helman (1990) points towards the need to consider the cultural factors that influence disease. In order to understand why a particular individual gets a particular disease at a particular time, one has to consider a wider range of factors such as sociocultural and psychological ones and the interrelationships between them. Examples include economic situation, family structure, gender roles, marriage patterns or population policy. This multifactorial explanation of ill health can, in some circumstances, be

more useful than postulating a simple cause and effect relationship between one type of risk factor and one type of disease.

Other researchers have emphasised factors that are less readily measured and quantified, such as the more subjective experiences of health. Social interactionist approaches are so called because meanings are constructed out of interactions that people have with each other in day-to-day life such as conversations and encounters (Aggleton, 1990). In health geography these perspectives have often been referred to as humanistic since they address implicitly human beliefs, values, meanings and intentions (Curtis and Taket, 1996). Such approaches concern themselves with examining human awareness, agency and creativity which, for example, may include researching beliefs about health or health related behaviour. They may, of course, differ in the particular type of social theory they draw on. Humanist writers have focused on the agency of individual actors and observed that through the situated nature of their behaviour, particular experiences of place are built up (Buttimer, 1980). Examples of this approach include anthropological and ethnographic studies of the experiences of health services in Bethnal Green, East London (Cornwell, 1984), in-depth interviews about the experiences of local residents near waste disposal sites in southern Ontario (Eyles *et al.*, 1993), or take a 'humanistic understanding' of place in order to examine the way in which health care in an isolated part of New Zealand contributes to a sense of place (Kearns, 1991).

The structural, material and critical turns are informed by a variety of different perspectives on the organisation of social life, sharing perhaps a common concern with highlighting the effects of broad social forces. Structuralist approaches do not

seek explanation at the individual level but rather suggest that the underlying causes of disease are embedded in political and economic systems (Gatrell, 2002). Research that has been informed by structuralist theory suggest that within regions, social and political structures operate to constrain opportunities and shape the experience of residents (Kearns, 1993). Much of the grounding of structuralist theories is derived from Marxist theories of oppression, domination, and class conflict where inequalities are embedded in society. From a health geographer's perspective, economic relations and structures that underpin health and access to health care and hence human agency is absent from classically structuralist accounts. Structuralist approaches include explicitly Marxist interpretations as well as other forms of structure which are embedded in society that are based on conflict and power. This includes the role played by male power (patriarchy) in structuring women's health or the relationship between those owning the means of production and those employed in the labouring classes (Gatrell, 2002). Examples of structuralist approaches include studies that examine the effects of colonisation on the health of the indigenous population (e.g. Arnold, 1988; Watts, 1997) or studies of the fear of rape which represents one manifestation of patriarchy (e.g. Pawson and Banks, 1993; Madge, 1997). The humanistic and structural critiques have challenged traditional spatial analysis and have helped to refocus health research to incorporate the social context of health and disease (Eyles and Woods, 1983).

Structuration theory seeks to provide a middle ground between social interactionist (humanist) approaches that focus upon the individual and human agency, and structuralist approaches that invoke the broader social, economic and political structures. Therefore, structuration theory recognises the duality of structure and

agency whereby structures shape social practices and actions whilst at the same time practices and actions can transform social structures. Examples of structurationist approaches taken by health geographers would be studies that have examined how both human agency and social structure shape women's health behaviour (Young, 1996) or explored the social processes involved in 'deinstitutionalising' the mentally ill (Dear and Wolch, 1987). In the first example, women's own health care decisions can become constrained by the economic structure (division of the labour market) and the caring responsibilities in the household. In the second example, the reciprocal relationship between structure and agency is manifested in several ways. For example, the macro level political and economic influences, which are evidenced in specific state policies, will influence processes such as the gentrification of urban areas which limits the housing options for the formerly institutionalised. At the same time the forces of skilled and knowledgeable actors involved in the social welfare system can help to limit or enable the actions of the mentally ill.

Most recently, some health researchers have begun to engage with what have been labelled post-structuralist theories or what other social scientists refer to as postmodernist theories. These perspectives are concerned with how knowledge and experience are constructed in the context of power relations (Gatrell, 2002), criticising the assumptions on which much of public health research is based. Much post-structuralist work calls upon the work of Foucault which suggests that control or power is not so much exercised by repression but rather through 'expert knowledges' about human beings and societies, which serve to channel or constrain thinking and action' (Peterson and Lupton, 1996). One way in which this is manifested is through subtle forms of 'surveillance' that have seen changes in

legislation that control our health behaviour (for example wearing of seat belts or controls of smoking in public places: Poland, 1998) or the direct 'gaze' of the medical professional on the individual's body to survey the health landscape of the population (Parr, 2002). Kearns (1993) discusses how the location of birth (home or hospital) can be seen as a contested terrain, the principal protagonist being women (the birthers) and medical men (the birth controllers). Secondly, there is an emphasis on difference or 'otherness' and how, for example, the diseased are:

> "defined, disempowered and controlled through metamorphic associations of place and affliction, inscriptions of contagious space, and the restructuring of purportedly diseased environments" (Craddock, 1995, p. 958).

Examples include how the Chinese in San Francisco were made the scapegoats for a smallpox outbreak in the late nineteenth century, with the Chinatown area of the city becoming a metaphor for disease and the Chinese becoming the 'other' (Craddock, 1995).

It is the appreciation of the wider social, cultural, economic and behavioural factors that has led to a relabelling of traditional Medical Geography as the Geography of Health (Kearns, 1993). Kearns and Gesler (1998) suggested that this is much more than a matter of semantics but reflects a shift in norms of content, epistemology, and research practices. This has probably been a reflection of the more critical view that has been taken of the biomedical and positivist views of health that had dominated health research in many disciplines including geography (Curtis and Taket, 1996).

The shift also reflects the developing concerns of mainstream geography, and most particularly, the theoretical developments in the social theory used within geography. According to Brown and Duncan (2000, p. 363) this has '(re)connected medical geography to developments happening elsewhere within the geographical corpus'. Kearns and Gesler (1998) considered the implications of this name change to be three-fold. Firstly, it more faithfully names the interests of a community of scholars. Secondly, it symbolically and practically places geographers closer to social scientists, policy analysts, and planners who evaluate population health and the adequacy of services. Thirdly, and most fundamentally, it legitimises geographers' shift beyond the driving metaphors of medicine and disease to embrace issues generated by emerging models of health and disease.

It is worth reiterating that this classification of health geography research is not intended to provide a tight framework for pigeon holing this work. As Philo (1996) pointed out:

> "Any attempt to categorise the theoretical approaches taken in medical geography is surely doomed to be flawed and partial, to illuminate some aspects of the intellectual landscape whilst obscuring others, and in so doing so be just one possible way of telling the story among many" (Philo, 1996, p. 36).

Rather, this discussion is intended to reveal the range and diversity of approaches that are taken whilst making it clear that there are multiple interpretations and explanations of health. Each of the approaches has much to offer our understanding

of health, whether it be the examination of the effects of place and space using epidemiological methods, or a structuralist interpretation of the impact of air pollution that considers the world industrial economic system as the root cause of the problem (Gatrell, 2002). In short, there are many philosophical approaches that geographers can take to examine health issues but it seems clear that there is no wrong or right approach but rather a set of options where the implications of using the selected standpoint require careful consideration.

Whilst it is tempting to conclude that the developments that grew out of criticisms of quantitative methods, which leant on the biomedical philosophies, have set up a duality between the 'old' and the 'new' medical / health geography, this would be a misrepresentation of the developments in this field. Rather, they should be interpreted as complementary rather than competitive (Curtis and Taket, 1996). Although there have been a number of welcome developments in the geography of health, many authors have suggested that it is important to maintain the more traditional approaches alongside the so-called 'new health geographies' to preserve its position within mainstream geography and within the wider health research field (Kearns and Moon, 2002). Quantitative methods have much to offer to the key research themes in health geography that have been identified in recent progress reports and introductory commentaries on major conferences (Kearns 1995; Kearns 1996; Kearns 1997; Moon *et al.*, 1998; Cummins and Milligan, 2000; Earickson, 2000; Kearns and Moon, 2002). Methods such as generalised linear modelling (particularly multilevel modelling) and GIS and spatial analysis have much to offer studies examining issues such as health inequalities (e.g. Boyle *et al.*, 1999), epidemic modelling (e.g. Smallman-Raynor and Cliff, 1999), and the relationship

between health and the environment (e.g. Sabel *et al.*, 2000). Multilevel modelling is another quantitative approach which allows the notion of hierarchies and the nesting of people within places to examine variations at different levels and hence is useful for considering the effect of 'place' (Jones, 1991). These approaches have provided theoretical insights into a range of issues that could not be achieved using alternative guiding theoretical principles. It therefore seems clear that quantitative methods are appropriate for many, but certainly not all, studies of health. Therefore, health geographers should not necessarily be distancing themselves from quantitative methods, but rather, should acknowledge the importance of alternative philosophies in the contemporary geography of health.

This discussion has highlighted the important themes that contribute to a new contemporary health geography. In the specific context of this research, many of these philosophical standpoints are appropriate for studying particular components of lung cancer research. For example, qualitative methods can be used to address issues such as the gendered patterns in smoking (Jacobson, 1981), the effects of community disadvantage on smoking behaviour (Stead *et al.*, 2001) or the perceptions of asbestos risk among migrant workers (Capelletto and Merler, forthcoming). However, the broad aim of this research is to examine local area variations in lung cancer incidence in Scotland in order to gain insights into the risk factors that influence incidence of the disease. This aim is best met using a quantitative approach because methods can be developed to directly compare the incidence of lung cancer between small areas in Scotland and also to estimate the exposure to the key risk factors that operate in Scotland.

This approach has advantages from a public health perspective, because it allows areas to be identified that have a high rate of lung cancer compared to the rest of Scotland. This is a key component of the health promotion plan in Scotland, which aims to improve the health of the population in the country. The Health Education Board for Scotland is responsible for the implementation of a health education program that is responsive to the needs and priorities of the population and acknowledges variations in health status between sub-groups of the population (HEBS, 2003). More specifically, the Board aims to enhance positive health, comprising of well-being and fitness, while at the same time preventing ill-health and tackling health inequalities (HEBS, 2003). An important part of this strategy is to target vulnerable groups of the population, such as the young or the old, or specific areas that are susceptible to particular types of ill health. This is in line with a recent UK Government Green Paper that detailed a strategy to improve the health of the worst off in society and to narrow the health gap (Department of Health, 1998a). In the context of this study, if areas can be identified that have a particularly high incidence of lung cancer, once the age and sex structure of the population are accounted for, then resources can be channelled into public health campaigns in these areas to help to reduce its prevalence.

1.3 Aims of Thesis

A range of quantitative methods are used to examine the geography of lung cancer incidence in Scotland in this thesis. The aims of this research are to:

1. Use exploratory quantitative techniques to examine the geographical distribution of lung cancer in Scotland at a range of scales. Comparisons of lung cancer incidence between areas are made by age-standardising and then mapping the data. This has useful public health implications because it allows for areas with high rates of lung cancer to be identified and targeted.
2. Identify whether there are certain areas that have significant clusters of cases of lung cancer in Scotland.
3. Estimate the probability of smoking for different age and sex groups in small areas across Scotland using socio-economic data at different scales that are thought to influence smoking behaviour. These estimates allow for the relationship between smoking and lung cancer to be examined and then controlled for as other causes of lung cancer are investigated.
4. Develop methods to estimate pollution levels in small areas across Scotland for 1971 and 1991. The two estimates are used to examine the effects of lagged and non-lagged variables in the analysis of lung cancer.
5. Develop methods to estimate exposure to the geological causes of lung cancer for small areas in Scotland, including the effects of radon levels, coal mining and quarrying activity. Each factor has been suggested as a potential cause of lung cancer in distinct geographical areas. However, these factors are poorly understood, partially due to the difficulties in measuring exposure to them.
6. Examine how the potential risk factors described above influence the incidence of lung cancer in Scotland. This is achieved by incorporating all of the potential risk factors into a series of regression models with the count of cases of lung cancer as the dependent variable, and investigating which factors are significant in explaining lung cancer incidence.

7. Identify areas of particularly high and low incidence of lung cancer once the established risk factors have been controlled for. This is achieved by calculating residual values from the regression models and then mapping them for different age and sex groups. This will provide the information necessary to identify neighbourhoods with a higher than expected incidence of lung cancer for spatial targeting by health care planners.

1.4 Data Considerations

In order to fully understand how the risk factors associated with lung cancer influence the incidence of the disease it would require individual-level data on lung cancer patients that included detailed historical information such as details of their smoking habits and exposure to air pollution and radon gas. However, individual data are limited because they do not usually include a full residential and lifestyle history of the patient and therefore it is not possible to be precise about the level of exposure to particular risk factors. Furthermore, many individual-level data sets are often limited by issues of confidentiality that prohibit data on individuals being released (especially medical data) and instead aggregated data are released for larger administrative areas. In addition individual-level data that are compiled for a specific study are expensive to collect and would still suffer from recall bias problems (Moon *et al.*, 2000).

An alternative approach is to explore the incidence of lung cancer in small geographical areas where information on risk factors can be calculated and associated with the records of the lung cancer patients living there. By considering

lung cancer in small areas the analysis is sensitive to localised variations in the risk factors which may explain important variations in incidence. This study uses individual-level data supplied by the Information and Statistics Division of NHS Scotland on patients who were diagnosed with lung cancer in Scotland between 1988 and 1991. Each record identifies the age and sex of the individual, the type of lung cancer that they were diagnosed with and the census output-area (OA) in which they lived when they were diagnosed. No additional information such as data on the individual's smoking behaviour or socio-economic information about the areas in which they lived was collected. However, the OA identifier allows for ecological estimates of the key risk factors to be associated with the individual-level lung cancer records. These data are used to explore the geographical distribution of lung cancer cases and examine the extent to which the important risk factors associated with lung cancer explain the spatial variations of the disease.

Two complications arise when using ecological estimates of risk factors that are potentially associated with lung cancer. Firstly, the effect of the lag period between the exposure to risk factors and the diagnosis with lung cancer and secondly, the effects of migration during this latency period. A number of studies have suggested that studies of lung cancer incidence are complicated by the influence of the lag period that has been identified between the exposure to these risk factors and the initiation of lung cancer. This arises because exposure to the potential risk factors among lung cancer patients can often precede the symptoms and diagnosis of the disease (Williams, 1992). In the case of lung cancer, a lag period of approximately 20 years is generally agreed (Williams, 1992) although it is possible that a period of up to 40 years may be relevant (Williams and Lloyd, 1991). Therefore, measures of

the risk factors that coincide temporally with the diagnosis of lung cancer data may be inappropriate if the risk factor has changed substantially during the latency period. This problem is not only relevant to studies of lung cancer but also to many cross-sectional studies of ill health such as studies of limiting long term illness (Bentham *et al.*, 1995; Gould and Jones, 1996), mortality (Dorling, 1997; Senior *et al.*, 2000), morbidity (Boyle *et al.*, 1999; Saul and Payne, 1999) and particular illnesses such as coronary heart disease (Shewry *et al.*, 1992). These, and other studies, have tended to explore the relationship between a particular type of ill health and specific explanatory variables such as occupation type or deprivation, without considering the potential lag effects between the exposure to the risk and the diagnosis of the illness. Similarly, many studies that have considered the risk factors associated with lung cancer have not incorporated a lag effect into the analysis. For example, studies have examined the relationship between lung cancer and smoking (Colby *et al.*, 1994), occupation type (Boffetta *et al.*, 1992), socioeconomic status (Vacchino, 1999) and passive smoking (Jee *et al.*, 1999) without addressing the latency issue.

Given the importance of the lag effect in studies of lung cancer, it may seem sensible to use estimates of the risk factors from 20 years ago. However, ecological studies are further complicated by the effect of migration, because an individual living in an OA who is diagnosed with lung cancer, did not necessarily live in the same OA 20 years previously. Therefore, if the residential address of the lung cancer patient changes between the critical period of exposure to the risk factor(s) and the time at which they were diagnosed with the disease then the measure of exposure will be inaccurate. Therefore, the importance of using lagged variables depends upon the

absolute and geographical changes in the risk factor during the period between exposure and diagnosis. If there has been a large increase or reduction in the risk factor, or a change in its geographical extent during the latency period, then it may be important to use lagged variables, despite the methodological difficulties posed by migration. However, if the risk factor has changed only slightly during the lag period then it may be more sensible to use contemporary measures of exposure. Consequently, the explanatory variable either incorporates the lag effect but ignores the migration effect or alternatively it represents contemporary measures of the risk factor but ignores the possible lag effect. Therefore, because the implications of this decision are specific to each risk factor, the importance of the lag effect will be discussed in relation to each factor in subsequent chapters.

1.5 Structure of Thesis

The thesis begins with a brief definition of cancer and a review of the causes of lung cancer that have been explored in the literature. In particular, there is a discussion on the debates surrounding each risk factor in order to establish the importance of each one.

Chapter three demonstrates that it is important to appreciate the scale at which analysis is being undertaken by examining the geography of lung cancer at different geographical levels. The chapter begins with a discussion of the global rate of lung cancer but then focuses on the international variations in lung cancer, lung cancer in the UK and lung cancer in Scotland using existing geographical studies of the disease and national and international data.

A mixture of methods is used in chapter four to examine the geography of lung cancer in Scotland. The aim is to explore whether lung cancer rates are consistent across Scotland using a range of exploratory techniques at different scales. The distribution of lung cancer is compared using Poisson probabilities and age-standardised rates. These results have potential public health implications because they identify areas with a high incidence of lung cancer compared to the rest of Scotland. Methods of cluster detection are also used to test the spatial dependence of lung cancer in order to explore whether significant clusters of the disease exist.

Chapters five, six and seven present methods and results for estimating exposure to three sets of risk factors: smoking, air pollution and geology. For each risk factor, estimates of exposure are calculated for small areas across Scotland and then mapped and evaluated. The smoking probabilities are calculated using multilevel modelling techniques to examine the factors that explain lung cancer at different geographical scales. Individual-level smoking behaviour data from the Scottish Household Survey, that are distributed in a subset of OAs in Scotland, are modelled using areal census data. The parameters from these models are then used to calculate smoking probabilities for different age and sex groups for all OAs in Scotland. Pollution estimates are calculated using data on the types of industry in which workers in the area are employed and on the road network to model air pollution measurements that were collected in 1971 and 1991. Methods of exposure to radon are estimated for geological units in Scotland and used to estimate exposure in all OAs. The exposure to coal mining activity is estimated using employment data from the 1971 census. Finally, exposure to different types of quarries is estimated using a type of spatial

statistic called kernel estimation, using point data on individual quarry locations and types in Scotland.

Chapter eight uses Poisson regression as a framework for drawing together the estimates of the risk factors to evaluate whether they are significant or not in explaining the variation in lung cancer incidence. Different models are used to examine the relationships between the risk factors and the occurrence of lung cancer. The variations between age and sex groups, between urban and rural areas and between lung cancer types are discussed and the residuals from the optimum regression model are used to identify where the incidence of lung cancer is higher or lower than may be expected, having controlled for the key risk factors.

1.6 Conclusion

To conclude, global lung cancer incidence is rising and lung cancer risk is particularly high in Scotland. Furthermore, the potential risk factors in Scotland are poorly understood. There is a need to explore these risk factors, and epidemiological texts would point to undertaking an individual level study to explore individual exposure to different risk factors and ultimately how they influence lung cancer. However, it is not always feasible to obtain individual level data because of restrictions that are placed on data access in order to maintain patient confidentiality. In addition, individual data from medical records do not often contain historical data on the level of exposure to the risk factors experienced by each person. Furthermore, one of the key aims of this research is to explore the local geographical variations in lung cancer, which cannot be examined using individual-level data exclusively.

Therefore, this study uses individual level lung cancer data in association with ecological estimates of potential risk factors to examine the geography of lung cancer for small areas in Scotland. This raises methodological difficulties because the key period of exposure to the risk factors tends to be approximately 20 years prior to the diagnosis of the disease. This points to the employment of variables that incorporate the lag effect. However, the residential histories of the lung cancer patients are not known, which complicates the use of lagged variables. Therefore, subsequent chapters discuss the appropriateness of using lagged variables for examining each of the potential risk factors.

Several philosophical frameworks that are used to research issues of health and ill health have been outlined and it has been suggested that the approach that is used should be suitable for the research question that has been raised. These approaches include the biomedical, social interactionist, structuralist, structurationist and post-structuralist and each one has been shown to be appropriate for addressing particular research questions. Some authors have suggested that this has led to a duality in the subdiscipline, with traditional biomedical approaches on the one side and the new developments that claim to consider the wider social context of health on the other. However, this chapter has argued that quantitative methods are important for addressing many of the new research themes that health geographers have placed on the research agenda, and that the new avenues of research should not be interpreted as a paradigm shift but rather a diversification of this agenda. This is by no means an attempt to denigrate more qualitative approaches but rather to suggest that quantitative approaches are appropriate for addressing the aims of this research.

2. BACKGROUND TO LUNG CANCER

2.1 Introduction

Cancer is a condition that can be traced back as far as the early Egyptians (King, 1996). However, the type and number of cancers have evolved dramatically in response to two modern factors. First, populations now live longer and as cancer is a disease that is more common in old age, it is of little surprise to find that the incidence of cancer has risen. Second, lifestyles and patterns of consumption have altered greatly, resulting in different diets and exposure to new and different risk factors such as pollution, radiation and infections that are likely to affect all types of health outcomes, including cancer incidence.

It is estimated that in the year 2000 approximately 10 million people had some form of cancer. Of these approximately 5.3 million were male and 4.7 million female (GLOBOCAN, 2000). Each year, five million people die of cancer of which approximately 50 percent of these deaths occur in industrialised countries, despite the fact that they have just 25 percent of the world's population (Murray and Lopez, 1997). The aim of this chapter is to briefly define cancer before focusing more specifically on lung cancer. The majority of this chapter will review the aetiology of lung cancer and the biological risk factors that are associated with it.

2.2 Defining Cancer

A cancer cell can be defined as a cell that has escaped the normal controls regulating its growth and division (Abercrombie *et al.*, 1992). This produces a clone of dividing daughter cells that invade adjacent tissues and may interfere with their activities. This is a process known as metastasis (King, 1996). In normal development and growth, there is a very precise mechanism that allows individual organs to reach a specific size. If the tissue becomes damaged then the surviving cells in the organ begin to grow and replace the damaged cells. In cancer cells this balance is lost and the abnormal cells replace and destroy healthy tissue resulting in a lump or mass called a tumour which can be classified as benign or malignant. Benign tumours do not grow and spread and they are not usually life threatening. Malignant tumours are cancerous and are made of cells that divide without control or order which enables them to invade and damage nearby tissue and are potentially far more dangerous (American Cancer Society, 2001).

Carcinogenesis is the process by which cancers are generated and it is a multi step process that initiates from a single cell or clonal origin (King, 1996). This cell then multiplies and acquires additional changes that give it a survival advantage over its neighbours and these altered cells are then amplified to generate the billions of cells that constitute a cancer. A cancer producing agent (carcinogen) is necessary to initiate the process but this does not lead to the immediate production of a tumour (Franks, 1997). This first step is very rapid but the initiated cells may persist for a considerable time, perhaps the lifespan of the individual, remaining latent until acted upon by the promoting agents which may induce tumour development. The latency

period between initiation and the appearance of the tumour is one of the least understood processes of cancer development (Williams, 1992). Following the carcinogenic event metastasis can occur which is the ability to spread from the site of origin to distant tissues which can eventually lead to the development of a malignant tumour.

Key to this study is the fact that carcinogenesis can be initiated by numerous factors. Different agents have different carcinogenic effects on different parts of the body. Therefore the causes of different cancers can vary widely and often there may not be a single causal factor. It is these factors that it is important to isolate and to understand in order that their threat can be reduced and both biological cancer research and epidemiological studies have been used for this purpose.

2.3 Introduction to Lung Cancer

At the beginning of the twentieth century, lung cancer was rare but it is now the most common cancer in the world and the epidemic is still continuing (Haugen, 2000). This development reflects both demographic and lifestyle changes during this period. This section provides an introduction to the key factors that have been suggested as causes of lung cancer. First, the biological properties of lung cancer are described and then the major causes of lung cancer that have been identified in the literature, mainly through epidemiological studies, are considered. The focus here is on the aetiology and physical causes of lung cancer. The geographical variations in lung cancer incidence as well as the social and cultural aspects of the causal factors are considered in subsequent chapters.

Cancer of the lung has proven to be one of the most life threatening forms of cancer due to the difficulty in detecting it early enough. The trachea brings air down into the lungs and it then divides into tubes called bronchi, which divide into smaller branches called bronchioles. At the end of these small branches are tiny air sacs known as alveoli. Most lung cancers start in the lining of the bronchi but they can also begin in other areas such as the trachea, bronchioles or alveoli. The development can take many years and the cancer cells can break away and spread to other parts of the body. This spread can initially take place with few side effects and so much of the damage is often already done by the time of detection (American Cancer Society, 2001).

There are two main types of lung cancer: small cell and non-small cell. Each has its own specific features that affect symptoms and treatment. Approximately 25 percent of all lung cancer cases are of the small cell type, which consists of small round cells that form white fleshy lumps that usually start in larger airways. They grow very rapidly and tend to spread to other parts of the body, especially the liver, at an early stage. There are three types of non-small cell lung cancer that collectively account for the remaining 75 percent of lung cancer cases; squamous carcinoma, adenocarcinoma and large cell carcinoma. Squamous carcinoma is the most common type of lung cancer accounting for 40 percent of all cases of non-small cell lung cancer although new cases have been declining over the past 30 years. New cases of adenocarcinoma have been increasing over the past 30 years and now account for approximately 30 percent of non-small lung cancer. Large cell carcinomas account

for 10 to 15 percent of all non-small cell lung cancers and this rate has been consistent over the past 30 years.

A final type of lung cancer is mesothelioma, which is slightly different because it effects the coverings rather than the lining of the lungs and is usually occupationally linked with high rates among workers who have been exposed to asbestos (CancerBACUP 2001). The rate of mesothelioma is thought to have peaked around the year 2000 with rates expected to drop by about 75 percent over the next 50 years in response to reduced use of raw asbestos and a reduction in workplace airborne asbestos levels (Price, 1997).

2.4 The Causes of Lung Cancer

Although an appreciation of the physical and carcinogenic processes involved in the development of cancer is an important prerequisite to a study of cancer incidence, much of the understanding that has been achieved is a direct result of observational and epidemiological studies. This section reviews the main causes of lung cancer, the relative importance of which continue to be disputed. The focus here is upon the physical causes of lung cancer that have been identified in the literature and subsequent chapters provide a closer examination of the cultural factors that influence behaviour and some methods for examining and modelling these factors in Scotland. The risk factors that are identified here are all causes of each type of lung cancer (small cell and the three types of non-small cell) although there are some subtle differences in how important each cause is to the various types. Most

importantly squamous carcinoma is strongly linked with a history of smoking, which is less true for adenocarcinoma.

This section will use the framework provided by Schottenfeld (1996) who identifies eight key risk factors. These are tobacco, environmental tobacco smoke (passive smoking), air pollution, indoor radon, occupational carcinogens, nutrition, previous lung disease, and genetic predisposition.

2.4.1 Smoking

Smoking has been established as a major cause of many cancers, and especially lung cancer, by a variety of studies (e.g. Doll and Hill, 1950; Shairer and Schöniger, 2001). It is estimated that in the UK in 1995, 120,000 deaths were caused by smoking, that is one in five of all deaths and 40 percent of all cancer deaths (Peto *et al.*, 1994; Department of Health, 1998b). Smoking is estimated to cause 85 percent of lung cancer deaths (Schottenfeld, 1996; Mannino *et al.*, 1998).

Smoking was first identified as a potential cause of lung cancer by two separate studies that took place in the late 1940's. Doll and Hill (1950) undertook a case-control study in which the life histories of several hundred patients with lung cancer were compared with those of several hundred people without the disease. At the time, cigarette smoking was only one of several possible causes that were being investigated. The study found that the only big difference between those who had lung cancer and those who did not was that almost all those with lung cancer had been smokers. In the same year Wynder and Graham (1950) reached similar

conclusions from their study in the USA. These two studies provided the earliest convincing evidence of tobacco being a major cause of lung cancer death.

A number of studies followed up the initial findings of this work. Doll and Hill (1954) began a study in 1951 that still continues, using British Medical Association records to question all doctors in Britain about their smoking habits. Concurrent American studies (Hammond and Horn, 1954) have tended to be undertaken over a shorter time period but included a larger sample size. The publication of these prospective studies demonstrated the strong associations between smoking and death not only from lung cancer (and later from other cancers) but also from respiratory disease, particularly chronic bronchitis, and vascular disease, particularly heart attacks (Doll and Hill, 1964; Doll and Hill, 1966). Tobacco accounted for most deaths from lung cancer and, in addition, it caused an even larger number of deaths from other diseases. The continued importance of this work was shown in follow up work (Doll *et al.*, 1994) which demonstrated that until then the hazards of prolonged smoking had been greatly underestimated. During the second half of the 40 year follow up (1971 to 1991), the death rate from all causes in middle aged smokers was three times that of non smokers. These studies were the first to assess the true hazards of prolonged smoking (Peto, 1994).

Tobacco smoke contains over 3,000 different chemicals and of these it is thought that around 40 have carcinogenic properties (Hoffman and Hoffman, 1997). The sites that become affected as a consequence of smoking are those that come into contact with carcinogens that are inhaled in the smoke and enter the bloodstream through the lungs and are then excreted via the kidneys and the bladder. The

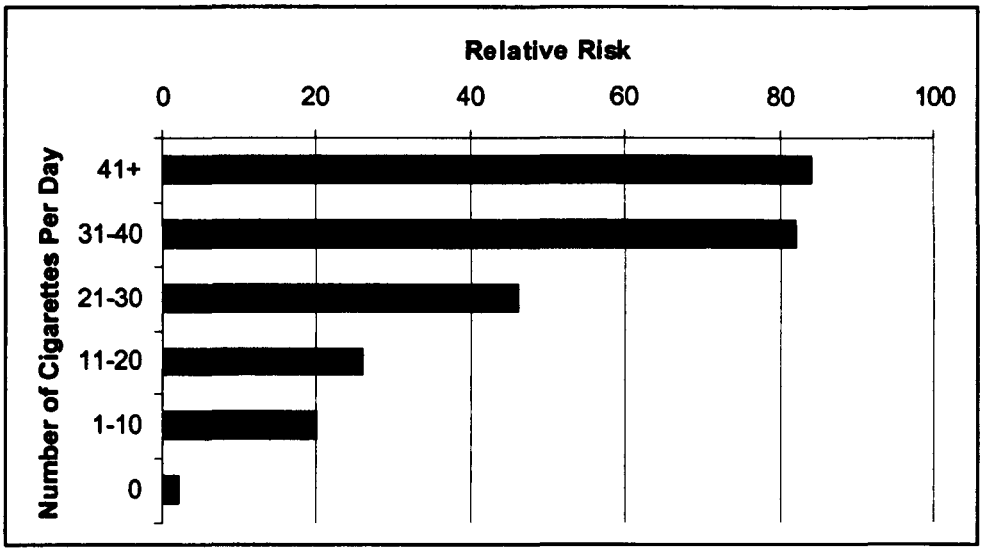


Figure 2.1. Lung cancer risk increase with the number of cigarettes smoked (King, 1996)

carcinogens contained in tobacco smoke include the polynuclear aromatic hydrocarbons (PAHs), nitrosamines, aromatic amines, other organic (e.g. benzene and acrylonitrile) and inorganic (e.g. arsenic, acetaldehyde) compounds and metals such as polonium-210. Of these, PAH compounds and nitrosamines are of major importance to lung cancer induction although many others have the potential to initiate the process of carcinogenesis (Haugen, 2000). The precise composition of the smoke which influences the carcinogenic potential depends on the ambient conditions of smoking, the blend of tobacco leaf, filtration, additives and paper wrapping. The main carcinogenic agents are thought to be in tar contained in the cigarette smoke that settles in the trachea, bronchi or lung. Lower tar content and the use of filters may reduce lung cancer risks in those who smoke. Garfinkel and Stellman (1986) claimed that doubling cigarette tar yield would result in a 40 percent increase in the relative risk of dying of lung cancer. King (1996) demonstrates the dose-response relationship that exists between the number of cigarettes smoked and the increased risk of lung cancer. Figure 2.1 shows that the maximal increased risk

of contracting lung cancer occurs at greater than 30 cigarettes per day but that even less than 10 cigarettes per day results in a 20-fold risk increase.

Rates of smoking and associated smoking-related deaths show important geographical variations at a number of scales. At a global scale there are clear but evolving differences between developed and developing countries. It is estimated that at present one in three adults or one billion people smoke worldwide. Of these, about 80 percent live in low and middle income countries. Partly because of the growth in the adult population, and partly because of the increased consumption, the total number of smokers is expected to reach about 1.6 billion by 2025 (World Bank, 1999). Peto (1994) claims that two million people a year are killed by smoking in developed countries and, if current smoking patterns continue, when the children of today reach middle age the annual number of deaths will have increased from two million to about three million. Of these two million, half die in middle age (35 to 69) and half in old age. Smoking accounts for one sixth of the 11 million adult deaths each year in these populations.

However, in less developed countries the increase will be far larger, from about one million deaths to about seven million around 2025, leading to a world total of about 10 million deaths a year from tobacco (Peto *et al.*, 1992; Peto *et al.*, 1994). The variations between developed and developing countries extend beyond frequencies of deaths but show demographic and gender differences. Liu *et al.* (1998) studied the increasing number of deaths from tobacco in China during the period from 1986 to 1988. It was estimated that there were one million deaths in total during this period. They interviewed the surviving family members of those who died during

this period to establish if they were smokers and estimated that tobacco caused 13 percent of deaths in men but only about three percent of deaths in women. If the present trend continues then this figure will eventually rise to approximately 33 percent for men but fall to approximately one percent for women. This translates into 0.6 million deaths in 1990, at least 0.8 million in 2000 (0.7 million in men) and about three million by the middle of the century. The study also showed that two thirds of men now become smokers before the age of 25, few give up and about half of those who persist will be killed by tobacco in middle or old age. This observation is reiterated in Figure 2.2 which shows the per capita consumption of cigarettes in developed countries slowing and then beginning to fall whilst in developing countries it continues to rise steadily (WHO, 1997).

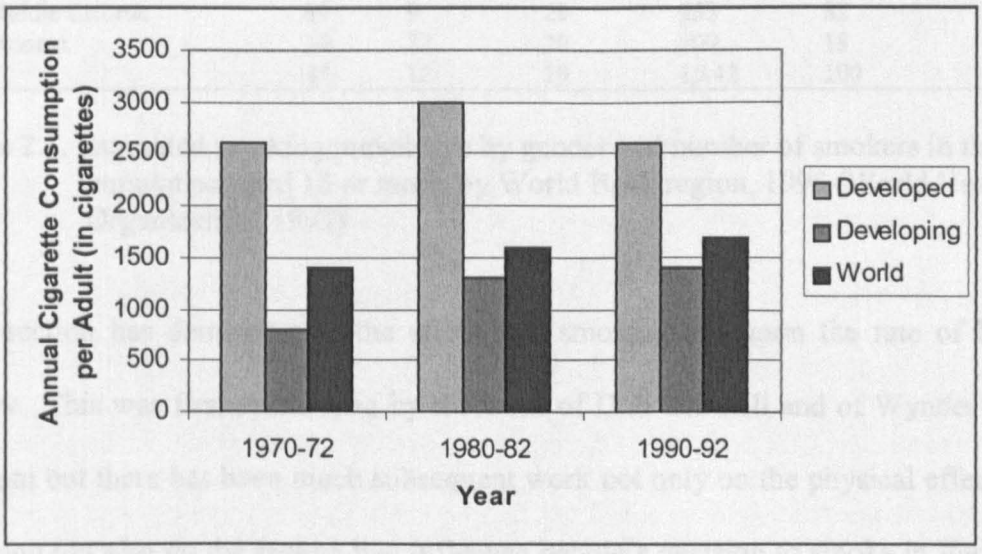


Figure 2.2. Trends in per capita adult cigarette consumption (World Health Organization, 1997)

This pattern is broken down further in Table 2.1 which shows the patterns of smoking in different regions of the world. This table was compiled by the World Health Organisation using more than 80 separate studies and it is clear from this that

there are important regional differences in smoking behaviour. For example, in eastern Europe and central Asia 59 percent of men and 26 percent of women smoked in 1995, more than in any other region whilst in Asia and the Pacific where the percentage of male smokers is equally as high, just 4 percent of women are smokers. This demonstrates the different smoking behaviour between regions of the world and between males and females.

World Bank Region	Smoking prevalence (%)			Total smokers	
	Males	Females	Overall	(millions)	(% of all smokers)
East Asia and Pacific	59	4	32	401	35
Eastern Europe & Central Asia	59	26	41	148	13
Latin America & Caribbean	40	21	30	95	8
Middle East & North Africa	44	5	25	40	3
South Asia (cigarettes)	20	1	11	86	8
South Asia (*bidis*)	20	3	12	96	8
Sub-Saharan Africa	33	10	21	67	6
Low/Middle Income	49	9	29	933	82
High Income	39	22	30	209	18
World	47	12	29	1,142	100

Table 2.1. Estimated smoking prevalence by gender and number of smokers in the population aged 15 or more, by World Bank region, 1995 (World Health Organization, 1997)

This section has demonstrated the effect that smoking has upon the rate of lung cancer. This was first established by the work of Doll and Hill and of Wynder and Graham but there has been much subsequent work not only on the physical effect of smoking but also on the factors that influence people's decision to smoke in light of the well established health risks. Although smoking is the single most important causal factor of lung cancer, research has also shown that the effect of smoking is not confined to the smoker. Other studies have considered the effect that smoking has on non-smokers who breathe in the smoke.

2.4.2 Passive Smoking

A second factor that is identified in the literature as being a potential cause of lung cancer is passive or involuntary smoking. Although this certainly causes fewer cases of lung cancer than smoking, it is much more controversial because its effects are less well understood. Passive smoking can be defined as the consumption of smoke that is released through the burning of tobacco between puffs and from the exhaled smoke of the smoker (Schottenfeld, 1996). There is a wide literature considering a large number of health effects from passive smoking (e.g. Lam *et al.*, 1998; Jousilahti *et al.*, 2002) but this section focuses on the link between passive smoking and lung cancer.

Exposure to passive smoking has become widely accepted by scientific committees and national organisations as a cause of lung cancer (National Research Council, 1986; US Department of Health and Human Services, 1986; Australian National Health and Medical Research Council, 1987; UK Department of Health and Social Security, 1988) even if the size of the risk is poorly understood. Much of the early work considered the influence of passive smoking on the spouses of partners who smoke and a number of studies showed that the risks in wives increased with the amount smoked by husbands (e.g. Hirayama, 1981; Trichopolous *et al.*, 1983; Trichopolous, 1984; Stockwell *et al.*, 1992; Jee *et al.*, 1999).

Table 2.2 demonstrates the wide range in relative risk estimates that have been calculated by different studies. For example among the non-smoking wives of light smoking husbands the relative risk ranges from 0.8 to 1.9 while in other studies the

relative risks have increased by between 50 percent and 150 percent amongst the most heavily exposed. Schottenfeld (1996) calculated a weighted average of the various study-specific relative risks and calculated that there was an elevated risk of 25 to 30 percent among non-smoking wives of smoking husbands, when compared with non-smoking wives of non-smoking men. This is supported by the report of the National Research Council (1986) that concluded that 20 percent of cancers that occurred in non-smoking women and men might be attributable to exposure to passive smoking.

		Husband's Smoking Status	
Investigators	**No. of Lung Cancers**	**Light**	**Heavy***
Hirayama *et al.* 1981	201	1.4	1.9
Trichopolous *et al.* 1983; 1984	77	1.9	2.5
Garfinkel 1981	153	1.3	1.1
Correa *et al.* 1983	22	1.2	3.5
Koo *et al.* 1984	88	1.9	1.2
Wu *et al.* 1985	28	1.2	2.0
Garfinkel *et al.* 1985	134	1.1	2.0
Akiba *et al.* 1986	94	1.4	2.1
Pershagen *et al.* 1987	67	1.0	3.2
Lam *et al.* 1987	199	1.9	2.1
Gao *et al.* 1987	406	1.2	1.7
Janerich *et al.* 1990	191	0.8	1.1
Fontham *et al.* 1991	420	1.1	1.3
Brownson *et al.* 1992	432	0.9	1.3
Stockwell *et al.* 1992	210	1.5	2.4

*Definitions of heavy smoking varied but typically included those who smoked 20 or more cigarettes per day

Table 2.2. Relative risk of lung cancer among non-smoking women according to the level of husband's smoking (Schottenfeld, 1996)

Hackshaw *et al.* (1997) attempted to provide a more precise estimate of the risk of lung cancer in life-long non-smokers exposed to environmental tobacco smoke. They reviewed 37 published epidemiological studies of the risk of lung cancer in non-smokers who did and did not live with a smoker (4,626 cases). The risk estimate was compared to that from linear extrapolation of the risk for smokers using

seven studies of biochemical markers of tobacco smoke intake. From this information, they calculated an excess risk of lung cancer of 24 percent amongst non-smokers who lived with a smoker after adjustments for effects of bias and dietary confounding variables. They also showed that there is a clear dose response relationship between a non-smoker's risk of lung cancer and the number of cigarettes and years exposure to the smoker. This study appears to provide strong evidence that breathing other people's tobacco smoke is a cause of lung cancer.

This is contradicted by other studies such as Lee *et al.* (1986), Lee (1992) and Copas and Shi (2000). Lee *et al.* (1986) challenged the evidence that passive smoking is an important factor that causes lung cancer. They suggested that studies have been limited by their failure to overcome the methodological difficulties that studies of passive smoking need to address. This is because these studies firstly suffer from weak exposure data as they tend not to obtain objective data by measurement of ambient levels of smoke constituents in the home or workplace or in body fluids of non-smokers who have been exposed. This has been further complicated by changes in exposure that have occurred in response to increased restrictions on smoking in public places and changing smoking habits in the home (Jarvis *et al.*, 2000). For instance there has been a substantial decrease in the number of children exposed to passive smoking since the late 1980s (Jarvis *et al.*, 2000). Secondly, they suggest that studies fail to take into account the possibility of misclassification of active smokers as non-smokers as this may have consistently biased relative risk estimates upwards as a small proportion of active smokers deny smoking. These two factors are further strengthened by dosimetric studies suggesting that, in cigarette-equivalent terms, passive smoking only results in a relatively small exposure to non-smokers.

They concluded that the marked increases in risks that have been noted in some studies are more likely to be a result of bias in the study design than of a true effect of passive smoking. This is supported by Copas and Shi (2000) who suggested that the Hackshaw *et al.* study, in particular, may suffer from publication bias. Published studies, particularly smaller ones, will be biased in favour of more positive results (Hugod *et al.,* 1978; Jarvis *et al.,* 1985). They demonstrated that even a modest degree of publication bias leads to a substantial reduction in the relative risk and to a weaker level of significance. This suggests that the relative risk is probably exaggerated and at the very least needs to be interpreted with caution.

It is apparent from this discussion that it is still debateable whether passive smoking does have an important effect on the rate of lung cancer. There are a large number of studies that support this observation but as has been pointed out they seem to have methodological flaws. It is likely that even if there is an increased risk, it may have been exaggerated in many studies.

2.4.3 Air Pollution

A number of pollutants in the air have been identified as containing carcinogens that can potentially initiate lung cancer. The products produced by the combustion of fossil fuels, emissions from motor vehicle exhausts, power plants, industrial and residential emissions have all been considered. Although these substances are known carcinogens it is still debated whether they actually do cause lung cancer and if they do how important they are in increasing the risk. There are many difficulties in attempting to understand the influence of pollutants upon lung cancer incidence

rates due to the problems of distinguishing the cancers caused by pollutants from those caused by smoking and other causes of lung cancer. Often the cause may be a complex mixture of more than one influence (Doll, 1978; Friberg and Cederlof, 1978).

A number of different constituents of various types of pollutants have been suggested as having the potential to raise the incidence of lung cancer. Cohen (2000) identified five key categories of atmospheric pollution that have been identified in the literature as causing lung cancer; Polycyclic Organic Matter (POM), Particles, Diesel, 1,3-Butadiene and Aldehydes. The role that each pollutant plays in affecting the incidence of lung cancer varies because a compound can be shown to have carcinogenic properties but its effect upon the health of a population can be more difficult to prove. Those studies that have considered the effects of air pollution in a non-occupational setting have tended to focus upon the effects of particles (Arden Pope and Dockery, 1999) particularly those particles with an aerodynamic diameter of less than 10 μm (PM_{10}). This is because particles of this size can be inhaled into the lung and carry carcinogenic substances on their surfaces. Most other work has focused upon the health effects of POM but these tend to be limited to an occupational setting (Doll *et al.,* 1972; Redmond, 1983; Boffeta *et al.,* 1997).

There is a wide epidemiological literature that has examined the relationship between outdoor air pollution and lung cancer. The studies that have been undertaken can be considered to fall within one of five categories: migration studies that compare the lung cancer risks of the native and destination country of migrants;

studies of urban and rural differences in lung cancer rates; studies of populations living near to a particular pollution point source; and case-control and cohort studies of lung cancer occurrence.

The results from studies of migrants provides only weak evidence of a link between air pollution and lung cancer. It has been noted that those migrants that come from countries with higher rates of lung cancer and higher levels of air pollution to countries with lower levels of lung cancer tend to have lung cancer rates higher than those of the new country. However, these studies tend not to control for factors such as rates of smoking and occupational exposure (Speizer and Samet, 1994). This approach has some clear potential if methods can be devised to adequately control for confounding variables and to measure exposure to pollution.

A number of studies have pointed towards higher rates of lung cancer prevalence in urban areas compared to rural areas (Dean *et al.*, 1978; Haynes, 1988; Friis and Storm, 1993; IARC, 1997a; Kadafar *et al.*, 1996) which suggests that the risk factors associated with lung cancer are greater in urban areas. Although air pollution levels can be expected to be higher in urban areas and may be one reason for these observations, this may be confounded by other risk factors, patterns of migration or occupational exposure also varying between urban and rural areas (Haynes, 1988; Cohen, 2000). Little detail is known about differential changes in smoking behaviour in urban and rural areas which makes it difficult to interpret differences in lung cancer incidence between these areas (Lawther and Waller, 1978).

Studies of lung cancer incidence in populations residing in close proximity to point sources of air pollution have often been limited by similar factors. Studies have focused on the emissions from fossil fuel electrical power plants (Natusch, 1978; Cohen and Pope, 1995), nonferrous metal smelters (Pershagen *et al.* 1977) and municipal solid waste incinerators (WHO, 1988; Elliott *et al.*, 1996). Pershagen (1990) provided a review of the studies that have considered the potential link between residential proximity to industrial point sources of air pollution and lung cancer incidence. These consisted of five ecological studies and six case-control studies. The ecological studies all found an elevated risk but were not able to satisfactorily control for confounders at the individual level (e.g. smoking and employment at the plants themselves). The case-control studies provided conflicting results with many showing no association but they also suffered from the problem of adequately controlling for smoking behaviour and employment at the plant.

Other studies have focused more generally on the proximity to a group of steel and chemical industries rather than an individual point source. One such study examined the impact of industrial air pollution on poor health, including lung cancer, in Teeside, England (Bhopal *et al.*, 1998; Pless-Mulloli, 1998). The study compared two areas with similar social and economic circumstances but with different proximity to major industries. The authors grouped people in similar social and economic groups (and therefore assumed that smoking habits would be similar) with varying proximity to major industries and tested the health gradient between each. Amongst their findings, they demonstrated higher levels of lung cancer in the areas closest to industrial areas with the association being clearest in women aged 0 to 64. Again, difficulties arose in estimating precise levels of exposure but the results lend

support to the theory that air pollution plays a role in influencing the lung cancer rate in localised areas. Even so, some doubt must remain as it has proven difficult to be fully confident in the methods of measuring exposure to air pollution and any confounding factors.

Attempts have been made to address the reoccurring issue of confounding from factors such as smoking in cohort and case-control approaches. A number of studies have found a positive association between lung cancer and air pollution (e.g. Jedrychowski *et al.*, 1990; Dockery *et al.*, 1993) although the studies used different methods for measuring exposure to air pollution. One such study undertaken by Barbone *et al.* (1995) investigated the relationship between air pollution and histologic type of cancer amongst men who had died in Trieste, Italy from 1979 to 1981 and from 1985 to 1986. They identified 755 cases of lung cancer and 755 controls and obtained information on smoking habits, occupation, and place of residence from the subject's next of kin. Air pollution levels were estimated from the average value at the nearest monitoring station. Logistic regression was used to evaluate the effect of residence and air pollution on lung cancer after adjustment for age, smoking habits, likelihood of exposure to occupational carcinogens, and social group. They showed that there was an increase in the risk of lung cancer with increasing levels of air pollution and that the risk was elevated for city dwellers compared to residents in rural areas.

Xu *et al.* (1996) examined lung cancer incidence in cities of Liaoning in northeastern China, an area with heavy concentrations of industry. They undertook a case-control study involving 1,249 lung cancer patients and 1,345 population-based controls

during the period 1985 to 1988. Although they found that smoking was the principle cause of lung cancer, they also found a significant increase in lung cancer risk associated with an overall index of indoor air pollution due to coal burning emission. They suggest that smoking and environmental pollution combine to account for elevated rates of lung cancer.

Tango (1994) compared the incidence of lung cancer and other control diseases for which smoking is not a risk factor (ischemic heart disease and cerebrovascular disease) in a number of wards in Tokyo with various levels of air pollution (SO_2 and NO_2 levels). He hypothesised that if there was a long term effect from air pollution upon lung cancer incidence then it is likely that there would be an increased incidence of lung cancer with higher levels of air pollution but that the control diseases would not mirror this rise. This assumes that tobacco consumption is not correlated with pollution levels. Using Poisson regression techniques he found that the long term effects of air pollution did raise the incidence of lung cancer at levels above the control diseases.

Although cohort and case-control studies provide a very useful and efficient means of interpreting the effects of air pollution on lung cancer incidence a number of issues remain that may potentially confuse any results (Cohen, 2000). First, an individual's exposure to pollution may occur in a large number of microenvironments as people migrate at different temporal and spatial scales. Therefore, it may be difficult to characterise the exposure of an individual for use in an epidemiological study. Second, even if an individual remains relatively stationary, pollution levels may vary over time. For instance the clean air acts of

1956 and 1968 in the UK have resulted in lower levels of air pollution in urban areas (Dunn and Kingham, 1996). Finally, misclassification of exposure and of the confounders can lead to spurious results that may elevate or diminish the estimates of the effect (Birkett, 1992; Hatch, and Thomas 1993). This is especially true for the confounding effect of smoking behaviour, which is often not adequately controlled for (Katsouyanni and Pershagen, 1997).

2.4.4 Radon Gas

A number of studies have examined the potential links between exposure to high levels of radiation and the incidence of various types of cancer. These include studies of workers exposed to the radium used in the painting of watch dials (Martland, 1929; Martland and Humpheries, 1929; Carnes, *et al.* 1997), radiation exposure amongst nuclear industry workers (e.g. Checkoway *et al.*, 1985; Rinsky *et al.*, 1988; Wiggs *et al.*, 1991), exposure close to weapons testing sites (Conrad *et al.*, 1966; Simon and Graham, 1997; Takahashi *et al.*, 1997), exposure to the detonation of atomic bombs (Stewart, 1997; Shimizu *et al.*, 1999), potentially higher levels of exposure close to nuclear installations and reprocessing plants (Black, 1984; Gardner 1987a, Gardner 1987b; Openshaw, 1987; Pobel and Viel, 1997), x-ray exposure leading to thyroid cancer amongst children (Ron *et al.*, 1988; Neuberger *et al.*, 1991) and fetal exposure to x-rays (Stewart *et al.*, 1956; Knox *et al.*, 1987; Doll and Wakeford 1997).

Radon gas provides by far the largest source of exposure to radiation that the general public come into contact with and is natural in origin (Darby, 1999). It has long

been identified as a potential risk factor in cancer incidence especially in cases of lung cancer. Radon naturally decays into lead as a part of the uranium-238 series emitting alpha particles, which have potentially carcinogenic properties. This section concentrates upon the epidemiological evidence of the potential links between radon and lung cancer. Subsequent chapters consider the physical properties of radon that make it a potential hazard and the approaches that have been taken to model radon levels.

A number of studies have investigated the potential links between radon and cancer. They include rather confined and specialised cases such as studies of underground miners, experimental studies on rats and also much larger and more dispersed studies such as case control studies of populations. Despite the range of studies, the relationship between radon and lung cancer is still not fully understood and the degree to which radon is a health risk is still debated especially in residential settings. Even so, it has been estimated that there are 2,000 deaths from radon caused cancer each year in the UK and 5,000 in the United States (Peto and Darby, 1994). Darby (1999) has shown that radiation doses that are illegal in the nuclear industry are not uncommon in high radon areas such as Devon and Cornwall. If these figures are accurate then radon may be the most serious environmental cause of cancer in Britain.

The first observations of lung cancer incidence due to radon exposure were among miners. The links between high death rates from lung diseases and mining can be traced back to the middle ages for miners in the Ore Mountains in what is now Germany and the Czech Republic (NRPB, 2000) and was first reported by Härting

and Hesse (1879). Although the cause of these high death rates was not known at the time, it is assumed that radon was to blame. The studies that have since been conducted on underground miners who are occupationally exposed to radon have consistently demonstrated an increased risk of lung cancer. Lubin *et al.* (1994) combined information on 60,000 miners to show that the fatal lung cancers far exceeded the expected number that would have been found in the appropriate general population. This study also confirmed observations made by other authors (Hornung and Meinhardt, 1987; Ševe *et al.*, 1988; Darby and Doll, 1990; Xuan *et al.*, 1993) that the exposure response relation declined with increasing exposure duration, implying that for equal total exposure a higher risk of lung cancer is experienced by those who accumulate exposure at lower exposure rate and over a longer period of time. Most of the studies of miners seem to show that even allowing for confounding factors such as smoking, there still seems to be a strong causal relationship (NRPB, 2000) and many of the studies undertaken on cohorts of miners support a multiplicative relationship between smoking and radon exposure (Hornung and Meinhardt, 1987; Whittemore and McMillan, 1993). Lubin (1994) in his review of lung cancer and radon feels that the information collected from miner data is clear: radon in mines can cause lung cancer and a significant excess is observed. As there is no threshold below which radon cannot increase risk it seems unlikely that this phenomena can be considered to be exclusive to the mine environment.

The results from studies of underground miners are supported by experimental animal studies undertaken by Cross (1992, 1994). These studies have the advantage of allowing the animals to be exposed to a variety of agents under carefully controlled conditions. The work has shown that there is evidence of lung cancer

development at low levels of exposure and that tumours can develop due to exposure to radon alone and does not require exposure to other irritants such as uranium dust or cigarette smoke (Brooks *et al.*, 1997). Importantly, the studies showed that protraction of exposure in rats produced significantly higher incidence of multiple primary lung tumours and fatal primary lung tumours. The results can potentially allow a greater understanding of the dose response relationship between radon exposure and lung cancer, assuming results from animals translate to humans.

Although the link between radon and lung cancer was first recognised through studies of underground miners, more recent ecological investigations have shown that there is a link between residential radon and lung cancer incidence in the wider population. Such ecological studies consider the association between disease occurrence in groups and the assessed exposure of the groups (usually the people residing in a geographical unit). Etherington *et al.* (1996) investigated the relationship between domestic radon levels and the incidence of 14 major cancers in Devon and Cornwall. The counties can be divided into postcode sectors that each represent between 3,000 and 10,000 people. Without controlling for any other risk factors, each postcode sector was categorised into one of ten categories based upon the average radon levels, but no significant difference in cancer incidence between postcode sectors with high or low radon level was found. Magnus *et al.* (1994) studied the association between indoor radon exposure and lung cancer in 427 municipalities in Norway. They used information gained from radon detectors sent to 10,000 households aggregated to the municipality unit along with surveys on smoking habits and data on asbestos exposure. They found a consistent increase in the incidence of lung cancer with higher tobacco consumption, but as radon exposure

increased, the lung cancer incidence was higher only amongst females. The authors themselves point out that systematic errors cannot be excluded when an ecological approach such as this one is taken.

Further examples of ecological approaches to this problem were reviewed by Stidley and Samet (1993). They examined 15 ecological studies that considered the influence of radon on lung cancer. They found that of these 15 studies, seven found increasing lung cancer risk with increasing indoor radon levels, six found no significant relation, and two found negative associations between indoor radon and cancer. They suggest that ecological studies are limited by errors in radon measurements which can occur for a variety of reasons such as current radon exposure measurements being assumed to be biologically relevant indicators of exposure (i.e. ignoring the latent period between exposure to indoor radon and the onset of lung cancer); the use of indirect measurements of radon as an index of exposure; and the ecological fallacy that assumes individual exposure to be adequately estimated by group indicators. The authors conclude that the shortcomings of the ecological studies render them uninformative on the lung cancer risk associated with indoor radon. Lubin (1994) agreed with this conclusion and suggested that using geographical units to summarise radon values is too imprecise and that such ecological approaches are unlikely to be informative and are of little use in addressing radon risk. This is supported by Lagarde and Pershagen (1999) who performed concurrent analysis on individual and aggregated data from the nationwide case-control study of residential radon and lung cancer in Sweden. They concluded that ecological studies may be misleading. However, most of these studies have been carried out for relatively broad geographical units and there is little

discussion of ecological studies that undertake analysis for small geographical units. This is one of the aims of this study.

Lubin (1994) suggested that a more informative approach has been to use in-home measurements of radon to estimate radon exposure. However, even if this is possible then there are no guarantees that the measure of exposure will be more successful. This is demonstrated by Blot *et al.* (1990) who found no association between lung cancer and radon exposure measured using in-house detectors in Shenang, China, which is an industrial city with one of the worlds highest rates of lung cancer in women. However, the study did not fully trace the residential history of the women diagnosed with cancer and the controls, and it was thought that the high levels of indoor pollution could possibly have produced artificially low radon concentration readings. This is reiterated by a study undertaken by Alavanja *et al.* (1999) who used both indoor radon detectors and CR-39 alpha-particle detectors (surface monitors) in their study of Missouri women (aged 30 to 84 years) newly diagnosed with primary lung cancer during a one year period in 1993 to 1994. They found that when surface monitors were used there was a significantly elevated risk of lung cancer at or above the maximum level that is considered safe but not when the standard radon dosimetry was used. These studies demonstrate the importance of both accurate and representative measurements of radon levels and this may be a crucial factor in explaining why conflicting results emerge from epidemiological studies of the effects of radon on lung cancer.

Pershagen *et al.* (1994) attempted to address some of these issues in their Swedish study of 586 women and 774 men aged 35 to 74, diagnosed with lung cancer

between 1980 and 1984 and 1467 female and 1380 male controls. Radon was measured in 8,992 dwellings that were occupied at some time since 1947 whilst potentially confounding information was obtained from questionnaires. They found that residential exposure to radon was an important cause of lung cancer in the general population. Darby *et al.* (1998) took radon measurements at the addresses at which subjects of their study had lived over the previous 30 years in order that time-weighted average indoor radon concentrations could be included. They suggest that as many as 1 in 20 cases of lung cancer in the UK are caused by residential radon, most being caused in combination with smoking. They undertook a case-control study in south-west England in which 982 subjects with lung cancer and 3,185 control subjects were interviewed. They found that the increased risk was consistent with past studies (i.e. 1 in 20 cases) and that there was a dose response relationship in which the risk of lung cancer increases with increasing radon concentrations. People who lived in homes where the radon concentration was at 200 Bq m^3 had lung cancer rates that were 20 percent higher than people living in homes with radon concentrations of 20 Bq m^3 (the average for the whole country). Where the concentrations reached 400 Bq m^3 (which it does for around 50,000 people in the UK) the estimated increase is 40 percent. It should be noted that in 30 percent of the cases radon measurements were not known and the authors had to rely upon radon models. Nonetheless, this study is perhaps the most comprehensive assessment of the effect of residential radon exposure in the UK.

Despite these results, other studies that have followed the long-term exposure of lung cancer subjects and their controls have found that there is either no relationship or a rather weak one. Auvinen *et al.* (1996) in their study of 1,973 lung cancer patients in

Finland diagnosed between 1986 and 1992 and 2,885 controls found that there was no elevated risk due to indoor radon exposure. Alavanja *et al.* (1994) used a sample of 538 non-smoking white women diagnosed with lung cancer and 1,185 age matched control subjects in Missouri. They obtained year long radon measurements in dwellings occupied for the previous five to thirty years. They found that there was little evidence that lung cancer incidence increased with higher radon concentrations. They concluded that the magnitude of the lung cancer risk from radon levels commonly found in US dwellings appears low. Ruosteenoja *et al.* (1996) examined the link between radon and lung cancer in Finland and found that the increase in risk was not statistically significant.

Similarly Létourneau *et al.* (1994) found that there was no increase in relative risk in relation to cumulative exposure to radon. They studied 738 individuals with lung cancer and 738 controls in Winnipeg, Manitoba and took radon measurements in the residences of subjects where they had lived for more than one year. They found no increase in the relative risk of lung cancer with increased cumulative exposure to radon. It is possible that some of this discrepancy can be accounted for by the random errors that occur when estimating radon exposure and this was explored by Lagarde *et al.* (1997). They suggested that the errors are potentially significant and should be taken into account in studies that use a risk estimate. At the very least, there would seem to be important regional variations in the lung cancer incidence where the cause can be attributed to radon exposure.

Some studies have focused upon the differential effect of radon upon smokers and non-smokers. This offers the opportunity to examine the risk associated with radon

whilst controlling for the most important confounding variable and also to explore any multiplicative effect that exists. It has been claimed that radon exposure among non-smokers carries no risk at all but evidence from miners who have never smoked suggests that this is not the case (BEIR VI, 1998) and this is supported by animal experiments (BEIR IV, 1988). The precise nature of this synergistic relationship at various levels of tobacco smoke and radon is controversial (Lee *et al.*, 1999) but NRPB (2000) concluded that the relationship between smoking and radon exposure may be mulitplicative and this may be especially important where the radon levels are low and not otherwise considered dangerous. This is a conclusion that is supported by Pershagen *et al.* (1994) in their Swedish study who found that the interaction between radon exposure and smoking exceeded additivity and was closer to a multiplicative effect. This notion is also supported by Lee *et al.* (1999) and Lichtenstein *et al.* (2000) who developed novel ways for encouraging smoking cessation amongst residents who were exposed to the 'radon-smoking synergy'. They suggested that smokers, and perhaps non-smoking residents, of smoking households are at increased risk from lung cancer even when radon levels are relatively low (below the action level) and hence used radon risk to try and motivate smoking households to reduce their consumption of cigarettes.

The studies of uranium miners and research using experimental studies have shown that radon and its decay products are potentially a major health risk in lung cancer incidence. What is less certain is the role that radon plays in the residential environment as radon gas accumulates in houses in areas with high radon potential. Ecological studies have demonstrated the difficulty of controlling for other confounding factors such as smoking, whilst all types of studies have come up

against the problem of estimating exposure to radon. This is due to the high level of uncertainty that is inherent in the models and the amount of variability that is possible over small areas and time spans. This is a key issue that is important to this study and subsequent chapters will attempt to address this by suggesting that it is more sensible to use geological units as the units for predicting radon levels. In addition some of the problems can be solved by undertaking the analysis for smaller areas than previous studies have used.

2.4.5 Occupational Carcinogens

Occupational carcinogens refer to the chemical and physical agents that particular groups of workers come into contact with in the workplace which have been identified as causing cancer. Some groups of workers have an elevated lung cancer incidence when other factors, such as smoking, have been controlled for, and a number of chemical and physical agents have been identified in the workplace as potential causes. There are some obvious difficulties involved with estimating the relative importance of occupational carcinogens on the lung cancer rate and the estimates vary widely. Table 2.3 shows some of the most important chemicals and industrial processes associated with human cancer which include arsenic, asbestos, bis, ether, chloromethyl ether and chromium and its compounds, each of which has been identified in the epidemiological literature as a potential cause of lung cancer. Mossman (1988) suggests that occupational respiratory carcinogens may account for between 5 and 20 percent of lung cancers occurring among men and women of different cultures and nations and Steenland *et al.* (1996) estimated that

approximately 9,000 to 10,000 men and 900 to 1,900 women develop lung cancer in the United States due to past exposure to occupational carcinogens.

Agent	Human Target Organs	Background
Arsenic	Lung Skin Urinary tract	By-product of copper, lead, zinc and tin ore smelting. Also some pesticides.
Asbestos	Lung Mesothelium or serosa of pleura, pericardium, and peritoneum	Workers in industries such as miners, millers, textile insulation, shipyard, cement. Synergistic relationship with smoking Asbestos exposure accounts for 5 percent of lung cancer deaths in men in US.
Bis (chloromethyl) ether (BCME) and chloromethyl methyl ether (CMME)	Lung	Used in manufacture of resins, polymers and plastics.
Chromium and compounds	Lung Nasal and paranasal sinuses	Used in metal alloys, electroplating, lithography magnetic tapes, paint pigments, cement, rubber, photoengraving, composition floor covering, oxidant in synthesis of organic chemicals

Table 2.3. Chemicals and Industrial Processes Associated with Human Lung Cancer (adapted from Schottenfeld, 1996)

Pezzutto and Poletto (1999) found that occupational exposure partially accounts for the high lung cancer mortality rate among male residents in Rosario City, Argentina. Risks were especially high for drivers, construction workers and agricultural workers. Droste *et al.* (1999) undertook a case-control study to investigate the relationship between lung cancer and exposure to occupational carcinogens in the industrialised region of Antwerp, Belgium. They demonstrated that there was an excess risk among manufacturers of metal goods, transport equipment (other than automobiles), and transport support services (using a census-based classification of occupation). It has been suggested that changes in cancer rates may, in part, be due to changes in exposure to occupational carcinogens and could help to account for the

larger increases in lung cancer incidence among women than among men (Travis *et al.*, 1995). This is supported by Levi *et al.* (1999) when attempting to explain the levelling off in lung cancer mortality in North America and Western Europe. Boffetta and Kogevinas (1999) made the point that the increasing importance of new industries, mainly in the service sector, are likely to alter occupational exposure to many hazards, although the outcomes are poorly known.

An important and well documented occupational hazard that has been identified as being a respiratory carcinogen is asbestos. It cannot be considered to be exclusively an occupational hazard but occupational exposure dominates the literature. Asbestos is a general term that is used to describe a variety of naturally occurring hydrated silicates that produce mineral fibres in mechanical processing. Prolonged exposure to asbestos results in the accumulation of macrophages and inflammatory cells in the alveoli, which is accompanied by the release of oxygen free radicals, the peroxidation of cell membranes, and damage to DNA and other macromolecules. The shape, length, and persistence of fibres may be important in eliciting cellular responses intrinsic to carcinogenesis. Exposure to asbestos has been identified by a number of researchers, through epidemiological and experimental animal studies, as being associated with lung cancer. Schottenfeld (1996) suggested that since the beginning of World War II, up to eight million people in the United States have been exposed to asbestos in the workplace and it accounts for five percent of the lung cancer deaths amongst men. Albin *et al.* (1999) reviewed the current asbestos related cancers in Europe and demonstrated that studies have in the past estimated the population attributable risk to asbestos exposure to be in the range of 2 and 50 percent. They estimate that between 15 and 30 percent of the male population in

Europe have ever been exposed to asbestos, mainly in construction or in shipyards, and conclude that 10 to 20 percent of all lung cancer cases can be attributed to asbestos.

Doll (1955) was one of the first researchers to examine the link between lung cancer and asbestos exposure in an occupational setting. He examined the cause of death among 105 persons who had been employed at one asbestos works and found that 18 died of lung cancer. He then followed this up by studying 113 men who had worked for at least 20 years in places that they were liable to be exposed to asbestos dust and compared the mortality among them to the mortality that would have been expected based on the mortality experience of the whole male population. He found that 39 deaths occurred in this group compared to 15.4 that were expected. He suggested that this excess was entirely due to deaths from lung cancer (11 against 0.8 expected) and from other cardiovascular diseases (22 against 7.6 expected). This study provided the first convincing evidence that lung cancer was a specific industrial hazard for certain asbestos workers and he quantified this by stating that being employed for 20 years or more increased the risk of contracting lung cancer by 10 times. Importantly, this risk has become progressively less as regulation has improved the conditions in workplaces. Since the publication of this ground-breaking study, other studies have examined links between asbestos and lung cancer in a number of different contexts. For example, the work undertaken by Muscat *et al.* (1995) in the US found exposure to asbestos for one year or more to be significantly related to all types of lung cancer incidence. They calculated an odds ratio of 2.2 for smokers who had been exposed to asbestos compared to those who had not been exposed. Similarly, deKlerk *et al.* (1996) examined asbestos exposure

among a group of workers from the Wittenoom asbestos industry in Australia. They questioned 2,928 former asbestos workers and, after adjusting for smoking behaviour, they found a positive relationship between asbestos exposure and asbestos-induced lung cancer.

Selikoff *et al.* (1968) reported on the combined exposure of asbestos and smoking among insulation workers. They concluded that the combined risk exceeded the risk that would have been expected if each exposure were to have acted independently, i.e. a multiplicative effect. This was supported by a study of the combined effects of asbestos and tobacco smoking among Quebec miners and millers, amosite asbestos factory workers, and Finnish anthophyllite miners and millers (Saracci, 1987). Although most studies have concluded that the resulting relative risks are multiplicative, McDonald and McDonald (1980) when studying Canadian chrysotile miners and millers claimed that the effect of each agent was independent and additive.

Other studies have concentrated on the relationship between specific occupations, rather than particular agents, and lung cancer incidence. Partanaen and Boffetta (1994) reviewed 20 epidemiological studies that considered the cancer risk for asphalt workers and roofers in various countries and the raised risks suggest that exposure to bitumen fumes may have carcinogenic effects. Pirastu *et al.* (1998) considered cancer mortality among art glass workers employed in 17 industrial facilities in Tuscany, Italy. They found that that there was a consistently higher mortality from lung cancer in the cohort of 3,390 workers. Kogevinas *et al.* (1998) reviewed the recent epidemiological evidence of cancer risk for workers in the

rubber industry, finding an excess risk of lung cancer in most studies which was approximately 1.5 in about half of them. They suggest that because the relative risks were so high in many cohorts, it cannot be due to confounding by smoking or other lifestyle factors. Rodriguez *et al.* (2000) examined the risk of lung cancer among workers in iron and steel foundries in Asturias, Spain. They used 144 male lung cancer cases and 558 controls selected from 24,400 workers employed in the industry between 1952 and 1995 and concluded that a higher risk was observed for workers who were exposed to a variety of carcinogenic chemicals. Acquavella *et al.* (1993) found an elevated lung cancer rate for a cohort of 3,630 workers at a metal components manufacturing facility. The rate was elevated for the workers that were exposed to metal dusts, cutting oils/fluids, metal fumes, and solvents but elevated mortality was restricted to those workers first exposed during the period 1950 to 1959 which the authors suggest is due to an unidentified short-lived occupational carcinogen or an unspecified confounding factor.

In 1997, the International Agency for Research on Cancer classified crystalline silica (quartz) as a human carcinogen (IARC, 1997b). Many industrial sand plants process sand into silica flour, which is composed of fine particles of respirable crystalline silica. Inhalation of crystalline silica can result in chronic inflammation and fibrosis of the lungs (Donaldson *et al.*, 1992) and these processes may result in an increased risk of lung cancer (Souter *et al.*, 2000). This is supported by studies such as the one by Steenland and Sanderson (2001) who studied 4,626 industrial sand workers employed in 60 plants in the US to test whether crystalline silica influences lung cancer. The study suggested that the substance explained some of the excess cases of lung cancer that occurred among these workers. Similarly, McDonald *et al.*

(2001) found an increased SMR for lung cancer (139) among a cohort of 2,670 men employed before 1980 for three years or more in one of nine North American sand-producing plants. However, other authors consider results of studies that have examined the link between crystalline silica and lung cancer to be inconsistent because other methodologically strong, negative studies have not been considered and several studies supporting the carcinogenicity of quartz have significant methodological weaknesses (Holland, 1990; Miller *et al.*, 1997; Hessel *et al.*, 2000; Soutar *et al.*, 2000). Donaldson and Borm (1998) suggested that this confusion may arise because the carcinogenicity of quartz may be dependent on inherent characteristics of the quartz and/or on external factors affecting its structure. Other studies have suggested that the effects of quartz may extend beyond the sand industry and include others such as glass, ceramics, bricks, metal and quarrying (Kvåle, *et al.*, 1986; Partanen *et al.*, 1995). If these industries are causes of lung cancer then it is possible that some of them, especially quarrying activity, could have an effect on the nearby resident population due to the generation of dust into the atmosphere by industrial activity. The effect of these factors has not been explored in a residential setting.

There are a number of studies in the literature that have examined the links between lung cancer incidence and coal mining. There is no consensus as to whether there is an established link as a number of studies have found that there is little evidence for a link between the two (e.g. Miller and Jacobsen, 1985; Leigh *et al.*, 1994; Kuempel *et al.*, 1995; Lewis *et al.*, 1996; Wang *et al.* 1997). However, many of these studies have identified raised incidence of other lung diseases such as pneumoconiosis and emphysema amongst coal miners. Studies in Japan examined the relationship

between lung cancer and coal mining among 5,818 Japanese males (Une *et al.*, 1995; Miyazaki and Une, 2001) and compared the mortalities between coal miners and non-coal miners using Cox's proportional hazards model. They found that the males who had worked as coal miners for at least 15 years had a higher risk ratio incidence of lung cancer whilst coal miners with less than 15 years' experience had almost the same risk as non-coal miners. Although some authors suggest that the excess of lung cancer for coal miners is probably due to smoking (e.g. Ng *et al.*, 1990) there is some evidence to suggest that there may be a link between coal mining and lung cancer incidence.

There are a wide range of occupational carcinogens that workers have in the past been, and in some cases still are, exposed to, which may be related to lung cancer. Some, such as asbestos, are well researched and fairly well understood. Others are less well understood and therefore protection against them in the workplace maybe inadequate. Furthermore, there is a call in the literature for developing the methodology used in occupational epidemiology in the future to understand the mechanisms of cancer. This requires the improvement of the assessment of quantitative exposures, genetic factors and non-occupational exposures (Blair *et al.* 1999).

2.4.6 Nutrition

A number of epidemiological studies have pointed towards the influence that dietary considerations have upon the risk of contracting lung cancer, with a particular focus on the intake of foods that are rich in fat and cholesterol, including whole milk and

eggs. Although case-control studies have supported this link, cohort studies have tended not to find increases of lung cancer amongst people with elevated cholesterol (Schottenfeld, 1996). Some studies have demonstrated a positive association between lung cancer and dietary fat but have not found a link with increasing body mass. Alavanja *et al.* (1993) focused on non-smoking women in Missouri with lung cancer conducting 429 interviews between 1986 and 1991. They used logistic regression to analyse the responses and found that lung cancer risk was far greater for women with higher saturated fat consumption. The relative risk for the highest consumption quintile was six time higher than the lowest quintile which led them to suggest that observations such as these have been masked in earlier studies by the prevalence of smokers.

Nyberg *et al.* (1998) examined dietary risk factors associated with lung cancer among never smokers in a case-control study in Stockholm during the period 1989 to 1995. They found an increased risk associated with higher consumption of cultured milk products which may support previous observations that a high fat intake increases lung cancer risk. Sinha *et al.* (1998) demonstrated that higher red meat consumption increases the risk of lung cancer and that this is increased further if the meat is well done and/or fried. Breslow *et al.* (2000) studied records from 20,195 participants in the 1987 National Health Interview Survey in the United States which they linked to the National Death Index. They studied different food groups (e.g. fruit, vegetables, red meats, processed meats etc.) and found that the intake of red meats was positively associated with lung cancer mortality while the intake of dairy products was inversely associated. They suggested that whilst other factors such as smoking are a far greater risk factor, diet may have a contributory role.

Other studies have approached the effect of diet on lung cancer from a different angle and instead considered the potentially preventative effects of some foods. These studies have consistently found that an increased consumption of fresh vegetables and fruits lowers the risk (e.g. Chow *et al.*, 1992; Nyberg *et al.*, 1998). A number of antioxidants are considered chemopreventative nutrients, particularly β-carotene which inhibits carcinogenesis due to the quenching or neutralising of free radicals. These agents probably have both complementary and overlapping mechanisms of action (Steinmetz and Potter, 1991). A number of cohort and case-control studies have investigated the influence of dietary intake of vitamin-A upon lung cancer risk. Mayne *et al.* (1994) found that an increased consumption of raw fruits and vegetables was associated with significantly reduced risk from lung cancer. Brennan *et al.* (2000) studied the joint effect of diet and environmental tobacco smoke on the risk of lung cancer among non-smokers and found that a number of dietary variables provided a protective effect against lung cancer including fruit, lettuce, tomato, carrot, and cheese consumption as well as intake of carotenoids, ß-carotene, and retinol.

However, an experimental study undertaken by the Alpha-Tocopherol, Beta Carotene Cancer Prevention Study (1994) evaluated whether a daily supplement of α-tozopheral, ß-carotene or both would reduce the incidence of lung cancer for male cigarette smokers from south-west Finland who were aged 50 to 69 years. The group concluded that previous benefits of the vitamins may have been overestimated. This is partially supported by Willett (2001) who discusses the relationship between diet and cancer and demonstrates that the evidence is, at best, unclear. Although

there are some clear correlations between factors such as dietary fat and common cancers it is thought that this may be, at least in part, due to confounding by other factors of western lifestyles.

2.4.7 Previous Lung Disease

A further potential cause of lung cancer that has been identified in the literature is the increased susceptibility in people who have had some form of lung disease in the past. Various studies report upon the susceptibility of people with a previous history of tuberculosis, silicosis, or chronic bronchitis emphysema to contracting lung cancer. Zheng *et al.* (1987) conducted a population-based case-control study of lung cancer in Shanghai. They reported that the age, sex and smoking-adjusted relative risk of lung cancer was increased by 50 percent among all survivors of tuberculosis and by 100 percent among those diagnosed with tuberculosis within the previous 20 years. They claim that as much as nine percent of lung cancer cases may be attributed to prior infection.

Wu *et al.* (1995) undertook a case-control study of lung cancer in nonsmoking women in five metropolitan areas of the United States between 1985 and 1990. They found that a history of lung disease was associated with a significantly increased risk of lung cancer. Mayne *et al.* (1999) conducted a population-based case control study of lung cancer in male and female non-smokers in New York State, during the period 1982 to 1984. They found statistically significant associations for emphysema, chronic bronchitis, and the combined effects of emphysema, chronic bronchitis and asthma. This is supported by Brownson and

Alavanja (2000) who undertook a study to consider specifically the association between previous lung disease and lung cancer risk. They studied 676 incidents of lung cancer in women who contracted lung cancer in 1993 and 1994 and 700 controls. Elevated risks were found for chronic bronchitis, emphysema, pneumonia, and for all previous lung diseases.

Although, once again, the precise importance of this risk factor is disputed, there is evidence that for people who have contracted lung disease in the past there is an increased risk of lung cancer. This factor may have an effect both in isolation and in tandem with the other identified risk factors.

2.4.8 Genetic Predisposition

It is thought by some authors that certain genetic combinations can cause a person to be more susceptible to contracting lung cancer compared to somebody else who has been exposed to identical risk factors. Inherited genetic combinations are not in themselves sufficient for the development of cancer, and so the genes themselves are better described as predisposing to cancer (Bench and Rabbitts, 1994). Other common solid tumours such as breast, colon, kidney and ovary have shown a clear and important familiar inherited component but the evidence is less clear with lung cancer. Some epidemiological studies have suggested that inherited genetic changes do have a role in the development of lung cancer but this has been difficult to prove because of the competing effects of environmental factors.

In the past, family history has been relatively overlooked; in a review of cancer studies published between 1982 and 1984, only 25 percent of the studies included a family history component (Phillips *et al.*, 1991). Tokuhata and Lilienfeld (1963) claimed that there was a significant excess in lung cancer mortality in their study of familial aggregation. This translated into a two fold increase in smoking relatives and a four fold amongst non-smoking relatives. Sellers *et al.* (1990) described a pattern of autosomal condominant inheritance and hypothesised that segregation of this gene locus would account for 69 percent of the lung cancers diagnosed in people up to 50 years of age. Samet *et al.* (1986) claimed that the personal risk of cancer was increased more than five fold if at least one parent had lung cancer.

Osann (1991) considered 217 smoking women and found that after adjusting for smoking, the risk of lung cancer was increased by 1.9 in women with a family history of lung cancer. It is interesting to note that the study also observed a significant interaction between smoking and family history. Similarly, Kreuzer *et al.* (1998) studied lung cancer incidence amongst young adults aged 45 or under in Germany between 1990 and 1996 and observed that if a first degree relative had developed lung cancer there was a 2.6 fold increase in the risk of lung cancer amongst the young age group.

Bromen *et al.* (2000) investigated familial aggregation of lung cancer by using a case control study in Germany that was undertaken between 1988 and 1993. They compared lung cancer prevalence in first degree relatives of 945 patients and 983 controls, finding that lung cancer among parents or siblings was associated with a 1.67 fold increase in lung cancer risk. Wu *et al.* (1996) interviewed 646 female lung

cancer patients and 1,252 population controls about the history of cancer in their first degree relatives. They found that there was a 30 percent increased risk of lung cancer when there was a history of lung cancer in parents or siblings after adjustment for exposure to environmental tobacco smoke and a threefold increased risk of lung cancer when the patient's mothers and sisters had been diagnosed with lung cancer.

Despite this evidence, there is still much debate about the effect of any inherited susceptibility to lung cancer due to the difficulty that many of the studies have had in matching controls with similar environmental exposure to other risk factors. Relatives of patients with the disease are also more likely to smoke (Tokuhata and Lilienfeld, 1963) and are more likely to be subjected to the effects of passive smoking. Therefore, relatives share a similar environment as well as genetic make-up, with lung cancer patients (Bench and Rabbitts, 1994) and as a result, some studies have failed to convince that they have controlled for confounding factors adequately. Once again, the importance of genetic inheritance compared to the smoking behaviour is poorly understood but it seems that at best it plays a very minor role.

2.5 Conclusion

This section has demonstrated that there are a number of risk factors that influence the rate of lung cancer. Of these, smoking has been shown to be the most important cause whilst passive smoking, air pollution, radon, occupational carcinogens, nutrition, previous lung disease and genetic predisposition have also been shown to be significant in some studies. However, there is doubt from some authors about

these causes with the exception of smoking and asbestos because there are also spatial dimensions to each of the risk factors so that in specific areas some of the more peripheral factors can become more important. For example, in areas where radon levels are high this risk factor could be almost as important as smoking behaviour.

This study will attempt to address the issues that are thought to be important in Scotland by considering the geography of the key risk factors and examining how they influence lung cancer incidence in small areas. Many of the risk factors are not fully understood due to the difficulties of controlling for smoking in epidemiological studies. A key component of the research is therefore to develop methods to accurately estimate smoking behaviour.

3. GEOGRAPHY OF LUNG CANCER

3.1 Introduction

Lung cancer is the most common cancer worldwide both in terms of incidence and mortality (Parkin *et al.*, 1999). However, the incidence of lung cancer varies greatly between different regions of the world. There are also important variations at other geographical scales such as between countries that are within the same region and in different subdivisions of the same country. In addition, there are variations between males and females, between social classes, ethnic groups and areas with different levels of deprivation. How each of these relates to lung cancer depends on the scale of analysis, so the rate of lung cancer for males and females and the ratio of these rates will depend on the geographical area being considered. These differences reflect variations in exposure to the most important risk factors considered in the previous chapter, particularly smoking.

This chapter describes the geography of lung cancer at a range of scales but also examines how differences in rates vary with sex, social class, ethnicity and deprivation at each of these scales. It will firstly consider the recent global trends in lung cancer incidence. It then considers some of the broad differences in lung cancer incidence that exist between regions and countries of the world. There will then be a discussion of how rates of lung cancer can vary within individual countries with a particular focus on the differences within the UK. Finally, this chapter will review lung cancer rates and patterns in Scotland and discuss how they compare to the wider

geographies in which they fall, providing a context for the rest of this study which considers lung cancer for small areas within Scotland.

3.2 Global Incidence

In the year 2000, lung cancer accounted for 12.3 percent of all cancers worldwide which broke down into 17.0 percent of all cases among males and 7.1 percent for women (GLOBOCAN, 2000). Furthermore, the global incidence of lung cancer was increasing at approximately 0.5 percent per year (Haugen, 2000). Accurate global comparisons of cancer incidence are difficult because comparable high quality data are not always available (Boyle *et al.*, 2000). However, attempts have been made by the World Health Organisation (WHO) to address this issue using a package named GLOBOCAN (Ferlay *et al.*, 1998). The GLOBOCAN database was set up using the cancer incidence data collected by the Unit of Descriptive Epidemiology of the International Agency for Research in Cancer (IARC) from national cancer registries. However, the degree of detail and quality of data vary considerably. GLOBOCAN 2000 uses these data to provide global and national estimates of cancer incidence in the year 2000 based upon the most recently available data from each country (generally 3 to 5 years earlier). It has been estimated that in the year 2000, there were 1.25 million cases of lung cancer. Of these, 900,000 were male and 350,000 female. Globally, 900,000 deaths were thought to be caused by lung cancer each year (Murray and Lopez, 1997). The age-standardised lung cancer incidence rate was estimated to be 34.92 cases per 100,000 for males which was the highest rate of all the cancer types. For females, the age-standardised rate was estimated to be

11.22 cases per 100,000[2] which was the fourth highest after breast cancer, cervix uteri, and colon / rectum cancer (GLOBOCAN, 2000). Although global rates provide a useful context, they ignore important variations at a range of geographical scales. As will be demonstrated, there are significant variations in lung cancer rates between continents, countries and also within countries.

3.3 International Variations

The rates of lung cancer incidence vary at a range of geographical scales, reflecting broad variations in wealth and more localised cultural and environmental differences that result in an unequal exposure to the key risk factors. Therefore, it is only possible to appreciate differences in lung cancer rates by considering how the rates vary at different geographical scales. There are also variations in the rates between different social groups such as males and females, social classes and ethnic groups and differences in the rate between these groups also varies according to the scale of analysis. For example, the ratio of lung cancer rates between males and females varies dramatically depending on whether it is examined at a global scale or for local authorities in Scotland.

The age-adjusted lung cancer incidence rates for all regions of the world and a selection of countries are shown in Table 3.1 (GLOBOCAN, 2000). The data in the table use the most recent figures that were available in the year 2000. The table highlights that there are clear differences in the rates of lung cancer between the

[2] All subsequent ASRs in this chapter are expressed as rate per 100,000

Region / Country	Males			Females		
	Cases	Crude	ASR (W)[3]	Cases	Crude	ASR (W)[2]
World	901746	29.57	34.92	337115	11.22	11.05
More developed countries	470836	81.49	55.62	175392	28.74	15.62
Less developed countries	430919	17.44	24.79	161719	6.75	8.44
Eastern Africa	1711	1.39	3.08	1284	1.04	2.13
Middle Africa	1289	2.72	5.65	220	0.46	0.76
Northern Africa	7881	9.01	15.41	1658	1.93	2.76
Southern Africa	2861	12.43	23.81	1184	4.96	7.32
Western Africa	1157	1.05	2.15	231	0.21	0.35
Caribbean	4798	25.29	28.76	1832	9.56	9.70
Central America	9326	13.89	22.71	4047	5.94	8.44
South America	31493	18.41	25.28	12656	7.25	8.34
North America	119664	78.39	58.20	85994	54.75	33.59
Canada	11997	77.84	55.08	7589	48.24	30.24
United States of America	107618	78.45	58.56	78320	55.48	33.97
Eastern Asia	293392	38.60	39.41	122452	16.89	15.01
South-Eastern Asia	47456	18.34	27.83	18368	7.07	9.07
South Central Asia	59478	7.76	11.61	12666	1.75	2.33
Western Asia	19384	20.15	31.21	3270	3.56	4.80
Eastern Europe	126653	87.18	69.70	24420	15.10	8.77
Hungary	6526	136.07	95.45	2100	40.07	22.62
Northern Europe	33976	73.47	44.32	18063	37.52	18.85
Denmark	1999	76.35	46.82	1336	49.96	27.71
Estonia	606	92.30	69.86	138	18.66	9.50
Finland	1443	57.18	36.83	450	16.94	8.86
Iceland	58	40.53	31.52	47	33.00	23.79
Norway	1231	55.67	35.07	649	28.82	16.58
Sweden	1737	39.31	21.41	1058	23.54	12.09
United Kingdom	23708	82.07	47.61	13423	44.83	21.76
Southern Europe	67518	95.90	58.75	11227	15.22	7.95
Italy	29937	107.67	59.41	5689	19.29	9.02
Yugoslavia	5842	110.49	80.89	1156	21.60	13.79
Western Europe	75350	83.90	53.21	18183	19.44	10.68
Belgium	6256	125.71	76.43	1055	20.34	11.15
France	22910	79.55	53.52	3833	12.66	7.44
Germany	33568	83.37	50.25	9403	22.41	11.37
Switzerland	2698	73.89	48.52	745	19.94	11.61
Australia/New Zealand	6607	58.56	42.10	3231	28.18	18.18

Table 3.1. Lung cancer incidence in all regions and a selection of countries in the world (GLOBOCAN, 2000)

[3] Age Standardised Rate using world standard population

more and less developed countries. The more developed countries have an age-standardised rate (ASR) of 55.62 for men and 15.62 for women. The less developed countries having an ASR of 24.79 for men and 8.44 for women. This demonstrates that the sex differences that were highlighted at a global scale persist at a broad regional scale as well. Lung cancer rates are therefore higher in the more wealthy parts of the world, and are thought to be a response to lifestyle choices that are more available or affordable in the developed world. More specifically, the lung cancer rates are likely to reflect the stages of cigarette smoking in a population. Cavelaars *et al.* (2000) explain that smoking spreads through populations like an epidemic with different stages relating to the percentage of male and female smokers. Developing countries are likely to be in the early stages where smoking is the exception rather than the rule and mainly a habit of higher socio-economic groups. The developed countries are probably in the latter stages which are typified by declining prevalence rates among men and then later for women. Present day lung cancer trends are likely to reflect smoking habits 20 to 30 years previously, which in developing countries will have been low but in developed countries were high (typically 50 to 80 percent among men).

Table 3.1 also demonstrates that there are some substantial differences in lung cancer incidence between different regions of the world. For example, in some of the poorer regions the rate of lung cancer is very low; in Western Africa the ASR is only 2.15 for men and 0.35 for women. This compares to 58.20 for men and 33.59 for women in North America. However, regions that are close by and have relatively similar levels of wealth can have markedly different rates of lung cancer incidence. For example, Northern Europe has an ASR of 44.32 for men and 18.85 for women

whereas in Southern Europe the rates are 58.75 for men and 7.95 for women. This firstly confirms that lung cancer rates tend to be substantially different between males and females but secondly that the differences between the sexes can vary. In this example, the rate for men compared to women is approximately 2.5 times greater in northern Europe but approximately seven times greater in southern Europe. Furthermore, the rate for men is higher in Southern Europe than it is in Northern Europe but for women it is higher in Northern Europe than it is in Southern Europe. Once again this is likely to reflect differences in smoking behaviour but instead of being a direct result of large differences in wealth, and hence the stage in the smoking epidemic, it is more closely related to specific cultural differences between countries that influence the smoking habits of males and females.

It can also be noted from Table 3.1 that there are important variations between the countries within a region. For example, within Northern Europe, Estonia has an ASR of 69.86 for men and 9.50 for women. In Sweden the ASR for men is much lower (21.41) and a little higher (12.09) for women. This highlights two points, firstly that the overall ASRs vary markedly between countries which in this example may reflect differences in income and the stage in the smoking epidemic. However, if one compares the Swedish data to lung cancer prevalence in another Scandinavian country, Denmark, it can be seen that there are large differences here too. In Denmark, the ASR for men is 46.82 and for women it is 27.71 which demonstrates that large differences in rates can occur between seemingly similar countries. Secondly, it mirrors the observation that was made when regions were compared that the difference in the relationship between males and females can vary widely. The example of Estonia highlights this and shows that there can be very large differences

in the ASR between males and females and that the ratio is not necessarily the same between countries. This is also demonstrated in the example of Italy, where the ASR for men is 59.41 for men but only 9.02 for women. These variations reflect some of the cultural differences between men and women which result in unequal exposure to the risk factors, especially smoking behaviour.

In the UK, the ASRs of lung cancer are relatively high compared to most regions in the world. The ASR for men is 47.61 and for women it is 21.76, which compares to the world rates of 34.92 for men and 11.05 for women. The world's highest rates are 95.45 for men (Hungary) and 33.97 for women (United States). The UK rate for men falls below those found in North America, Southern Europe, Western Europe and Eastern Europe but it is higher than the ASR for other countries in Northern Europe, New Zealand and Australia. The ASR for women in the UK is one of the highest in the world and the rate is exceeded by very few countries. The countries that do exceed this rate include the United States, Canada, Iceland and Denmark. It is therefore clear from these figures that, in an international context, the rate of lung cancer incidence is relatively high in the UK and the rate is particularly high for females.

3.4 National Variations

There are not only important differences in the lung cancer rate between countries but also within individual countries. For example, in the United States lung cancer is the leading cause of cancer mortality among both men and women, accounting for 28 percent of all cancer deaths each year (Schottenfeld, 1996). Seventeen percent of

cancer incidence is attributed to lung cancer for men and 12 percent for women. The ASR for lung cancer is 58.56 for men and for women it is 33.96 which is the highest rate of all types of cancer for men and second only to breast cancer for women (WHO, 2002). However, there are pronounced geographical variations in lung cancer rates in census divisions in the United States, and these patterns have changed substantially over time (Devesa *et al.*, 1999). For example, there are consistently low rates in mountain and plains states whilst there are elevated rates across the south-east and south-central areas. Similar trends are found in other more wealthy countries such as Canada where the ASR is 55.08 for men and 30.24 for women and once again there are important variations within the country. For example, lung cancer incidence rates are consistently higher in the Northwest Territories than those reported in any other province and this disparity is particularly pronounced for females (de Groh, 2002). Although the rates tend to be lower, geographical differences can also be observed in less developed countries such as India where the highest ASR is 14.6 in Bombay and the lowest is 2.0 in Barshi (Pandey *et al.*, 1999).

In the previous section we saw that the UK has a high incidence of lung cancer compared to most other countries. However, the ASR values vary dramatically within the UK in a similar way to the variations discussed in the United States and Canada. In 1997, there were approximately 37,000 new cases of lung cancer in the UK and, of these, 24,000 cases were males and 13,000 females (GLOBOCAN, 2000). The difference in the rates between males and females is comparable to those in the US and Canada although the values are higher in the north American countries. Figure 3.1 shows the change in the incidence of lung cancer during the period 1969 to 1997. The graph shows that rates of lung cancer have been

consistently higher for males than for females with the former declining and the latter increasing during this period. In 1969 the ASR for men was 73 and 12 for women. By 1997 the rates had decreased to 42 for men and increased to 20 for women. There is some suggestion in the literature that the rates may now have peaked among women in other developed countries (e.g. in Canada (de Groh *et al.*, 2002)) and this may be reflected in the UK where the incidence rates for women were consistent during the period 1990 to 1997.

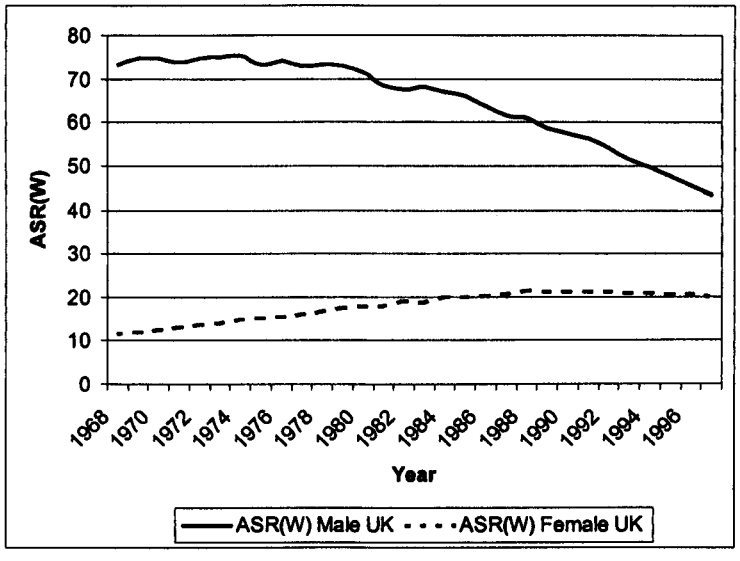


Figure 3.1. Age Standardised Rate for trachea, bronchus and lung cancer incidence 1968-1997 in the UK per 100 000 person-years at risk (world standard population) (WHO, 2002)

There are also important regional variations in the incidence rate across the UK. Approximately 80 percent of all the cases occurred in England, 12 percent in Scotland, five percent in Wales and two percent in Northern Ireland (Babb *et al.*, 2001). Figure 3.2 shows the Comparative Incidence Ratios (CIRs) for lung cancer in countries and regions of the UK. In this instance, the CIR is the ratio of the ASR for the country or region to ASR for the UK multiplied by 100 (i.e. it is expressed as a percentage) and the value is calculated for both males and females. The bars on the chart below 100 represent areas with low rates of lung cancer and the bars above 100

represent high rates. Among the regions in England, the North East has the highest incidence of lung cancer for both males and females, with levels similar to those in Scotland. The rate is about 30 percent higher than the average for males and 45 percent higher for females. The lowest incidence of cancer is in the East of England, the South West and the South East where rates are about 20 percent below the UK average.

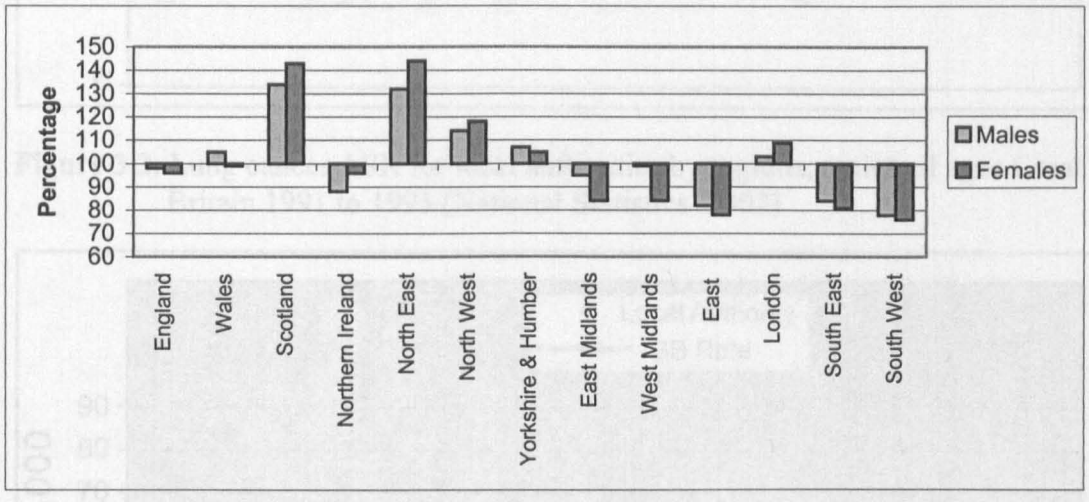


Figure 3.2. Lung cancer CIR by country and region (United Kingdom = 100), all ages, United Kingdom 1991 to 1993 (National Statistics, 2002)

This is broken down further in the boxplots in Figures 3.3 and 3.4 which show the ASR by local authorities within countries and regions for males and females with their 95 percent confidence limits. Figure 3.3 shows the ASR values for males and it can be seen that some regions have ASR values that are almost entirely below the rate for Great Britain. For example, the South East and South West regions have very few values above the GB rate and many that are less that 66 percent of the GB rate. There are also some regions where many of the ASR values are generally

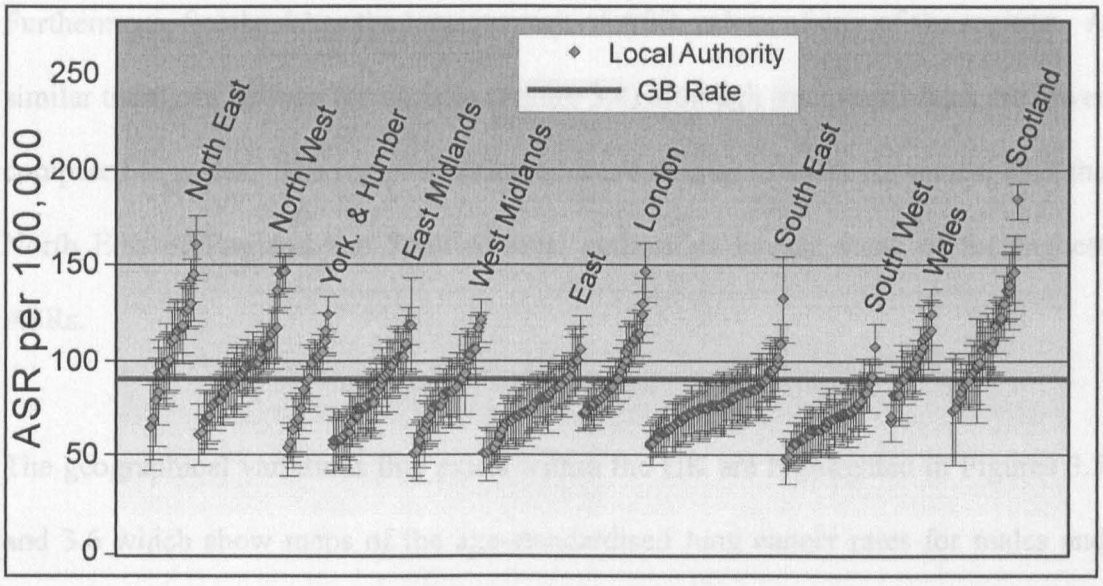


Figure 3.3. Lung cancer ASR for local authorities by regions, males all ages Great Britain 1991 to 1993 (National Statistics, 2002)

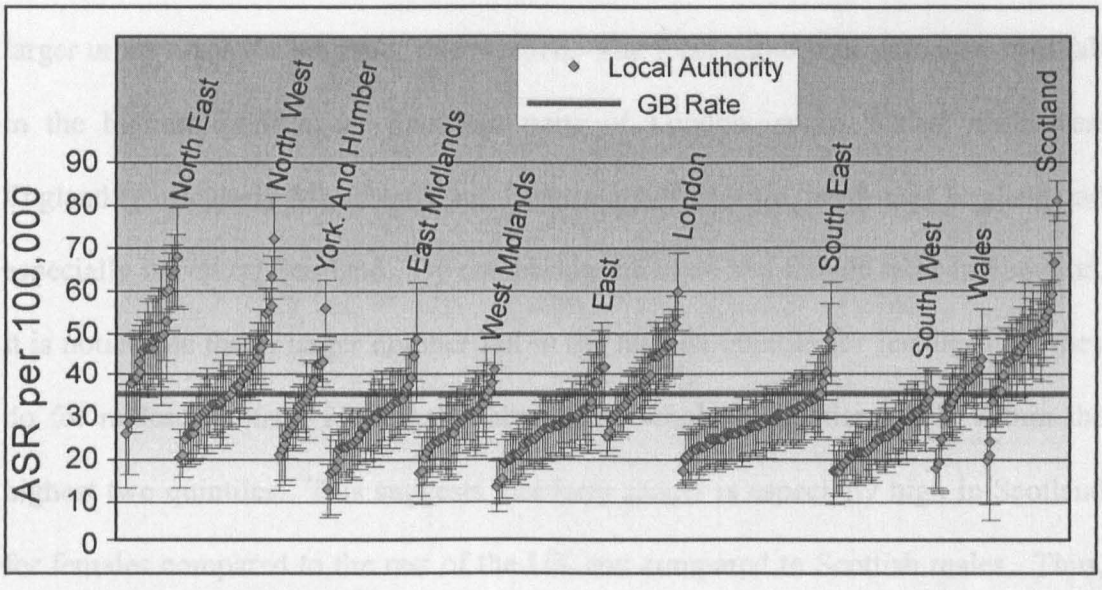


Figure 3.4. Lung cancer ASR for local authorities by regions, females all ages Great Britain 1991 to 1993 (National Statistics, 2002)

above the GB rate especially the North East region and Scotland. Importantly, the boxplots demonstrate that there is a considerable geographical variation within regions. In Scotland although the ASRs tend to be high, there is a significant handful of local authorities that fall below the GB rate and there are some areas that have extremely high values over 150 (i.e. 50 percent above the GB rate).

Furthermore, Scotland has the largest range of ASR values of any of the regions. A similar trend can be seen for females (Figure 3.4) although the overall rates are lower compared to males. The regional differences are similar to those for males, with the North East of England and Scottish local authorities having some of the highest ASRs.

The geographical variations that exists within the UK are highlighted in Figures 3.5 and 3.6 which show maps of the age-standardised lung cancer rates for males and females in local authorities in the UK. There are clear variations in the rates across the UK for both males and females, although the highest rates tend to be found in the larger urban areas for both men and women. The local authorities with rates that fall in the highest quintile are found in parts of London, south Wales, north-west England (particularly Manchester and Liverpool), Yorkshire, north-east England and especially in central Scotland. By comparing the male and female rates in Scotland, it is noticeable that a larger number fall in the highest quintile for females than they do for males. In fact, for females almost the whole of Scotland falls within the highest two quintiles. This suggests that lung cancer is especially high in Scotland for females compared to the rest of the UK and compared to Scottish males. Thus, the consideration of regional variations in lung cancer rates has identified Scotland as one area of the UK that has consistently higher levels of lung cancer than most other parts of the UK.

Although the major differences in lung cancer rates within a country are between males and females, there are also some differences in the rates of the disease

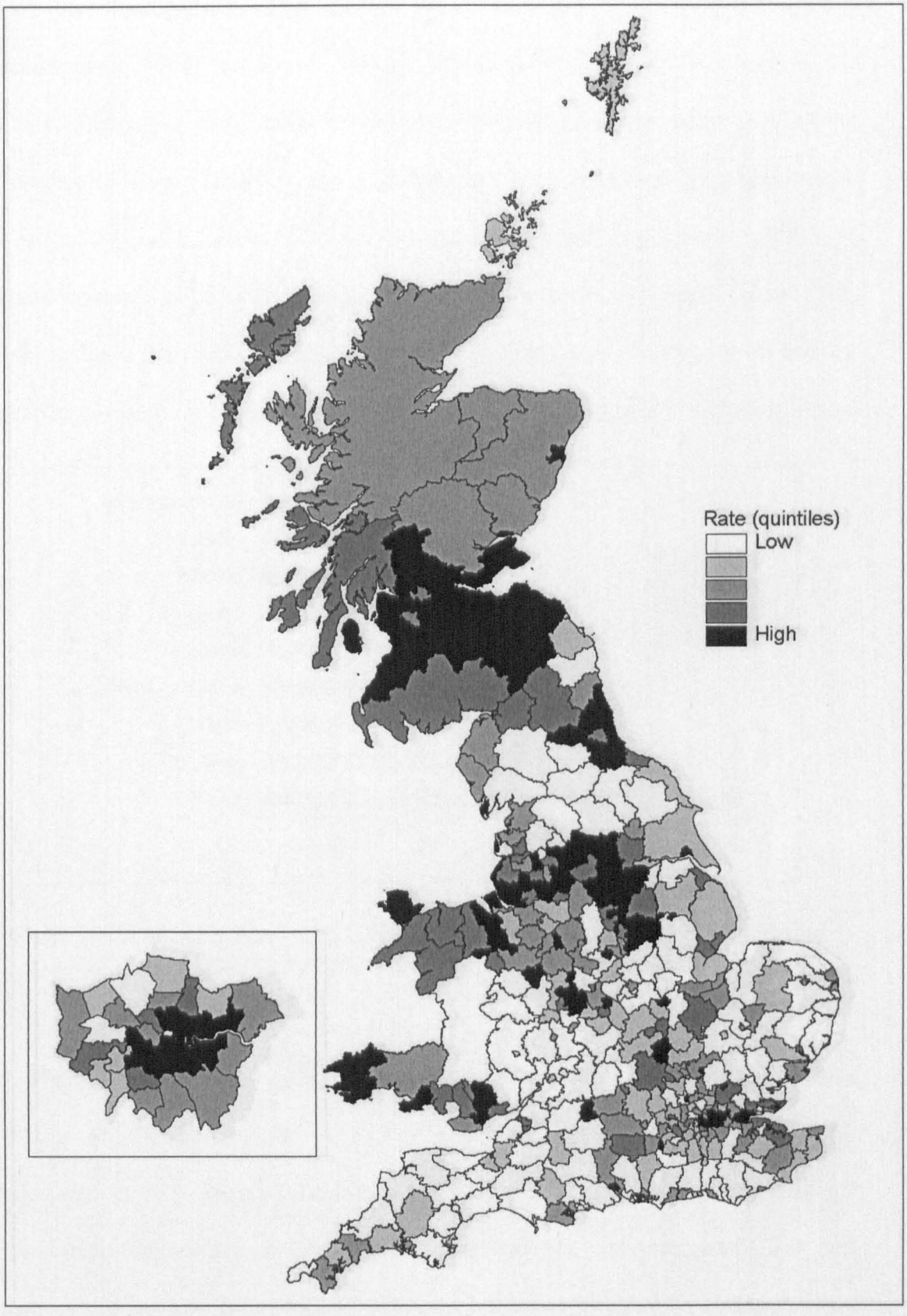


Figure 3.5. Age-standardised lung cancer rates by local authority grouped in quintiles, males all ages Great Britain 1991 to 1993 (National Statistics, 2002)

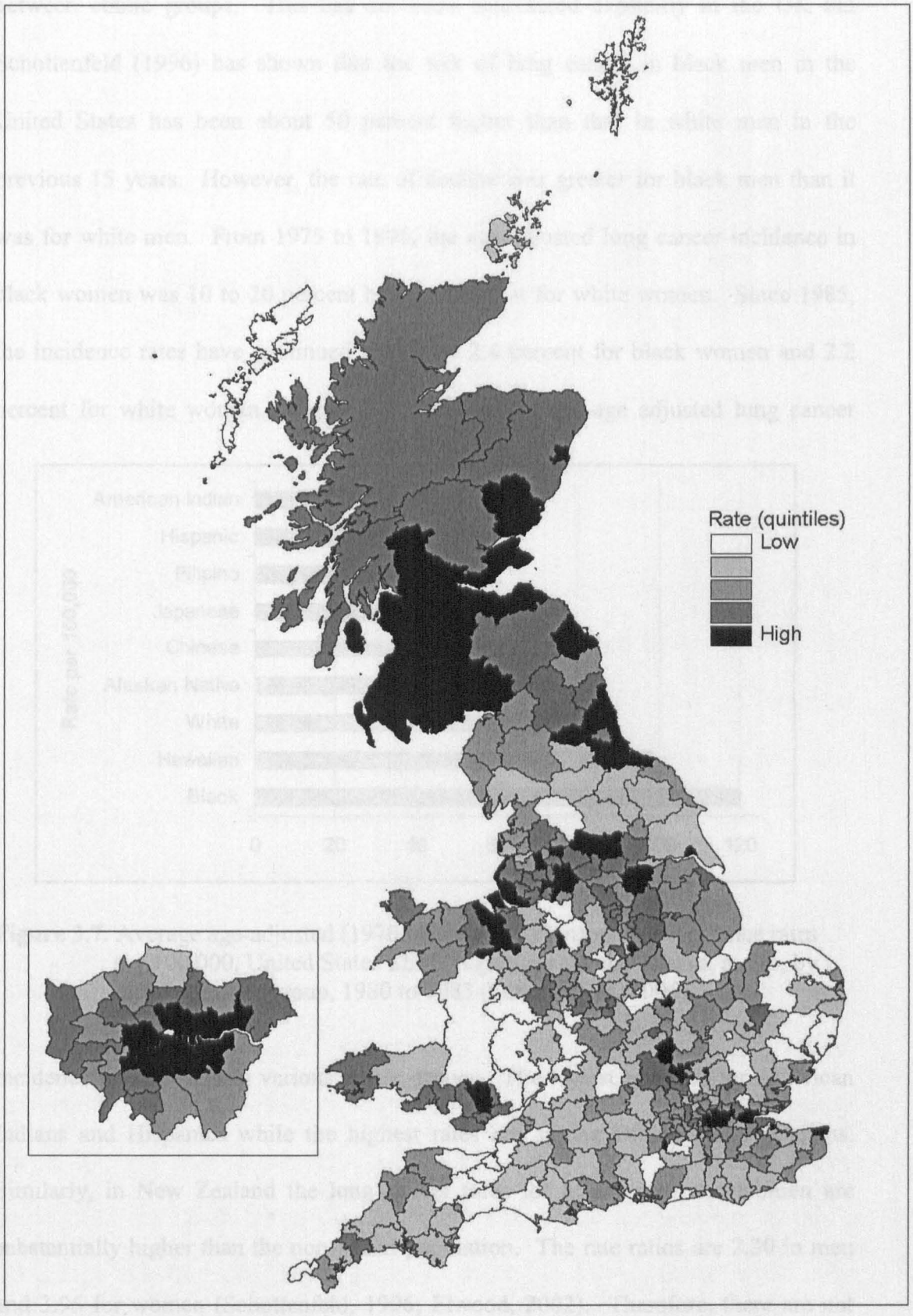


Figure 3.6. Age-standardised lung cancer rates by local authority grouped in quintiles, females all ages Great Britain 1991 to 1993 (National Statistics, 2002)

between ethnic groups. This has not been considered explicitly in the UK but Schottenfeld (1996) has shown that the risk of lung cancer in black men in the United States has been about 50 percent higher than that in white men in the previous 15 years. However, the rate of decline was greater for black men than it was for white men. From 1975 to 1990, the age adjusted lung cancer incidence in black women was 10 to 20 percent higher than that for white women. Since 1985, the incidence rates have continued to rise by 2.4 percent for black women and 2.2 percent for white women. Figure 3.7 demonstrates the age adjusted lung cancer

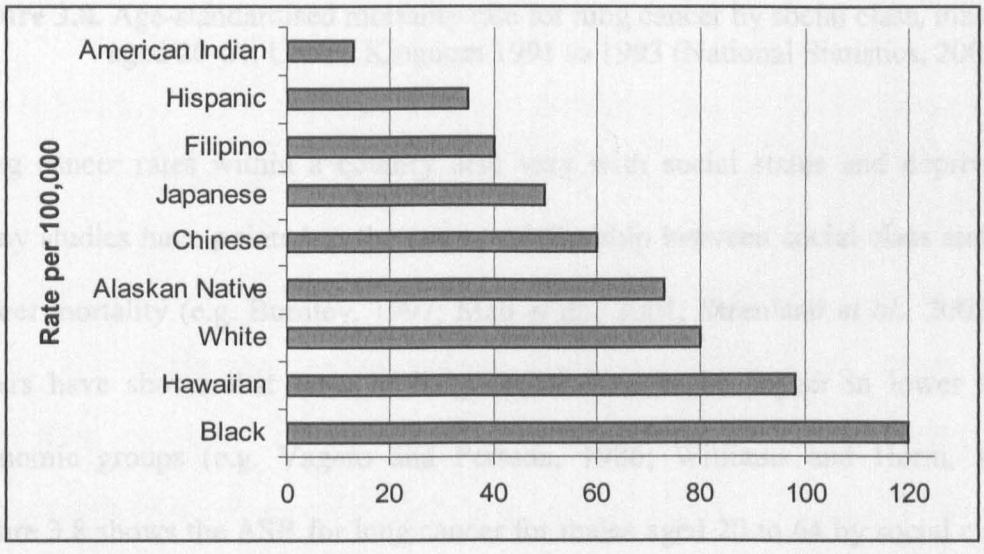


Figure 3.7. Average age-adjusted (1970 US standard) lung cancer incidence rates per 100,000, United States SEER registry areas and Alaska, males, by racial or ethnic group, 1980 to 1983 (Schottenfeld, 1996).

incidence for males from various ethnic groups. The lowest rates are for American Indians and Hispanics while the highest rates are among Blacks and Hawaiians. Similarly, in New Zealand the lung cancer rates for Maori men and women are substantially higher than the non-Maori population. The rate ratios are 2.30 in men and 3.96 for women (Schottenfeld, 1996; Elwood, 2002). Therefore, there are not only differences in lung cancer rates between ethnic groups but also differences between men and women in the same ethnic group (Payne, 2001).

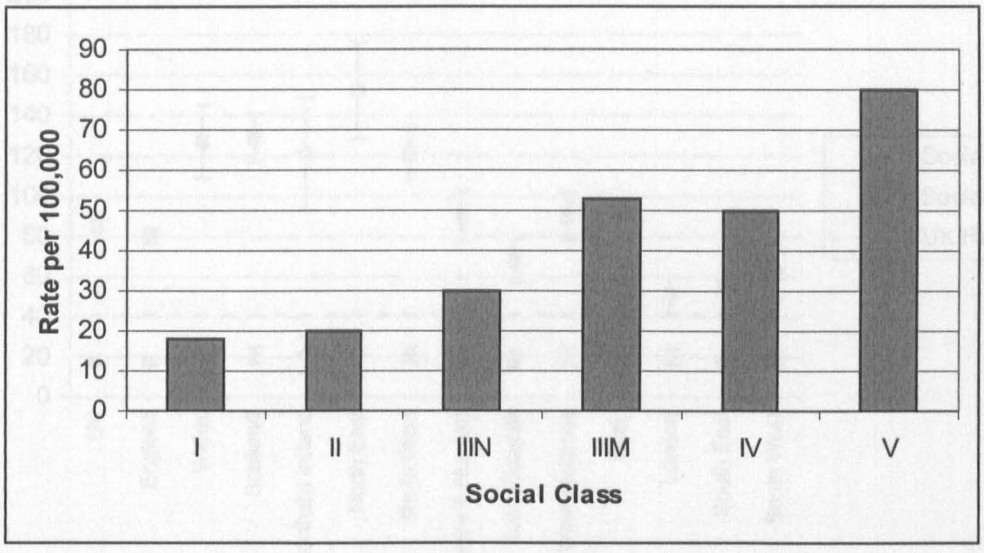


Figure 3.8. Age-standardised mortality rate for lung cancer by social class, males aged 20-64, United Kingdom 1991 to 1993 (National Statistics, 2002)

Lung cancer rates within a country also vary with social status and deprivation. Many studies have pointed to the strong relationship between social class and lung cancer mortality (e.g. Burnley, 1997; Mao *et al.*, 2001; Steenland *et al.*, 2002) and others have shown that rates of lung cancer tend to be higher in lower socio-economic groups (e.g. Vagero and Persson, 1986; Williams and Horm, 1997). Figure 3.8 shows the ASR for lung cancer for males aged 20 to 64 by social class in the UK. The ASR increases from Social Class I (professional) to Social Class V (unskilled) and the ASR in Social Class V is five times higher than it is in for Social Class I. However, Figure 3.9 demonstrates that there is a sharp geographical gradient in Social Class V with the lowest rate in the East of England and the highest rate in the North East. Scotland has a far higher rate than England and the UK as a whole and is marginally higher than Wales and Northern Ireland, although the rate is not as high as in the North East. This can be compared to Social Class I where there is little significant difference in rates between the regions and countries of the UK.

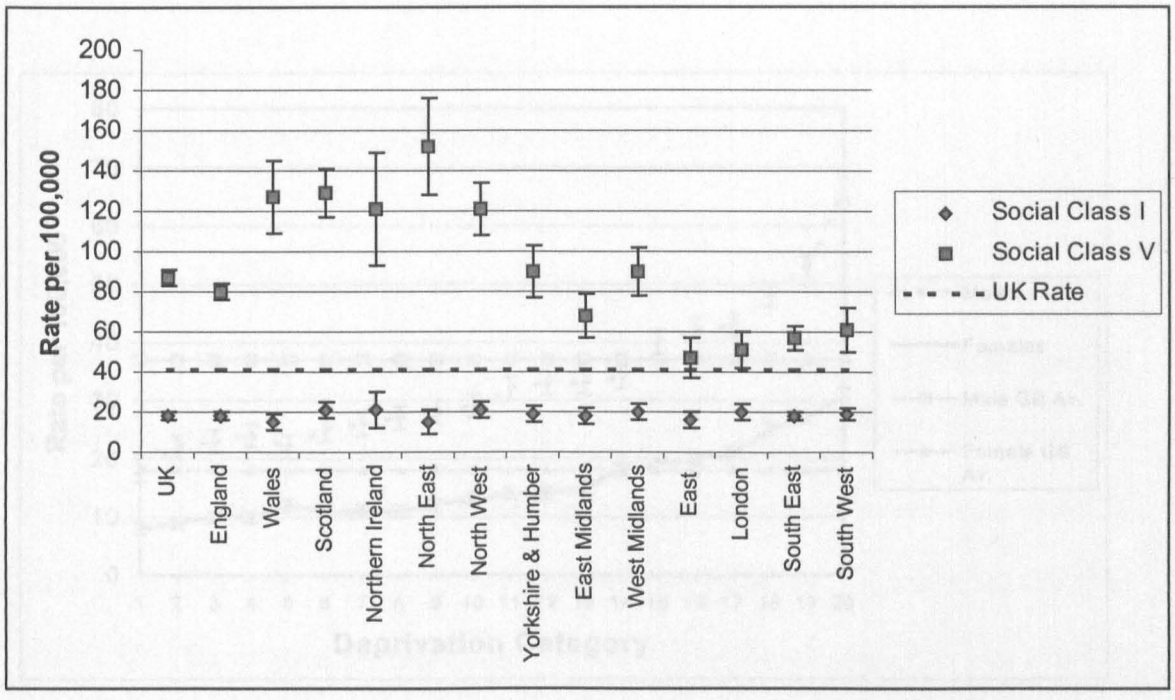


Figure 3.9. Age-standardised mortality rate for lung cancer for Social Classes I and V, lung cancer by country and region, males aged 20-64, United Kingdom 1991 to 1993 (National Statistics, 2002)

A similar pattern can be found when deprivation and lung cancer are compared. Figure 3.10 shows the age-standardised mortality rates for lung cancer by 20 deprivation classes in Great Britain. There is a clear gradient in mortality with higher levels of deprivation for males and females, a steady rise up to category 15 and then a sharper rise (especially for males) in the five highest categories. This is especially true for the highest level of deprivation where the difference in ASR between the nineteenth and twentieth category is approximately 15 for males and 7 for females when it is typically 4 or 5 for men and 2 or 3 for women.

Although there is a strong relationship between socio-economic status, deprivation and lung cancer, it is likely that the measures of socio-economic status and deprivation serve as surrogates for other risk factors such as smoking behaviour, occupation, diet and ambient air pollution. They may also influence the quality,

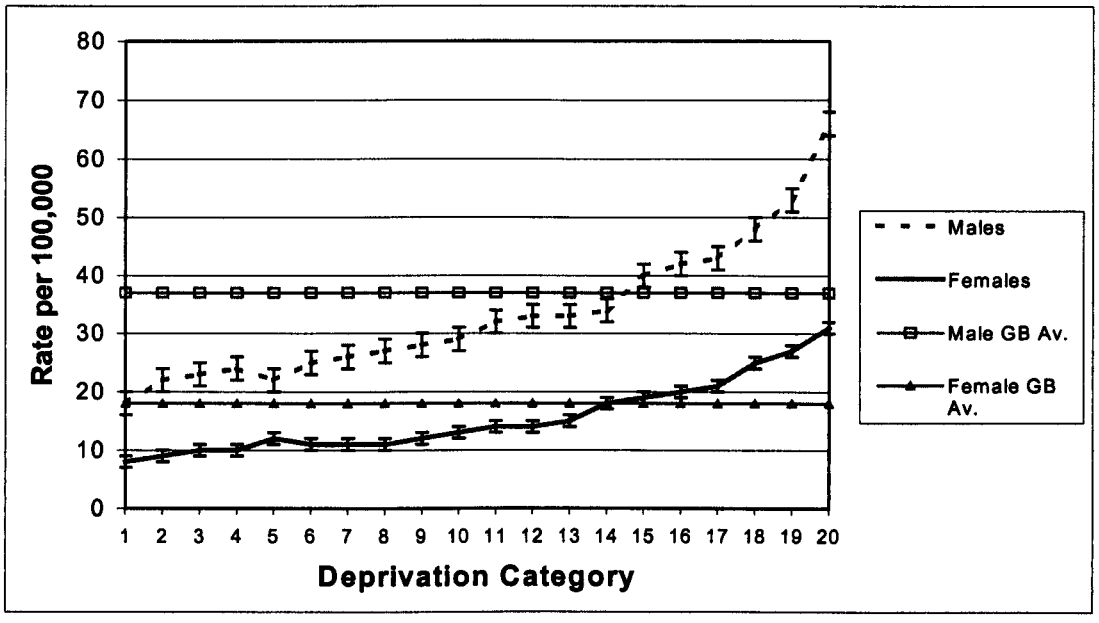


Figure 3.10. Age-standardised mortality rate for lung cancer by deprivation, ages 15-64, Great Britain, 1991 to 1993 (National Statistics, 2002)

access, and use of health care services (Schottenfeld, 1996). However, authors such as Hart *et al.* (2001) have suggested that the different rates of lung cancer between the social classes cannot be explained purely by the levels of smoking and the other known risk factors. Indeed, many studies of the relationship between lung cancer and deprivation have suggested that a relationship persists even once the risk factors are controlled for (e.g. Law and Morris, 1998; Hart *et al.*, 2001). This will be explored later in this study.

3.5 Lung Cancer Incidence in Scotland

It is of little surprise that some of the UK's highest lung cancer rates are in Scotland when one considers that it is faced with some of the greatest health challenges of any region of the UK using a range of measures (Carstairs and Morris, 1991). This is

especially true of lung cancer because the rates of incidence are not only the highest in the UK but among the highest in the world (WHO, 2002).

Lung cancer trends in Scotland have many parallels with the patterns that are found in the UK as a whole. It is the most frequently diagnosed cancer in men in Scotland accounting for 22.1 percent of new cases of invasive cancer in 1998. For women, it is the second most frequent cancer, accounting for 14.7 percent of new cancers (ISD, 2002). In 1997 there were a total of 2,696 new registrations of lung cancer for men and 1,887 for women. Figure 3.11 shows the ASRs of cancer of the trachea, bronchus and lung from 1975 to 1998 in Scotland. The incidence of lung cancer is

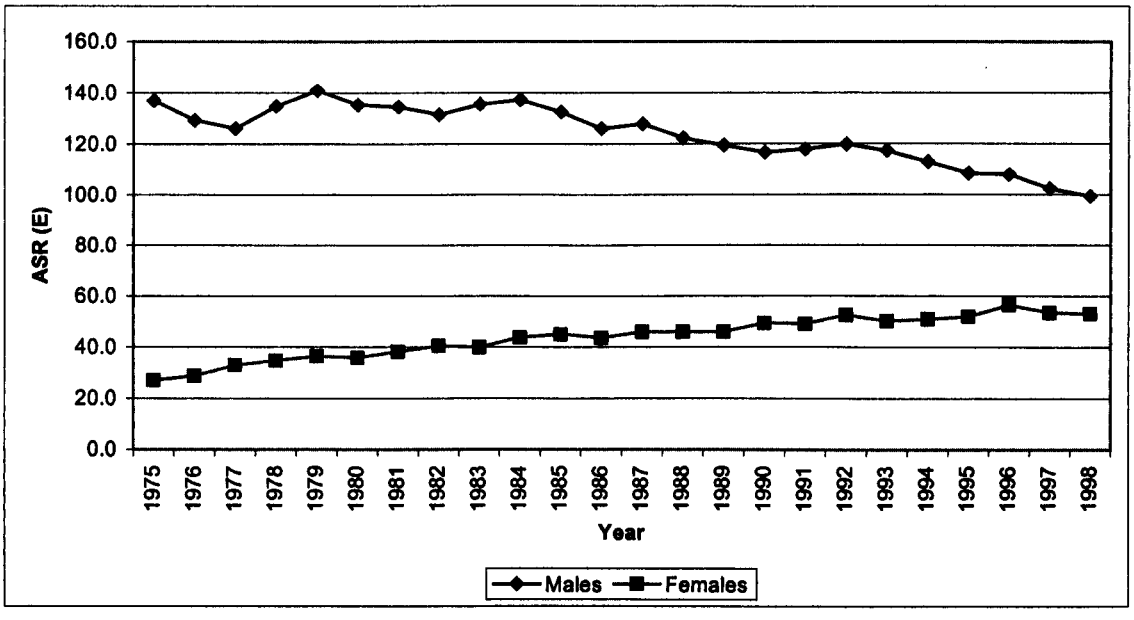


Figure 3.11. Incidence of cancer of the trachea, bronchus and lung in Scotland 1975-1998 - Age Standardised Incidence Rate per 100,000 person-years at risk (European standard population) (ISD, 2002)

consistently higher for men than it is for women throughout this period. The percentage risk of contracting lung cancer in Scotland (assuming no other cause of death is in operation) up to the age of 74 (using data from 1986 to 1995) is 9.8 percent for men and 4.4 percent for women. It is important to note that there has

been a consistent decline in lung cancer incidence in men and gradual increase in incidence for women in Scotland. The ASR for men has fallen from around 140 in 1975 to about 58 in 1998. This decline has been seen through all age groups up to the age of 74 with a particular decrease in the 40 to 44 age group (Williams and Lloyd, 1991). On the other hand, the ASR for women has risen from about 30 in 1975 to around 55 in 1998. In Scotland during the period 1989 to 1998, the change in the lung cancer rates saw a 16.1 percent decrease in men and a 21.4 percent increase for women (ISD, 2002). It is likely that this is a reflection of changing smoking habits among both men and women.

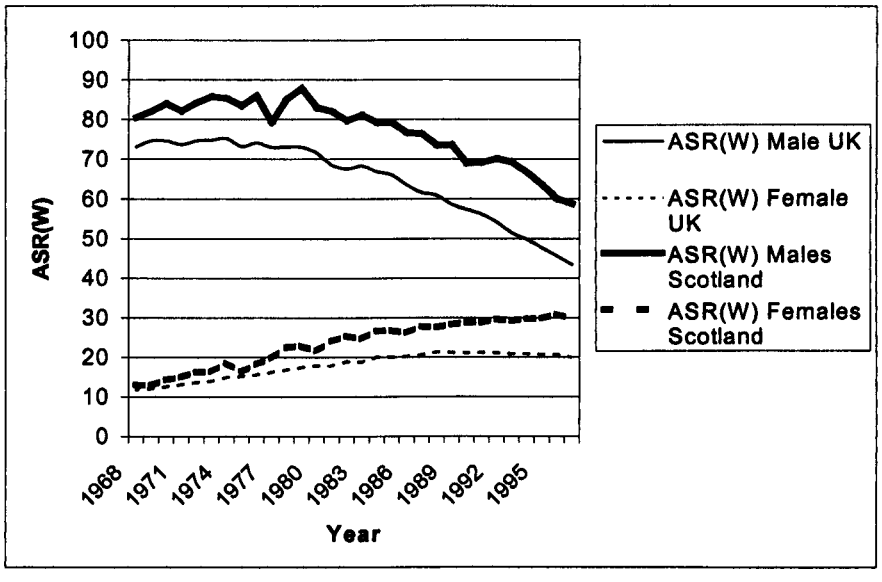


Figure 3.12. Age Standardised Rate for trachea, bronchus and lung cancer incidence 1968-1997 in the UK per 100 000 person-years at risk (world standard population) (WHO, 2002)

The trend in lung cancer incidence rates is mirrored in the changes that have been taking place in the UK as a whole. Figure 3.12 shows the change in ASR from 1968 to 1998 for men and women in the whole of the UK and also for Scotland on its own. Again it can be seen that the incidence of lung cancer is consistently higher in men that it is for women and also that the rates seem to be converging as the rate for men

declines whilst for women it has increased. It can also be noted that the ASRs are consistently higher in Scotland than they are for the entire UK for both men and women throughout the time period. This is reiterated in Table 3.2 which shows the lung cancer incidence rates for the UK, England, Wales, Northern Ireland and Scotland in 1996 (WHO, 2002). The table provides the crude incident rate for males and females (number of cases per 100,000 people) and the ASR rate using the world population as the standard population which allows lung cancer incidence rates to be compared between countries. Lung cancer incidence is consistently higher in Scotland than in any other country in the UK and for the UK as a whole using either measurement. This makes an examination of the incidence of lung cancer in Scotland especially pertinent.

	Males		Females	
	Crude Rate*	**ASR (W) ****	**Crude Rate***	**ASR (W) ****
Scotland	98.40	60.09	63.54	30.86
England and Wales	77.38	44.22	41.71	19.60
Northern Ireland	65.05	46.00	33.42	19.22
UK	78.86	45.62	43.40	20.61

* Rate per 100,000
** Age standardised rate (ASR) using world standard population

Table 3.2. Rates of lung cancer in the UK in 1996 amongst males and females (WHO, 2002)

The rate of lung cancer in Scotland is not only the highest in the UK but is also amongst the highest in the world. Tables 3.3 and 3.4 place the incidence of lung cancer into an international context by showing the highest rates of lung cancer in the world (World Health Organisation, 2002). The tables show the crude incidence rate and the ASR for the five countries with the highest lung cancer rates and for Scotland in 1996, ranked according to their lung cancer ASR. The year 1996 was

chosen because although data for subsequent years are available in Scotland, it is not available in all countries. For males, the crude incidence is one of the highest in the world and once the incidence rates are age-standardised Scotland comes tenth in the rankings behind a set of countries that are dominated by the states of the former Soviet Union. For women, Scotland has the highest incidence of lung cancer of any country in the world regardless of whether the incidence is measured using a crude incidence rate or ASR.

Country	Crude Rate (per 100,000)	ASR (W)
1. Hungary	120.47	85.78
2. Croatia	91.37	70.48
3. Poland	82.38	70.32
4. Czech Republic	89.40	65.91
5. Russian Federation	75.73	65.48
10. Scotland	98.40	60.09

Table 3.3. The highest rates of lung cancer amongst males in 1996 (WHO, 2002)

Country	Crude Rate (per 100,000)	ASR (W)
1. Scotland	63.54	30.86
2. Denmark	48.55	27.61
3. USA	44.58	26.65
4. Iceland	38.81	25.85
5. Canada	38.67	24.58

Table 3.4. The highest rates of lung cancer amongst females in 1996 (WHO, 2002)

Although it is clear that Scotland has a high incidence of lung cancer compared to the rest of the UK and also compared to many other countries, the rates are not uniform across the country. Figures 3.13 and 3.14 show the ASRs of lung cancer incidence in Health Board Regions in Scotland for both males and females. There are clear regional variations in the incidence of lung cancer in Scotland and for both males and females the higher incidence rates are in the central Health Board Regions of Greater Glasgow, Lothian, Argyll and Clyde, Ayrshire and Arran and Lanarkshire. For males, the rates of lung cancer are lower in the Borders region,

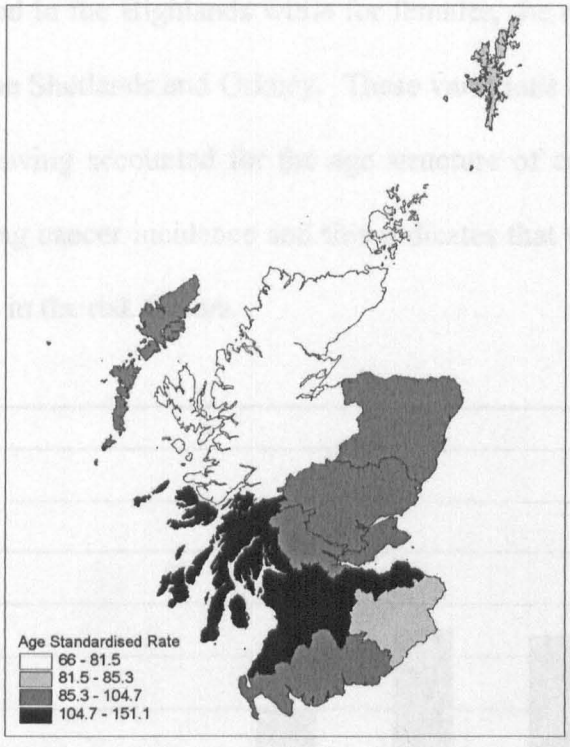


Figure 3.13. Age Standardised Incidence Rate of Cancer of the Trachea, Bronchus and Lung amongst males (European standard population) in Health Board Regions in Scotland (ISD, 2002)

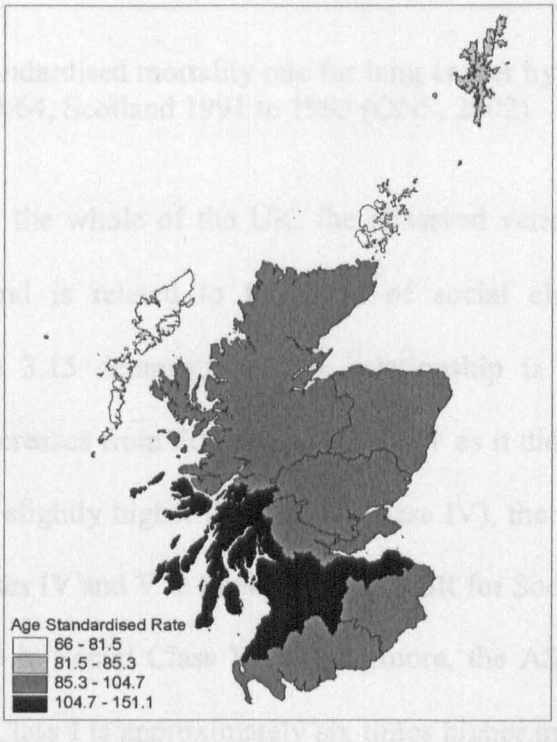


Figure 3.14. Age Standardised Incidence Rate of Cancer of the Trachea, Bronchus and Lung amongst females (European standard population) in Health Board Regions in Scotland (ISD, 2002)

Shetland, Orkney and in the Highlands while for females, the rates are lower in the Western Isles and the Shetlands and Orkney. These variations are important as they indicate that even having accounted for the age structure of each region, there are still variations in lung cancer incidence and this indicates that there are likely to be important variations in the risk factors.

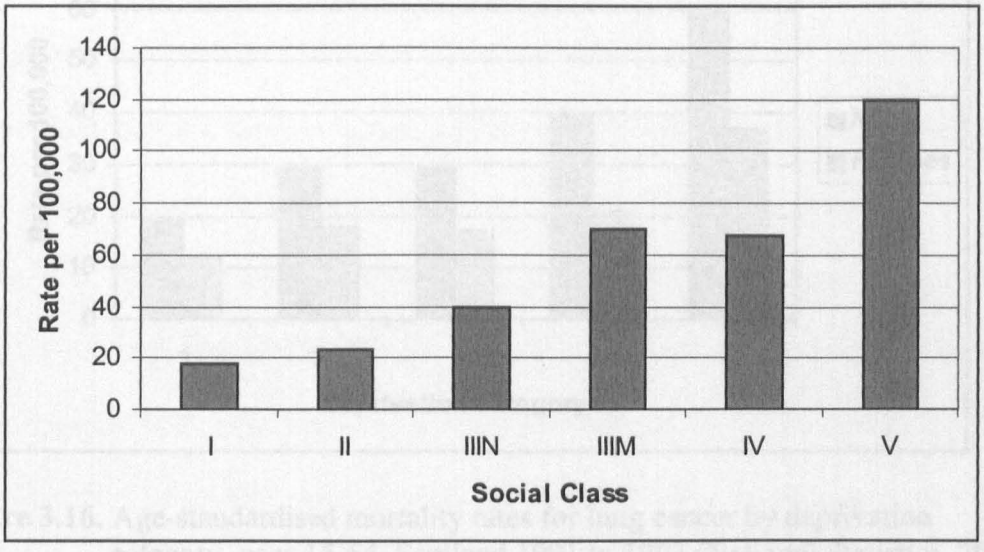


Figure 3.15. Age-standardised mortality rate for lung cancer by social class, males aged 20-64, Scotland 1991 to 1993 (ONS, 2002)

As was the case for the whole of the UK, the observed variations in lung cancer incidence in Scotland is related to the level of social class and deprivation. However, as Figure 3.15 demonstrates, the relationship is slightly different as although the ASR increases from Social Classes I to V as it did in the UK (although Social Class IIIM is slightly higher than Social Class IV), there is a larger increase between Social Classes IV and V in Scotland. The ASR for Social Class IV is 68 but this increases to 120 in Social Class V. Furthermore, the ASR in Social Class V compared to Social Class I is approximately six times higher in Scotland whereas in the entire UK it was only four times higher. This suggests that the unskilled in Scotland are more likely to suffer from lung cancer than the unskilled in the rest of

the UK. Similarly, there is a subtly different relationship between lung cancer and deprivation when it is considered for Scotland rather than for the whole of Great Britain. Figure 3.16 shows that the ASR for males and females in Scotland increases through five categories of deprivation and this effect is

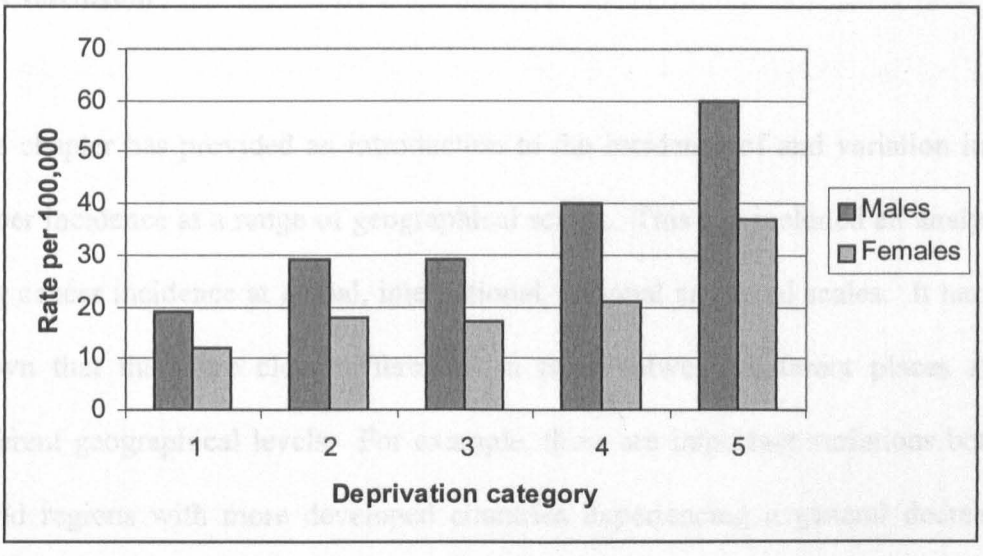


Figure 3.16. Age-standardised mortality rates for lung cancer by deprivation category, ages 15-64, Scotland 1991 to 1993 (National Statistics, 2002)

even greater for men than it is for women. For women, the ASR steadily increases from 12 to 36 per from deprivation categories one to five. For men, the increase is more dramatic as the ASR is 18 in the lowest category and 60 in the highest category. These observations are similar to the pattern that was observed for all of Great Britain except that the rates are higher, especially in the highest deprivation category where there is a large increase between deprivation categories four and five. The ASR for males increases from 40 in deprivation category four to 60 in deprivation category five. These graphs support the evidence found at a national level of a strong relationship between social class, deprivation and lung cancer. However, they also demonstrate that the relationship is different when considering Scotland rather than the whole of the UK. At both scales, it is not clear whether this

relationship can be explained by a compositional effect that is a product of the individual characteristics of the people living in these areas. This will be explored more thoroughly later in this study.

3.6 Conclusion

This chapter has provided an introduction to the incidence of and variation in lung cancer incidence at a range of geographical scales. This has included an analysis of lung cancer incidence at global, international, national and local scales. It has been shown that there are clear differences in rates between different places and at different geographical levels. For example, there are important variations between world regions with more developed countries experiencing a general decrease in lung cancer rates whilst less developed countries have experienced a general increase. At the same time it has been shown that there are more localised variations within countries such as between the regions of the UK.

However, certain features of lung cancer incidence are common across all scales such as that rates of the disease are higher for men than women, and higher for non professional social classes and more deprived areas. It has also been demonstrated that these features have specific attributes in different geographical areas. For example the ratio between the male and female rates varies geographically with some places having relatively similar rates for both sexes whilst others have markedly different rates. Similarly, there is a clear relationship between lung cancer and deprivation for all regions of the UK but here too there are interesting differences depending on the study area. This highlights how important it is to

consider the geography of lung cancer as there are important variations in lung cancer from place to place that will reflect geographical differences in the associated risk factors.

This chapter has also focused upon lung cancer incidence in Scotland and shown that the rates of lung cancer in Scotland are the highest in the UK and among the highest in the world. Furthermore, there are clear geographical variations within Scotland, with a greater prevalence of lung cancer in the more urban areas and the south-central region more generally. It has been shown that the key social descriptions of lung cancer such as sex, class and deprivation have properties that are unique to Scotland. The next chapter will develop the discussion of lung cancer incidence in Scotland using a geographically specific dataset on lung cancer cases. This will provide further insight into the pattern of lung cancer in Scotland.

4. THE GEOGRAPHY OF LUNG CANCER IN SCOTLAND

4.1 Introduction

This chapter provides an introduction to the geography of lung cancer incidence in Scotland. The intention is to compare the levels of lung cancer incidence in small areas in order to gain an insight into the geographical variability of the disease across Scotland using data that have been specially provided by the Information and Statistics Division (ISD) of NHS Scotland for the period 1988 to 1991. First, the lung cancer incidence dataset is described. Second, some descriptive statistics are used to describe the data. Third, the data are mapped and some comparisons between the incidence of lung cancer and deprivation are made.

The remainder of the chapter will develop a set of exploratory techniques to move beyond description to provide some further insights into the nature of lung cancer in Scotland. First, age standardised rates are constructed to allow different areas to be compared. This is suitable for areas where the population is sufficiently large and when big differences in the results are not a product of small populations. Therefore, in addition, Poisson probabilities are used to examine lung cancer incidence in smaller areas where the population is lower. This identifies areas of higher then expected lung cancer incidence and is an appropriate technique when the counts are small and it is unreliable to calculate age-standardised rates. The two methods are complementary because each is appropriate at different scales and may result in different conclusions.

The final section of this chapter uses some cluster detection techniques to identify clusters of cases of lung cancer once the age structure of the population is accounted for. Cluster detection offers a different method of exploratory analysis as it considers the relationship of lung cancer incidence between areas that are close by. This is useful because it helps to identify areas where there are many high rates of lung cancer close together and this has the potential to provide useful insights into some of the factors that may influence lung cancer incidence. Furthermore, this combination of exploratory techniques has useful implications for public health policy as the results can be used to identify areas, and clusters of areas, that have high rates of lung cancer having controlled for the population structure of each area.

4.2 Lung Cancer Data in Scotland

The previous chapter provided an introduction to the nature of lung cancer incidence over the past 30 years and considered it within a national and international context. This section focuses upon data on lung cancer incidence in Scotland for a specific period and in greater geographical detail. The dataset on lung cancer incidence has been provided by the Information and Statistics Division (ISD) of NHS Scotland for the period 1988 to 1991 and this is the first exploration of such geographically detailed data. The ISD collect, validate, interpret and disseminate health related data from Unified Boards, NHS trusts and General Practices. The data on lung cancer patients include information on their age at diagnosis, their sex, the census unit in which they lived, the tumour morphology and the tumour site. The date of diagnosis is defined as being the date of the first consultation or admission to hospital for the

cancer (ISD, 2002). There were 18,632 lung cancer patients in this period, of which 12,073 cases were male and 6,559 female.

Although it was not possible to gain access to individual postcodes for each patient, due to confidentiality restrictions, the census output-area identifier was provided. The census in Scotland is disseminated at four different levels: the region, the district, the pseudo postcode sector (pseudo-sector) and the output-area (Dale, 1993). The largest geographical unit is the region / island area of which there are 12 across Scotland. Each region represents around 150,000 households. These divide into 56 local government districts, each representing around 35,000 households. The next level down is the pseudo postcode sector, which is commonly treated as being equivalent to the ward in England and Wales. These are not necessarily true postcode sectors as they were specially created by the census office to recognise local government district boundaries and hence they are referred to as 'pseudo' postcode sectors (PPSs). There are 1,003 PPSs in Scotland each of which represents approximately 2,000 households. The smallest unit of dissemination in the Scottish census is the output-area which aggregate neatly into PPSs. There are 38,254 output-areas (OAs) each of which contains approximately 50 households. This hierarchy is shown in Table 4.1. Therefore, the association of the OA identifier with each record allowed the individual to be placed in a relatively specific geographical area for which census data were available. As Table 4.1 demonstrates, the PPS code can be identified from the first six characters of the OA code and hence data that are disseminated for PPSs can also be associated with each record.

Level	No areas	Av. No households	Example Code	Example Name
Region	12	150,000	57	Central Region
District	56	35,000	5705	Clackmannan District
Pseudo- Postcode Sector	1003	2,000	5705AC	Postcode Sector 'FK1 01'
Output-Area	38,254	50	5705AC10A	n/a

Table 4.1. Hierarchy of units of dissemination in the Scottish Census

At the time of diagnosis, the 18,632 lung cancer patients lived in 13,434 different OAs out of the total of 38,254 OAs that encompass Scotland. The highest number of cases in any OA was 12 whilst the lowest was zero. The cases fell within 897 of the 1003 PPSs and the highest count of cases in a PPS was 90. Figure 4.1 shows that the distribution of the count of cases of lung cancer in each PPS was positively skewed.

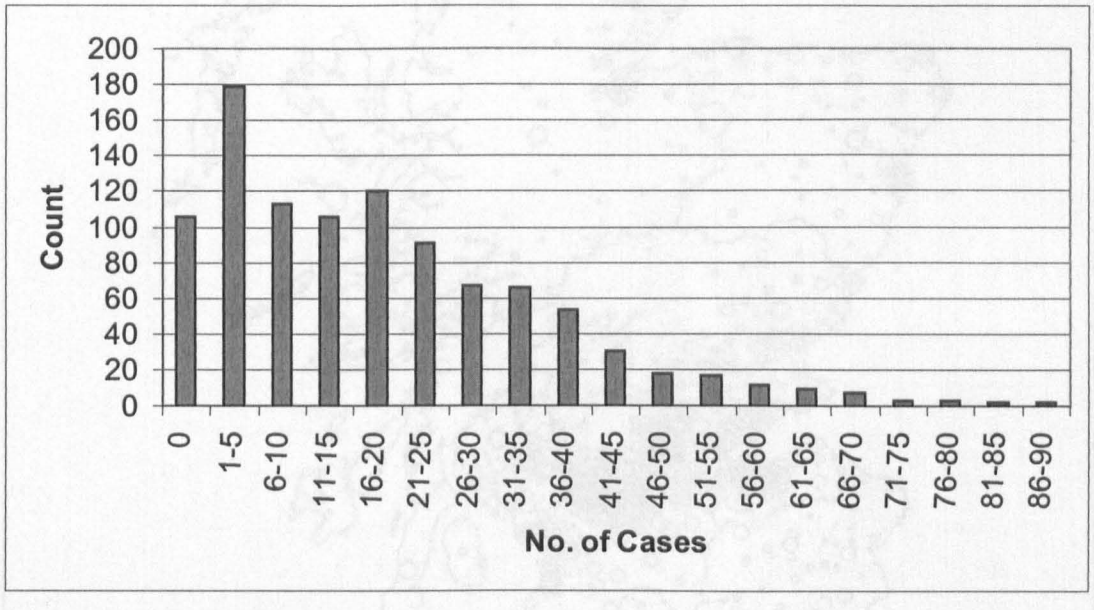


Figure 4.1. Count of cases of lung cancer in each PPS in Scotland.

Figure 4.2 maps the count of lung cancer in each PPS in Scotland during this period using circles drawn proportional to the count of cases. The highest numbers of lung cancer cases were in PPSs in Glasgow, Edinburgh, Dundee and Aberdeen. There were also some high counts across the central belt. The more isolated larger circles

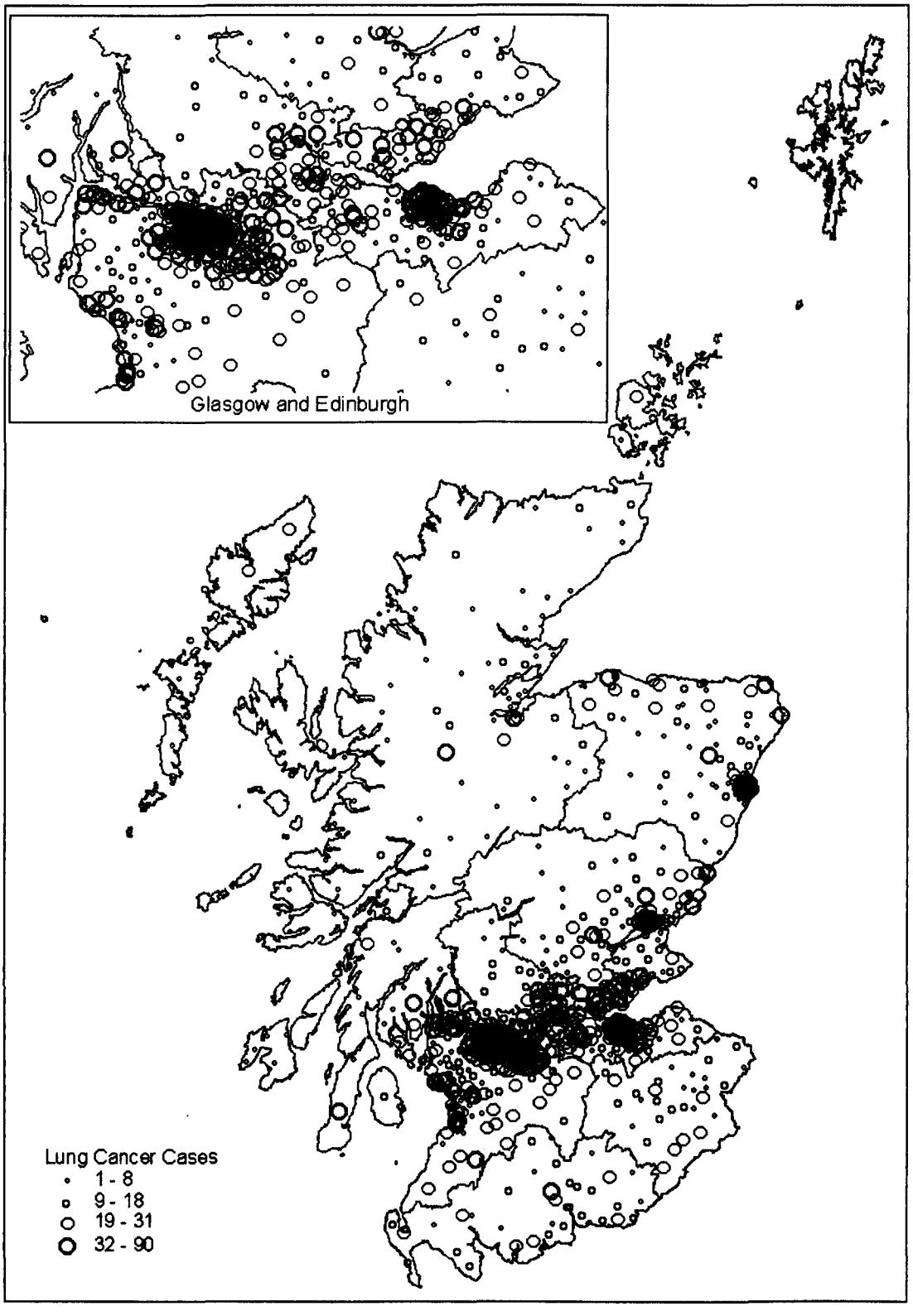


Figure 4.2. Count of cases of lung cancer in PPSs in Scotland during the period 1988 to 1991.

tended to identify the medium sized towns. Clearly, this map simply reflects the fact that the majority of people live in urban areas which hence have the majority of lung cancer cases. Therefore, the map is not very useful in terms of drawing any conclusions about the nature of lung cancer but does demonstrate that most lung cancer patients live in urban areas.

Of course, lung cancer rates do not only vary spatially and temporally but also within subgroups of the population. The number of cases of lung cancer in a PPS will not just be dependent upon the total population and exposure to any causal factors but also upon the demographic profile of the population. This can be seen in Figure 4.3 which shows the age distribution of the lung cancer patients within the dataset. Lung

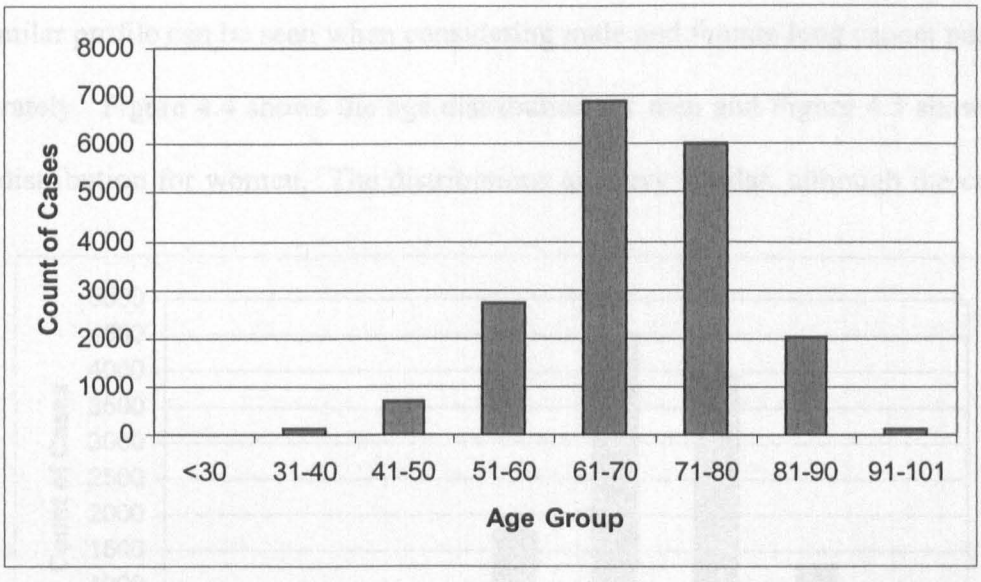


Figure 4.3. Age distribution of lung cancer cases in Scotland between 1988 and 1991.

cancer is a disease of older age with relatively few cases under 50 and the majority occurring among those aged 61 to 70 and 71 to 80. Table 4.2 shows that of the 18,632 cases in this period, only 827 (4.3 percent) of the patients were aged 50 or

under. A total of 12,901 of the patients were aged between 61 and 80, which is 69.2 percent of the total cases. The youngest lung cancer patient was 16 and the oldest was 101. The fact that there were fewer cases in the older categories is in part due to the smaller populations in these age groups.

Age	Male	Female	Total
Under 30	10	7	17
31-40	59	52	111
41-50	423	276	699
51-60	1771	968	2739
61-70	4502	2374	6876
71-80	3955	2070	6025
81-90	1303	743	2046
91-101	50	69	119
All Ages	12073	6559	18632

Table 4.2. Count of cases of lung cancer (1988-1991) in age-sex groups.

A similar profile can be seen when considering male and female lung cancer patients separately. Figure 4.4 shows the age distribution for men and Figure 4.5 shows the age distribution for women. The distributions are very similar, although the counts

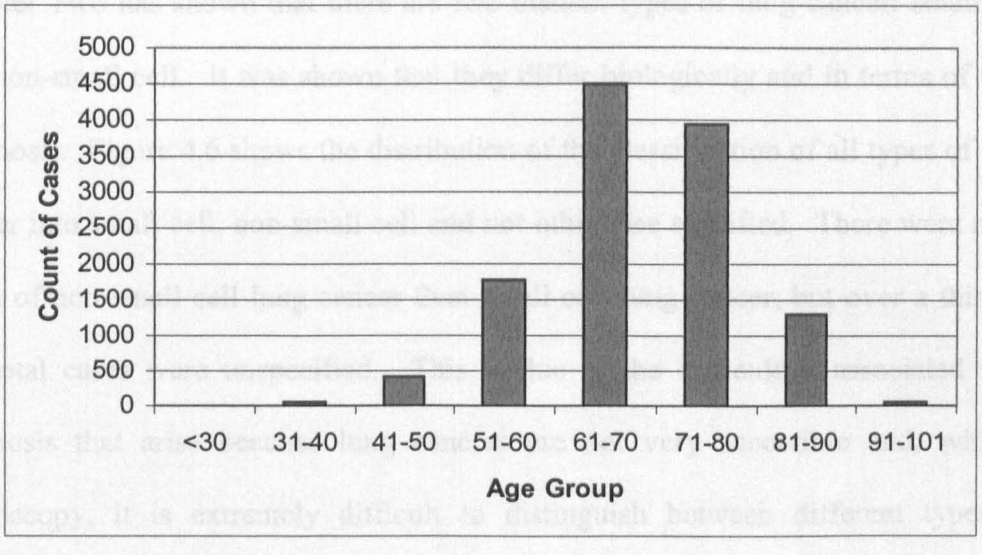


Figure 4.4. Age distribution of male lung cancer cases in Scotland between 1988 and 1991.

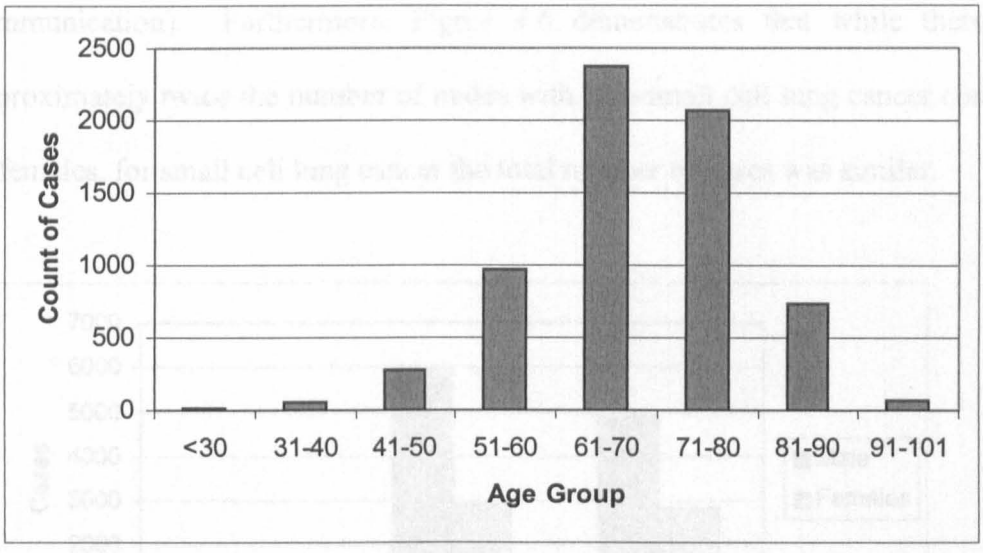


Figure 4.5. Age distribution of female lung cancer cases in Scotland between 1988 and 1991.

in each group are higher for the men. The age profiles of the lung cancer patients demonstrate the importance of controlling for age in a study of lung cancer; clearly, there is likely to be a higher incidence of lung cancer in areas that have higher numbers of older people, especially those aged over 60.

Chapter Two has shown that there are two distinct types of lung cancer: small cell and non-small cell. It was shown that they differ biologically and in terms of their prognosis. Figure 4.6 shows the distribution of the classification of all types of lung cancer into small cell, non-small cell and not otherwise specified. There were more cases of non-small cell lung cancer than small cell lung cancer, but over a third of the total cases were unspecified. This is due to the difficulties associated with diagnosis that arise because lung cancers are not very accessible and, without microscopy, it is extremely difficult to distinguish between different types of tumours. Even when microscopic verification is used, the specimen is often so poor that it is still impossible to classify the specimen (Brewster, personal

communication). Furthermore, Figure 4.6 demonstrates that while there were approximately twice the number of males with non-small cell lung cancer compared to females, for small cell lung cancer the total number of cases was similar.

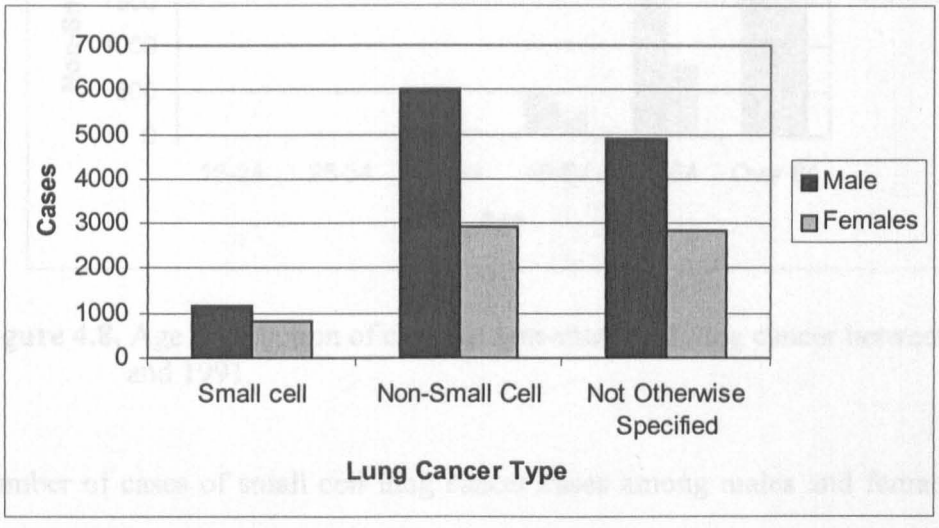


Figure 4.6. Distribution of small cell, non-small cell and not otherwise specified cases of lung cancer in Scotland between 1988 and 1991.

Figures 4.7 and 4.8 show the age distribution of male and female cases of small cell and non-small cell lung cancer. These graphs reiterate that there was a similar

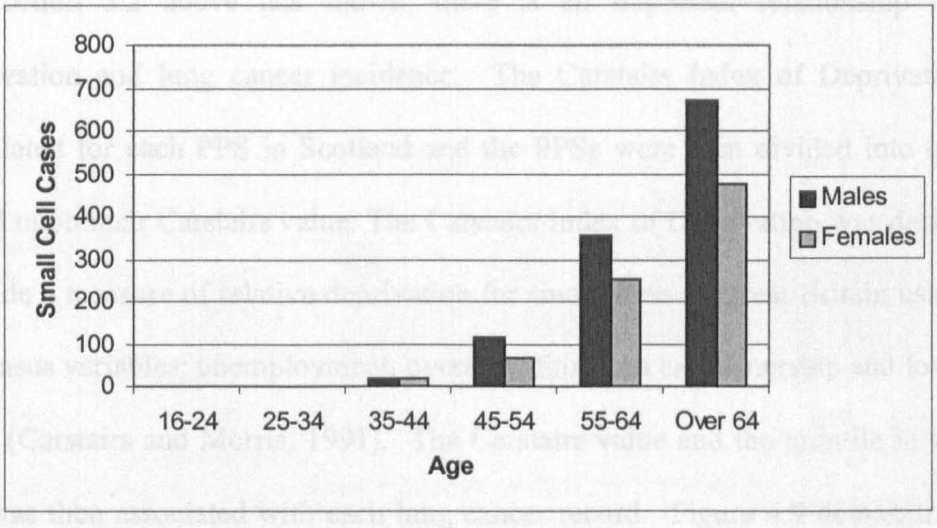


Figure 4.7. Age distribution of cases of small cell lung cancer between 1988 and 1991.

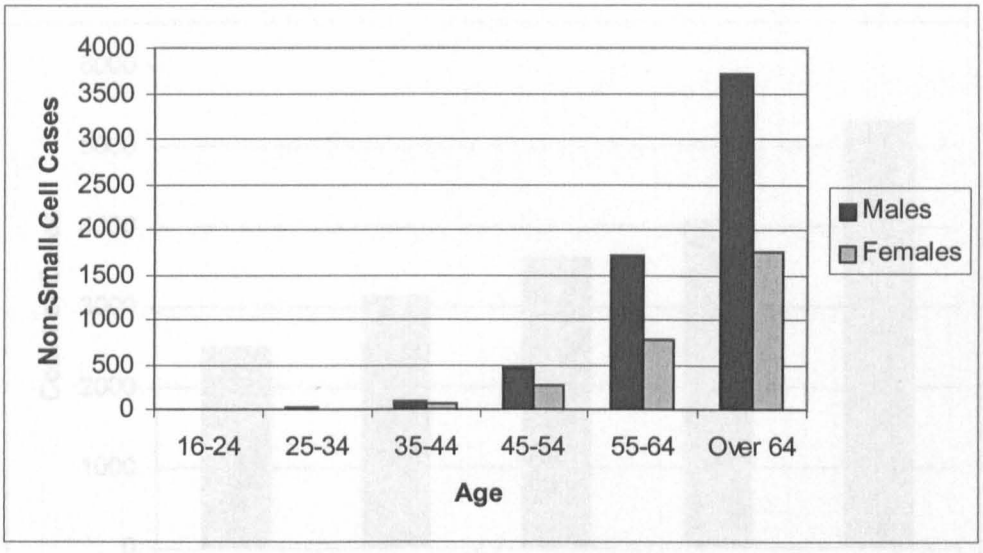


Figure 4.8. Age distribution of cases of non-small cell lung cancer between 1988 and 1991.

number of cases of small cell lung cancer cases among males and females but that there were many more males that are diagnosed with non-small cell lung cancer than there were females. Furthermore, the total number of cases of non-small cell lung cancer was higher than for small cell lung cancer. The two graphs also demonstrate that these observations were consistent across all age groups.

As section 3.2 above has shown, there is an important relationship between deprivation and lung cancer incidence. The Carstairs Index of Deprivation was calculated for each PPS in Scotland and the PPSs were then divided into quintiles based upon their Carstairs value. The Carstairs Index of Deprivation was designed to provide a measure of relative deprivation for small areas in Great Britain using a set of census variables: unemployment, overcrowding, non car ownership and low social class (Carstairs and Morris, 1991). The Carstairs value and the quintile in which it fell was then associated with each lung cancer record. Figure 4.9 demonstrates the relationship between deprivation and the count of the number of cases of lung

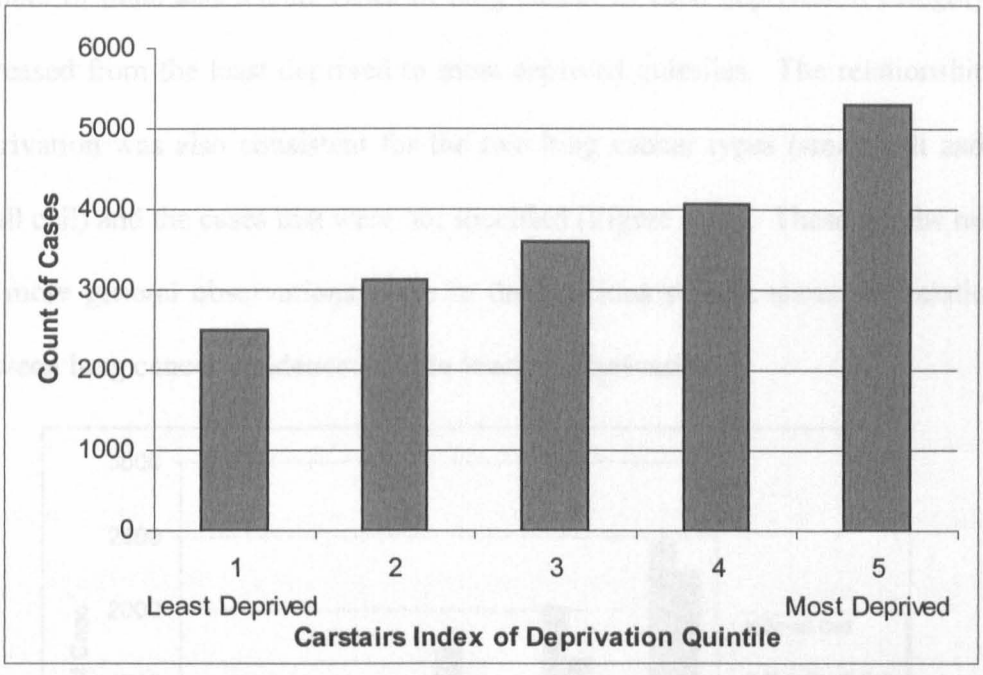


Figure 4.9. Count of cases of lung cancer in five Carstairs Index of Deprivation categories.

cancer. With each increase in Carstairs quintile from the least deprived to the most deprived there was a gradual increase in the count of the number of cases of lung cancer. There were 2,512 cases in the lowest quintile and 5,302 in the highest quintile. Figure 4.10 shows that this relationship was consistent for both sexes as the

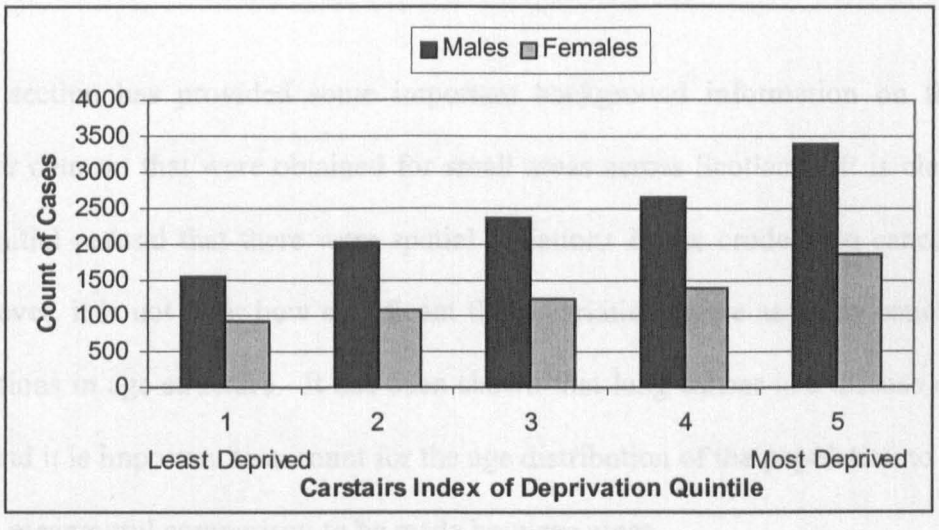


Figure 4.10. Count of cases of lung cancer in five Carstairs Index of Deprivation categories for males and females.

number of male and female cases of lung cancer in each deprivation category both increased from the least deprived to most deprived quintiles. The relationship with deprivation was also consistent for the two lung cancer types (small cell and non-small cell) and the cases that were not specified (Figure 4.11). These graphs reiterate the more general observations made in the previous section about the relationship between lung cancer incidence and the level of deprivation.

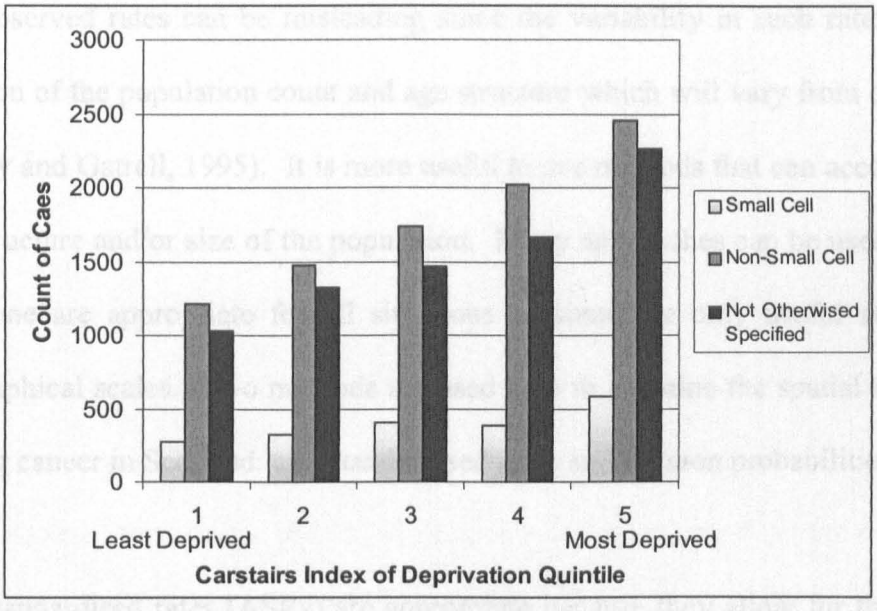


Figure 4.11. Count of cases of lung cancer in five Carstairs Index of Deprivation categories for small cell, non-small cell and not otherwise specified cases of lung cancer.

This section has provided some important background information on the lung cancer data set that were obtained for small areas across Scotland. It is clear from the initial perusal that there were spatial variations in the crude lung cancer rates. However, it is not clear how significant these variations were as crude counts mask variations in age structure. It has been shown that lung cancer is a disease of older age and it is important to account for the age distribution of the population to allow a more meaningful comparison to be made between areas.

4.3 Analysis of Lung Cancer Incidence

4.3.1 Introduction

Although the mapping of disease incidence data to display geographical variability is a useful first step in gaining an understanding of the geography of lung cancer, the raw observed rates can be misleading since the variability in such rates will be a function of the population count and age structure which will vary from area to area (Bailey and Gatrell, 1995). It is more useful to use methods that can account for the age structure and/or size of the population. Many approaches can be used to do this but none are appropriate for all situations as some are only useful at particular geographical scales. Two methods are used here to examine the spatial distribution of lung cancer in Scotland: age-standardised rates and Poisson probabilities.

Age-standardised rates (ASRs) are appropriate because they allow for the structure and size of the population to be taken into account when the disease and population data are available for the same areas. Furthermore, they provide a measure that health researchers are familiar with and can easily interpret. However, when the disease of interest is rare or the analysis is undertaken for small areas with low populations the methods can be less appropriate as the results can be absurdly variable. One method to counter this is to calculate Poisson probabilities which allow the observed count to be compared to that which would be expected given the probability for the entire dataset. Although Poisson probabilities are more difficult to interpret they are more suitable when the numbers are small.

4.3.2 Standardised Lung Cancer Rates

The previous section demonstrated that there are important differences in the lung cancer incidence rates between different age and sex groups. In Figure 4.2 the clusters of circles represented centres of population because there were more cases of lung cancer where there was a greater number of people. However, in order to compare the incidence of lung cancer between different areas in a more meaningful way, it is important to adjust the data so that the confounding effect of age is removed.

Two methods of calculating age standardised rates exist: direct standardisation and indirect standardisation (Plane and Rogerson, 1994). With direct standardisation, the proportions in each age group of a standard population (e.g. world, European, national) are applied to the age-specific death rates of the population being compared. These values are then summed for each age group to provide the directly age-standardised rate. Indirect standardisation is similar except that a standard population is used to provide age-specific death rates rather than the proportions of the population in different age groups. This provides an expected number of deaths or incidents using the age-specific death or incident rates of a standard population to each of the populations that are being compared. The ratio between the observed deaths and the expected deaths provides a measure called the Standardised Mortality Ratio (SMR), which is a useful summary measure for comparing death rates between different population groups. In the case of disease incidence, such as the dataset used here, the Standardised Incident Ratio (SIR) can be calculated using the same method but using incidence data instead of mortality data (Moon *et al.*, 2000).

The choice of whether to use direct or indirect methods of standardisation is important as it has been demonstrated that there can be substantial differences in the ranking of areas when the two approaches are applied using identical data in the same study region (Julious *et al.*, 2001). Direct standardisation has the advantage of being able to precisely adjust for the effects of age. In addition, comparisons can be drawn between all population groups which have had their rates directly standardised to the same base (e.g. regional, national and world). The disadvantage of direct standardisation is that there can be considerable instability in age-specific rates if the calculations are based upon a small number of cases. Indirect methods on the other hand have the advantage of being less sensitive to small numbers in the population subgroups (Foster, 1993). The method also has the advantage that it is not necessary to know the age distribution of cases in the study population. The main disadvantage of the indirect method is that it is considered to be less precise in adjusting for age than the direct method. However, this issue is only considered crucial when the age structure of the study population is dramatically different from the standard (e.g. military personnel). In the context of lung cancer in Scotland it seems sensible to use both methods because a comparison of the results from both approaches will help to assess the advantages and disadvantages of both methods.

ASRs of lung cancer were therefore calculated for males and females using both direct and indirect methods. The procedure was carried out for OAs and also for the data aggregated up to PPS level but only the maps for the PPS calculations are shown due to the large number of OAs in Scotland that were difficult to represent in one map and the problem of statistical insignificance in the smaller OAs. Figures

Figure 4.12. Age Standardised Rates of lung cancer in Scotland for males (direct standardisation)

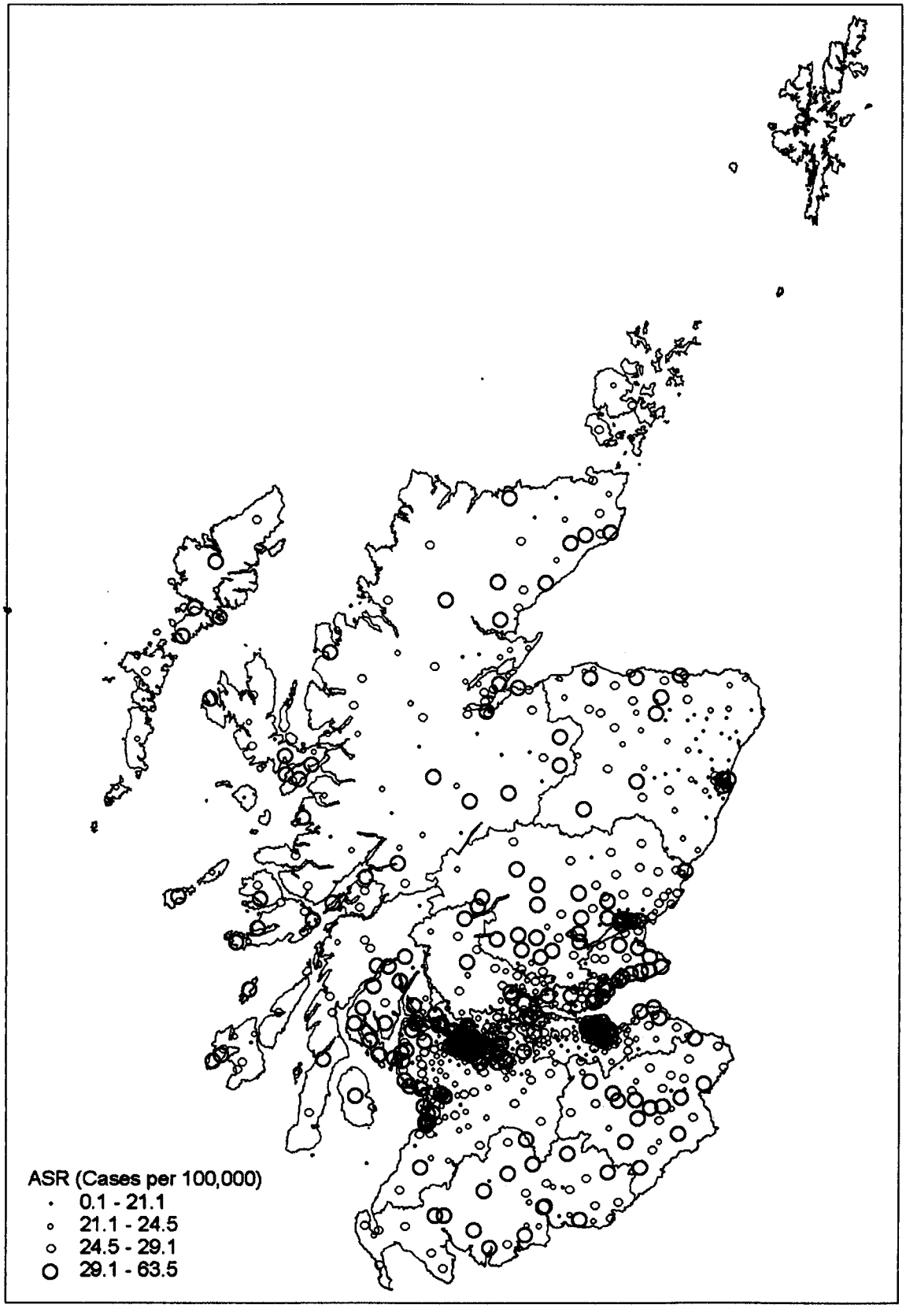


Figure 4.13. Age Standardised Rates of lung cancer in Scotland for females (direct standardisation)

4.12 and 4.13 show the results of direct standardisation and the ASR values were divided into quartiles with the largest circle representing the values in the highest quartile. The rates varied from 8.3 to 142.3 cases per 100,000 for men and 0.1 to 63.5 cases per 100,000 for women. There was a clear urban bias in the male ASRs with a large number of high ASR values in Glasgow, Edinburgh, Dundee and Aberdeen. It is also noticeable that there were some high values in more rural areas. This was especially true in southern Scotland in the Borders and Dumfries and Galloway regions. There were also a number of high values in the rural parts of Tayside and in the Grampians as well as the far north of the Highlands. The map of the ASR values for women showed a broadly similar pattern to that for men. There was a concentration of high values in the major urban areas but also some high values in the more rural areas in southern Scotland. A clear limitation of the observations for both males and females is that they have not been tested for statistical significance. Therefore, it is likely that some of the high ASR values may be based on only a small number of cases.

The SIR for lung cancer was calculated for males and females in each PPS in Scotland and the chi-square value was calculated for each SIR value to test its significance. If the chi-squared value exceeded 3.8 then the result was considered significant (Moon *et al.*, 2000). Figures 4.14 and 4.15 show the results of mapping the SIR values for males and females. A SIR value of more than 100 indicated that the observed number of cases were greater then the expected number of cases, given the demographic profile of these areas. A SIR of less than 100 indicated that the observed number of cases were less than the expected. The maps represent the PPSs

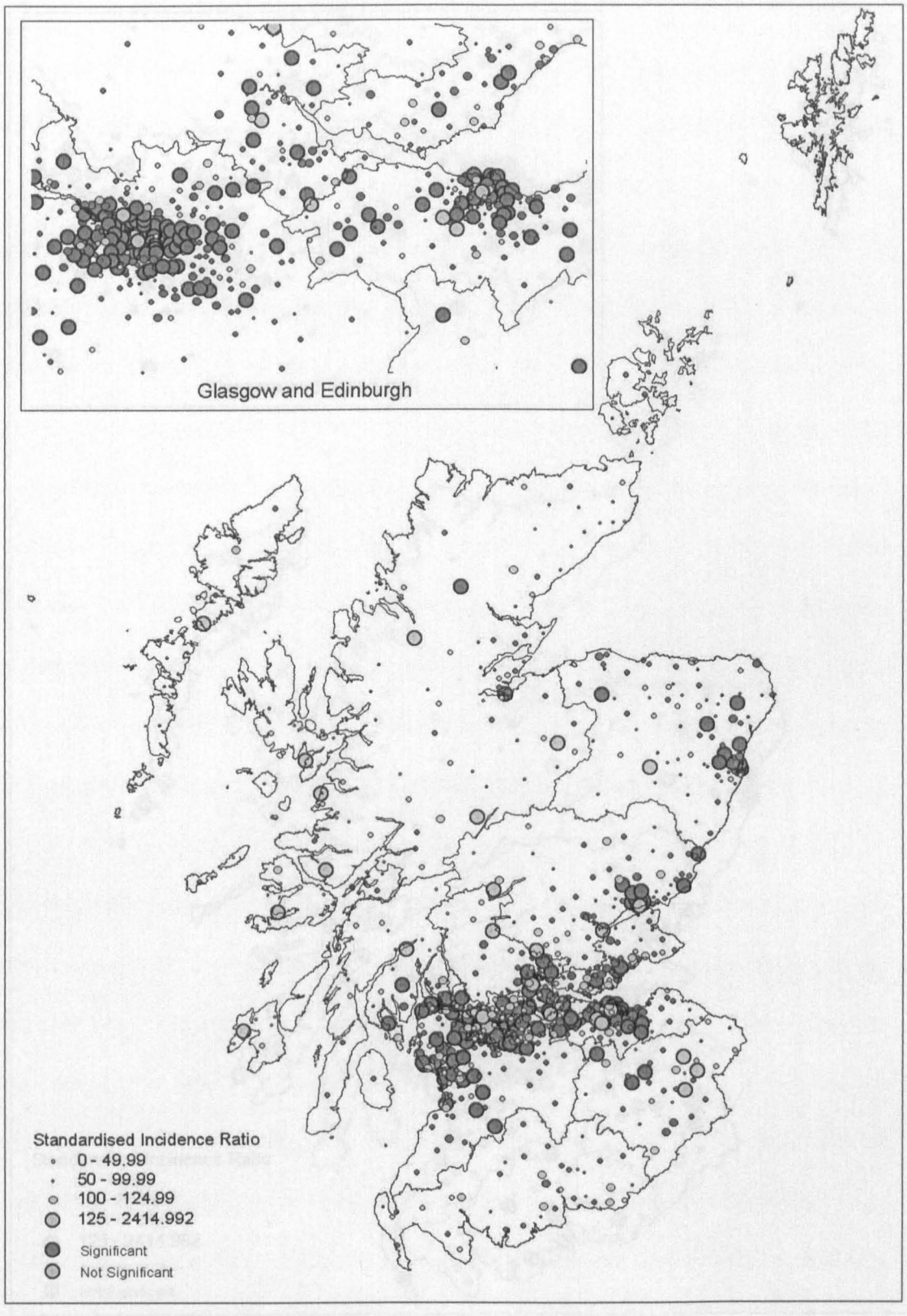


Figure 4.14. Age Standardised Rates of lung cancer in Scotland for males (indirect standardisation)

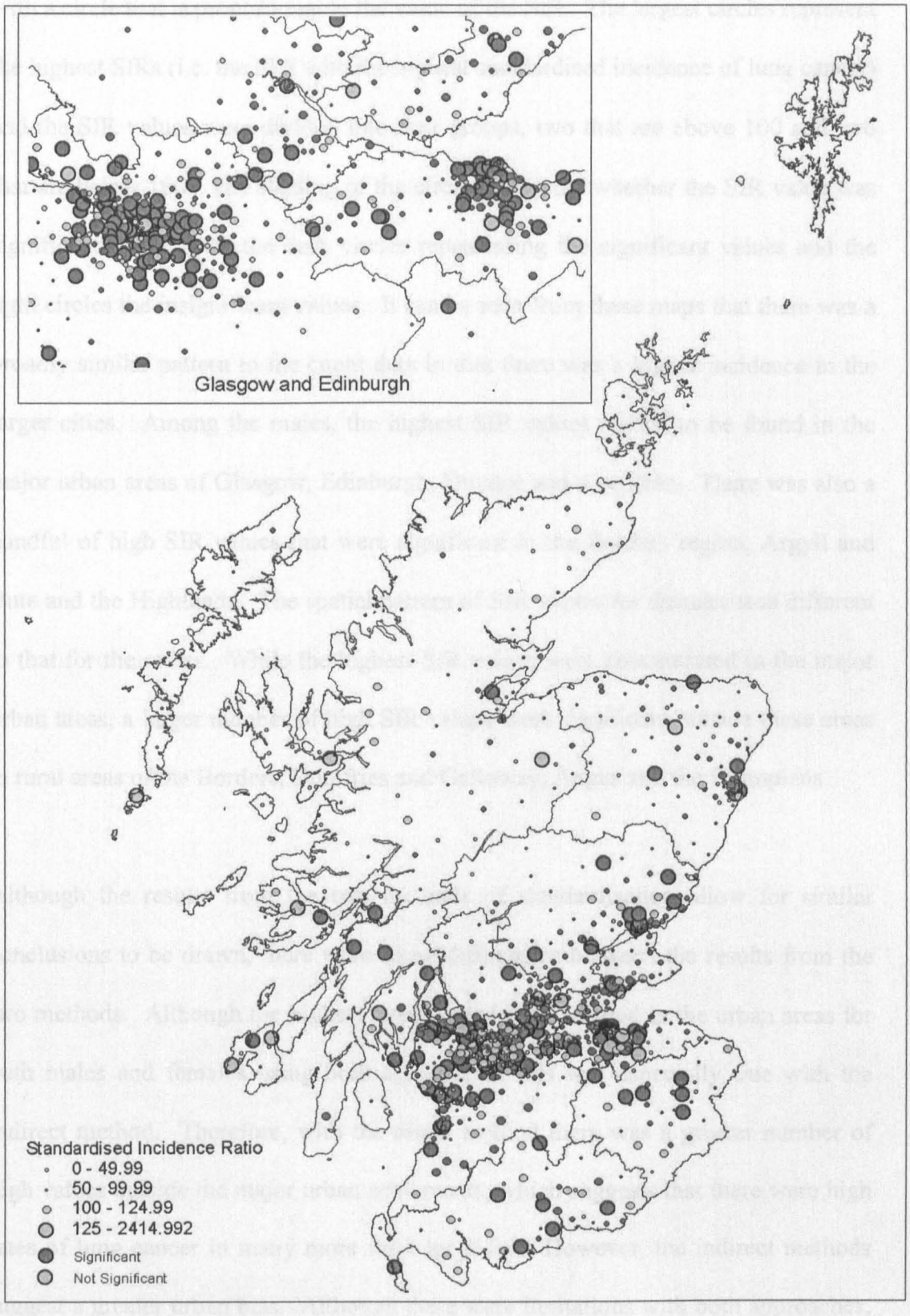


Figure 4.15. Age Standardised Rates of lung cancer in Scotland for females (indirect standardisation)

with a circle that is proportional to the value of the SIR. The largest circles represent the highest SIRs (i.e. the PPS with the highest standardised incidence of lung cancer) and the SIR values were divided into four groups, two that are above 100 and two that are below 100. The shading of the circle represents whether the SIR value was significant or not with the dark circles representing the significant values and the light circles the insignificant values. It can be seen from these maps that there was a broadly similar pattern to the count data in that there was a higher incidence in the larger cities. Among the males, the highest SIR values tended to be found in the major urban areas of Glasgow, Edinburgh, Dundee and Aberdeen. There was also a handful of high SIR values that were significant in the Borders region, Argyll and Bute and the Highlands. The spatial pattern of SIR values for females was different to that for the males. While the highest SIR values were concentrated in the major urban areas, a larger number of high SIR values were significant outside these areas in rural areas of the Borders, Dumfries and Galloway, Angus and the Grampians.

Although the results from the two methods of standardisation allow for similar conclusions to be drawn, there were some differences between the results from the two methods. Although the highest rates tended to be located in the urban areas for both males and females using both approaches, this was especially true with the indirect method. Therefore, with the direct method there was a greater number of high values outside the major urban settlements, which suggests that there were high rates of lung cancer in many more rural locations. However, the indirect methods suggest a greater urban bias. Although there were limitations with both approaches, it seems sensible to place more confidence in the results from the indirect approach because of the problems associated with small counts which the direct approach is

less successful at coping with. In addition the indirect method allows for the testing of the statistical significance of each result.

Although there are some differences between males and females, overall there was a clear correlation between the two sets of rates. This is demonstrated in Figure 4.16 which is a scatterplot that summarises the relationship between the SIRs calculated for males and females in all PPSs in Scotland. The log of the SIR was taken for both due to the effect of a small number of extreme values. There is a broadly positive relationship, with the SIR of one increasing proportionally with the SIR of the other. However, it is also clear that there are a number of points that fall on one of the axes which is because a number of PPSs had a relatively high SIR for one sex but a SIR of zero for the other sex. This reinforces the importance of considering both males and females separately in an analysis of lung cancer.

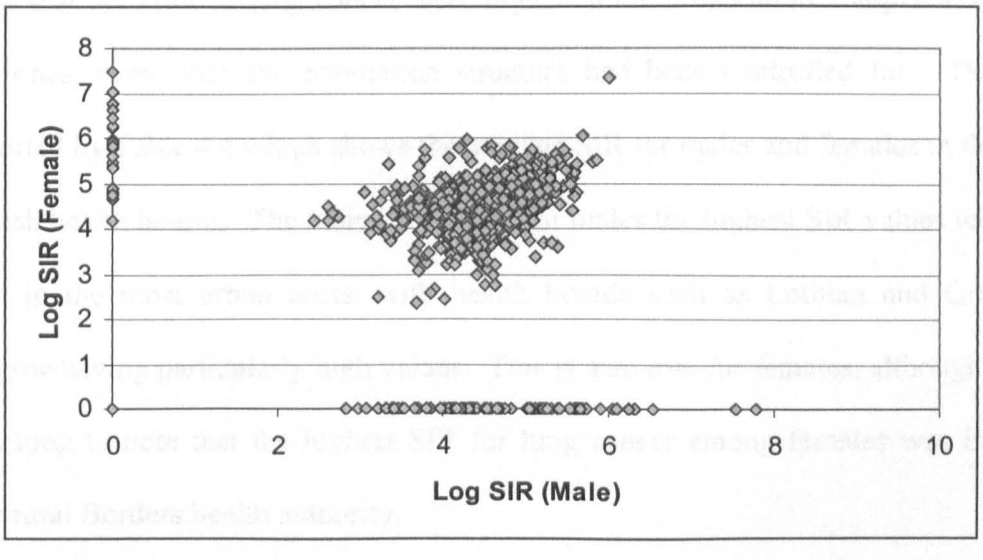


Figure 4.16. Graph comparing SIR values for males and SIR values for females in PPSs in Scotland.

The results of using both the direct and indirect methods of standardisation have demonstrated that, regardless of the approach, there is a clear urban bias in the lung

cancer incidence rates. This suggests that even when the population size and age structure are taken into account, there is still a higher incidence of lung cancer in the larger urban areas. This is confirmed in Table 4.3 which shows the average SIR in five urban/rural categories. The population density was calculated for each PPS and

Urban – Rural	Males	Females
Most Rural	46.40	61.42
	88.55	71.93
Intermediate	88.85	82.77
	98.00	106.20
Most Urban	125.46	127.94

Table 4.3. Mean ASR in five urban/rural categories

the values divided into quintiles to provide five urban/rural categories. The average SIR for each of the quintiles was calculated for males and females. The SIR increased from the most rural to most urban category for males and females which shows that the rates of lung cancer were higher in the urban areas compared to the rural ones, even once the population structure had been controlled for. This is supported by Table 4.4 which shows the average SIR for males and females in the 15 Scottish health boards. The table shows that for males the highest SIR values tended to be in the most urban areas, with health boards such as Lothian and Greater Glasgow having particularly high values. This is also true for females, although it is interesting to note that the highest SIR for lung cancer among females was in the more rural Borders health authority.

Table 4.5 shows the average SIR in PPSs that have been divided into quintiles according to their Carstairs value. The table demonstrates that the average SIR was lowest in the least deprived category and increased through most quintiles for both

Males		Females	
Health Board	**ASR**	**Health Board**	**ASR**
1. Shetland	43.83	1. Western Isles	33.87
2. Highland	58.13	2. Shetland	33.90
3. Orkney	58.68	3. Highland	35.47
4. Western Isles	59.11	4. Orkney	47.18
5. Dumfries & Galloway	64.95	5. Grampian	58.19
6. Argyll and Clyde	78.62	6. Fife	82.48
7. Fife	79.84	7. Ayrshire and Arran	82.77
8. Forth Valley	82.31	8. Tayside	82.82
9. Lanarkshire	84.19	9. Forth Valley	94.12
10. Borders	87.30	10. Dumfries & Galloway	95.77
11. Tayside	94.94	11. Argyll and Clyde	96.23
12. Ayrshire and Arran	95.06	12. Lothian	108.15
13. Grampian	101.31	13. Lanarkshire	111.21
14. Lothian	101.77	14. Greater Glasgow	137.00
15. Greater Glasgow	139.57	15. Borders	138.09

Table 4.4. Mean SIR in Scottish Health Boards.

Urban – Rural	Males	Females
Least Deprived	80.26	82.04
	79.49	68.05
Intermediate	76.15	81.22
	99.44	99.02
Most Deprived	132.52	137.49

Table 4.5. Mean ASR in Carstairs Index of Deprivation quintiles

males and females (although the intermediate quintile was lower that the previous category for males and the least deprived category was higher than the subsequent category for females). The average SIR increased from 80.26 for males and 82.04 for females in the least deprived category to 132.52 for males and 137.49 for females in the highest deprivation quintile. This result suggests that the relationship between deprivation and lung cancer that was discussed earlier is maintained once the age and sex structure of the population are controlled for. These results also suggest that there was a greater exposure to the causal factors associated with lung cancer such as smoking and air pollution in the cities and deprived areas than there is in rural and more prosperous areas.

This section has outlined two methods for standardising lung cancer data in Scotland. Both methods show that the highest rates of lung cancer for both men and women tend to be found in urban areas. They also show that there are some important differences in the rates between males and females. It is thought that the indirect method is more appropriate in this context because it is more appropriate for data that contain a large number of low counts. This section demonstrates that there is an urban bias in the incidence of lung cancer which suggests that there is an urban bias in the factors that influence lung cancer.

4.3.3 Poisson Probabilities

Although age-standardisation is useful for comparing how a disease varies between geographical areas, it is less appropriate when the population size of the areas are small. Poisson probability analysis provides a method for identifying areas of higher or lower than expected incidence of a particular phenomenon when the data have a Poisson distribution. The Poisson distribution is associated with events that have a small probability of occurring and therefore small differences in the count may be a very important observation. The observations forming the data set have the following characteristics: they are counts; they are not restricted in their range of values; they occur independently of each other so that the occurrence of one event does not affect the occurrence of another[4]; they occur irregularly without exhibiting a predetermined pattern; they occur at a constant average frequency; and they are not restricted to a given number prior to analysis (O'Brien, 1992). Therefore, the

[4] Although this assumes that there is no genetic effect

Poisson distribution is used to describe patterns of rare events such as the occurrence of particular diseases and is appropriate for considering lung cancer incidence when the population counts are small, such as in OAs. It is appropriate to undertake this analysis at the OA level because this may reveal important localised differences and the methodology is capable of addressing the issue of small counts.

The probability of an event (X) occurring can be calculated as:

$$P(X) = \frac{e^{-u}u^{x}}{X!}$$

where e is a constant (2.781) and u is the mean number of events. This provides the Poisson probability for a given value. Low Poisson probabilities indicate a larger observed than expected value. For example, a Poisson probability of 0.05 indicates that there is a significant difference at the 95 percent level.

Usually the global mean of the number of cases is used to represent u. In the context of this work it was possible to improve upon this and calculate an expected number of lung cancer cases based upon the age structure of the population because the number of cases of lung cancer in the OA and the adult population of the OA were known. The expected number of cases had been calculated in the previous section to provide the denominator in the SIR ratio. The expected number of cases E, in OA i is calculated by:

$$E_i = \frac{\Sigma O_i.P_i}{\Sigma P_i}$$

where P is the population aged 16 or over and O is the observed number of lung cancer cases. This calculation can then be used instead of the standard mean to calculate the Poisson probability.

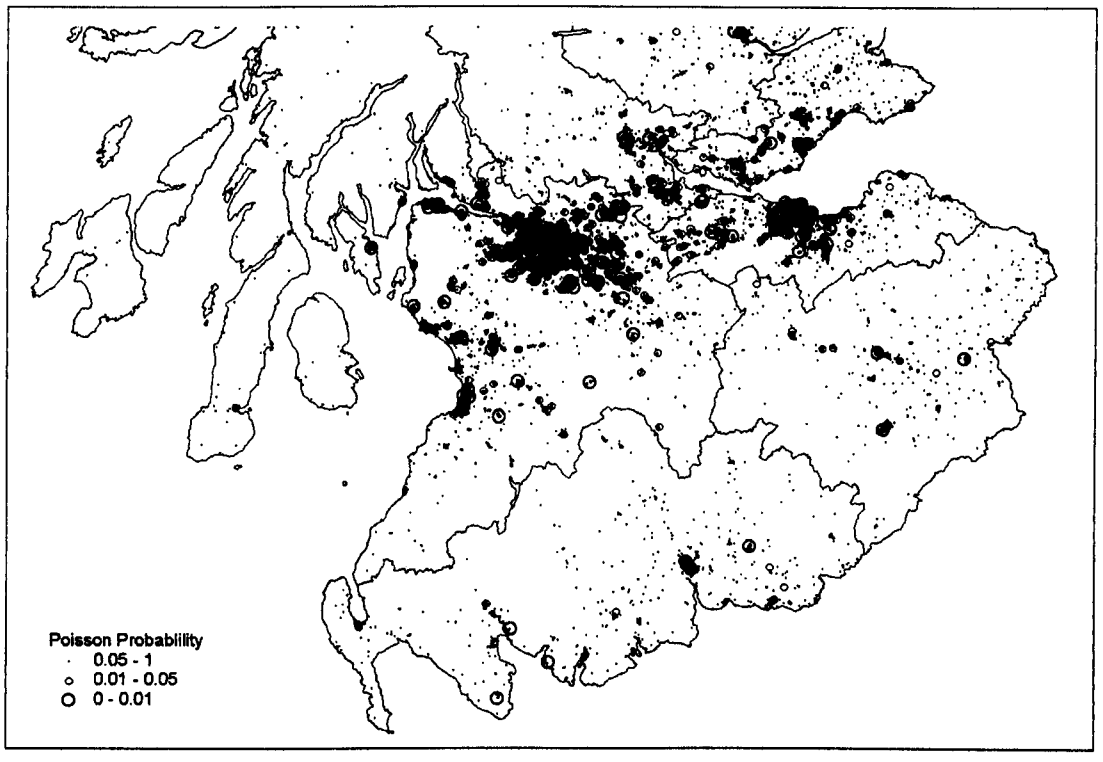


Figure 4.17. Poisson probabilities in southern Scotland.

Poisson probabilities have been calculated for all cases in the period 1988 to 1991 in 1991 census OAs. A low Poisson probability indicates a significantly higher rate of lung cancer than would be expected given the structure of the population. The results for south, central and northern Scotland are shown in Figures 4.17, 4.18 and 4.19. The majority of Poisson probability values were between 0.05 and 1. This occurs because the majority of the counts were zero (i.e. no cases of lung cancer). All Poisson values below one indicate that there is a count of more than one and the closer the Poisson value is to zero, the smaller the likelihood that it would arise by

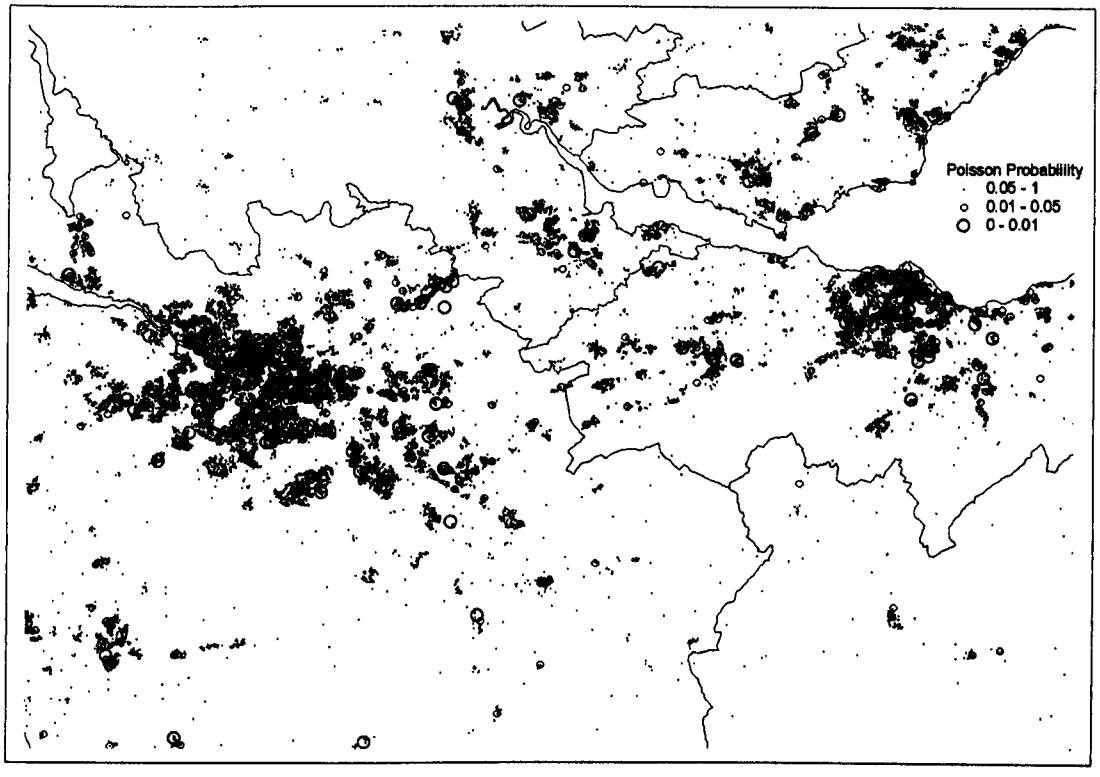


Figure 4.18. Poisson probabilities in central Scotland.

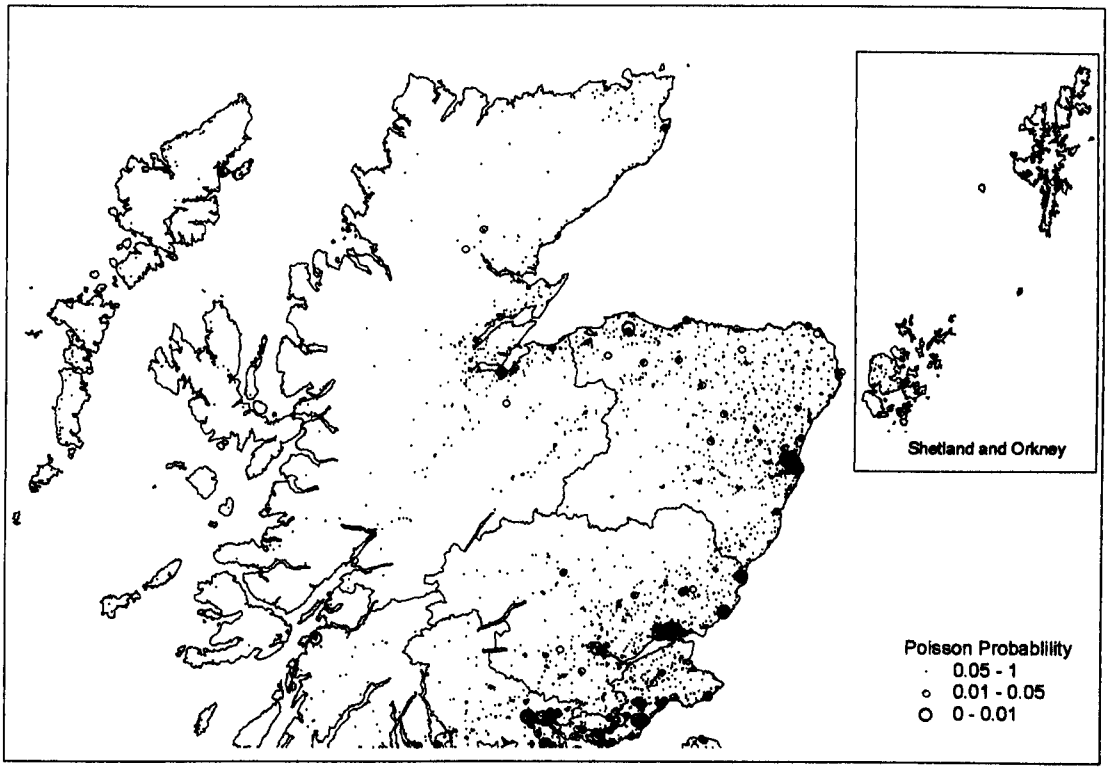


Figure 4.19. Poisson probabilities in northern Scotland.

chance alone. This is to say that a lower Poisson value indicates that there is a significant difference between the observed and expected that is unlikely to be due to chance.

Figure 4.20 shows just the most significant OAs and shows that there were some interesting spatial differences in these values. There were a large number of especially significant values in the major cities such as Edinburgh, Glasgow, Dundee and Aberdeen. Interestingly, there were also some significant values in the Borders around Dumfries, south of Edinburgh and south-west of Glasgow and very few in the north and west of Scotland. Isolated significant OAs were apparent near Oban and Glen Coe as well as the north coast of Aberdeenshire. This method is therefore useful in identifying locations that have a higher (or lower) incidence of lung cancer than one would expect based on the observed count compared to the expected count of lung cancer. Clearly, these differences are likely to be due to one or more of the well-known risk factors associated with lung cancer, such as smoking. This is considered in subsequent chapters.

4.3.4 Comparison of the Results

A comparison of the results between the ASR calculations (direct and indirect) at PPS level and the Poisson probabilities for OAs shows that they both lead to similar conclusions. Although each approach has suggested some subtle differences between the results, all of the methods point towards an urban bias in lung cancer incidence. This is regardless of the methodology used and of the scale at which the analysis is undertaken. The highest rates and probabilities tend to be in the urban

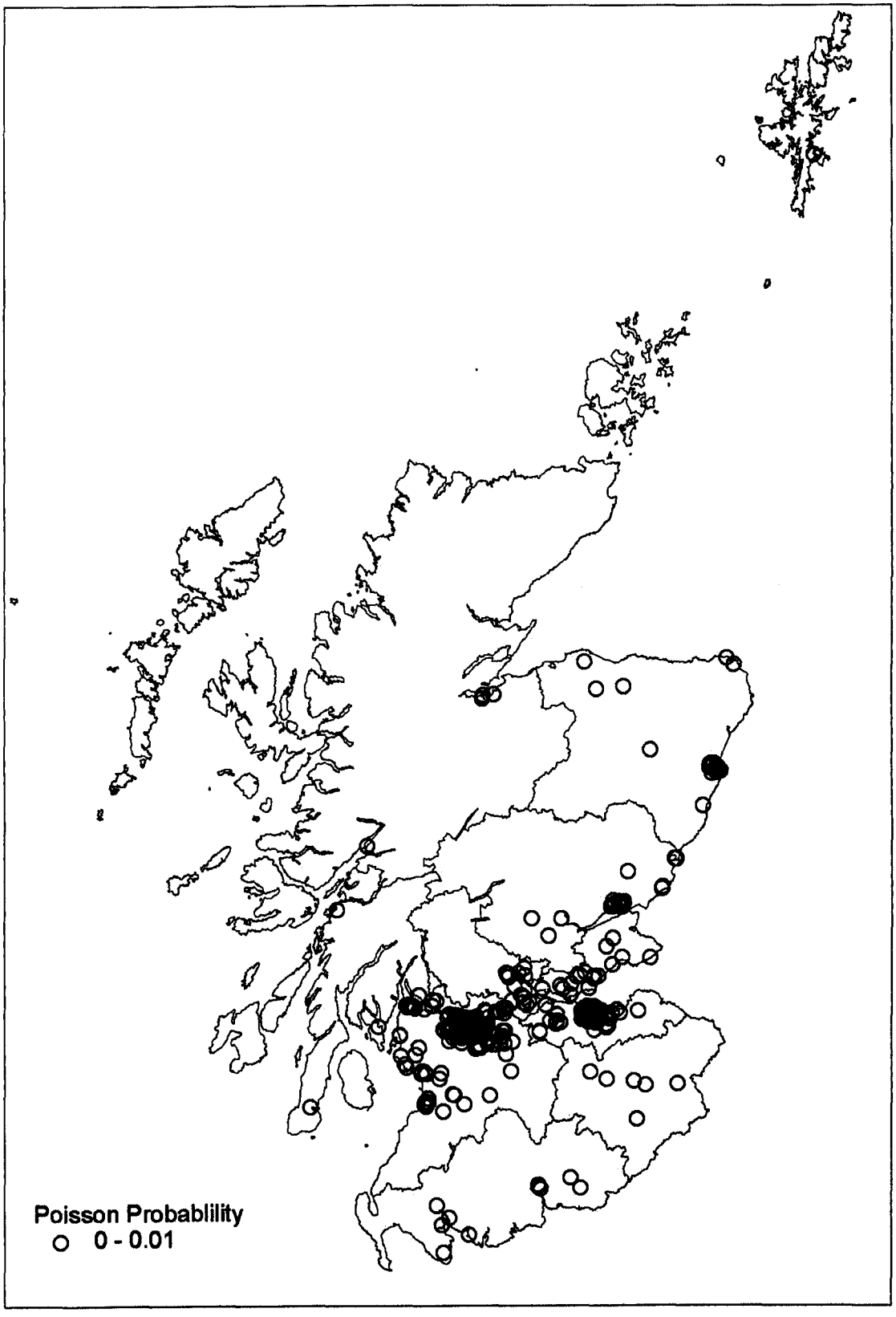


Figure 4.20. Most significant Poisson probabilities in Scotland

areas, especially the large urban areas, and it is rare to find high values in rural areas. This is similar to the maps of the count of lung cancer which showed that the highest numbers of cases were in the urban centres. Therefore, although each of the approaches has limitations, all of them draw us to similar conclusions.

The consistency of the results suggests that the risk factors associated with lung cancer are greater in the urban areas so smoking rates, air pollution and the other risk factors that have been identified also have an urban bias. Subsequent chapters will explore this further. However, this analysis compares the lung cancer rates between areas and does not consider whether the rates are similar in nearby areas, which is more difficult to interpret when one is trying to identify risk factors. Therefore, it is also useful to consider whether there is a similarity in lung cancer rates in areas that are close to each other as this may provide a clearer insight into the major risk factors because it identifies broad areas that have in common higher incidence of lung cancer. Consequently, it is useful to develop methods of cluster detection to examine if there are clusters of high incidence of lung cancer.

4.4 Detecting Clusters of Lung Cancer Incidence

The advances in the statistical analysis of spatial point processes have received renewed interest in Geography largely due to the developments of GIS (Bailey and Gatrell, 1995). There are a number of methods that have been developed ranging from simple point pattern analysis to the more complex spatial and space-time clustering. The focus on spatial point processes, especially in the disease context, can provide useful insights into the possible causes of the disease. The aim of this

section is to develop methods for examining whether there are clusters of cases of lung cancer incidence across Scotland. This will allow initial hypotheses about potential risk factors associated with lung cancer to be developed that can be explored further in later sections.

There has been a particular focus in the spatial analysis literature on methods of cluster detection. A cluster can be defined as a group of occurrences that are related to each other through some social or biological mechanism, or have a common relationship with some other event or circumstance (Kelsall, 1996). The occurrence of a cluster may be a useful indicator of an elevated risk factor at a limited geographical scale. Cluster analysis techniques have been used frequently in health research for a variety of diseases. However, there are also examples in other fields, most noticeably in the analysis of crime data (Brunsdon, 1991; Chainey, 1997).

A number of different cluster detection methods have been developed that provide methods for detecting spatial and space-time clustering. One of the first packages that identified the location and tested the significance of specific clusters was provided by Turnbull *et al* (1990). This used overlapping circles with a fixed population size. Another group of methods that are available to test for clustering throughout a study region are known as global clustering tests (e.g. Diggle and Chetwynd 1991, Grimson 1991, Walter 1994). For example, the K function developed by Diggle and Chetwynd (1991) has been particularly influential and widely used. The estimate of the K function is simply the average number of events within a specified distance of a specific event taken over all events and then divided by the number of further events per unit area. This is not suitable when it is

necessary to adjust for the inhomogeneity of another spatial point process (such as the population distribution) which is important in disease clustering. Furthermore, none of the global clustering tests have the capability to locate specific clusters and instead test for clustering throughout the study region.

The Geographical Analysis Machine (GAM) was developed as an exploratory cluster detection approach (Openshaw *et al.*, 1987). The GAM method examines a large number of overlapping circles at a variety of scales and assesses the statistical probability of the number of events occurring by chance given the denominator population in each circle. If the number of observations is statistically significant then a circle is plotted. This provides a visual representation of where a cluster may occur. However, GAM suffers from the problem of multiple hypothesis testing which arises because a separate significance test is made for each circle and therefore in almost any dataset there will be a multitude of 'significant clusters' as defined in this way. A similar method has been developed with the SatScan software (Kuldorf *et al.*, 1998). However, in SatScan the circular window is flexible both in location and size and is able to resolve the issue of multiple hypothesis testing because the software allows the circle size to vary continuously. Instead of testing the significance of each circle, a single value for the test of the null hypothesis is provided. If the null hypothesis is rejected then the location of the cluster(s) that caused the rejection is specified. The SatScan method was used in this study because of its ability to locate and test the significance of individual clusters using areally aggregated data, and the capacity it has to allow standardisation and deal with multiple hypothesis testing.

The SatScan software is designed to evaluate reported spatial or space-time clusters and to see if they are statistically significant; to test whether they are randomly distributed over space and/or time; and to detect areas that have significantly high or low rates (Kuldorf *et al.*, 1998). It has been used in a range of applications including around a solid waste incinerator (Viel *et al.*, 2000), childhood leukaemia (Hjalmars *et al.*, 1996) and breast cancer (Kuldorf *et al.*, 1997).

The SatScan software was used here to examine whether any clusters of lung cancer incidence existed, once the age and population structure of each area had been controlled for. This differs from the exploratory analysis in the previous section because it considers the spatial dependence of the incidence of lung cancer rather than treating each area independently. The detection of contiguous areas of high rates may reflect the geography of a particular risk factor. In this case, the software assumes the number of cases in each area to be Poisson distributed. If no covariates are included then the number of cases within each area can be expected to be proportional to its population size. However, in this instance categorical covariates were included that controlled for both the population size and the age distribution of the population. Therefore an adjusted expected number of cases was used.

The groups analysed were males and females aged 16 to 24, 25 to 34, 35 to 44, 45 to 54, 55 to 64, 65 to 74 and over 75. The population count data for each age-sex group were taken from the 1991 census. Because lung cancer tends to be a disease of old age, the inclusion of these covariates avoids a scenario where the clusters that are detected are merely clusters of older people. One problem with this is that the lung cancer dataset covers the period 1988 to 1991 and so the overlap with the population

data was not exact. It was therefore possible that if the patient was diagnosed prior to the 1991 census then they could have moved into a different age group or they could have died and therefore would not have been included in the census. The date of diagnosis was not known so it was not possible to calculate the exact age of the lung cancer patient in 1991. Because the population counts in OAs are small, or even zero, in each age-sex group this has some implications for running the analysis at this level as the software is not able to cope with a population of zero. Therefore, in the OA population file the zero population counts were recoded to one. Although this may have some minor implications for the analysis, this was considered acceptable and allowed the algorithm to run to completion.

The basic procedure that SatScan uses is for each point to have an observed and an expected value assigned to it. The observed number is taken directly from the user defined file while the expected value is calculated internally using the covariates. The software will then search for clusters beyond those which are expected due to the covariates. For the observed cases, a circle is created around each centroid that is big enough to include a user-defined percentage of the total population (e.g. if 10 percent were chosen in this Scottish example then this would equate to approximately 500,000 people). The circle is maximised until it reaches this threshold. The software then randomly assigns cases to the coordinates and the analysis is run a further 9,999 times (plus the one 'real' run totals 10,000 runs). The results from each are then compared to see if the one real has been exceeded by any of the 9,999 centroids. If the real case has never been exceeded then it is very significant and if it is within the five percent highest then it is significant at the 0.05 level.

In this example, population-weighted centroids were used for PPSs and then for the smaller OAs. The centroids of these small areas were used as the geographical reference for the disease as the precise location of the lung cancer patients was not known. The population-weighted centroids for OAs were available from the 1991 census data whilst the PPS population centroids were calculated from the OA coordinates. The 'special OAs' (e.g. military bases) that did not have a coordinate were excluded from this analysis. There is no consensus about which percentage of the total population should be included in the window but Kulldorf *et al* (1998) suggest that when in doubt to choose a high percentage since the program will then look for both small and large clusters without any preselection bias in terms of the cluster size. Therefore, a 50 percent window was used. The clusters that were identified in the output file were joined to the boundary file and mapped using ArcView GIS.

The results of running the analysis at the PPS scale are shown in Figure 4.21 and all the clusters that were significant are circled. It can be seen that there were clear and significant clusters in around the major urban areas of Glasgow, Edinburgh, and Dundee whilst the clusters that were found in Aberdeen were not significant. This analysis indicates that even having controlled for the age and sex structure of an area there is an urban bias in the lung cancer incidence. This is not surprising as this observation is probably a reflection of the smoking behaviour in Scotland as it has already been shown that lung cancer incidence is strongly correlated with smoking (Doll and Hill, 1950) and as subsequent chapters will discuss, there are higher rates of smoking in urban areas compared to rural areas.

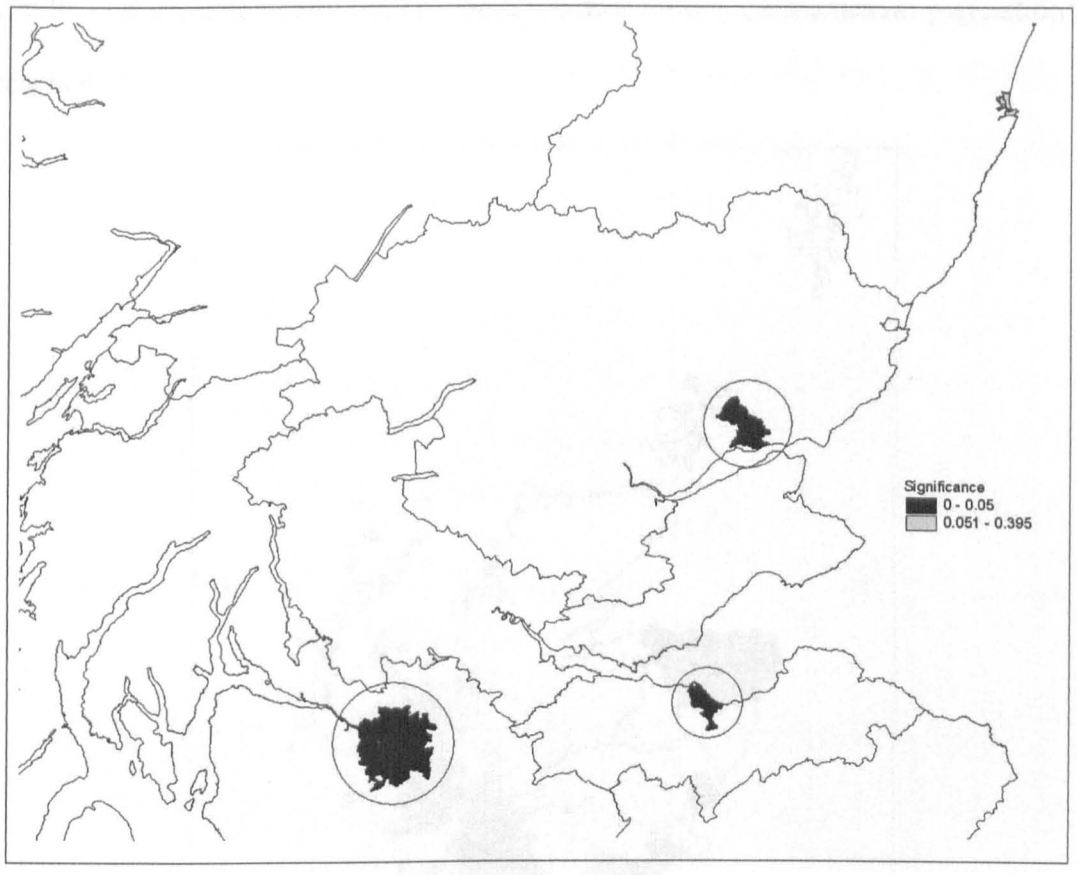


Figure 4.21. Clusters of lung cancer incidence PPSs (all cases) using SatScan with a 50% window.

An identical set of runs were done but with Edinburgh, Glasgow, Dundee and Aberdeen removed in order to identify clusters that occur when one ignores these large urban areas. The results from each run were similar and Figure 4.22 shows the results from the run with a 50 percent window. It can be seen from this figure that the significant clusters tended to be located in the more populated areas especially in the central belt, southern Fife, the Perthshire coast and the west coast south of Glasgow. There were also some interesting clusters in northern Scotland around Banff and Elgin and on some of the islands such as Mull. The size of the clusters tended to be larger in area because of the smaller populations in the rural areas. This

results in a geographically larger search window before the maximum population is reached.

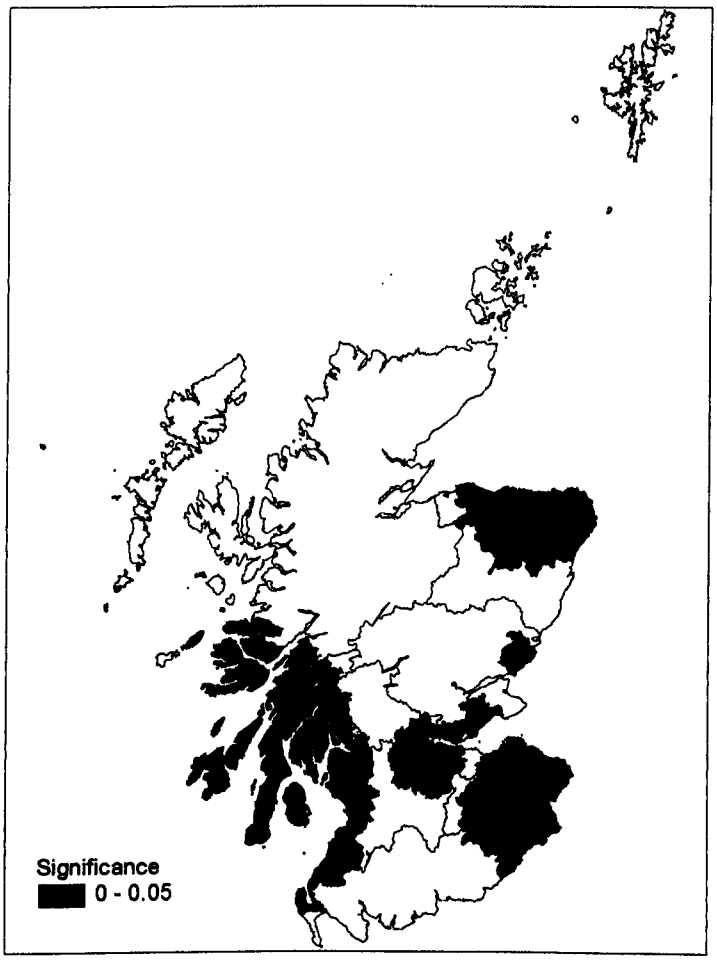


Figure 4.22. Clusters of lung cancer incidence in PPSs (large urban areas excluded) using SatScan with a 50% window.

The analysis was also run at the smaller OA level, using population-weighted OA centroids as the geography. However, running the software at OA level is more problematical due to the processing power required. It was found that it was not possible to run the analysis with the maximum size of the circular window set to anything greater than 15 percent even on a high specification PC so the analysis was run for a 15 percent window. Each run took approximately three days. Figure 4.23 shows the results of running the analysis with a 15 percent window. The results were similar to the analysis at PPS level because the clusters in Glasgow, Edinburgh

City and Dundee remained significant but in addition there was a significant cluster in Aberdeen.

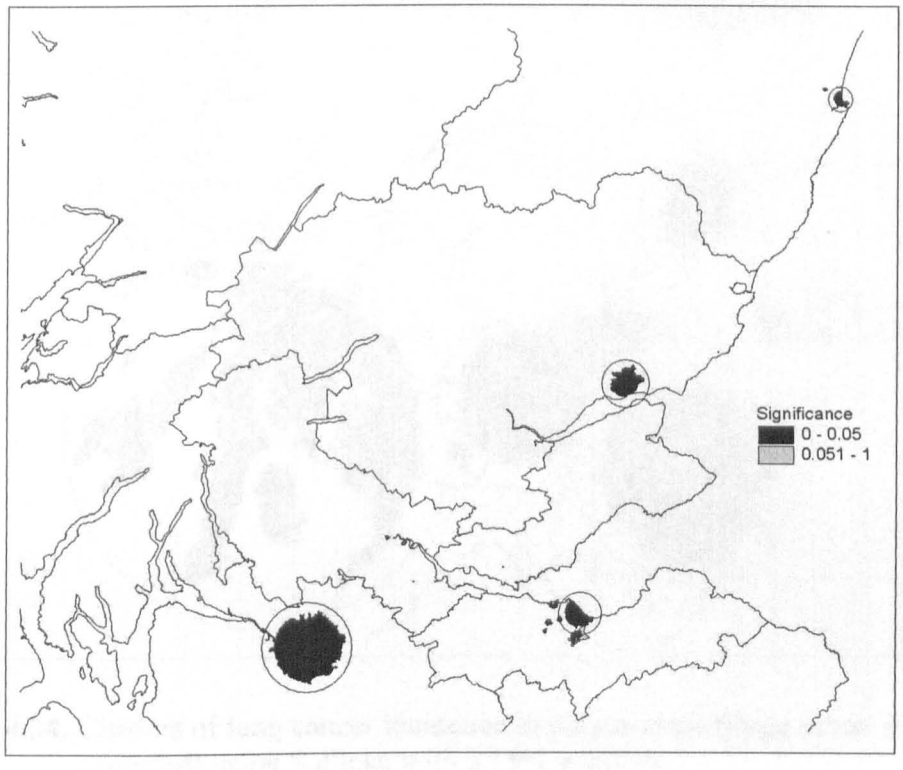


Figure 4.23. Clusters of lung cancer incidence in output-areas (all cases) using SatScan with a 15% window.

The analysis was run again, this time with the major urban areas (Glasgow, Edinburgh, Dundee and Aberdeen) excluded. The results of running the analysis with a 15 percent window are shown in Figure 4.24. It can be seen that there were seven significant clusters: a large area in Argyll; Arran; Bute and the Kintyre peninsula; East Kilbride and Hamilton; the mid part of the central belt including Cumbernauld and Falkirk; a large part of the Borders; Dunfermline and Kirkcaldy; Angus; and in Inverness. These results are interesting because these areas represent some of the more urban areas outside of the largest cities which suggests that even having removed the largest urban areas from the analysis there remains a consistent urban bias in the clusters of lung cancer incidence.

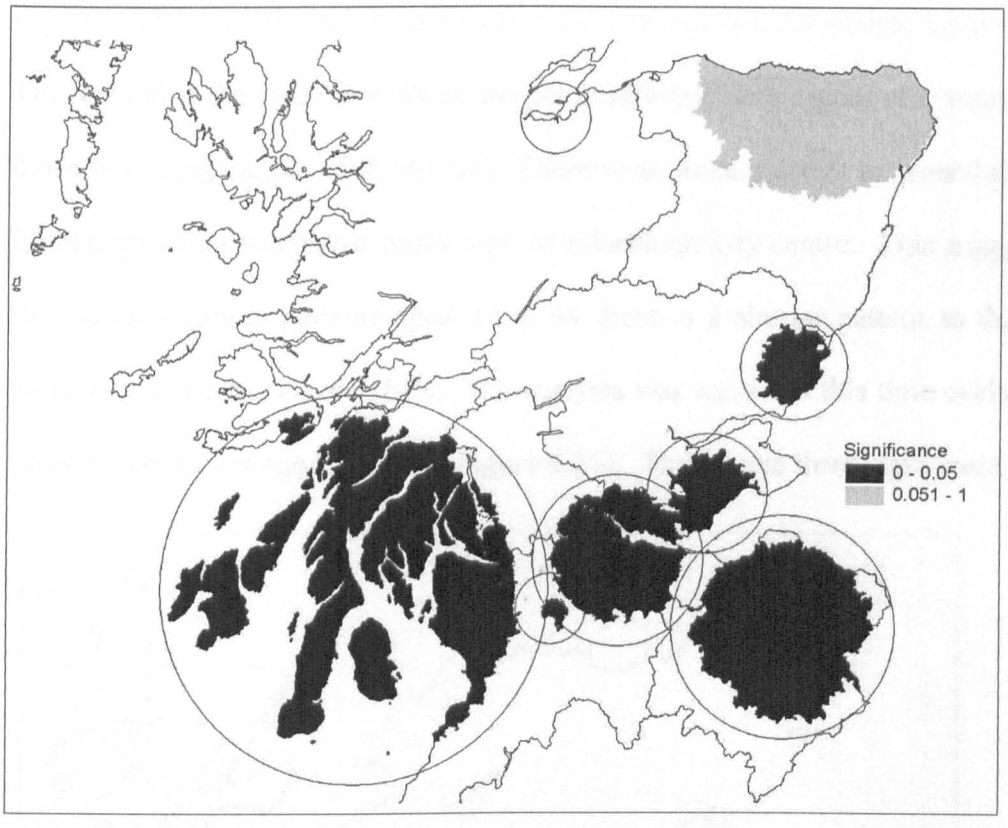


Figure 4.24. Clusters of lung cancer incidence in output-areas (large urban areas excluded) using SatScan with a 15% window.

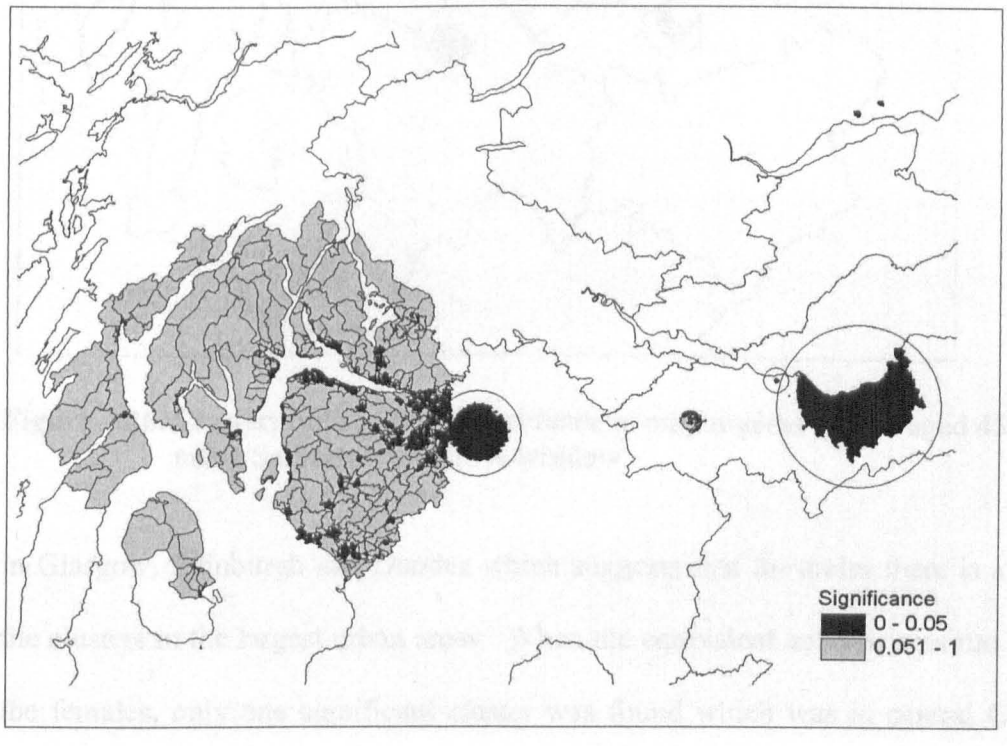


Figure 4.25. Clusters of lung cancer incidence in output-areas (all cases aged 45 to 64) using SatScan with a 15% window.

Figure 4.25 shows the results of running the analysis just for people aged 45 to 64. This excluded the rare cases where people contracted lung cancer at a young age or those with lung cancer at an old age. There were three clusters in central Glasgow, Edinburgh and a small area north-west of Edinburgh city centre. This suggests that for the lung cancer patients aged 45 to 64 there is a similar pattern to that of all patients, i.e. there is a urban bias. The analysis was rerun but this time excluding all cases except males aged 45 to 64 (Figure 4.26). There were three significant clusters

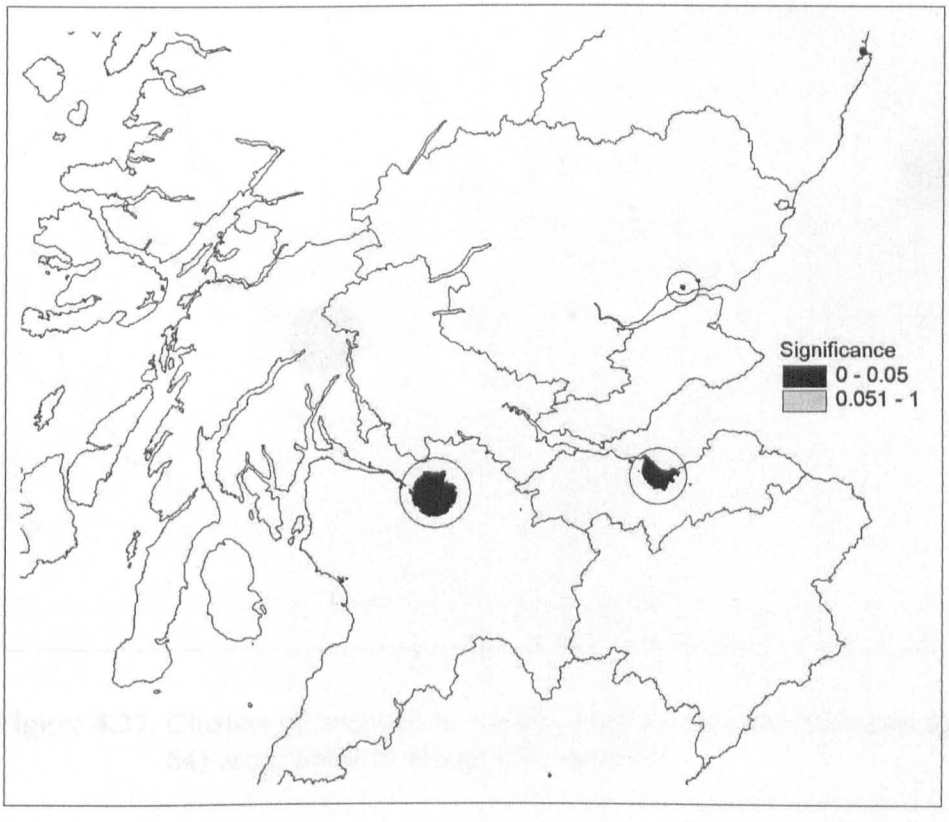


Figure 4.26. Clusters of lung cancer incidence in output-areas (males aged 45 to 64) using SatScan with a 15% window.

in Glasgow, Edinburgh and Dundee which suggests that for males there is a bias in the clusters in the largest urban areas. When the equivalent analysis was run for just the females, only one significant cluster was found which was in central Glasgow (Figure 4.27). This may reflect the lower levels of lung cancer among females. In order to explore the effect of urban bias, the analysis was run for all people aged 45

to 64 who lived outside of the largest urban areas of Glasgow, Edinburgh, Dundee and Aberdeen. It was found that the clusters that were identified were similar to the analysis that was done for all people outside of the major urban areas. The difference was that the clusters that were identified previously in western Fife and Angus were not significant.

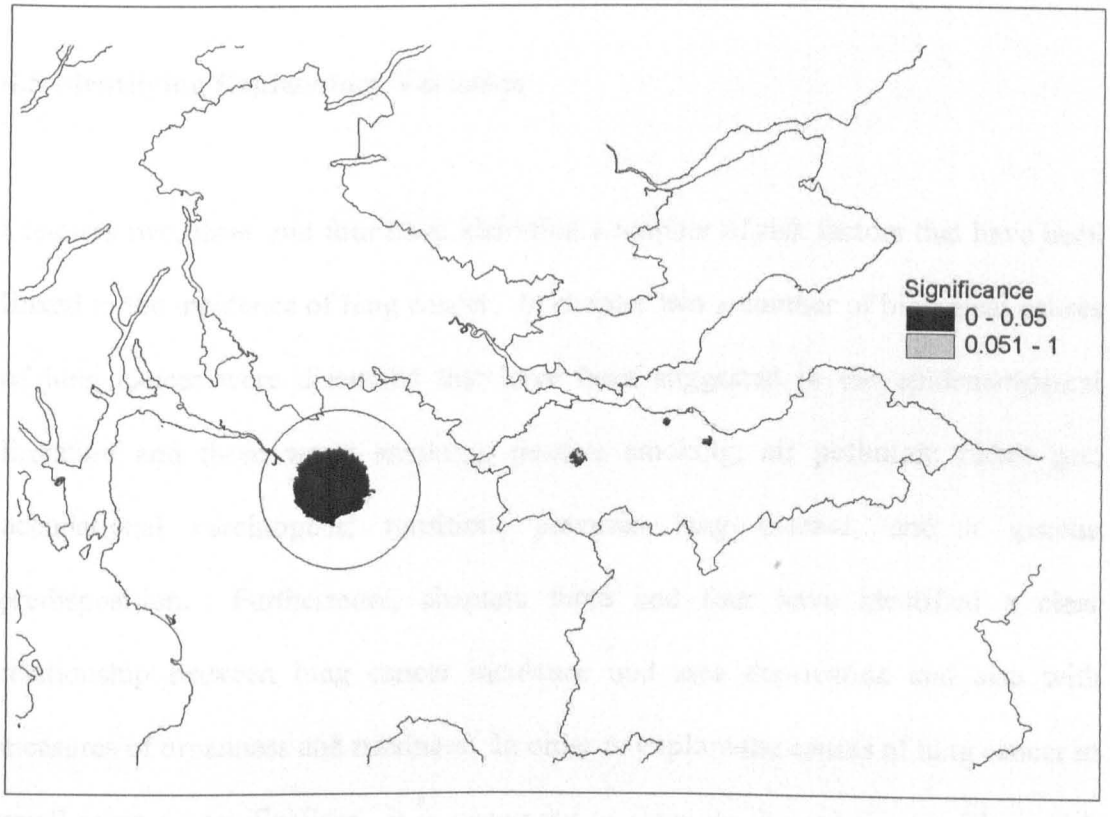


Figure 4.27. Clusters of lung cancer incidence in output-areas (females aged 45 to 64) using SatScan with a 15% window.

A consideration of individual age-sex groups has shown that some areas consistently provide part of a cluster for all age-sex groups. This is the case for central Glasgow which shows evidence for clusters of cases of lung cancer in all age sex groups that have been considered. Equally, other clusters were only identified in particular age-sex groups which would, at this stage, suggest that there are some processes operating that are unique to those particular groups. For example, a small area to the north-west of Edinburgh was only significant among the population aged 45 to 64.

The analysis of clusters of cases of lung cancer has demonstrated that regardless of the approach taken, the clusters tend to be located in the largest urban areas. This is also true for particular age and sex groups. Therefore, this section has reinforced observations that were made in earlier sections that there is an urban bias in the geography of lung cancer.

4.5 Identifying Explanatory Variables

Chapters two, three and four have identified a number of risk factors that have been linked to the incidence of lung cancer. In chapter two a number of biological causes of lung cancer were discussed that have been suggested in the epidemiological literature and these were: smoking; passive smoking; air pollution; radon gas; occupational carcinogens; nutrition; previous lung disease; and a genetic predisposition. Furthermore, chapters three and four have identified a clear relationship between lung cancer incidence and area deprivation and also with measures of urbanness and ruralness. In order to explore the causes of lung cancer in small areas across Scotland, it is necessary to estimate the exposure to these risk factors for the same geographical units. Therefore, subsequent chapters will, where necessary, develop methods to estimate the exposure to the key risk factors that have been highlighted in earlier chapters. The calculations can then be used within a series of regression models to examine which variables are significant in explaining the geographical distribution of lung cancer incidence in Scotland.

4.5.1 Derived Variables

There is a distinction between the explanatory variables that are readily available and those variables that have to be derived from various sources of data. Many of the risk factors that have been identified as potential causes of lung cancer cannot be calculated directly from comprehensive data sets such as the census, rather methods have to be developed to derive the estimates. The following three chapters therefore develop methods to estimate the exposure to some of the key risk factors associated with lung cancer that have been identified in earlier chapters.

In light of the crucial role that smoking plays in explaining lung cancer incidence, it is particularly important to have accurate and geographically-specific smoking estimates. Although information on smoking behaviour is available from sources such as the General Household Survey, the data are only available for broad geographical regions. Therefore, chapter five uses multilevel models to calculate the probability of smoking for different age-sex groups using smoking data from the General Household Survey and explanatory data from the 1991 census. The magnitude of the effect of air pollution on the incidence of lung cancer is less well understood. Furthermore, there has been a national reduction in pollution levels during the 20 years prior to the lung cancer dataset. Therefore, chapter six estimates air pollution levels in 1971 and 1991 for small areas in Scotland. Similarly, it was demonstrated in chapter two that the potential geological causes of lung cancer are still debated. Therefore, chapter seven develops methods to estimate the exposure to three potential geological causes of cancer: radon gas, coal mining and quarrying activity.

4.5.2 Non-Derived Variables

In addition to the derived variables, there are some factors that are potentially related to lung cancer that can be directly obtained or calculated from the 1991 census. Chapter two has highlighted the importance of different occupations as causes of lung cancer. Many of the industries, such as steel and rubber, are not major industries in Scotland and for others, such as those working in the asbestos industry, it is not possible to gain specific variables on the types of employment that have been associated with lung cancer. Therefore, the occupational groups from the 1991 census are used to explore the effect of occupation on lung cancer and these are the percentage of people in each OA who are in work and employed in the following industries: agriculture, forestry and fishing; energy and water; mining; manufacturing of metals etc.; other manufacturing; construction; distribution and catering; transport; banking and finance; and other services. These variables provide the best available measure of employment type and they coincide temporally with the data on lung cancer incidence. Important differences in lung cancer incidence have been noted between males and females and, in addition, there is a disparity between the types of occupations that males and females have traditionally been employed in. Therefore, the corresponding sex-specific occupational variables are also included so that the influence of the occupation can be examined separately for males and females.

Chapters three and four highlighted a clear relationship between lung cancer incidence and area deprivation. It was demonstrated that lung cancer rates tended to

be higher in areas with high levels of deprivation. To some extent, this is likely to be a reflection of the higher rates of smoking in more deprived areas. However, a number of studies have suggested that deprivation has an independent effect on morbidity in general (Macintyre *et al.*, 1993; Boyle *et al.*, 1999). To date, studies have not tended to examine the links between deprivation and lung cancer, once other risk factors, such as smoking, have been controlled for. Several methods of categorising the populations of small geographical areas in the UK into deprivation classes have been developed. The methods in most common use are those of Jarman (1984), Townsend *et al.* (1988) and Carstairs and Morris (1991). All of the methods combine census variables to generate a summary score which, it is argued, reflects the socioeconomic status of a locality relative to the distribution of scores obtained for all localities (McLoone and Boddy, 1994). Although the variables chosen by the authors are different, the statistical methods are similar. The variables used to construct the Carstairs Index were the proportions of the population: in households without access to a car; in overcrowded households; with the head of household in social class IV or V; and in households with unemployed men. The Carstairs Index was chosen because it closely captures the concept of material disadvantage which the Jarman score does not (Morris and Carstairs, 1991) and it was developed from the Townsend score specifically to capture deprivation in Scotland (Carstairs and Morris, 1991). Furthermore, it has been used in many other studies of health and mortality (e.g. McLoone and Boddy, 1994; Shah *et al.*, 2001). Therefore, the Carstairs Index of Deprivation was included as an explanatory variable in order to test whether the level of deprivation in a PPS influenced lung cancer once the other risk factors had been controlled for.

Similarly, chapters three and four have demonstrated a positive relationship between lung cancer incidence and how urban an area is, but again this is likely to reflect higher levels of smoking in more urban areas. Even so, previous studies have identified an urban bias in lung cancer rates, even when other risk factors, such as smoking, have been controlled for (Haynes, 1988). Therefore, the population density was calculated to provided an urban-rural proxy (Martin *et al.*, 2000) in order to test whether the urban-rural bias in lung cancer incidence identified in chapter four remained once the various explanatory factors had been controlled for. The population density was calculated for OAs by dividing the total population in each OA by its geographical area.

Other variables that have been identified as potential causes of lung cancer are more difficult to capture. Passive smoking is likely to be strongly correlated with smoking probabilities (Jee *et al.*, 1999) and it would be very difficult to model this effect independently of smoking. Furthermore, there is not a consensus on whether passive smoking has a significant effect upon lung cancer incidence. Similarly, nutrition, previous lung disease and genetic predisposition are considered to be responsible for only a very few cases of lung cancer and often only in tandem with other environmental risk factors such as smoking. Furthermore, it would be difficult to model these factors due to the lack of geographically specific data relating directly to them and the absence of further information about the individuals to make predictions about their personal history. For example, the specific aspects of diet thought to contribute to or prevent lung cancer, such as red meat or fresh fruit, would be extremely difficult to estimate because geographically-specific and comprehensive data are not readily available.

The combination of derived and non-derived variables will be calculated and obtained for small areas throughout Scotland. They can then be associated with data on lung cancer incidence that have been introduced earlier in this chapter. Finally, chapter eight will use regression analysis to examine the extent to which the risk factors explain the incidence of lung cancer in Scotland, having controlled for the age and sex structure of these areas. This will provide new insights into the influences of different risk factors upon lung cancer incidence in small areas across Scotland.

4.6 Conclusion

This chapter has provided an introduction to the geography of lung cancer in Scotland using a number of exploratory methods. The methods include simple mapping of lung cancer incidence, age-standardised rates, probability mapping and methods of cluster detection. The various approaches were used to gain different insights into the geography of lung cancer and reflected the fact that certain methods are only suitability at particular scales. The simple mapping, ASRs and Poisson probabilities considered the incidence of lung cancer in isolation from their surrounding areas and so gave an insight into how each area compared to the rest of the country. The cluster detection methods on the other hand gave an insight into how lung cancer incidence related to the incidence in adjacent areas by identifying if there were any clusters of high cases.

It has been seen in the previous chapter that the scale of analysis at which lung cancer rates are considered can be important and the combination of methods used here has allowed for the role of scale to be investigated. This is because it is only suitable to calculate the ASRs for the larger PPS level whilst the Poisson probabilities were more appropriate at the OA level. Each of the exploratory approaches has consistently demonstrated that there is an urban bias in lung cancer incidence, regardless of whether the data are age-standardised or not and whether the areas are considered in isolation from each other or considered in relation to their neighbours. The different maps of lung cancer have shown that the rates are consistently higher in the urban areas and the cluster detection work has shown that clusters of high incidence of cases also tend to be located in the urban areas. Furthermore, this observation is true for both males and females. This allows us to be confident that there is an urban bias in lung cancer incidence as the observation is not dependent upon the methodology, the demographic group being studied or the scale of analysis. Furthermore, this information is useful for public health policy workers as it identifies areas and clusters of areas where the incidence of lung cancer is particularly high.

The higher rates of lung cancer in urban areas suggest that the risk factors associated with the disease have an urban bias, which is consistent with previous studies (Haynes, 1988). Therefore, risk factors such as the rates of smoking, levels of air pollution etc may be higher in urban areas. The following chapters will develop methods to predict the key factors and then explore whether or not they are able to explain this urban effect.

5. PREDICTING SMOKING RATES IN SCOTLAND

5.1 Introduction

Smoking has been identified as the most significant cause of cancer, responsible for 40 percent of all cancer deaths and 18 percent of all deaths (Peto, 1994). It has been estimated that smoking kills two million people a year in developed countries, half of whom die aged between 35 and 69. This accounts for about one-sixth of all deaths in these countries (Peto, 1994). Studies of smoking behaviour (e.g. Balarajan and Yuen, 1986; Blaxter, 1990; Graham and Der, 1999) have generally focused upon national trends, regional trends or trends within certain groups of a population, with less work on the geographical patterns of smoking behaviour in small areas. This is despite suggestions that smoking is the single most important factor that determines geographical variations in mortality. For instance, it has been shown that mortality is 15 percent higher in the most deprived districts compared to the most affluent districts and that differences in smoking behaviour account for 85 percent of the excess (Law and Morris, 1998). Only through focusing on relatively small geographical areas can a better understanding of smoking behaviour be achieved (Twigg, 1999).

For the purposes of this study of lung cancer, smoking is especially relevant. For example ecological studies that consider the influence of radon upon the incidence of lung cancer must control for smoking (Stidley and Samet, 1993). Therefore, the aim of this section is to make predictions about the probability of smoking among people in different age/sex groups in small geographical areas across Scotland where 30.9

percent of people aged over 16 declared themselves as smokers (Anderson and Hope, 2000). Such geographical estimates will allow smoking behaviour to be controlled for in an ecological study of lung cancer incidence so that other risk factors can be examined. This is the first time that such geographically detailed estimates of smoking have been made available. Furthermore, this section will examine the relationship between estimates of smoking behaviour and the incidence of lung cancer in the same geographical areas. Given that these estimates need to be reasonably reliable, it is important that the model used is as accurate as possible.

The estimates of smoking behaviour will be used to help explain lung cancer across Scotland for the period around 1991. Of course, it may be argued that it would be preferable to have smoking estimates for a period some time ago, because it can be anticipated that there should be a lagged relationship between environmental risk factors, such as smoking, and lung cancer. This is impossible to achieve, as historical data on smoking are not available for small geographical areas. Also, given that the study is ecological, even if lagged data on smoking rates were available there would be no guarantee that those persons who developed lung cancer resided in the same place during the lag period. This problem would be exacerbated if the migration rate between the areas during the lag period is high. On the other hand, it is also possible that the relative geographical differences in smoking behaviour today are likely to resemble historical patterns of smoking quite strongly as the current smoking status of an individual is likely to be a reliable estimate of whether they smoked in the past. These combined concerns mean that it is difficult to know whether crude lagged smoking data would represent an improvement

compared to contemporary estimates that are likely to be more accurate for small areas.

5.2 Variations in Smoking Rates

In the UK, men's cigarette consumption rose from 800 per person per year in 1905 to 4,420 per person in 1945, then fell sharply after the war to 3,320 in 1949, after which it rose again to 4,030 in 1960. Since 1962, cigarette consumption has again shown a downwards trend to reach 2,380 in 1985. Women's pattern of consumption is rather different. In 1921 consumption was 13 cigarettes per person, then the trend steadily increased for the next 50 years reaching 1,250 per person per year in 1945 and peaking at 2,630 in 1974. Since then, women's cigarette consumption has also fallen, but more slowly than that of men, to 1,930 a person in 1985 (Wald *et al.*, 1988). The General Household Surveys (Bridgwood *et al.*, 1998) have included a question on smoking behaviour since 1972 and the percentage of smokers between 1974 and 1998 are shown in Figure 5.1. The prevalence of smoking among men has

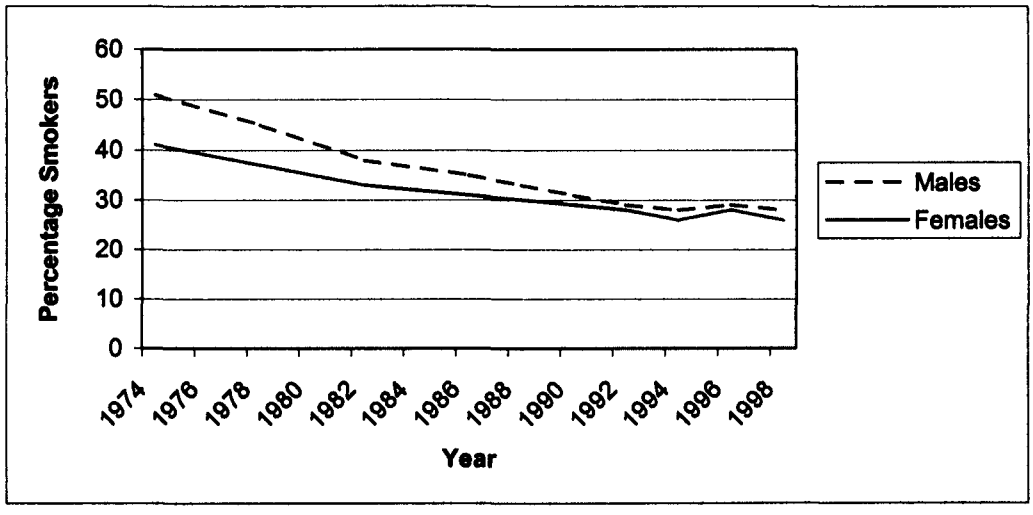


Figure 5.1. Prevalence of smoking by and sex: 1974 to 2000 (GHS)

decreased from 52 percent in 1972 to 31 percent in 1990 and from 41 percent in 1972 to 29 percent in 1990 for women. The differences in smoking behaviour between men and women have narrowed so that by 1999 the figures were similar for both sexes at 27 percent for men and 26 percent for women (General Household Survey, 1999). Figure 5.2 shows the rates of smoking among different age groups in the GHS in 1973 and 1999. It is clear from this graph that rates of smoking were consistently higher for all age groups for both males and females in 1973. The only exception to this was among the females aged over 65 who had a similar, low rate of smoking in both years. In 1999 the rates of smoking tended to be higher for males than for females except for those aged 35 to 45 where females smoking rates were slightly higher than those of males.

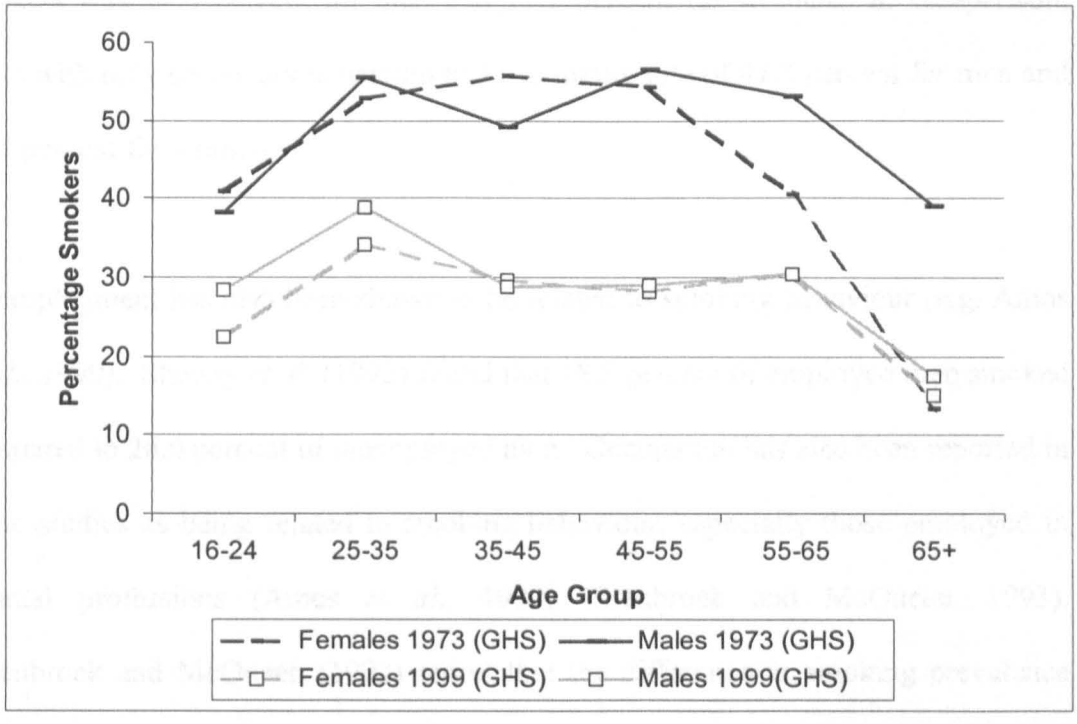


Figure 5.2. Percentage of smokers in Scotland in i) 1973 GHS ii) 1999 GHS

Raw data on cigarette consumption and smoking prevalence can conceal two important factors (Charlton, 1994). Firstly, smoking prevalence among young

women under 25 has increased in recent years and secondly that most of the cigarette smokers are in less affluent socio-economic groups. In fact, the decline in smoking prevalence has been heavily concentrated in older age groups and in the more affluent socio-economic groups.

A number of individual characteristics other than age and sex have also been shown to influence smoking behaviour. Educational attainment has been shown in a number of studies to be associated with smoking behaviour (e.g. King *et al.*, 1990; Shewry *et al.*, 1992; Escobedo and Pedicord 1996; Martikainen *et al.*, 2001). Shewry *et al.* (1992) used the Scottish Heart Health Study to examine coronary risk factors and they found that among those with a degree the percentage who were smokers was 14.6 percent for men and 12.1 percent for women. In comparison, those with only secondary education had a smoking rate of 45.3 percent for men and 42.4 percent for women.

Unemployment has also been shown to be related to smoking behaviour (e.g. Amos *et al.*, 1990). Shewry *et al.* (1992) found that 18.8 percent of employed men smoked compared to 26.0 percent of unemployed men. Occupation has also been reported in some studies as being related to smoking behaviour, especially those employed in manual professions (Amos *et al.*, 1990; Uitenbroek and McQueen, 1993). Uitenbroek and McQueen (1993) report that the difference in smoking prevalence between those in manual and non-manual occupations seemed to be increasing among females and decreasing among males. Although smokers were historically more likely to be affluent than poor, the affluent have tended to abandon tobacco in recent decades whilst the poorer have not done so. In the UK, for instance, only 10

percent of women and 12 percent of men in the highest socio-economic group were smokers. However in the lowest group 35 percent of women and 40 percent of men smoked (World Bank, 1999).

Studies have also indicated that marital status may be associated with smoking (Amos *et al.*, 1990; Shewry *et al.*, 1992). Amos *et al.* (1990) found that single people had the lowest smoking rates whilst divorced/separated and widowed men and women had the highest. For women, the smoking rate was 34 percent among singles and increased to 47 percent for the married, and 58 percent for those who were divorced, widowed or separated.

Other studies have attempted to examine how smoking rates are affected by the geographical context in which people live. Barnett (2000) focused on the contextual factors affecting smoking rates in Christchurch, New Zealand. He showed that smoking levels varied according to the level of deprivation and degree of social mixing in an area, with deprived populations having the highest smoking rates. Similarly, Diehr *et al.* (1993) found that the prevalence of smoking varied significantly between 15 different communities even once they had controlled for demographic, health status, and other health-behavioural characteristics of the individuals in the communities. Duncan *et al.* (1999) used multi-level modelling techniques to investigate whether the character of a place influenced smoking behaviour independently of personal characteristics. They found that neighbourhood deprivation had an independent effect on individual smoking behaviour, suggesting that the characteristics of local areas were an important determinate of smoking behaviour.

Of most relevance to this study, Twigg *et al.* (1998) used multi-level models to predict smoking (and drinking) behaviour in England and Wales to explore whether factors that influence smoking behaviour operated at different spatial scales. Their multi-level model operated at three scales: the individual (from the Health Survey for England), the postcode sector, and the district health authority. Area-specific predictions were then derived for a fourth set of areas – wards. Their resulting model found a number of ecological effects such as car ownership, social class and home ownership to be significant, having controlled for individual characteristics. They found that on average 28 percent of a ward population in England were smokers. The highest values tended to be found in inner cities, while the lowest rates were in rural areas and the urban fringe of resort towns.

5.3 Data and Methods

This study uses a mixture of individual-level survey and area-level census data to estimate the probability of smoking among 12 age/sex groups in small geographical areas in Scotland. In an ideal situation, individual-level data would be available both on smoking behaviour, and other potentially relevant individual characteristics expected to influence smoking, for small geographical areas. However, much of the information that exists in surveys is considered to be confidential and is therefore not released below broad geographical regions. In the absence of individual-level information, it is necessary to undertake ecological analysis, involving the use of aggregate data sets. In this study the available information included a small number of variables from the Scottish Household Survey (SHS) and a larger number of

variables from the 1991 British Census. The data and the multi-level modelling approach that were used to estimate smoking behaviour across Scotland are described below.

5.3.1 Scottish Household Survey

The SHS was commissioned by the Scottish Executive in 1998 to provide household and individual data that were not otherwise available in Scotland and which would aid policy making and inform the Scottish Parliament (Dixon and Finlayson, 2000). The survey is intended to provide accurate, up-to-date information about the characteristics, attitudes and behaviour of Scottish households and individuals on a range of issues such as population, housing, economic circumstances, household resources and community involvement.

The SHS included individuals from a diverse set of geographical areas using the Postcode Address File (PAF) as its sampling frame. One of two sample structures was used depending on the population density of the local authority. If the population density was 500 or more persons per square kilometre, a systematic random approach was adopted, where a random sample of addresses was drawn from each strata of a geo-demographic indicator called the Scottish MOSAIC. The Scottish MOSAIC drew upon a large number of census variables in order to classify Scottish postcodes in terms of housing, housing densities and household characteristics to identify neighbourhoods with similar characteristics. The MOSAIC classification includes groups such as 'country dwellers' and 'singles and flats' and these groups were used to identify areas with similar geodemographic

characteristics for sampling (Experian, 2002). If the population of the local authority was less than 500 persons per square kilometre then a systematic random sample was not appropriate because of the extra resources required to carry out the fieldwork. For the local authorities below this threshold, census Enumeration Districts (EDs) were used as primary sampling units (PSUs) instead. The survey aimed to conduct eleven interviews per PSU and to have a minimum of 50 PSUs within each of the local authorities (Anderson and Hope, 2000).

Data on smoking behaviour are available from other sources and can be compared with the smoking information from the Scottish Household Survey. The results from the Scottish records in the 1999 General Household Survey and the 1999 Scottish Household Survey are shown in Figure 5.3. Both datasets demonstrated that

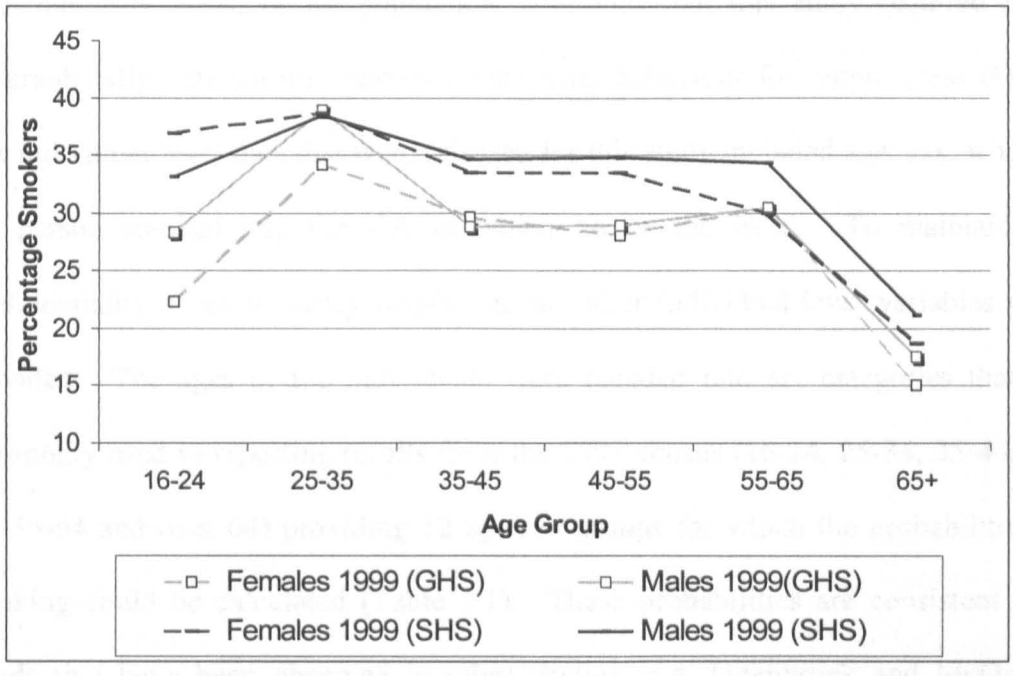


Figure 5.3. Percentage of smokers in Scotland in i) 1999 GHS; ii) 1999 SHS

smoking rates among males tended to be higher than those for females in the same age group. For both males and females, the rates of smoking rose after the youngest

age group and peaked for the 25 to 35 age group and then fell gradually until there was a dramatic decrease in the oldest age group. However, there are some noticeable differences between the two surveys with smoking rates in the Scottish Household Survey tending to be higher for both males and females. Among the four youngest age groups the female smoking rates are consistently higher in the Scottish Household Survey by between 5 and 10 percent. For males, the rates are consistently 5 to 7 percent higher in the Scottish Household Survey for all age groups except for the 25 to 35 age group where the rates are similar.

This study used data from the first year of the SHS, and included records on 13,784 individuals over the age of 15, distributed across Scotland. Usually information from the Scottish Household Survey is disseminated for broad geographical areas, such as local authority areas, or for population subgroups, but this study required more geographically detailed information on smoking behaviour for output-areas (OAs). The individual-level data that were released for this study included age, sex, whether the person smoked and the OA in which he or she lived. To maintain the confidentiality of each survey respondent no other individual-level variables were provided. The ages of the individuals were recoded into six categories that are commonly used in reporting results from the 1991 census (16-24, 25-34, 35-44, 45-54, 55-64 and over 64) providing 12 age/sex groups for which the probabilities of smoking could be calculated (Table 5.1). These probabilities are consistent with trends that have been observed in other studies (e.g. Uitenbroek and McQueen, 1993) with higher probabilities among younger women (especially ages 16-24) compared to younger men but lower rates among older women compared to older

	Raw Data	
Age	**Male**	**Female**
16-24	0.331	0.369
25-34	0.384	0.386
35-44	0.346	0.334
45-54	0.348	0.334
55-64	0.342	0.299
65+	0.212	0.186

Table 5.1. Probability of smoking using the raw data from the Scottish Household Survey

men. This reflects the historical gender bias in smoking behaviour. Although it was not possible to obtain any other variables from the SHS at the individual level, because the OA in which each individual lives was known it was possible to attach area-based information from the census. These data can be used firstly to model the geographical variations in smoking behaviour observed in the Scottish Household Survey and then the parameters from the model can be used to predict smoking behaviour for all OAs across Scotland.

5.3.2 Census Data

A number of 1991 census variables were extracted at the OA level and were used as explanatory variables in models designed to estimate the probability of smoking. The 13,784 individuals were distributed across 7,127 of the 38,254 Scottish OAs. A selection of census and census-derived variables were chosen that represented the characteristics of either individuals or households resident in that area. In addition, two variables were extracted for pseudo-postcode sectors (PPSs) which provide information on the types of areas in which these people live (Table 5.2). PPSs are

Person Variables
% unemployed*
% of males unemployed**
% of females unemployed**
% of unemployed individuals in 9 age groups**
% of individuals working in managerial and administrative occupations*
% of individuals working in professional occupations*
% of individuals working in associated professional occupations*
% of individuals working in clerical and secretarial occupations*
% of individuals working in craft and related occupations*
% of individuals working in personal and protective occupations*
% of individuals working in sales occupations*
% of individuals working in plant and machine operative occupations*
% of individuals working in other occupations*
% of individuals with further education qualification*
% of individuals who were widowed or divorced*
% of residents in households with an economically active head of household in a lower social class group (IV or V)*
Household Variables
% owner occupied households*
% privately rented households*
% of households renting from a housing association or local authority*
% of one person households*
% of households with dependent children*
% of lone parent households*
% of households without a car*
% of overcrowding households (1 or more persons per room)*
Population density (a surrogate urban/rural indicator)*
Place Variables
ONS ward classification
Carstairs index of deprivation*

* Each variable was compared to the log of the variable

* * Each variable was compared to the log of the variable and age-specific and sex-specific variables were also calculated and then applied to the individual depending on their age and sex.

Table 5.2. Census variables used in logit model

neat aggregations of OAs and there are 1,003 PPSs in Scotland. Because of their size, these areas better reflect the likely 'daily activity spaces' of people providing an indication of the 'place' in which people live. OAs, on the other hand, are much

smaller and homogenous, providing a good indication of the likely type of person who lives in a place.

The characteristics of people within OAs included information on unemployment, occupation, qualifications, marital status and social class. In addition to these variables both age-specific and sex-specific alternatives for each of the variables were included for the relevant age/sex group (for example, instead of using total unemployment the percentage of unemployed males and females in particular age groups were attached to individuals in the relevant age/sex groups). Household characteristics were also included, reflecting tenure, the number of people living in the household, car ownership and overcrowding. Population density was also included at the OA-level and this is an urban/rural proxy; it has been shown elsewhere that there are urban/rural differences in smoking behaviour (e.g. Twigg *et al.*, 1998).

The two variables extracted for PPSs were the Carstairs measure of deprivation and the ONS ward classification. The Carstairs measure of deprivation is based on the census variables of unemployment, overcrowding, non-car ownership and low social class. It was included because deprivation has been shown to influence health outcomes (e.g. Boyle *et al.*, 1999; Saul and Payne, 1999) and, more specifically, smoking status (Duncan *et al.*, 1999). The four component variables were also included individually as potential explanatory variables.

The ONS ward classification divides all wards (or postcode sectors in Scotland) into a set of geodemographic classes using a number of census variables (Wallace and

Denham, 1996). The classification has 14 groups, further subdivided into 43 clusters, which are intended to identify wards with similar housing and person characteristics. Here, the 14 groups were used which are: suburbia; rural areas; rural fringe; industrial areas; middling Britain; prosperous areas; inner city estates; established owner-occupier; transient populations; metropolitan professionals; deprived city areas; lower status owner occupied; mature populations; and deprived industrial areas.

The study therefore utilises a wide range of potential explanatory variables. Table 5.3 shows that multicollinearity is not a major issue as the correlation between the explanatory variables is relatively low in most cases. This is even the case for the Carstairs score and the four variables that are used to construct the score (although these were measured at the OA level).

5.4 Modelling Smoking

There were two stages in the modelling procedure. First, using a multi-level model the probability of smoking was estimated for the 13,784 individuals extracted from the SHS, who lived in 7,127 of the 38,254 OAs in Scotland. A number of models were examined and the optimum model was identified. The explanatory variables included those measured at the individual, OA and PPS levels. In the second stage, the parameter estimates from the optimum model were used to estimate the probability of smoking for the 12 age/sex groups in all 38,254 OAs in Scotland.

Table 5.3. Pearson correlations between variables

	% unemp.	% manage. & admin.	% profess.	% assoc. profess.	% clerical & secret.	% craft & related	% personal & protective	% sales	% plant & machine op.	% other occup.	% further educ. qual's	% widowed or divorced
% unemployed	1.00											
% managerial & admin.	-0.23	1.00										
% professional	-0.22	0.02	1.00									
% associated profess.	-0.12	-0.06	0.02	1.00								
% clerical & secretarial	-0.04	-0.13	-0.11	-0.10	1.00							
% craft & related	0.11	-0.22	-0.21	-0.16	-0.14	1.00						
% personal & protective	0.07	-0.15	-0.13	-0.11	-0.15	-0.12	1.00					
% sales	0.01	-0.11	-0.10	-0.09	-0.11	-0.07	-0.09	1.00				
% plant & machine op.	0.15	-0.20	-0.19	-0.15	-0.14	-0.10	-0.10	-0.07	1.00			
% other occupations	0.23	-0.20	-0.20	-0.17	-0.19	-0.09	-0.09	-0.10	-0.05	1.00		
% further educ. qual's	-0.33	0.21	0.58	0.31	-0.05	-0.24	-0.13	-0.08	-0.23	-0.26	1.00	
% widowed or divorced	0.54	-0.18	-0.09	-0.01	-0.03	0.02	0.07	0.00	0.07	0.14	-0.15	1.00
% social class (IV or V)	0.23	-0.25	-0.23	-0.17	-0.08	-0.01	0.14	-0.05	0.16	0.44	-0.29	0.17
% owner occupied	-0.64	0.30	0.33	0.18	0.08	-0.15	-0.11	-0.02	-0.22	-0.32	0.50	-0.51
% privately rented	-0.13	0.12	0.11	0.09	-0.03	-0.09	0.01	-0.05	-0.10	-0.05	0.22	0.17
% renting from HA / LA	0.65	-0.33	-0.33	-0.19	-0.05	0.18	0.08	0.04	0.25	0.30	-0.53	0.45
% one person households	0.22	-0.08	0.00	0.02	0.03	-0.02	0.02	-0.02	0.00	0.05	-0.08	0.51
% with depend. children	0.05	0.00	-0.06	-0.04	-0.06	0.04	0.03	0.03	0.04	0.02	0.00	-0.05
% lone parent	0.49	-0.18	-0.19	-0.10	-0.05	0.08	0.06	0.04	0.13	0.18	-0.24	0.44
% without car	0.72	-0.32	-0.27	-0.12	0.01	0.12	0.09	0.03	0.17	0.24	-0.43	0.67
% overcrowding	0.46	-0.19	-0.20	-0.08	-0.02	0.10	0.05	0.02	0.12	0.17	-0.25	0.39
Population density	0.28	-0.16	-0.09	-0.02	0.06	0.05	0.03	0.03	0.06	0.05	-0.11	0.32
Carstairs index	0.40	-0.22	-0.25	-0.11	-0.03	0.13	0.06	0.02	0.17	0.20	-0.35	0.29

	% social class (IV or V)	% owner occupied	% privately rented	% renting from HA / LA	% one person h'holds	% with depend. children	% lone parent	% without car	% over - crowding	Pop density	Carstairs index
% social class (IV or V)	1.00										
% owner occupied	-0.34	1.00									
% privately rented	-0.06	0.09	1.00								
% renting from HA / LA	0.33	-0.93	-0.40	1.00							
% one person households	0.07	-0.28	0.21	0.23	1.00						
% with depend. children	0.01	0.02	-0.20	0.01	-0.70	1.00					
% lone parent	0.17	-0.46	-0.12	0.47	-0.06	0.38	1.00				
% without car	0.28	-0.78	-0.09	0.78	0.51	-0.20	0.47	1.00			
% overcrowding	0.18	-0.43	-0.04	0.42	-0.03	0.24	0.37	0.45	1.00		
Population density	0.08	-0.26	-0.05	0.28	0.19	-0.02	0.21	0.42	0.23	1.00	
Carstairs index	0.22	-0.48	-0.10	0.48	0.10	0.03	0.32	0.49	0.33	0.16	1.00

In stage one, a logistic regression multi-level model was used which is appropriate where one wants to predict the presence or absence of an outcome (in this case smoking or non-smoking) using a set of predictor variables. By fitting a multi-level model it was possible to examine whether variables measured at the PPS level influenced smoking behaviour above and beyond the effects measured for individuals and OAs (Goldstein, 1995). Multilevel models are an appropriate tool to explore variations in a population when data are hierarchial (e.g. individuals who are residents in a neighbourhood that are nested within towns etc.) (Jones, 1991). This approach allows for the relationship between the variance within the population and within each level to be quantified.

Multilevel modelling techniques have been used in a range of contexts such as studying schools and their pupils (Goldstein, 1984; Aitkin and Longford, 1986) but also in health related studies. Examples include studies of geographical differences in diastolic blood pressure (Merlo *et al.*, 2001), frequency and timing of antenatal care (Magadi *et al.*, 2000) and chronic illness (Jones and Duncan, 1995). In addition multilevel modelling has been shown to be a useful way to consider the various factors that influence smoking behaviour (e.g. Duncan *et al.*, 1999; Twigg, 1999; Twigg *et al.*, 2000).

Given the nature of the data, a three-level model would be preferable, where individuals were nested within OAs, which themselves nest within PPSs. However, the number of individuals per OA (an average of less than two) meant that it was not possible to treat OAs as the second level. Consequently, the individual-level

variables were used (age and sex) along with the OA variables at level one and two PPS variables at level two.

Initially, a simple model was used that included only age and sex as explanatory variables to test whether there was significant random variation at the PPS level (level 2) before the census-based explanatory variables were included. Then, each of the remaining potential explanatory variables were tested individually, both in their raw and logged form, where applicable, and the most significant of the two was retained in the model. Interactions between age and sex and between age, sex and all other census-based variables were tested. If either the variable, the log of the variable or any of the interactions were found to be significant they were retained.

The next stage of the analysis was to use the results of the model to calculate the probability of smoking for all OAs in Scotland. The model results only relate to individuals distributed across 7,127 OAs, so the model parameters for each age/sex group had to be applied to all 38,254 OAs in Scotland. The probability of smoking could be calculated from the multi-level model parameters for each age/sex group by extracting the data for these parameters for all areas across Scotland and then applying the parameter estimates to these variables.

5.5 Results and Discussion

The parameters of the first multi-level logistic model are shown in Table 5.4 which shows that the probability of smoking was lowest for females, compared to males, and those older than 65. The highest probability of smoking was for those aged 25-

34. These results correspond with what would have been expected from the raw probabilities in Table 5.1. The random variation at level 2 is significant, suggesting that the probability of smoking varies between PPSs. The deviance value of the intermediate model is 16143.3, which has dropped from 16576.8 in the null model. However, the deviance values of a logit model must be interpreted with caution because of the non multi-level estimation procedure in multi-level software packages (Rodriguez and Goldman, 1995).

Variable	**Parameter estimate**	**Standard error**
Constant	-0.598	0.072
Level 1 fixed effects		
Female	-0.084	0.039
Age 25-34	0.158	0.078
Age 35-44	-0.014	0.079
Age 45-54	-0.004	0.081
Age 55-64	-0.108	0.082
Age 65+	-0.748	0.080
Level 2 random effects	0.195	0.025
-2 * Log Likelihood 16143.3		

Table 5.4. Logit model with smoking as the dependent variable and age and sex as independent variables

The model was then extended to include the census-based explanatory variables and a total of ten variables were significant in the final model, along with one significant interaction (Table 5.5). The deviance value of the final model has dropped to 15006.5, a reduction of 1136.8 from the intermediate model (Table 5.4). Both of the individual-level variables (age and sex) were highly significant, which is consistent with a number of previous studies (e.g. Amos *et al.*, 1990; Uitenbroek and McQueen, 1993; Twigg *et al.*, 2000). It was perhaps surprising that the interaction between age and sex was insignificant in this model in light of the raw probabilities in Table 5.1

and this suggests that the observed age/sex differences are explained by other characteristics that are accounted for in this model.

Of the OA-level variables derived from the census which relate to individuals, the log of the percentage of individuals classified as working in clerical and secretarial occupations and the percentage of individuals classified as working in plant and machine operative occupations were found to be positively significant. This is consistent with the review of smoking rates that found that smoking behaviour is related to occupation (particularly manual professions) and social class. Of the household variables the percentage of households renting from a housing association or local authority, the percentage of privately renting households, the percentage of households without a car and the percentage of lone parent households all increased the probability of smoking. The first two of these are consistent with other studies that have shown smoking status to be strongly related to housing tenure (Graham and Der, 1999). Smoking also increased with population density, suggesting that living in an urban area increases the probability of smoking, as was also found in Twigg *et al.* (2000).

It was perhaps surprising, in light of previous studies, that some of the other variables were not significant. For example, a number of studies have shown that educational attainment is associated with smoking (e.g. King *et al.*, 1990; Shewry *et al.*, 1992; Escobedo and Pedicord, 1996; Graham and Der, 1999) whereas in this study the percentage of individuals with further education qualifications was insignificant. Marital status has also been shown to be associated with smoking (Amos *et al.*, 1990; Shewry *et al.*, 1992) but the percentage of one person households

and the percentage of individuals who are widowed or divorced were insignificant here. Studies such as those by Amos *et al.* (1990) and Shewry *et al.* (1992) have also shown smoking to be related to unemployment but in this analysis the percentage unemployed was insignificant. The age-specific and sex-specific variables (for example, the percentage of unemployed males and unemployed females) were not found to be as significant as the non-specific variables (such as the overall percentage unemployed).

Note, however, that the studies cited above used individual level data and the variables used here are measured at the OA level, raising questions relating to the ecological fallacy. For example, it is not known whether the individual smokers and non-smokers work in clerical or secretarial occupations; only the percentage of people living in the same area who work in clerical or secretarial occupations according to the 1991 census is known. It is not possible, therefore, to state that working in clerical or secretarial occupations relates to the probability of smoking, simply that the percentage of people who work in clerical or secretarial occupations living in the same OA has a significant effect on smoking probabilities, controlling for the other variables in the model.

The Carstairs index, measured for PPSs, was also significant in the final model. It had a positive effect upon the smoking probability suggesting that an increased level of deprivation in an area increased the probability of smoking, having controlled for other characteristics at the OA level. The ONS ward classification was not significant. The random variation that was significant in the model that included only age and sex as explanatory variables (Table 5.4) became insignificant. The

Variable	Parameter estimate	Standard error
Constant	-1.407	0.101
Level 1 fixed effects		
Female	-0.146	0.041
Age 25-34	-0.025	0.110
Age 35-44	-0.181	0.108
Age 45-54	-0.040	0.113
Age 55-64	-0.219	0.117
Age 65+	-0.707	0.081
Population Density	6.00E-06	3.00E-06
Log of % of individuals classified as working in clerical and secretarial occupations	0.034	0.013
% of individuals classified as working in plant and machine operative occupations	0.004	0.001
% of households without a car	0.008	0.002
% of households rented from a housing association or local authority	0.006	0.001
Log of % of privately rented households	0.034	0.013
% of lone parent households	0.013	0.004
Level 1: 2-way interaction		
Age 25-34 * HA or LA	0.005	0.002
Age 35-44 * HA or LA	0.008	0.002
Age 45-54 * HA or LA	0.004	0.002
Age 55-64 * HA or LA	0.004	0.002
Level 2 fixed effects		
Carstairs index	0.026	0.007
Level 2 random effects	0.018	0.014
-2 * Log Likelihood 15006.5		

Table 5.5. Final logit model with smoking as the dependent variable

inclusion of the area-based variables in the model (Table 5.5) appears to have accounted for the random variation at level two.

The only significant interactions were for those four age groups between 25 and 64 and the percentage of households renting from a housing association or local authority. There was a greater probability of smoking among these age groups in OAs with a high percentage of households renting from a housing association or local authority. This was not found for those aged 65 and above, who also happen to have the lowest rates of smoking of all age groups.

The results for each age/sex group from calculating the probabilities using the final multi-level logistic model are shown in columns two and three of Table 5.6. These values represent the probability of smoking in different age/sex groups using the information from the SHS and the associated census data which would be expected to be slightly different to the raw probabilities in Table 5.1 because they only use data from 7,127 OAs. The table shows that the female smoking rates were lower than males in all age groups whereas in Table 5.1 this relationship was not apparent in all categories. Furthermore, the calculated smoking rates for males were higher than the raw probabilities in all age groups. For females, the calculated probabilities are lower than the raw probabilities in the younger age groups but higher in the older age groups.

The next step involved the calculation of the probabilities of smoking for each age/sex group at the OA level across Scotland. The parameter estimates from the final multi-level logistic model were applied to the corresponding census variables for all 38,254 OAs in Scotland. This was possible for 35,994 out of the total of

38,254 (94.1%). It was not possible to calculate probabilities where one or more of the explanatory variables was missing in a particular OA. The 'percentage of individuals classified as working in clerical and secretarial occupations', which is derived from the 10 percent sample from the census, was missing[5] in 1,699 OAs, for example.

Age	Male	Female
16-24	0.348	0.317
25-34	0.388	0.357
35-44	0.382	0.352
45-54	0.375	0.344
55-64	0.356	0.326
65+	0.213	0.191

Table 5.6. Logit model probabilities of smoking

The means of the resulting calculated probabilities for each age/sex group across all 38,254 OAs are shown in Table 5.7. These can be compared to the probabilities from the raw data (Table 5.1) and the probabilities derived from the multi-level logistic model (Table 5.6). In general, the calculated probabilities of smoking were higher in the younger age groups and among men. Because the age/sex interaction was insignificant in the model the predicted male smoking rate will always be higher than the female smoking probability unlike the results in the raw data (Table 5.1). This suggests that the difference between the men and women in Table 5.1 is due to other characteristics of men and women that are controlled for in the model (Table 5.5).

[5] Census data is suppressed for areas that do not meet a minimum population threshold

Age and sex group	Mean	Range	Minimum	Maximum
Males 16-24	0.355	0.622	0.170	0.792
Females 16-24	0.324	0.616	0.151	0.767
Males 25-34	0.396	0.693	0.167	0.860
Females 25-34	0.365	0.693	0.148	0.841
Males 35-44	0.390	0.730	0.146	0.876
Females 35-44	0.360	0.730	0.129	0.859
Males 45-54	0.384	0.680	0.165	0.845
Females 45-54	0.353	0.679	0.146	0.825
Males 55-64	0.364	0.706	0.142	0.848
Females 55-64	0.334	0.703	0.125	0.828
Males over 65	0.220	0.560	0.092	0.652
Females over 65	0.197	0.538	0.080	0.619

Table 5.7. Summary statistics for the predicted probabilities for all output-areas in Scotland

Table 5.7 also provides the range of predictions for each age/sex group across the Scottish OAs. A different probability of smoking has been calculated for each group in each OA, based on the characteristics of the OA. The range of probabilities for each of the groups is quite large. For example, the predicted smoking probability for males aged 16-24 has a maximum value of 0.792, suggesting that approximately 79 percent of men aged 16-24 smoke in that particular OA.

This OA falls in the PPS 'Edinburgh City' and has particularly high values for: population density (144,000 per km^2), Carstairs Score (4.74), households with no car (86.2%), households rented from a housing association or local authority (100%), and lone parent households (22.7%). The lowest probability that was calculated in an OA for males aged 16-24 was 0.17. This OA fell within the PPS of 'Renfrew' on the edge of Glasgow. It had low values for population density (1562.5 per km^2), Carstairs Score (-7.37), households with no car (0.74%), percentage households

rented from a housing association or local authority (0%), percentage of privately rented households (0%), percentage of individuals classified as having worked in clerical and secretarial occupations (0%) and the percentage of lone parent households (0%).

The probabilities for a selection of age/sex groups are mapped in Figures 5.4 to 5.6. Figure 5.4 shows the probability of smoking for people aged 16-24 for the whole of Scotland and there are higher smoking rates in urban areas, particularly those that were more deprived. Figures 5.5 and 5.6 show the smoking probabilities for women aged 16-24 and over 65, respectively, in an area around Dundee including parts of Fife and Perthshire. The geography of smoking is broadly similar for these two age groups, even though the overall levels of smoking are much lower among the elderly. This suggests that the geography of smoking is broadly consistent for these different age groups and this trend is consistent across the age and sex groups.

5.6 Calculating Missing Smoking Probabilities

It was not possible to calculate the probability of smoking for 2,259 of the 38,254 OAs in Scotland because the data for one or more of the parameters were missing. An alternative method for estimating smoking probability was therefore developed for those OAs where there were missing data for one or more of the model parameters. The population-weighted mean smoking rates of the adjacent OAs were calculated as an estimate of the missing probabilities for all the age/sex groups in each OA.

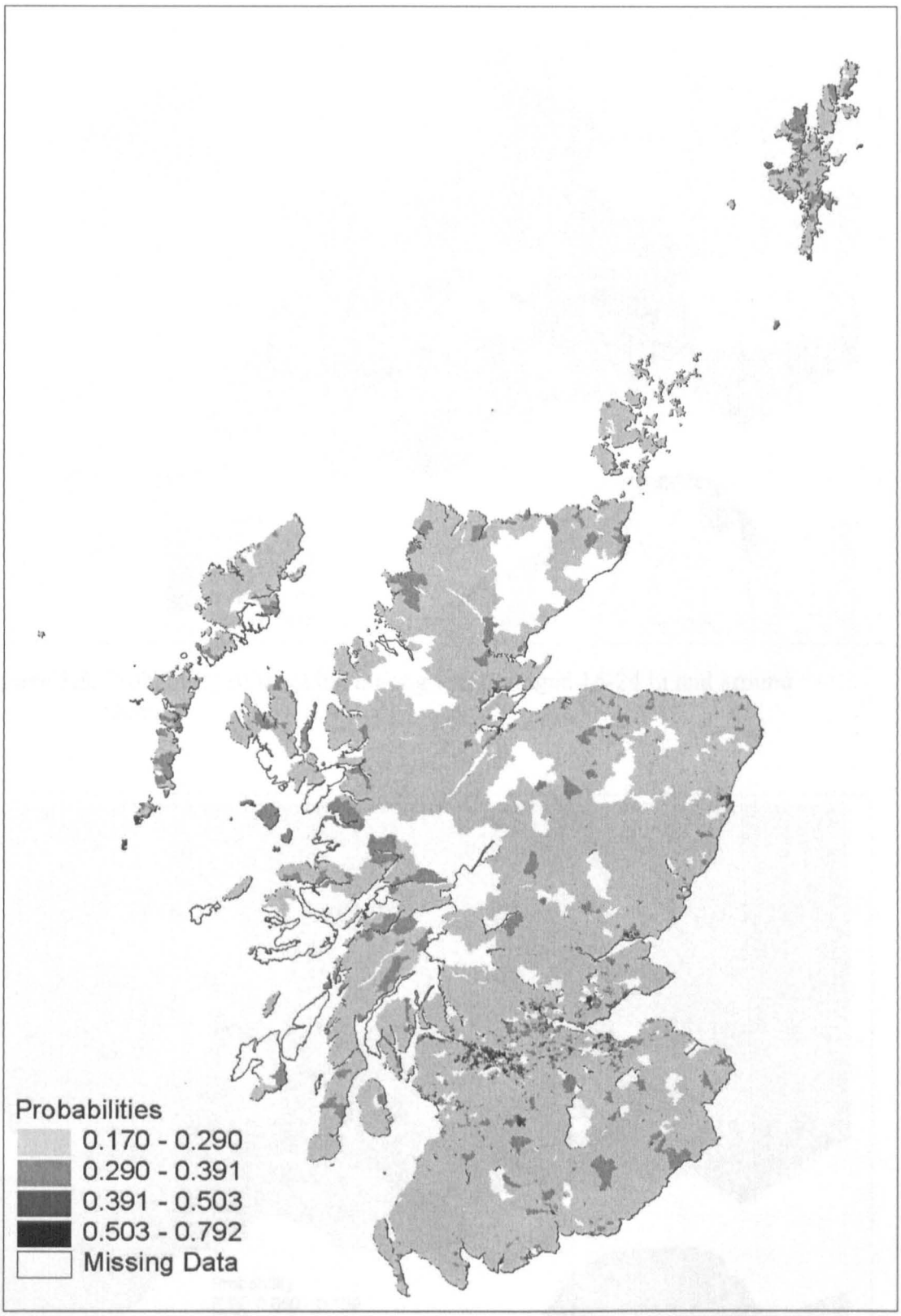


Figure 5.4. Probability of smoking among males aged 16-24 in Scotland

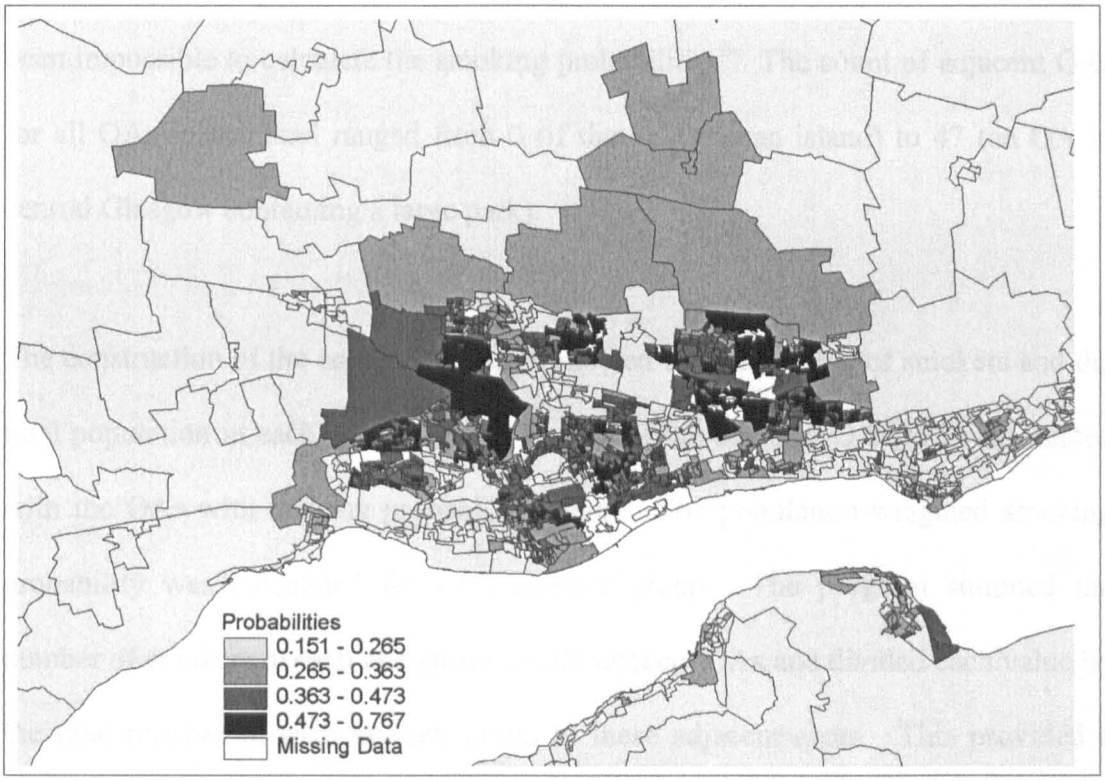


Figure 5.5. Probability of smoking among females aged 16-24 in and around Dundee

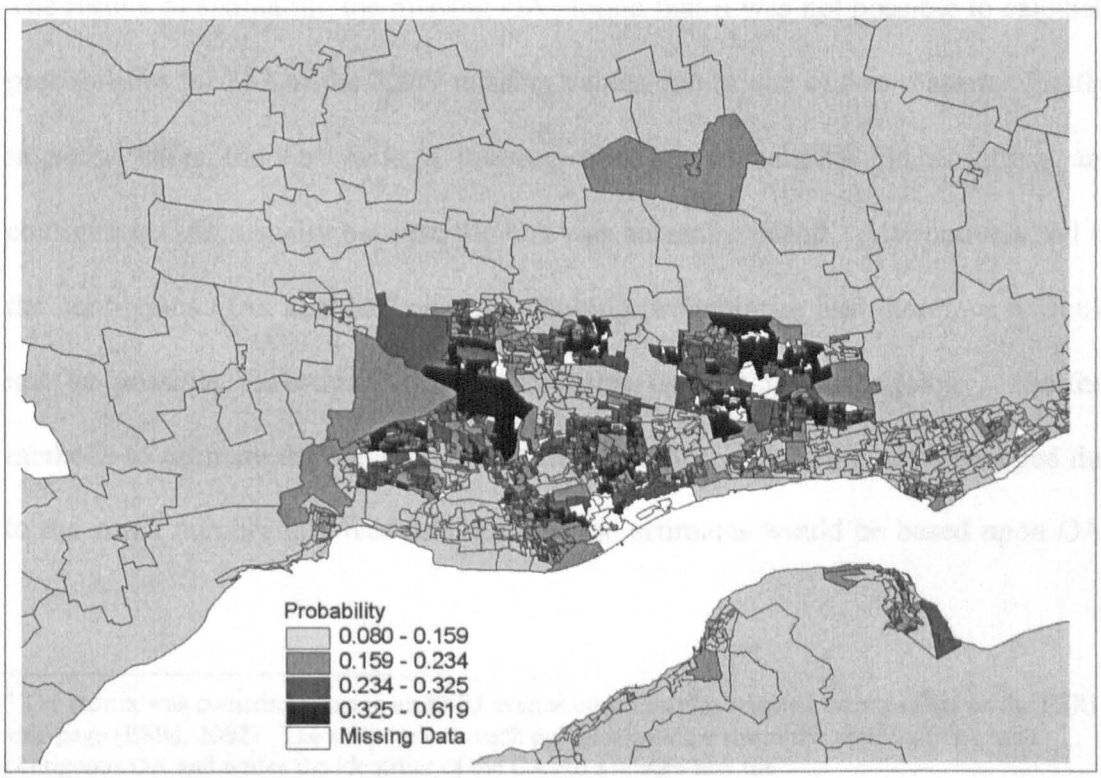


Figure 5.6. Probability of smoking among females aged 65+ in and around Dundee

The first stage was to create a contiguity matrix for each of the OAs where it had been impossible to calculate the smoking probabilities[6]. The count of adjacent OAs for all OAs in Scotland ranged from 0 (if the OA was an island) to 47 (an OA in central Glasgow containing a large park).

The construction of the contiguity matrix allowed for the number of smokers and the total population in each age/sex group in each of the adjacent OAs to be associated with the OAs with missing probability values. The population-weighted smoking probability was calculated for each age/sex group. The program summed the number of smokers in each age group for all adjacent OAs and divided each value by the total number in each age/sex group in these adjacent areas. This provided a population-weighted smoking probability for each age/sex group in the target OA.

The results of estimating the missing OAs found that it was not possible to calculate probabilities for 332 of the 2,259 missing values due to one of two reasons. Firstly, in some cases the OA with a missing smoking probability did not have any contiguous OAs, usually because the OA was an entire island. Alternatively, all of the contiguous OAs also had missing smoking probabilities and therefore it would not be possible to estimate the probabilities using this methodology. Further methods to estimate the smoking probabilities in these OAs were not developed due to the small number involved and because the estimates would be based upon OAs

[6] The matrix was constructed using an ESRI avenue code that was adapted from a script on the ESRI web page (ESRI, 2002). The script selects each polygon within a shape file and identifies each contiguous OA and writes the identifier of the OAs to a named text file

that were some distance away. However, the result of this procedure was that the smoking probabilities have been calculated or estimated for 12 age/sex groups in 37,921 of the 38,254 OAs in Scotland (99.13 percent).

5.7 Smoking and Lung Cancer

Table 5.8 shows the relationship between the smoking probability estimates and the age-standardised rates of lung cancer incidence. The estimated probabilities of smoking that were calculated above for all OAs were summarised by calculating a single value that represented the population-weighted average smoking rate. The estimates were categorised into five equal groups that represented different smoking probabilities and then the SIR for lung cancer was calculated for each group.

Table 5.8 demonstrates that the SIR for lung cancer increased with the smoking probabilities. In the lowest smoking band where the probabilities of smoking were 0.2037 or less, the SIR was 59.02 whereas in the highest band where the smoking probabilities were 0.4530 or more, the SIR was 156.45. The SIR was below 100 (the expected number of cases were greater than the observed number of cases) in the lowest three smoking categories but the SIR was over 100 in the highest two categories (the observed number of cases were higher than the expected number). Although it was established in Chapter Two that smoking is the most important cause of lung cancer, it is not possible to appreciate precisely how the two variables are related until they are examined in tandem with other potential risk factors. This is addressed in subsequent chapters.

Weighted Smoking Probability	Total Observed Cases	Total Expected Cases	SIR (per 100,000)
0.0000 - 0.2037	2224	3768.44	59.02
0.2038 - 0.2653	2512	3344.23	75.11
0.2654 - 0.3523	3953	4026.20	98.18
0.3524 - 0.4529	5013	4202.61	119.28
0.4530 - 1.0000	4701	3004.76	156.45

Table 5.8. Relationship between population-weighted smoking probability and lung cancer SIR.

5.8 Conclusion

This chapter presents a method for calculating smoking probabilities for OAs across the whole of Scotland using a multi-level logit modelling approach. The technique used individual-level data from the SHS in tandem with area-based census and census-derived variables. The mapped results from this model demonstrate the higher probability of smoking in urban areas with particularly high values in more deprived areas. Lower probabilities are found in more suburban and rural areas. The results seem to match those from studies in the past that have attempted to address this problem at coarser geographical scales (e.g. Twigg *et al.*, 2000).

Although the results appear broadly sensible there are a number of data issues to be aware of. First, 1991 census data were used to model individual level survey data from 2000. Even though the demographic and economic makeup of most OAs change only very gradually, some OAs may have changed markedly in the nine years between the 1991 Census and the SHS. It will be interesting to run a similar analysis using the results of the 2001 census, when they become available, to see how the results from the two analyses differ.

Second, there are confidentiality issues relevant to the release of the 1991 census data that may have had a small influence on the analysis. To maintain confidentiality, OA level counts have 1, 0 or –1 added randomly to them (Dale and Marsh, 1993). Clearly this may influence the percentage values of the explanatory variables, especially where the counts were small.

Third, in a number of OAs missing data made it impossible to calculate a probability because there was a small population. If data were missing for one or more of the variables when the parameters were applied across Scotland, then it is not possible to directly calculate the probability value. However, in most cases, population-weighted smoking probabilities were calculated for areas with missing values from those of contiguous OAs.

6. PREDICTING POLLUTION IN SCOTLAND

6.1 Introduction

Air pollution has been identified as an important cause of respiratory and lung disease as a consequence of both short-term, but severe, air pollution events and also longer term exposure (Cohen, 2000). A number of carcinogenic substances are found in air pollutants and some such as benzene and particulate pollutants are thought to increase lung cancer even at lower levels of exposure (Department of Health, 1997). This chapter presents methods for predicting the levels of air pollution in small areas across Scotland to investigate the relationship between air pollution and lung cancer but also allowing for air pollution to be controlled for when modelling lung cancer incidence. This is the first time that estimates of air pollution have been made for small areas in Scotland. In the analysis, levels of 'black smoke' were estimated, which is the pollutant that is most strongly related to lung cancer (Arden Pope and Dockery, 1999). Estimating the level of pollution across Scotland is problematical because air pollution data are only collected for a small number of locations and therefore methods have to be developed to estimate the likely levels of black smoke in areas for which no recordings have been made.

Models were developed for identifying the key variables that were related to pollution levels in areas for which pollution data were available using a conceptually similar model to that used to estimate smoking behaviour in areas for which no information on smokers and non-smokers was available. The parameters from this model were then used to make predictions of black smoke levels for all areas in

Scotland. Pollution values were calculated for 1971 and 1991, firstly to provide estimates of pollution exposure 20 years prior to lung cancer diagnosis which is the recognised lag period between exposure to a carcinogen and cancer development (Pike and Forman, 1991) and, secondly, to provide estimates that were consistent with other data sets in this study. Although 1971 estimates may seem to provide the most useful estimate of exposure to pollution in a study of lung cancer, this would be ignoring the migratory history of the lung cancer patients. As discussed in chapter one, the data on lung cancer incidence did not include a residential history of each patient and so it is not known where they were living in 1971. Equating the person's residential location in the early 1990s with the pollution in that area 20 years previously clearly ignores the fact that the person may well have resided elsewhere at that time. This represents the conundrum posed when attempting to model lung cancer. Therefore, both estimates will be tested and compared in a regression model of lung cancer incidence in chapter eight to help decide which is the most appropriate.

6.2 Background to Predicting Air Pollution

Levels of air pollution will vary at a number of geographical scales in response to the location of pollution sources and to the influences of environmental factors such as topography and meteorology (Dunn and Kingham, 1996). Therefore, in the absence of information on an individual's exposure to pollution, it is important to develop accurate methods to predict air pollution for small areas when trying to investigate the health effects of air pollution. There are a number of approaches that can be taken to predict air pollution although most have tended to focus upon atmospheric

models at small geographical scales (e.g. within a city) rather than estimates for larger areas across a region (e.g. Godzik, 1997; Davis *et al.*, 1998; Tayanc, 2000). However, small-scale predictions are not appropriate to this study as they usually rely upon a highly complex combination of localised meteorological and topographic factors as well as a large amount of geographically detailed pollution data that are inappropriate to apply to large areas such as the whole of Scotland. Instead, this study used an alternative approach that predicts pollution using data from the census, and on the road network, that are available for the whole study area.

Dunn and Kingham (1996) have identified four key issues that need to be addressed when measuring air pollution exposure: the type of monitoring; what to monitor; where to monitor; and when to monitor. There are two distinctive types of monitoring - statutory monitoring and non-statutory monitoring. Statutory monitoring relies upon an existing network of air pollution data collection whereas non-statutory monitoring uses data that have been especially collected, usually for small geographical areas, and the type of monitoring used depends upon the nature of the problem being addressed. If the research is focusing upon defining an individual's exposure then the use of personal monitors is appropriate, although this can be an expensive strategy. However, if the research is focused upon the impact of a point or linear source of emissions then the fixed air quality monitoring approach is more suitable and probably more practical. What to monitor should clearly be driven by the nature of the investigation. For example, in a study of lung cancer, data on emissions such as sulphur dioxide, benzene and particulate pollutants of different sizes would be most relevant. Ideally the monitoring program should be designed

and initiated by the researchers to reflect the needs of the research but in practice more pragmatic approaches are often taken which reflect statutory monitoring.

The choice of the location of monitoring sites is an important consideration especially for studies that do not rely upon statutory monitoring. Sites can be selected to represent different areas such as administrative regions or areas defined on the basis of epidemiological reasoning. A number of influencing factors need to be considered in designing a monitoring programme, such as the size of the study area, location and number of possible sources of pollution, topography, site features, including presence of buildings, and security of equipment (Laxen and Noordally, 1987). The choice of location is often restricted by the costs and practicalities of siting a specific station (Hewitt, 1991). Although continuous monitoring is clearly desirable, limitations on resources often restrict monitoring programs to taking snapshots of air quality. However, the length of time between measurements needs to be considered. If weekly or monthly averages are taken then this may conceal important peaks in pollutants which may be significant for health outcomes. Conversely, measurements taken 24 hours apart may be unrepresentative or ignore daily fluctuations. These issues are important because it is essential to identify both the precise pollutant that is of interest and also how best to gain a representative sample of the pollutant in terms of where and when it is measured.

Assuming that an appropriate and representative set of pollution data have been collected, some form of interpolation of the data often needs to take place, regardless of the quality and resolution, if estimates are required for a specified area. This is an approach that has been facilitated by the advent of GIS packages that readily include

algorithms for interpolation such as Thiessen polygons, trend surface analysis and kriging. These procedures, often in tandem with other statistical techniques, have been used to provide estimates of exposure to pollutants in order to help to understand their role in affecting health outcomes (Briggs *et al.*, 1997).

An example of this is provided by Holland *et al.* (2000) who fitted a generalised additive model (GAM) to airborne sulphur dioxide (SO_2) at each of 35 sites in the eastern United States to estimate the form and magnitude of the 'site specific trend' (or percentage total change) from 1989 to 1995. This was adjusted for the influences of meteorology and season and the results were compared to the results from an adapted kriging methodology that constructed spatially-smoothed estimates of airborne SO_2 levels for three large regions in the eastern US. The comparison demonstrated similar results in terms of regional trend and standard error. Briggs *et al.* (2000) modelled spatial patterns of Nitrogen Dioxide (NO_2) concentrations in four urban areas in the UK. They used data on road traffic, urban land use and topography as predictors of NO_2 levels. The model was calibrated and then validated using data taken from randomly selected sites. The model performed well in all cases with R^2 values between modelled and observed concentrations in the range of 0.58 to 0.76. Similarly, Diem and Comrie (2002) used spatial interpolation methods to predict ground level ozone concentrations in Tucson, Arizona. The authors used gridded estimates of emissions of ozone precursor chemicals and these provided the core predictor variables in multiple regression models. Cross-validation of the models revealed an overall R^2 of 0.90 and approximately a seven percent error.

It is clearly important to consider carefully both the nature of the data and the method of interpolation being used when attempting to estimate pollution levels. The spatial and temporal resolution of the data, the quality of the data and the method of interpolation employed have important implications for the estimated level of exposure to air pollution. This is especially problematical in broad regions, such as Scotland, where the number of monitoring sites is small.

6.3 UK Air Pollution Data

Air quality data in the UK are administered and disseminated by the National Air Quality Information Archive on behalf of the UK Department for Environment, Food and Rural Affairs (NETCEN, 2001). Data are collected from 1,500 sites across the UK and these sites are organised into networks that gather air quality information for particular purposes. Therefore, only a subset of these sites will collect information for a particular pollutant. The data collected include information on nitrogen dioxide, sulphur dioxide, ozone, particulate pollutants (including black smoke) and various types of hydrocarbons. The general aims of all the networks are outlined by NETCEN (2001) as:

- To understand air quality problems in order that cost effective policies and solutions can be developed
- To assess how far standards and targets are being achieved
- To provide public information on current pollution levels and forecast air quality
- To assist the assessment of personal exposure to air pollution

- Where appropriate, check compliance with existing European Community Directives

Understanding the geography of lung cancer requires that air pollution is considered, as many studies have pointed towards the influence that particles in the air have on its incidence (Doll, 1978; Cohen, 2000; Rushton, 2001). Particulate pollution is now measured and disseminated in a disaggregated form and the particle size of the pollutants are reported (e.g. PM10, PM2.5). However the collection of size-specific data only began in the UK in 1986 and it is only more recently that there have been more than a handful of sites collecting this information. Prior to this, 'black smoke' measurements were collected, which is a more aggregated measurement that has been implicated by numerous studies (e.g. McClellan, 1996). This measures all suspended particulate matter in the air (Muir and Laxen, 1995; Schwartz *et al.*, 2001) that is emitted mainly from fuel combustion. Data on black smoke have been collected since 1970 and were available from the NETCEN web page.

Black smoke data can be downloaded for each site in the monitoring network including the arithmetic mean of all daily measurements for the year, the standard deviation for all daily measurements for the year, the value that 98 percent of all values fall below and the maximum daily value recorded during the year. Information is also presented on the name and location of the site. Figure 6.1 shows the location and distribution of sites within the monitoring network in Scotland in

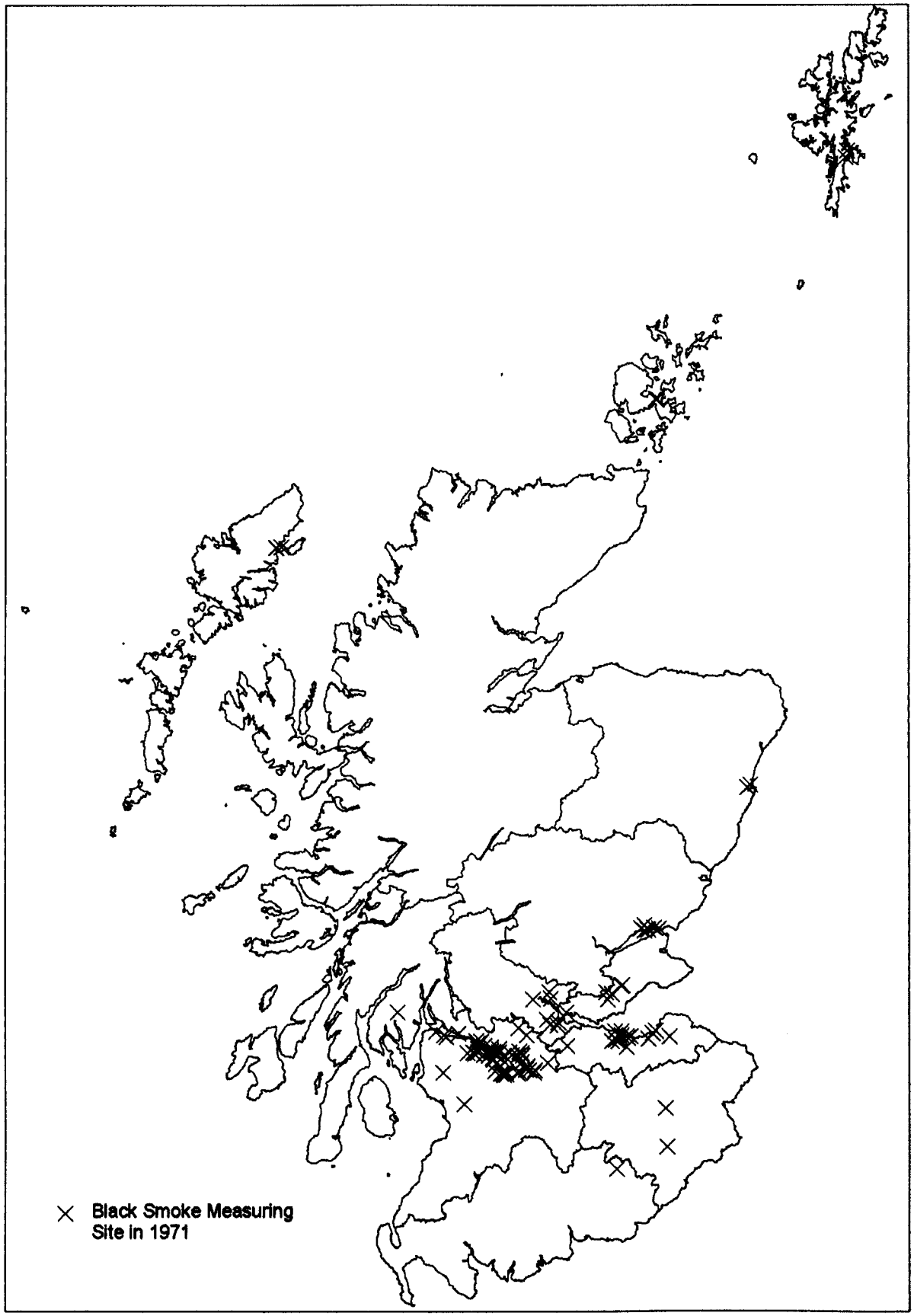


Figure 6.1. Location of Black Smoke Measuring Stations in Scotland in 1971.

1971 and demonstrates that pollution is not recorded in much of Scotland. Most of the pollution monitoring sites were within Glasgow and Edinburgh, although some smaller clusters were in Dundee, Aberdeen and Fife. The remainder of the sample sites were distributed unevenly throughout Scotland and there were no collection points in the Highlands. Figure 6.2 shows the monitoring network in 1991 and demonstrates that there were fewer sites than in 1971 and that the distribution of the sites was different. There were far fewer sites in the central belt, none in Dundee, but more in the north east of Scotland.

6.4 Predicting Air Pollution in Scotland

6.4.1 Introduction

The sparse distribution of air pollution monitoring sites means that a technique is required for estimating pollution across most of Scotland. Although there are a number of possible approaches to modelling air pollution levels, many of them were not appropriate for this study because they were only suitable for small areas, or relied on a dense set of measurements. However, in Scotland the density of measurements varies considerably as there were relatively large numbers in the central belt but fewer in the Borders and Highlands. Furthermore, the estimates had to be calculated for areas (census units) that were compatible with other data sets that were used in this study (e.g. lung cancer incidence and smoking behaviour). This made many of the spatial interpolation methods inappropriate.

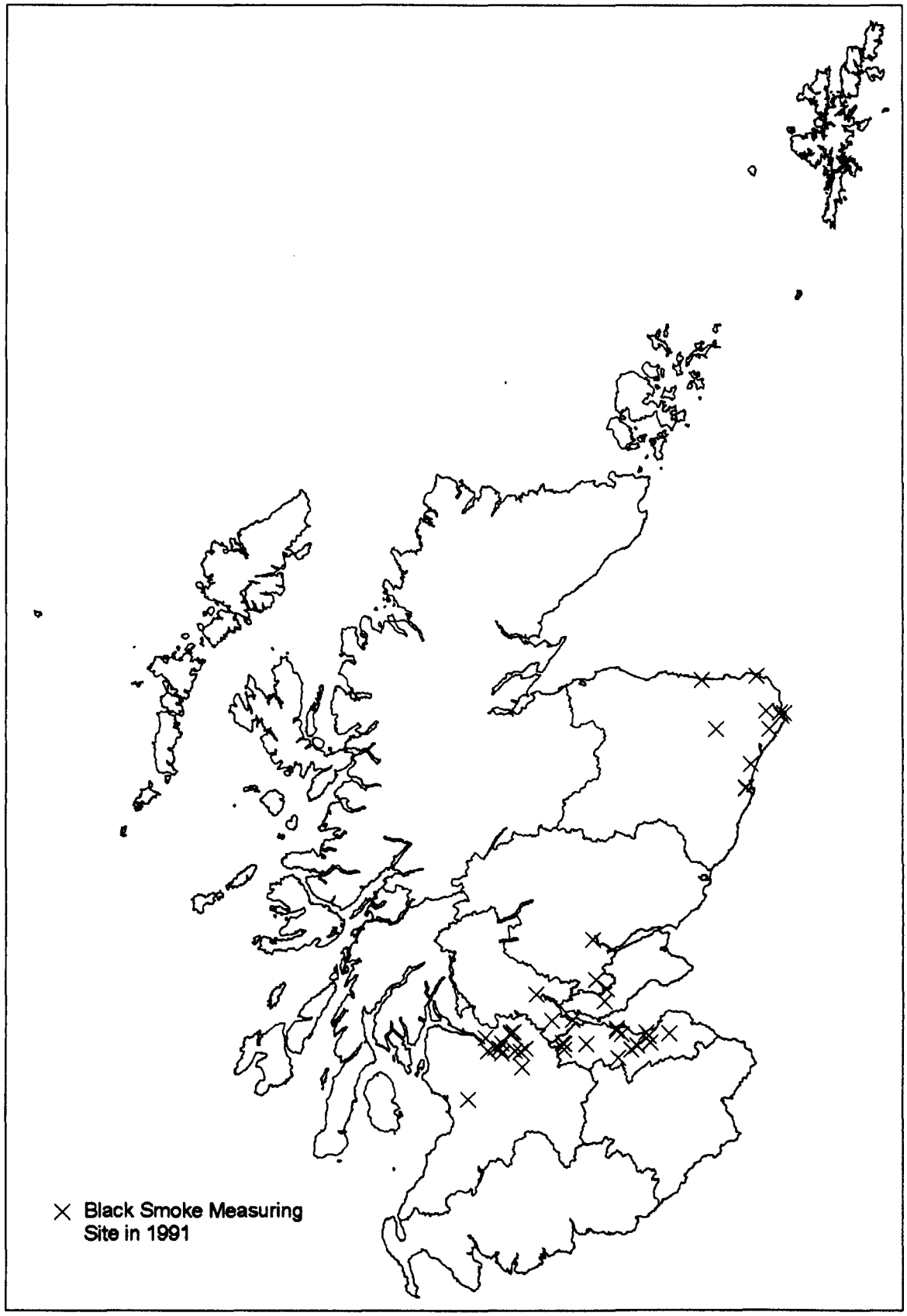


Figure 6.2. Location of Black Smoke Measuring Stations in Scotland in 1991.

Predictions were calculated for areas across Scotland using two sets of black smoke data for the years 1971 and 1991. 1971 was chosen as it would provide an estimate of pollution exposure 20 years before the date of the lung cancer registrations and 1991 was chosen because it coincided temporally with the data on lung cancer.

6.4.2 Methodology

Data

The pollution data for both years were downloaded from the UK National Air Quality Information Archive and where data were missing for a particular collection point in that year, it was supplemented by data from the previous year or, if that was also missing, from the subsequent year. The easting and northing of each monitoring station was also provided. In 1971 a total of 116 black smoke measurements taken at locations throughout Scotland were available (Figure 6.1). In 1991 there were just 53 black smoke measurements available and these were distributed unevenly across Scotland (Figure 6.2). The black smoke and SO_2 network was established in the 1960s when emissions from domestic and industrial coal burning were the main air pollution source. As this is no longer the case and it is now very rare to exceed the EC directives, the size of the network has gradually been scaled down (Lampert, private communication).

Data derived from the UK census can provide comprehensive information about the likely levels of pollution, particularly data on the occupation of workers (rather than residents) which are a good proxy for industry type. Information on where people

work will obviously provide a more valuable insight into the pollution than where they live, as workers often live and work in different PPSs. The Special Workplace Statistics (SWS) data set consist of a large set of matrices connecting origins and destinations that allows linkage to be made between people's home and workplace. These data were released in Scotland in 1991 for pseudo-postcode sectors (PPSs) (Flowerdew and Green, 1993). PPSs were the second smallest level of dissemination of Scottish census data (above the output-area) and so were compatible with other data sets used in this study. The SWS was therefore used to obtain variables on the types of employment undertaken by workers in all PPSs in Scotland and an additional variable was constructed representing the density of workers in each area as a proxy for the concentration of industry, which may be important even once the types of industries have been controlled for.

The SWS data are not available in Scotland for the 1971 or 1981 census' so 1991 data were used to predict pollution levels in 1971 and 1991. This had the advantage of maintaining consistent geographical boundaries for both of the estimates and with the data on lung cancer incidence. However, estimates of 1971 pollution levels based upon 1991 data were bound to be less reliable than if 1971 data had been available because of the industrial changes that have taken place over the 20 year period. It would have been preferable to use explanatory variables that coincided with the 1971 pollution data to model pollution levels, but these data were not available. In the absence of more suitable data the 1991 SWS was used instead.

Another important source of pollution is from road traffic emissions (Department of Health, 1997; Klaeboe *et al.*, 2000). Therefore, two further variables were calculated

representing the density of the road network in Scotland. The Ordnance Survey STRATEGI ntf tiles were downloaded from the Digimap web site at the Edinburgh Data and Information Access (EDINA) web site (EDINA, 2002) and the files were converted into ArcInfo coverages (Ordnance Survey, 2002). The data on roads were selected out from all the other available features and the 'line-in-polygon' functionality in ArcInfo was used to identify the PPS in which each arc of each road fell, allowing the calculation of the road density in each PPS. The density of all roads and that of just the major roads (motorways, A-Roads and B-Roads) were calculated for each PPS. Both variables were used because the total road density value will reflect the total length of residential streets which is highly correlated with population density. Residential streets have a relatively low traffic count and may not provide the biggest contribution to air pollution but may be important in explaining localised pollution levels in urban areas. Motorways, A-Roads and B-Roads, on the other hand, have a high traffic count and are likely to contribute more pollution per unit. Because it is not clear which measurement is the most appropriate, both were tested. Figure 6.3 shows that the density of all the roads is highest in the central belt, the main cities and the east coast of Scotland and lower in the rural areas. A similar pattern can be seen in Figure 6.4 which shows the density of motorways, A-roads and B-roads in each PPS in Scotland, although some parts of the major conurbations have a low road density because all of the roads within these PPSs were minor roads. A full list of the variables that were tested is shown in Table 6.1.

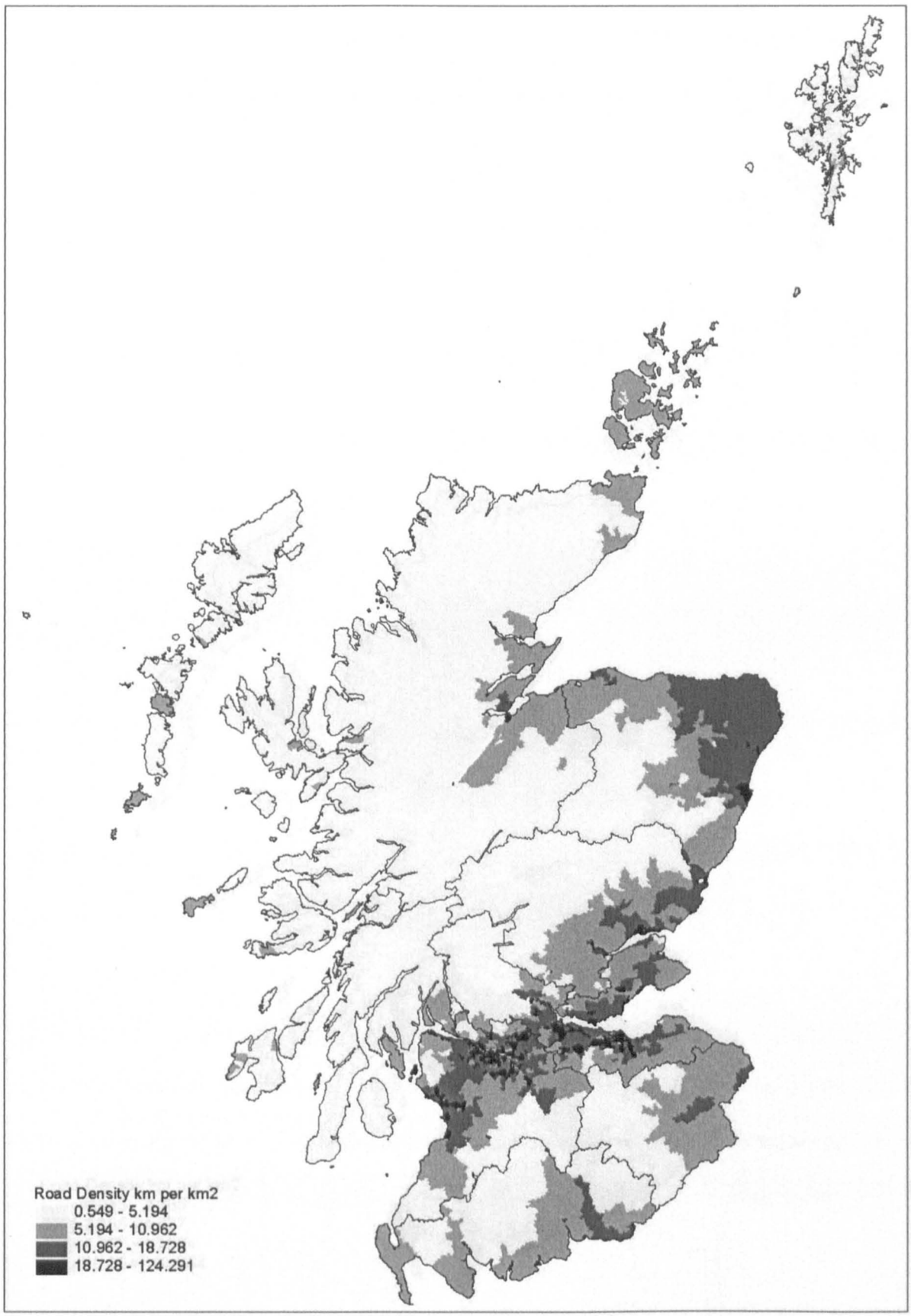


Figure 6.3. Density of all Types of Roads in PPSs in Scotland

Figure 6.4. Density of Motorways, A-Roads and B-Roads in PPSs in Scotland

1991 Pseudo-Sector Interaction Data (10% Sample)
% employed in a agriculture, forestry and fishing
% employed in energy and water supply industries
% employed in extraction of minerals and ores other than fuels: manufacture of metals
% employed in metal goods, engineering and vehicles industries
% employed in other manufacturing industries
% employed in construction
% employed in distribution, hotels and catering: repairs
% employed in transport and communication
% employed in banking, finance, insurance, business services and leasing
% employed in other services
Density of all Workers
Road Density
Density of All Roads
Density of Major Roads (Motorways, A Roads and B Roads)
Other
Log of Population Density

Table 6.1. PPS Variables Used in Regression Models

Methods

Using the GIS package ArcInfo, the 1991 PPS in which the measuring station fell was identified. This allowed 1991 census variables to be associated with each of the monitoring stations and this information was then used to predict the levels of pollution. Each black smoke measurement was linked to the corresponding SWS and road density variables.

Multiple regression methods were used to predict the black smoke level in 1971 and 1991 for each PPS. The mean black smoke reading was chosen as the dependent variable as this is the most representative measurement of pollution exposure over the course of a year. The log of mean black smoke was used because of the wide range of values. Each of the potential explanatory variables were tested individually, both in their raw and logged form, where applicable, and the most significant of the

two was retained in the model. A stepwise backward elimination approach was used so that all of the variables were entered in a block and then the least significant variables were removed one at a time until all the variables were significant. This found the balance between a high R^2 (hence explaining more of the variation in air pollution) whilst also deriving models that could be easily interpreted, which is sensible when the objective is to use the parameters for prediction.

Black smoke levels were only measured in 96 of the 1,003 PPSs in 1971 and 45 in 1991. Because pollution levels were not recorded in most PPSs in Scotland, they had to be estimated. Therefore, the parameters from the two models were applied to the variables in all of the PPSs in order to provide estimates of the level of pollution across Scotland.

6.4.3 Multiple Regression Results

Table 6.2 shows that multicollinearity was not a major concern as the correlation coefficients between the variables were not very strong in most cases. However, the strong correlation between the two road density measurements and the correlation of population density with some variables should be noted.

1971 Results

The final iteration from a backwards stepwise multiple regression model fitted for PPSs in Scotland is shown in Table 6.3 and the model had an R^2 value of 0.33. Five

	Agric	Energy	Extract	Metal	Oth Man	Construct	Distrib	Transp
% agric, forestry & fish.	1.00							
% energy & water	-0.10	1.00						
% extraction minerals etc	-0.08	0.01	1.00					
% metal goods, eng. Etc	-0.19	0.02	0.02	1.00				
% other manufacturing	-0.13	-0.01	-0.01	0.05	1.00			
% construction	-0.09	-0.03	-0.02	-0.07	-0.02	1.00		
% distrib, hotels etc	-0.20	-0.08	-0.08	-0.16	-0.14	-0.06	1.00	
% transport & comm'tion	-0.10	0.00	-0.06	-0.02	-0.12	-0.01	-0.04	1.00
% banking, finance etc	-0.16	-0.04	-0.07	-0.09	-0.14	-0.09	-0.01	-0.03
% other services	-0.35	-0.12	-0.13	-0.18	-0.27	-0.21	-0.26	-0.16
Log Of Pop Density	-0.61	0.03	-0.01	0.18	0.03	0.02	-0.03	0.06
Density of All Roads	-0.36	0.03	-0.05	0.09	-0.02	-0.07	0.01	0.09
Density of Major Roads	-0.37	0.02	-0.03	0.10	-0.02	-0.08	0.09	0.11
Density of all Workers	-0.13	0.00	-0.05	-0.04	-0.06	-0.11	0.00	0.03

	Banking	Other Serv	Pop Den	All Roads	Maj Roads	Workers
% agric, forestry & fish.						
% energy & water						
% extraction minerals etc						
% metal goods, eng. Etc						
% other manufacturing						
% construction						
% distrib, hotels etc						
% transport & comm'tion						
% banking, finance etc	1.00					
% other services	-0.09	1.00				
Log Of Pop Density	0.20	0.32	1.00			
Density of All Roads	0.33	0.13	0.66	1.00		
Density of Major Roads	0.26	0.11	0.63	0.85	1.00	
Density of all Workers	0.45	0.03	0.20	0.54	0.45	1.00

Table 6.2. Correlation Matrix of Variables Used to Predict Air Pollution Levels

R Square	Std. Error of the Estimate
0.33	0.78

Variable	B	Stand Error	Sig
Constant	4.805	0.891	0.000
Log of % employed in energy and water supply industries	-0.302	0.088	0.001
Log of % employed in distribution, hotels and catering: repairs	-0.440	0.168	0.010
Log of % employed in other services	-0.565	0.165	0.001
Log of Population Density	0.232	0.059	0.000
Density of Major Roads (Motorways, A Roads and B Roads)	0.020	0.009	0.025

Table 6.3. Multiple Regression Model: 1971 Pollution Data

variables were included in the final model: the log of the percentage of workers employed in energy and water supply industries; the log of the percentage of workers employed in the distribution, hotels and catering industries; the log of the percentage of workers employed in other services; the log of population density; and the density of major roads.

Each of the variables in the model seemed intuitively reasonable. For example, the coefficients for the log of population density and the density of major roads were positive and the log of the percentage of workers employed in the distribution, hotels and catering industries and the log of the percentage of workers employed in other services had negative coefficients. This was expected because both variables were likely to have higher values in the more rural areas where pollution levels were low. It is more difficult to explain why the coefficient for the log of the percentage of workers employed in the energy and water supply industries had a negative coefficient but this may be due to larger numbers of people employed in the water industry working in more rural locations with lower levels of pollution. It is clear that using 1991 census data and present day road data to predict 1971 pollution levels is problematical because there could have been many industrial changes

during that period of time. However, there was a strong relationship between the two sets of data and the socio-economic data explain much of the variation.

1991 Results

The final iteration from a backwards stepwise multiple regression model fitted for PPSs in Scotland with 1991 black smoke data as the dependent variable is shown in Table 6.4. This resulted in an R^2 value of 0.60 with four significant variables at the final iteration: the log of the percentage of workers employed in agriculture, forestry and fishing; the percentage of workers employed in other manufacturing industries; the percentage of workers employed in construction; and the log of the percentage of workers employed in energy and water supply industries. Most of these variables were intuitively reasonable. Thus, the log of the percentage of workers employed in agriculture, forestry and fishing had a negative impact on the pollution levels which seems sensible as the proportion employed in this industry will be higher in the rural areas with low levels of pollution. The percentage of workers employed in other manufacturing industries and the percentage of workers employed in construction had a positive effect upon the pollution level. This seems reasonable because they each represented heavily polluting industries and/or high concentrations of industry and so were likely to have larger values where pollution levels were high. Again, the log of the percentage employed in the energy and water supply industries had a negative estimate which lends support to the suggestion that this represented a large number of people employed in the water industry working in more rural locations. Therefore, the model parameters were generally sensible and the model explained much of the variation in the black smoke levels that were found in 1991.

R Square	Std. Error of the Estimate
0.60	0.50

Variable	B	Stand Error	Sig
Constant	1.691	0.156	0.000
Log of % employed in agriculture, forestry and fishing	-0.402	0.066	0.000
% employed in other manufacturing industries	0.020	0.009	0.027
% employed in construction	0.056	0.011	0.000
Log of % employed in energy and water supply industries	-0.157	0.058	0.010

Table 6.4. Multiple Regression Model: 1991 Pollution data

Comparison of the Models

There were some important distinctions to be made between the 1971 and 1991 black smoke regression models. The higher R^2 value that was derived from the 1991 model compared to the model that used the 1971 pollution data was not surprising given the temporal overlap between the data on levels of pollution (black smoke) and the explanatory variables. The analysis in 1991 is not susceptible to the industrial changes that have occurred during the 20-year period whereas the SWS data will less accurately represent the situation in 1971. For example, changes in employment types that would have been a good indicator of pollution in 1971 (e.g. coal mining) may not be represented in the 1991 census due to changes in employment patterns between 1971 and 1991 (e.g. closure of coal mines). Similarly, the road data use the current roads so this will better represent the road network in 1991 than 1971. Importantly, there will have been an overall reduction in heavy industry between the two years, which is difficult to model when the two sets of estimates use the same explanatory data.

Although the prediction of black smoke pollution is likely to be more accurate in 1991 than 1971, it remains desirable to examine the effects of both estimates on the

incidence of lung cancer because the 1991 prediction does not allow for the time lag between lung cancer initiation and development to be accounted for. It is possible that it is more important to incorporate the time lag into the analysis than to use the most accurate measure of prediction. A comparison of the effect that each estimate has when modelling lung cancer incidence will give an insight into how important the influence of the lag between exposure to pollution and cancer initiation is and whether it is useful to incorporate this lag at the expense of the more accurate predictions of black smoke levels.

6.5 Applying Model Parameters Across Scotland

Estimates of pollution levels for the two periods of interest (1971 and 1991) can be calculated for each PPS in Scotland using the parameters from the two regression models (Tables 6.3 and 6.4). Data were extracted for the parameters in the two models for all PPSs in Scotland and then applied to each set of variables to estimate pollution for 1971 and 1991. Because the log of black smoke was used as the dependent variable, the exponential of the prediction was used to provide the estimate in the original unit of measurement. It was not possible to calculate estimates for five PPSs in 1971 and 1991 due to missing data in those PPSs.

Figures 6.5 and 6.6 show the distribution of the estimated values. In 1971 the majority of the estimates fell within the 0 to 50 tonnes km^2 bands with 850 of the 1003 PPSs in the 0 to 50 tonnes km^2 band. There were 42 PPSs that exceeded 100 tonnes km^2, three of which were above 500 tonnes km^2, which may reflect the

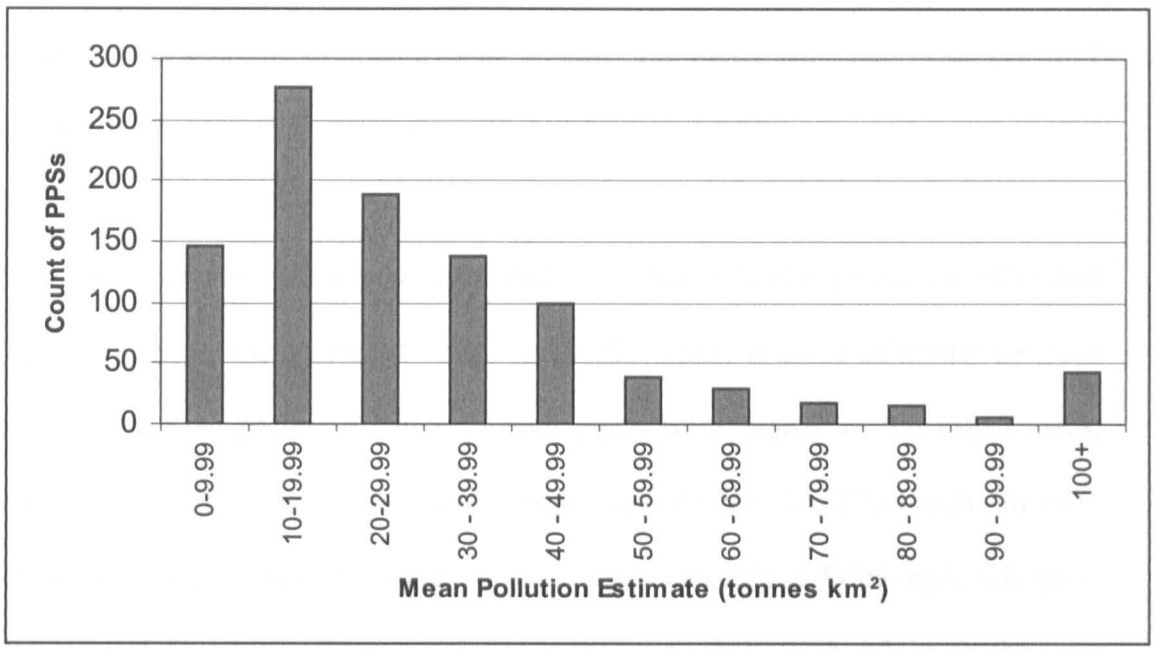


Figure 6.5. Distribution of Pollution Estimates in Scotland: 1971.

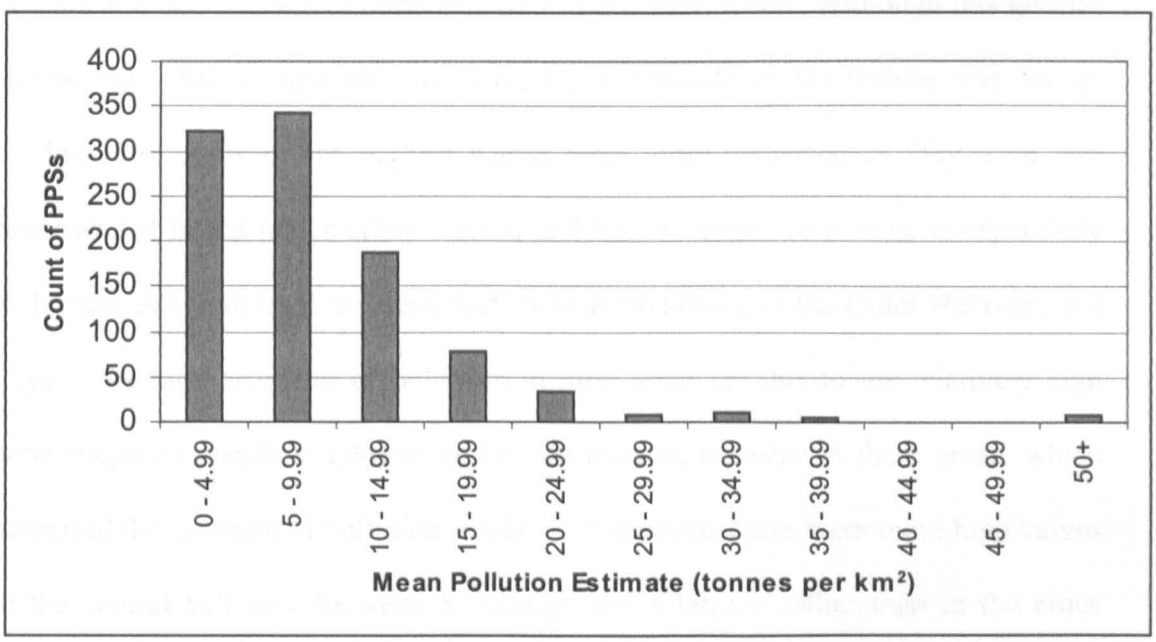


Figure 6.6. Distribution of Pollution Estimates in Scotland: 1991.

inaccuracies of the model. In 1991 the majority of values fall within the 0 to 20 tonnes km^2 (931 PPSs) band and only a handful (27 PPSs) exceeding 30 tonnes km^2, which reflects the national reduction in pollution during this period. The overall

reduction in the levels of pollution between the two years is also reflected in the difference between the estimated mean values. The mean estimated value for 1971 was 36.30 whilst in 1991 it had dropped to 9.44 tonnes km^2.

Figure 6.7 shows a map of the estimated mean black smoke values for 1971 and Figure 6.8 shows the estimated values for 1991. There is some similarity between the two maps as pollution levels in both years were higher in the urban areas, especially across the central belt that incorporates Glasgow and Edinburgh. In 1971 the highest values were in the major urban areas particularly Edinburgh, Glasgow, Dundee and Aberdeen. There were also some particularly high values on the west coast just west of Loch Lomond, the south part of the Isle of Mull and south of Mallaig and also in parts of north east Scotland around Keith. Although this general pattern was still recognisable in 1991, the magnitude of the values was lower. Furthermore, many of the highest values were more dispersed as they were less concentrated in the major urban centres and instead there were some unexpectedly high estimates in more rural areas such as near Inverness, in the Outer Hebrides and Skye. The high estimates of pollution in rural areas are due to the relatively high percentages of people employed in the construction industry in these areas, which increased the estimate of pollution levels. Furthermore, there were more high values in the central belt area between Edinburgh and Glasgow rather than in the cities themselves, reflecting the reduction of city centre heavy industry.

These comparisons were confirmed by the Pearson correlation between the 1971 and 1991 estimates which is 0.30, suggesting that there is a relatively weak degree of

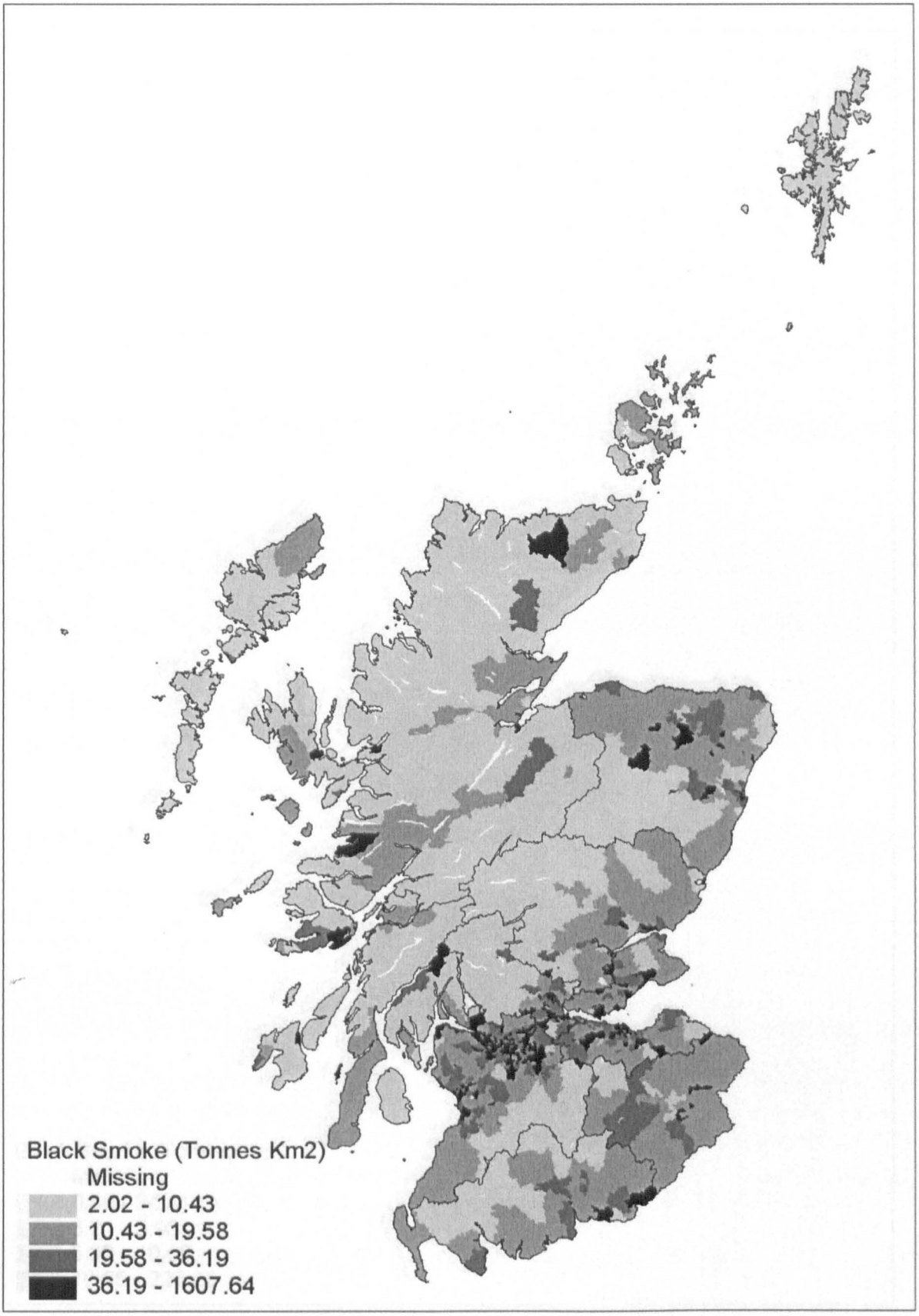


Figure 6.7. A Map of Estimates of Black Smoke Levels in Scotland in 1971.

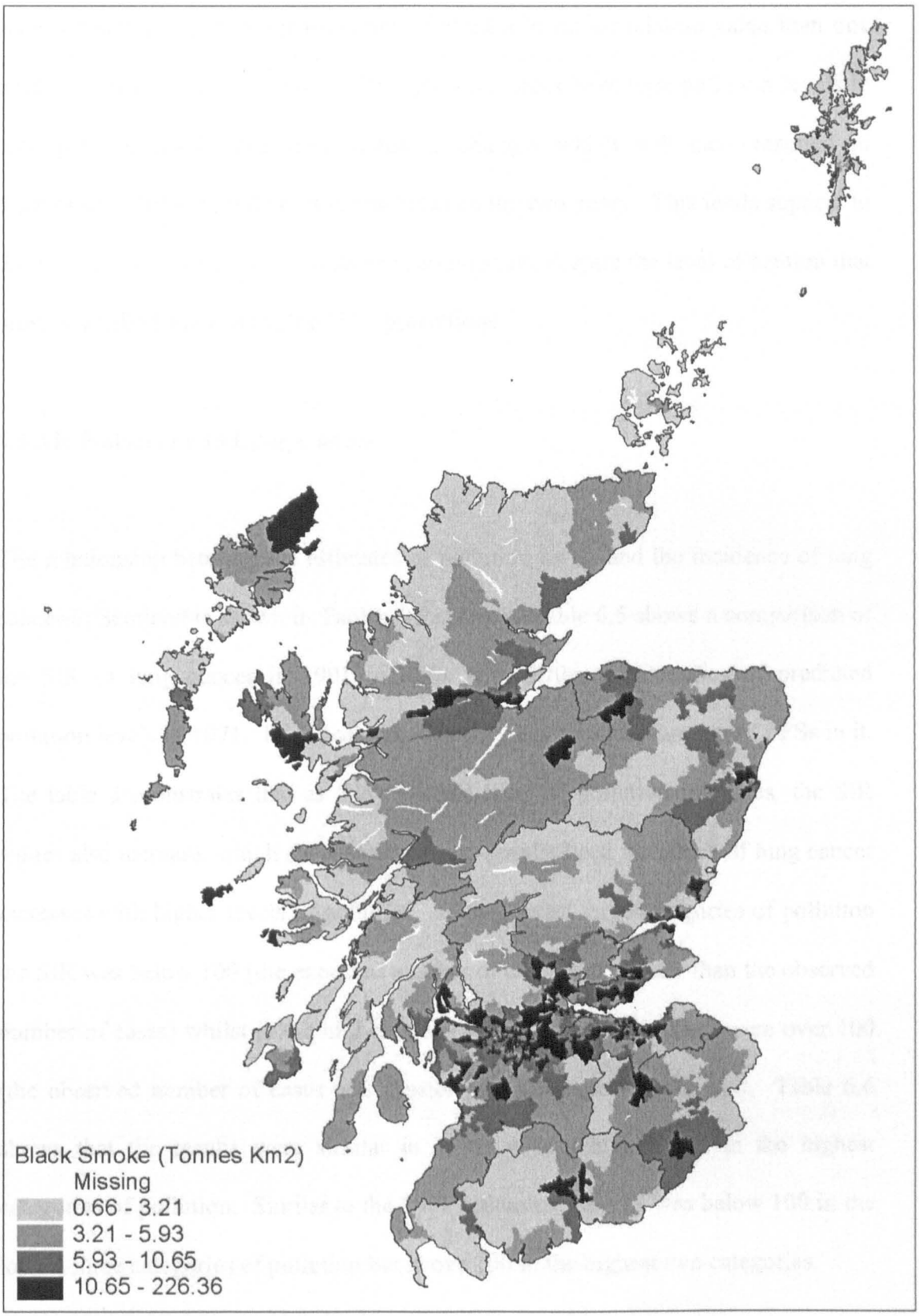


Figure 6.8. A Map of Estimates of Black Smoke Levels in Scotland in 1991.

similarity between the two sets of estimates. The industrial changes that have occurred between the two periods have caused a lower correlation value than one might otherwise have expected. Although some areas have high pollution levels in both periods many have seen industrial changes which will have resulted in dramatically different pollution levels between the two years. This lends support to the assertion that using both estimates is appropriate, despite the level of caution that must be applied when using the 1971 predictions.

6.6 Air Pollution and Lung Cancer

The relationship between the estimates of pollution levels and the incidence of lung cancer in Scotland is shown in Tables 6.5 and 6.6. Table 6.5 shows a comparison of the SIR of lung cancer in 1991 in PSSs with different categories of predicted pollution levels in 1971. Each pollution category has an equal number of PPSs in it. The table demonstrates that as the predicted level of pollution increases, the SIR values also increase, which shows that the age-standardised incidence of lung cancer increases with higher levels of pollution. In the lowest three categories of pollution the SIR was below 100 (the expected number of cases was greater than the observed number of cases) whilst in the highest two categories the SIR values were over 100 (the observed number of cases was greater than the expected number). Table 6.6 shows that the results were similar in 1991, with a higher SIR in the highest categories of pollution. Similar to the 1971 estimates, the SIR was below 100 in the lowest three categories of pollution but above 100 in the highest two categories.

Pollution Estimate 1971	Total Observed Cases	Total Expected Cases	SIR (per 100,000)
2.02 – 11.29	1832	2523.70	72.59
11.30 – 18.56	2649	3320.61	79.77
18.57 – 29.31	4483	4539.76	98.75
29.32 – 43.49	5050	4669.47	108.15
43.50 – 1607.64	4455	3580.33	124.43

Table 6.5. Relationship between pollution estimates in 1971 and lung cancer SIR.

Pollution Estimate 1991	Total Observed Cases	Total Expected Cases	SIR (per 100,000)
0.66 – 3.60	1303	1765.65	73.80
3.61 – 5.83	2506	3285.84	76.27
5.84 – 8.73	4151	4177.01	99.38
8.74 – 12.81	5467	4979.50	109.79
12.82 – 226.36	5042	4425.87	113.92

Table 6.6. Relationship between pollution estimates in 1991 and lung cancer SIR.

These results suggest that there is a positive relationship between lung cancer incidence and air pollution because the SIR for lung cancer tends to be greater in areas of high levels of pollution. However, it is important to realise that these results do not control for other risk factors associated with lung cancer. Therefore, it is possible that these observations may be confounded by other factors such as smoking rates, which were likely to be higher in urban areas where pollution levels were greater. In order to unravel these distinct risk factors it is necessary to incorporate all of the risk factors into a regression model of lung cancer incidence.

6.7 Conclusion

This chapter describes a method for predicting levels of pollution across Scotland in 1971 and 1991. The approach differs from those used by others in the past, because

the predictions use socio-economic data as explanatory variables rather than physical characteristics such as meteorology and topography. This is largely in response to the sparsity of pollution data in Scotland and the desire to provide estimates for small areas across the country.

The methodology that has been outlined here provides estimates of the levels of pollution in 1971 and 1991 for all 1991 PPSs in Scotland. The techniques use data from the 1991 census as the explanatory variables and 'black smoke' pollution data from NETCEN as the dependent variable. Models were constructed for both years that attempted to identify the key variables that explained pollution levels and this has highlighted some interesting parameters that were related to black smoke levels. The parameters from these models were then applied to all PPSs in Scotland to provide estimates throughout the country.

The resulting estimates were highest in urban Scotland, particularly Glasgow and Edinburgh, and lower in the more rural areas in both 1971 and 1991. However, the estimated pollution levels were higher in 1971 compared to 1991, reflecting the overall reduction in pollution emissions during the 20 year period. Although the general pattern was similar in both years, in 1991 the highest estimates were a little more dispersed than in 1971 and hence there were a handful of rural areas with high estimates. However, it is unlikely that levels of black smoke in many of these areas are actually very high and more likely that the high estimates in some rural areas are an artefact of the modelling process. Areas such as parts of the Western Isles, Skye and around Inverness had pollution estimates that were in the highest quintile. This discrepancy probably reflects the inclusion of the 'percentage employed in

construction' variable in the pollution model as the proportion of people employed in this industry is high in the rural areas that have high pollution estimates. Therefore, although the percentage employed in the construction industry is an important variable in explaining differences in air pollution levels, a discrepancy occurs in specific rural areas where there are small numbers of people who are employed in these areas.

In light of the unexpectedly high predictions of pollution in specific rural areas in Scotland, it would be useful to develop methods to validate the estimates. Given the absence of alternative datasets on pollution across the country that are geographically specific, validation would require the collection of field data on pollution levels using pollution censors. Measurements could be taken at urban and rural locations across Scotland and the mean annual measurement could be compared to the estimates calculated in this study. However, even this approach is limited as it will only provide contemporary measures of pollution rather than ones that coincide with the two sets of estimates (1971 and 1991).

The comparison between the air pollution estimates and the age-standardised lung cancer incidence has demonstrated that lung cancer rates were higher in areas that had greater levels of air pollution in 1971 and 1991. However, it is important to examine whether this observation is consistent when other key risk factors are controlled for such as smoking behaviour and radon levels.

This analysis raises some interesting methodological questions about the relative timing of explanatory variables in epidemiological studies. There were questions to

answer about whether researchers should attempt to incorporate the lag between exposure and initiation of disease but ignore the issue of migration, or instead use contemporary variables. It may have been possible to address some of these questions by using a longitudinal data set similar to the longitudinal study in England and Wales but it is not available in Scotland. However, even if the equivalent longitudinal data existed in Scotland it is unlikely that it would be available for small geographical areas. In this example of lung cancer, it would have been useful to use 1971 census variables to estimate 1971 pollution levels as this would have been likely to produce more accurate estimates. However, it was important to maintain the 1991 boundaries so that the estimates were compatible with other data sets in this study (especially the lung cancer data) and 1971 interaction data are not available for Scotland. The estimates will be used in this study as a control for air pollution but also to examine which of the pollution estimates (1971 or 1991) is the most appropriate to use in an examination of lung cancer.

7. PREDICTING EXPOSURE TO THE GEOLOGICAL CAUSES OF LUNG CANCER IN SCOTLAND

7.1 Introduction

Geological factors, both environmental and occupational, have long been considered to be a cause of lung cancer, with the most notable example being radon. Three potential geological influences can be identified: radon gas; coal mining; and quarrying / hard-rock mining activity. Although both radon and coal have previously been examined in some depth (e.g. Darby *et al.*, 1998), the influence of these factors is still disputed and neither have been examined in Scotland. The effect of expelling quartz into the atmosphere through quarrying activity has received very little attention despite there being evidence that quartz is a hazard to those working in particular industries such as in the sand extraction and sand blasting industries and among construction workers (Souter *et al.*, 2000). Thus, the health effects that quartz produced by the quarrying industry has on the wider population have not been considered before. Previous work has tended to focus only on those working in specific industries rather than on the local resident population. This section presents methods for estimating exposure to all three of these geological factors so that they can be used to examine how they influence lung cancer in Scotland.

The first part of this chapter discusses the physical properties of radon and outlines why it can be a potential health hazard. Then the sources of radon and the factors that account for the geographical variations in radon levels are outlined. Techniques for predicting exposure to radon levels in Scotland are then described and discussed,

and the accuracy of these estimates evaluated. In the second part of this chapter, methods are suggested that provide predictions of the levels of employment in the coal mining industry. Finally, methods to estimate the level of exposure to different types of rock quarrying such as granite and basalt are outlined. All of the estimates will be calculated for small areas in Scotland.

7.2 Chemical and Physical Properties of Radon

Naturally occurring radiation has been estimated to account for 85 percent of the effective dose equivalent to which the majority of the UK population is exposed (Hughes and O'Riordan, 1993). The main components are cosmic rays, terrestrial gamma rays, the long-lived radionuclides in the body from diet and the short-lived decay products of radon. Of these, naturally occurring radon provides by far the largest source and has long been identified as a potential risk factor in cancer incidence, especially in cases of lung cancer (Darby, 1999). Chapter 2 suggested that there is epidemiological evidence of a link between high levels of radon exposure and lung cancer incidence in specific areas. However, the effect of exposure to high levels of radon has received little attention in studies of lung cancer in a Scottish context.

Radon has three naturally occurring isotopes, radon-219, radon-220, and radon-222 although only the last two are of radiological consequence. Radon-222 is perhaps the more important because it has a longer half-life (3.824 days) and it is formed in the uranium-238 decay series (Adams and Cox, 1997). The uranium-238 decay series is outlined in Figure 7.1, which shows the transition from uranium to radium

to radon and subsequently to lead. It is environmentally important because, as part of this process, it emits alpha particles and decays to form daughters that are themselves radioactive. The decay products of radon are radioactive and they have a half-life of half an hour.

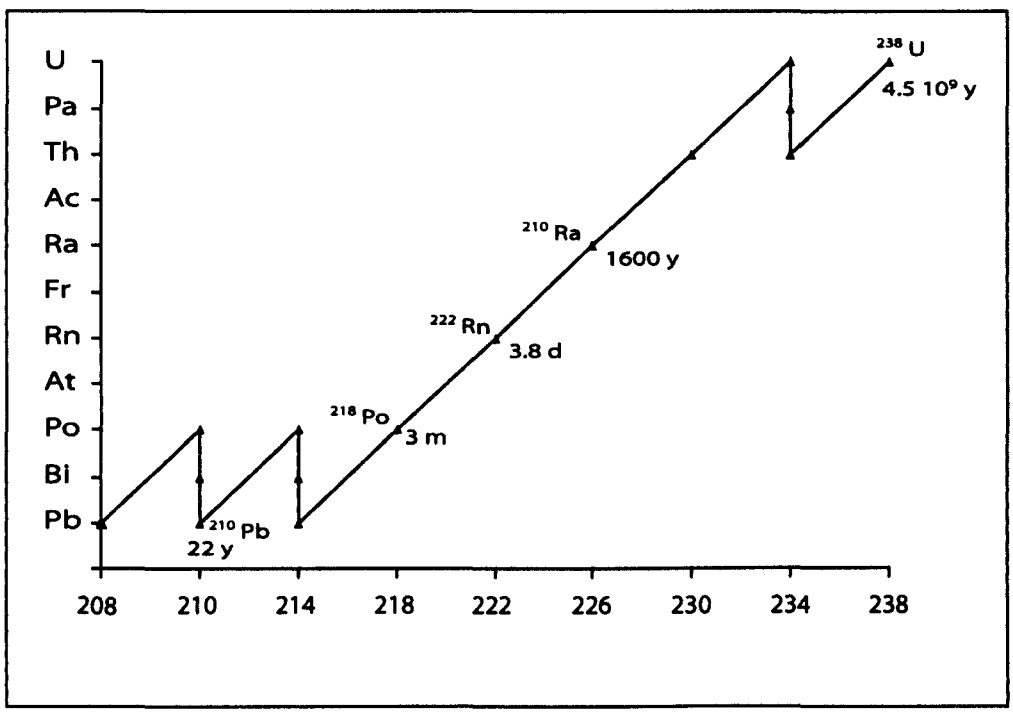


Figure 7.1. Decay scheme for ^{238}Uranium (NRPB, 2000).

The alpha particle doses have potentially carcinogenic properties which form small ion clusters with water molecules or react chemically with vapours present in the air which in turn become attached to aerosol particles and ultimately enter the body by inhalation. The attached particles can lodge in the lungs and airways but due to their short half-life there will usually be insufficient time for the natural clearance processes to have removed them before they have decayed and thus irradiated cells in the respiratory tract. Although it is possible for a single alpha particle to cause major genomic changes in a cell and cause irreparable damage, the probability of

such damage is higher with greater exposure. Therefore, in sufficient quantity, alpha particles that are produced from radon progeny are more likely to damage target lung cells and cause cancer. However, due to the 'single alpha particle theory' it is possible that there is also a relationship between lung cancer and radon exposure at very low levels (BEIR, 1998).

Radon levels are measured in Bq m^3. One Bq (or becquerel) corresponds, on average, to one nuclear disintegration per second. Therefore, the activity concentration (Bq m^3) is the activity in a cubic metre of air. It should be noted that in epidemiological terms this measurement may be misleading because the risk of developing lung cancer from radon depends on the total exposure over many years (NRPB, 2000). Radon concentrations can vary dramatically over a number of different time scales from hours to months and so a long-term average radon concentration is needed.

7.3 Natural and Human Sources of Radon

Radon is generated in rock, soil, and building materials and diffuses readily in the open air. However, it can attain relatively high levels, particularly in poorly ventilated rooms, because the insides of buildings have a slightly lower barometric pressure than the surrounding atmosphere so it has the effect of sucking soil gas through cracks or gaps in the floor (Green *et al.*, 1996). It has been estimated that in the UK the total average annual effective dose-equivalent from radon is about half of the radiation dose received from natural sources. However, this is misleading because of the great variation in domestic radon exposure that in some cases results

in annual doses that may be one to two orders of magnitude higher than the average (Adams and Cox, 1997).

As the decay series diagram in Figure 7.1 suggests, high levels of radon are found in local rocks and soils that contain uranium. Because granite tends to have a relatively large uranium content, radon levels tend to be very high. However, it is important to understand that the relationship is not straightforward because high radon levels are also found outside granite areas and, conversely, not all granite areas have high radon levels; rather the radon levels are a complex interaction of the bedrock geology and surface properties. If, for example, the underlying rock is overlaid with clay then radon is less likely to be able to escape than if it were covered by a more porous material. Furthermore, once radon gas is released from minerals, its migration to the surface is controlled by the transmission characteristics of the bedrock and soil, including porosity, permeability and moisture content, the nature of the carrier fluids including air, carbon dioxide and groundwater and meterological factors such as barometric pressure, wind, relative humidity and rainfall (Ball *et al.*, 1991). The complexity of these processes makes radon level prediction difficult and, in fact, accurate predictions are still not possible (NRPB, 2000). To further complicate this, the transport of radon from soil gas to room air is also complex and will depend on the details of construction (particularly on cracks and openings in the floor) and on the way it is occupied (temperature relative to the outdoors, and ventilation) (NRPB, 2000). Therefore, the level of radon not only varies between broad regions but can even vary between adjacent buildings.

Some of these complexities have been addressed in the construction of Figure 7.2 which shows the mean radon levels in houses for grid squares across the UK using data collected by the National Radiological Protection Board (NRPB). This shows that the greatest radon concentrations are in the granite areas of Devon and Cornwall but high levels can also be found in other areas such as the limestone of the Derbyshire Dales. Importantly, in the context of this study, the map suggests that there are also high radon levels in parts of Scotland, especially in the Grampian and Highland regions. It should be noted that this map is constructed from pixels at a five-kilometre resolution and that a large amount of local variation is possible within these. This reflects how difficult it is to model radon levels to provide predictions of the potential exposure to the gas. Developing accurate radon models is an important initial step to take before the effect that radon has on lung cancer can be explored.

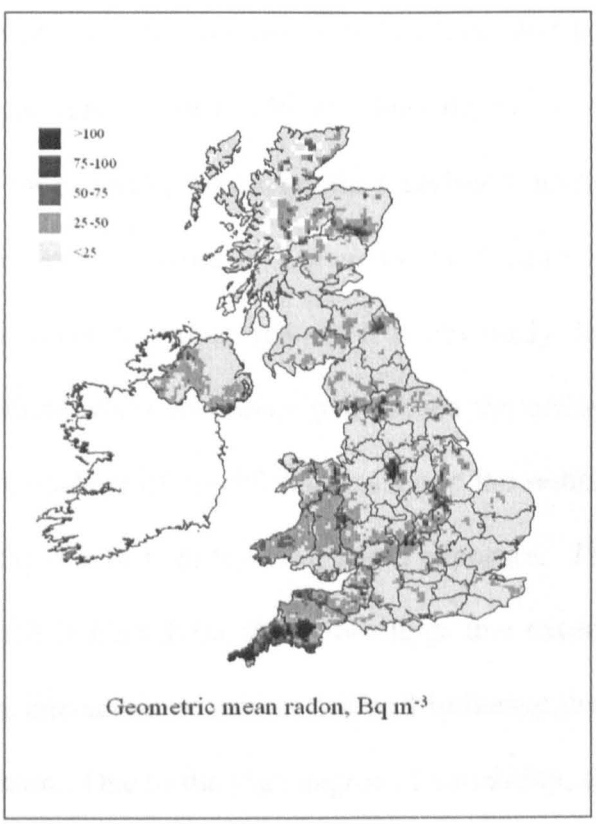


Figure 7.2. Radon map of the UK (NRPB, 2000)

7.4 Background to Modelling Radon Levels

Measurements of radon levels can vary widely between apparently identical buildings within the same geological unit and hence no map or model can predict the radon level of an individual building (Miles, 1998). However models can be used to calculate the geographical variations of the probability that a building will exceed a particular level. Such approaches have been used for a large number of different locations and for different purposes. Radon estimates are usually made by government agencies to identify areas that exceed a predefined acceptable level or by studies that wish to estimate radon levels in order to study the effects of these levels. This section reviews the various approaches that have been used to estimate radon levels in order to identify sensible methods for this particular project.

The maximum acceptable levels for radon in dwellings that have been adopted by various governments vary between 150 and 800 Bq m^{-3} (Colgan and Gutiérrez, 1996). In the UK, for example, the government advises householders that where the radon concentration exceeds what is known as the 'action level' (200 Bq m^{-3}), measures should be taken to reduce radon. It is obviously desirable to be able to identify those locations where dwellings will exceed the action level. Clearly, the accuracy of these models is of crucial importance to the policies that governments adopt to alleviate the risk of high levels of radon exposure. The difficulty with the action level approach is identifying those dwellings that exceed this defined level, due to the complex interaction of factors that will influence the reading, as outlined in the previous section. Due to the high degree of variability, the modelling process can only predict the potential and not absolute measures of radon and this will

fluctuate greatly with geographical scale. Nonetheless, radon modelling remains a key component in the understanding of radon fluctuations because it provides a relatively low cost means of delimiting areas of high radon potential.

Data on radon levels can be collected in five ways: from indoor radon measurements; radiometric and geochemical data; rock and soil permeability; the moisture content of rocks and soils; and soil gas data. The most accurate of these for mapping radon potential is indoor radon data because this is the only definitive method of estimating the hazard from radon in a particular building (Miles and Ball, 1996). In the UK, measurements are taken using passive integrating detectors over a period of three months (Miles, 1998), whereas in the US shorter term screening measurements taken over a period of up to a week are common (Cohen *et al.*, 1994). Uranium and radium concentrations in surface soils and rocks provide an alternative indicator of radon emissions and these can be measured using gamma spectrometry (Ball *et al.*, 1992). Although the correlation with radon measured in houses is high, this relationship is not always consistent because some areas with relatively high indoor radon levels are not associated with high concentrations of uranium in soil and rock (Talbot *et al.*, 1998).

Radon levels can also be measured by considering the permeability and moisture content of rocks and soils. Radon diffuses further in air than in water and so high radon readings are likely to result from more permeable and less saturated rocks for a given concentration of uranium and radium. However, this method is only useful in explaining localised variations in radon levels and should be used in tandem with a direct measurement of radon because the permeability and moisture content is only

one component of radon potential (Ball *et al.*, 1995). The final method of collecting data is by measuring the concentration of radon in soil gas due to the close correlation between bedrock geology and radon in overlying soils (Gates and Gunderson, 1992). However, the measurements can be difficult to interpret due to the effects of large diurnal and seasonal variations in soil gas radon close to the ground surface (Washington and Rose, 1992), causing significant variations between areas only a few metres apart (Oliver and Badr, 1995). Therefore, high quality indoor radon measurements are the most suitable data for constructing radon potential maps and all other measurements can be considered proxy indicators that should only be used in the absence of house measurements (Appleton and Ball, 1995). However, taking a sufficient quantity of radon measurements is a time consuming and often a prohibitively expensive exercise and consequently other sources of radon data are often used instead.

Two approaches to mapping radon prone areas using radon measurements can be identified. The first approach uses radon measurements in existing dwellings to map the variation of radon potential between administrative or postal districts, or grid squares. Cohen *et al.* (1994), for example, examined radon levels at the unit of the US County. This study used the average of a minimum of ten measurements in each county from three independent surveys. Marcinowski *et al.* (1994) reported on the National Residential Radon Survey that was conducted in the US in regions defined by the US Environmental Protection Agency and in subgroups of the housing stock. The aim of this survey was to develop a more rigorous estimate of the distribution of radon concentrations in the US housing stock and to assess regional differences in radon levels. The Radiological Protection Institute of Ireland determined the

percentage of dwellings above the reference level in ten kilometre grid squares of the Irish National Grid System using radon measurements in selected dwellings with a minimum sample size of five dwellings per grid square (RPII, 1996). Wrixon *et al.* (1988) carried out a program of house radon surveys in Great Britain which indicated high levels of radon in Cornwall and Devon together with above average levels in parts of Somerset, Derbyshire, Grampian and Highlands. Similarly, Miles *et al.* (1993) delimited potential radon affected areas in parts of the Grampian and Highland regions of Scotland by calculating the number of homes exceeding the action level in five kilometre grid squares. These approaches are often criticised because the units used are unlikely to reflect natural variations in radon levels due to the incorporation of administrative boundaries or grid squares rather than natural boundaries that delimit differences in radon levels (Appleton and Ball, in press). It is therefore usually more sensible to model radon within geological units that are more likely to represent variations in radon levels.

A second approach assesses the geological and pedological factors that influence the emission of radon at the surface (Appleton and Ball, 1995). This is based upon the interpretation of the concentration of mineralogical occurrence and chemical state of uranium and its radioactive daughter product radium in the ground; rock permeability; soil permeability; concentration of radon in soil gas; and radon concentrations in dwellings. Models have been developed using geological, radiometric and indoor radon data, the first of which were developed in Sweden (Åkerblom, 1987). This study used data from airborne radiometric surveys (covering 65 percent of Sweden) and ground radiometric surveys together with geological data, results from radon surveys in buildings, results from earlier

geotechnical investigations, field surveys and orientation soil gas radon measurements. The information that was gained from these sources was then used to establish simple maps that classified areas into three levels of risk, at a 1:50,000 scale. This approach has been adopted in other countries such as the Czech Republic (Barnet, 1991), Finland (Slunga, 1988; Markkanen and Arvela, 1992), the USA (e.g. Ettlinger, *et al.* 1987; Kline and Mose, 1990) and in Canada (Cocksedge *et al.*, 1993; Darnley *et al.*, 1986). For example, the United States Geological Survey (USGS) classified radon potential on the basis of a scoring system which considers five factors in a Radon Index Matrix: indoor radon; radioactivity; geology; soil permeability; and house architecture. The scoring system was used in combination with a Confidence Index Matrix to assess the overall radon potential of a specific geological unit (Gunderson *et al.*, 1992). This approach has been used to provide a detailed evaluation of the radon potentials for each state.

Other approaches to radon modelling have used soil gas radon, permeability and geology. Ball *et al.* (1992) demonstrated that the soil gas radon map can be used to delineate the variation in radon potential between contrasting geological units. They showed that in many cases the lithological control over the production of radon is high and hence provide the most efficient and least expensive survey method for radon distribution mapping. Ball and Miles (1993) suggested that, in the absence of radon measurements in the home, soil gas radon surveys are the most efficient method for identifying zones of high radon potential.

In the UK, the British Geological Survey (BGS) have used the data on indoor radon measurements and geology type to develop lognormal models to produce estimates

of the percentage of the housing stock above the UK Action Level for each geological unit within a map sheet, grid square or administrative district (Miles, 1998). Miles and Ball (1996) suggested that by using indoor radon measurements and grouping them by geological unit in this way, there is the capacity to measure radon potential at the spatial level of the geological map data. This has been developed further for the 1:625,000 and 1:250,000 radon potential maps of Great Britain (Appleton and Ball, 1995; BGS, 1998) where the radon potential is based not only on the solid geology but upon the drift units. Therefore, a geological unit can be assigned different radon potential classes in different parts of Great Britain depending on variations in the lithology (mainly permeability) and also the geochemistry within mapped chromostratigraphic units.

Despite these developments, there are a number of other factors that need to be incorporated before highly accurate radon potential models can be developed. This was recognised by Appleton and Ball (in press) who pointed out that the models are limited by the data that they use. For example, in the geological map, local features may not be represented and the boundaries can only be generalisations. Higher or lower radon potential may occur within classified areas due to the occurrence of units of contrasting lithology and permeability or unmapped shear and fracture zones. The models cannot give detailed information on a specific area because of the complex interaction of variables that determine radon potential, some of which are unique to the individual dwelling.

This section has identified two approaches to modelling radon levels that have been developed for different purposes and under different data limitations. The first

approach maps radon levels for administrative units or for grid squares and has the advantage of being cheap and relatively easy to provide comprehensive estimates for large areas. However, it provides insufficient resolution for a number of purposes and does not account for the fundamental controls of radon release (i.e. geology, ground permeability etc.). The second approach maps radon potential for geological units and, although more expensive, has the advantage of reflecting natural variations in radon levels rather than administrative boundaries. This allows accurate radon estimates to be made even when a limited number of measurements are available. More importantly it provides the most accurate estimate of radon potential for an area (Appleton and Ball, forthcoming). The development of accurate models of radon levels or radon potential is clearly an important component of identifying areas that require remedial action. However, the radon models may also be utilised in an epidemiological framework to aid predictions of the exposure of a population to radon in studies that have attempted to examine the carcinogenic properties of radon.

7.5 Radon Data in Scotland

7.5.1 Introduction

The policy on radon data collection in Scotland is not well established whereas England and Wales has a more thorough approach for monitoring radon levels which is coordinated by the NRPB. This is reflected in the disparity between the quantity of measurements on radon levels taken in Scotland compared to the number taken in England. In England and Wales there have been a total of 385,000 measurements

taken whilst in Scotland there have been only 6,600 (NRPB, 2000). This equates to 7.26 measurements per 1,000 people in England and Wales but only 1.32 measurements per 1,000 people in Scotland. This section reviews radon measurement data that have been collected in Scotland by the NRPB and the Leukaemia Research Fund.

7.5.2 National Radiological Protection Board Data

The most recent and comprehensive survey that has taken place in Scotland was carried out by the NRPB on behalf on the Scottish Office (Green *et al.*, 1996). In addition to the results of this survey, the final dataset contained all available data on radon in Scottish homes from work done by the NRPB including surveys previously undertaken on behalf of other organisations and individual households (Green *et al.*, 1996). The intention was firstly to encourage householders in the areas of greatest risk to measure levels in their homes and secondly to complete a radon map of Scotland. The houses that fell in areas that had previously been delimited as being radon affected in Aberdeenshire and the Highlands were identified from the Postcode Address File and offered a free radon measurement. Of the 2,598 households that fell within these areas, 335 had already had their radon levels measured and of the remaining 2,263, 32 percent chose to be included in the study. These measurements were complemented by private householders, private landlords and local authorities who independently requested a radon measurement. In addition, measurements were taken in other areas to complete a radon map of Scotland to identify any additional radon affected areas. This required a minimum of four radon measurements in each ten kilometre square of the Ordnance Survey

national grid containing a significant population. There are a total of 1,082 ten kilometre grid squares covering the whole of Scotland, 230 of which contain less than 16 dwellings.

Housing stock	1,900,000
Average radon level	16 Bq m^{-3}
Number of dwellings measured	5,726
Number found above Action Level	223
Highest level found	2,200 Bq m^{-3}

Table 7.1. Summary of domestic radon data for Scotland (adapted from Green *et al.*, 1996)

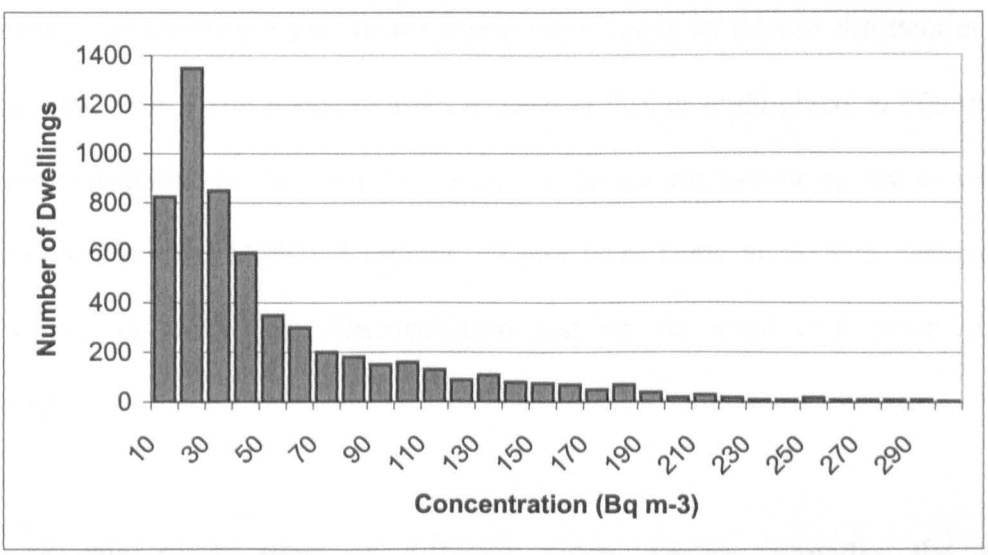


Figure 7.3. Distribution of indoor radon concentrations in Scotland (adapted from Green *et al.*, 1996)

A total of 5,726 households were surveyed by the NRPB and the results are summarised in Table 7.1 and Figure 7.3. Table 7.1 shows that the population-weighted average radon reading among the houses surveyed was 16 Bq m^{-3} and this is lower than the average in England and Wales which has been calculated at 21 Bq

m^{-3} (Wrixon *et al.*, 1988). Of these, 223 houses were above the action level of 200 Bq m^{-3}. The NRPB estimated that approximately 2,000 of the two million dwellings in Scotland would exceed the action level (Kendall *et al.*, 1994) and there are some clear regional variations in these levels that may be important to recognise in a study of lung cancer incidence. This is reflected in Figure 7.3, which shows that the vast majority of households had very low radon levels but that there was a significant number of households that exceeded the recommended limit. Importantly, there were clear regional variations within Scotland and this is demonstrated in Figure 7.4, which shows the radon concentrations in local postcode districts across Scotland. Low mean concentrations were found in southern and western Scotland and relatively high values in Aberdeenshire and the Highlands. There have not been sufficient measurements taken in the central belt areas to release the data at this geographical level. The geographical variation is further highlighted in Figure 7.5, which demonstrates the variation in average radon concentrations within two areas of the Grampian and Highland regions. There were some areas with particularly high values in south-west Aberdeenshire and on the north-east coast of the Highlands.

Although this combination of different radon surveys provides the most comprehensive set of readings that have been carried out in Scotland, it is important to note that the various surveys had different objectives. Some of the individual surveys were intended to be representative whilst others were targeting dwellings with potentially high radon levels. Therefore, the aggregated dataset should not be

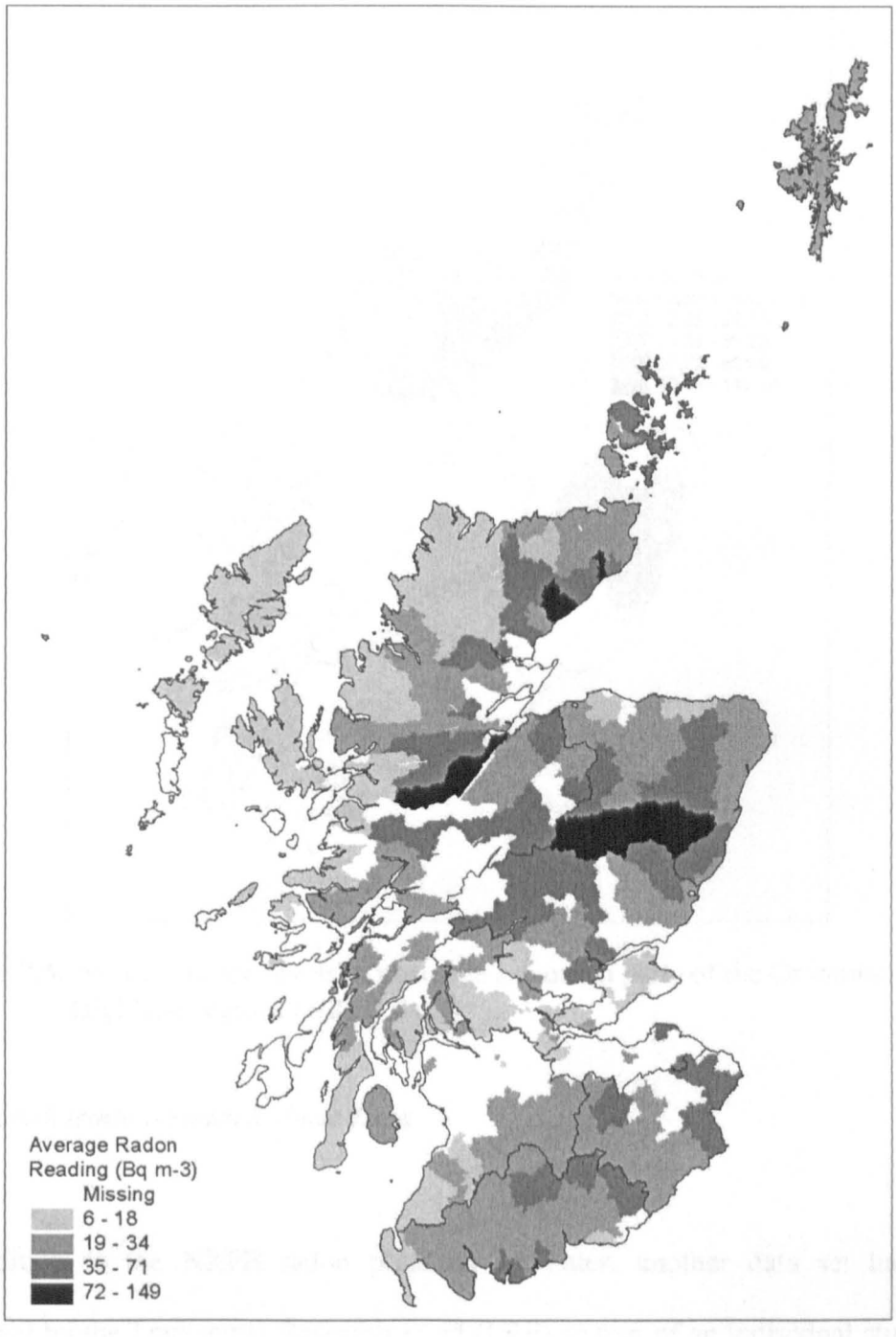


Figure 7.4. Average radon reading in postcode districts in Scotland (NRPB data)

considered representative of the overall radon levels in Scotland. Nonetheless, the dataset could be utilised to calculate radon estimates throughout Scotland if individual records were made available. However, the NRPB were unwilling to release the data to other organisations.

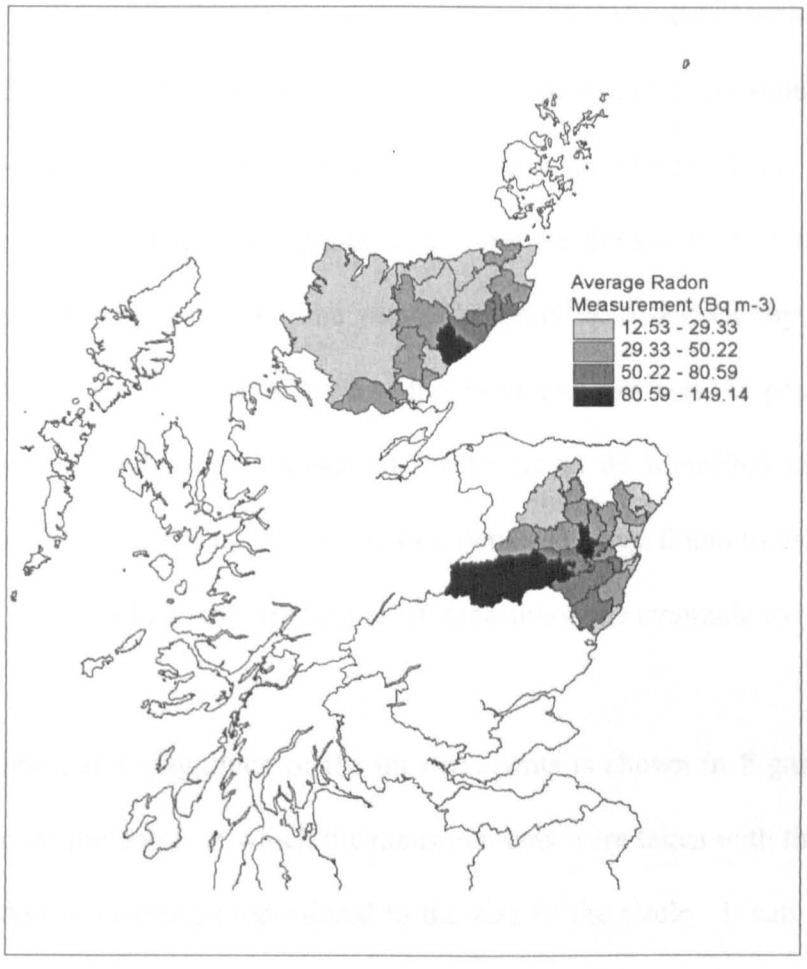


Figure 7.5. Average radon levels in postcode sectors in parts of the Grampian and Highland regions (NRPB data)

7.5.3 Leukaemia Research Fund Data

In addition to the NRPB radon potential estimates, another data set has been collected by the Leukaemia Research Fund (LRF) as part of an individual study that considered the effect of radon levels on the incidence of childhood cancer, especially leukaemia (UK Childhood Cancer Study Investigators, 2000). The data consisted of 1,329 seasonally adjusted measurements in houses of children living in Scotland. The children had either been diagnosed with cancer or been chosen as part of a control group that had been matched for sex, age and geographical area. For each child, a residential history was collected and all UK addressees lived in by the child

from birth to diagnosis were targeted for measurement of radon levels. Detectors were sent to each house with instructions to place one in the living room and one in the main bedroom. The detectors were collected six months later. The radon measurements were seasonally adjusted to account for the natural variation in radon emission that takes place through the year (Wrixon *et al.*, 1988; Gunby *et al.*, 1993; Pinnel *et al.*, 1995). Attached to each of the measurements was the postcode of the house, which enabled the coordinate of the house to be identified using a 1999 postcode lookup table. Of the 1,329 measurements, 11 were found to be incorrect or not present in the lookup table leaving 1,318 measurements available for analysis.

The distribution and magnitude of the measurements is shown in Figure 7.6 which shows a map of the points at which the measurements were taken with the seasonally adjusted radon level being proportional to the size of the circle. It can be seen that the measurements were distributed throughout Scotland with a greater number in the urban areas where most people live. There were high measurements in areas that were identified as having relatively high levels of radon by the NRPB, particularly in the eastern Grampian region and in the Highland area around Inverness. However, there were also some high measurements in central and southern Fife, around Dunoon, in the city of Edinburgh and the central Border areas.

There is no doubt that Scotland has been underrepresented in the UK's collection of radon measurements. This makes it more difficult to precisely estimate radon levels throughout Scotland. Due to this lack of data, it is necessary to not only estimate radon levels using information on radon itself but also to incorporate other factors

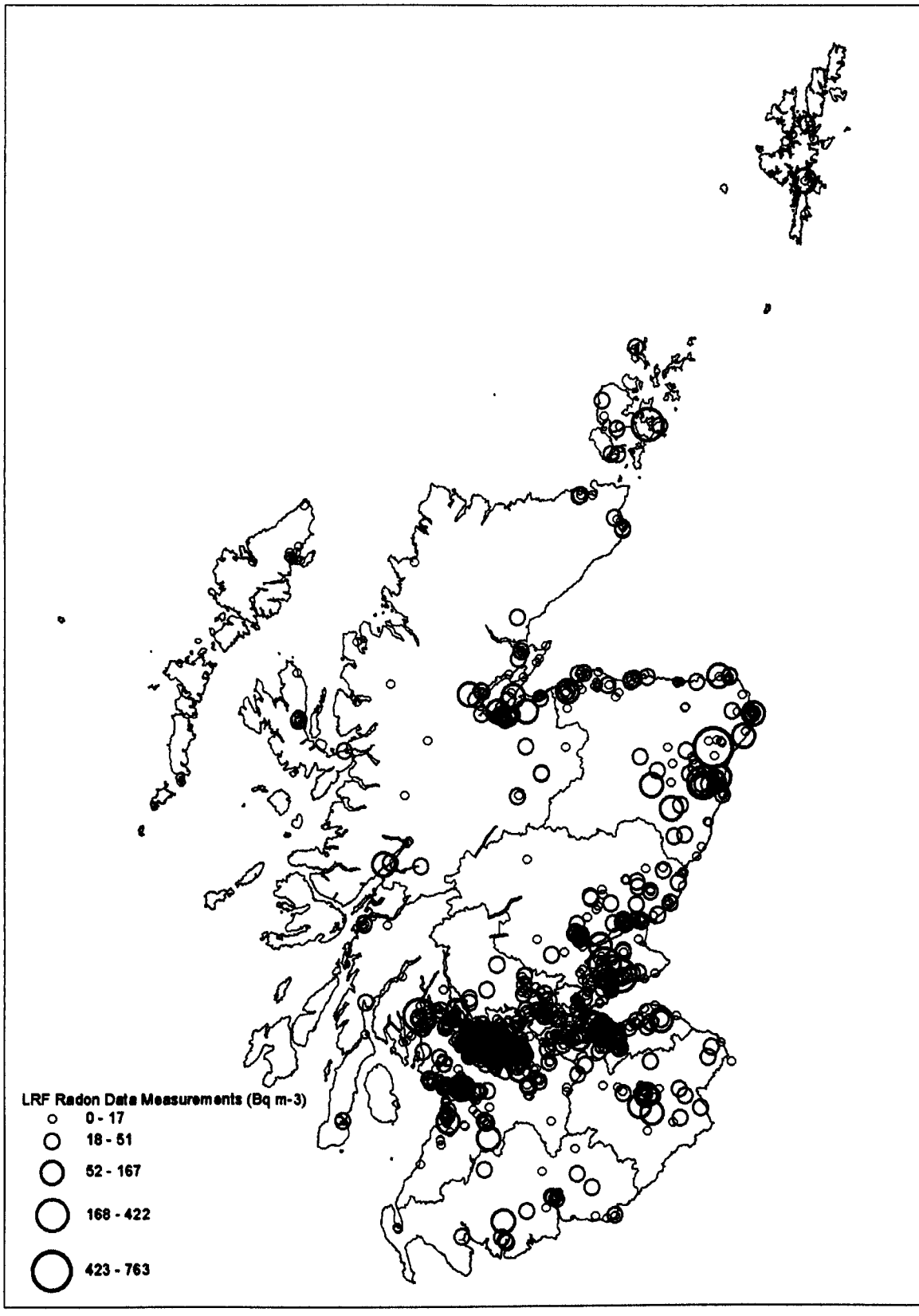


Figure 7.6. Distribution and magnitude of radon data collected by the LRF

that influence radon levels, especially bedrock geology and drift deposits. These estimates will reflect the natural variation in radon levels.

7.6 Methods for Predicting Radon Levels

This section describes methods for predicting radon potential across Scotland using the radon measurements that have been collected by the NRPB. The methodology reflects the influence of both bedrock geology and drift deposits by using geological boundaries to delineate differences in radon potential and calibrating the values to reflect the influence of the drift deposits. This is more reliable than using administrative boundaries that are unlikely to accurately reflect natural variations in radon emissions (Miles, 1998). The radon estimates are validated using the independent measurements that were collected by the Leukaemia Research Fund. Because the radon levels are likely to be largely dependent upon the combination of solid geology and drift, the levels are likely to remain relatively stable during the latency period of lung cancer. Therefore, in the absence of historical information on the migration history of the lung cancer patients, contemporary measures of radon potential are suitable for this analysis.

7.6.1 Radon Potential Estimates

Due to the general paucity of radon measurements that have been taken in Scotland and the inadequate geography for which these measurements are available, the most suitable option for calculating radon estimates was to utilise the radon potential estimates calculated by the British Geological Survey derived from data collected

through collaborative work by the NRPB and BGS (Appleton and Ball, 1995). The results of radon measurements in houses have generally been shown to be distributed lognormally for both large and small areas (e.g. White *et al.*, 1992; Nero *et al.*, 1994; Gundersen and Schumann, 1996) and Gunby *et al.* (1993) showed that data from a national survey in the UK fitted this distribution well. The properties of this distribution allow for the proportion of houses above a threshold level to be estimated even when there are insufficient data to calculate the proportion directly (Miles, 1998). A radon potential classification has been devised to produce a 1:625,000 radon potential map of Great Britain which classifies different groups of rocks and unconsolidated deposits into radon potential classes (Appleton and Ball, 1995). This is based upon the interpretation of one or more of the following data: radon concentrations in houses; concentration, mineralogical occurrence and chemical state of uranium and radium in the ground (radiometric and geochemical data); rock and soil permeability and moisture content; concentration of radon in soil gas; and building architecture (construction characteristics). The classification interprets radon measurements that were available to March 1994.

Log-normal modelling was used to estimate the percentage of dwellings exceeding the action level (200 Bq m^{-3}) for each geological unit. Where data conform to the lognormal distribution, the proportion of the distribution above a threshold can be estimated from the geometric mean and geometric standard deviation. This involved subtracting the average outdoor radon concentration from the measured indoor values, setting negative values to a small positive value, taking natural logarithms and calculating the mean and standard deviation (Miles, 1998). The reliability of the classification is relatively high in areas where a large number of soil gas and house

radon measurements have been made (e.g. Derbyshire, Northamptonshire and Somerset) (Appleton and Ball, 1995). The emission of radon will be modified if the underlying solid rocks are overlain by unconsolidated deposits. The levels of radon are dependent upon the thickness, composition, source, permeability and wetness of the unconsolidated deposits. This is summarised in Table 7.2 which shows that waterlogged or impermeable deposits such as peat and lacustrine clay tend to reduce the radon potential class to low, whereas permeable sand and gravel tend to enhance the radon potential in comparison.

Type and characteristics of unconsolidated deposits	Impact of unconsolidated deposits on radon potential class indicated by underlying rocks
Generally impermeable: peat and lacustrine clays	Reduces radon potential to low
Variable permeability: alluvium, boulder clay and morainic drift	Reduces, or has no significant effect, on radon potential, depending upon thickness, composition and derivation of unconsolidated deposit (no effect if radon potential is low)
Generally permeable: blown sand, river terrace, raised beach and marine, glacial sand and gravel	Increases or has no significant effect, on radon potential (normally no effect if radon potential is high or very high)

Table 7.2. Effect of unconsolidated deposits on radon potential (adapted from BGS, 1995).

The radon potential classification was modified to reflect the distribution of drift deposits. Because the radon potential map used the geological boundaries as its basis, the radon potential boundaries mirrored the boundaries in the 1:625,000 solid geology map produced by the BGS. Therefore, the solid geology was used as the boundary file and the radon potential estimate was manually associated with every polygon using the radon potential map sheet to classify each polygon into an appropriate class. The six categories are described in Table 7.3. Each category

Radon Potential Class	Description	Estimated % of dwellings exceeding Action Level (>200 Bq m^{-3})
Very High	Ground susceptible to very high levels of radon emissions	>10
High	Ground susceptible to high levels of radon emissions	3-10
Moderate	Ground susceptible to moderate levels of radon emissions	1-3
Low-moderate	Ground susceptible to low levels of radon emissions but with small sub-areas susceptible to moderate or high levels of radon emissions	<1
Low	Ground susceptible to low levels of radon emissions	<1
Unclassified	Insufficient information available to estimate radon potential	

Table 7.3. Classification related to underlying rocks (adapted from BGS, 1995)

corresponds to a different percentage of the dwellings that exceed the action level (200 Bq m^{-3}) and are classified as unclassified, low, low-moderate, moderate, high and very high. The distribution of radon potential values in Scotland is shown in Figure 7.7. The highest radon potential classes were in the Grampian and Highland regions, which is in line with the NRPB postcode district maps in Figure 7.4. Low radon potential levels were found in the Central Belt and most of the borders region. The map reflects the limitations of the data, as localised or anomalous features may not be represented and the boundaries shown are approximate (Appleton and Ball, 1995). This is reflected by the large area of Scotland that was unclassified due to a lack of radon measurements. This generally occurred in areas of low population density especially in the north-west Highlands, the Isle of Lewis and parts of the borders area. Due to the limitations of the methodology it is sensible to evaluate the

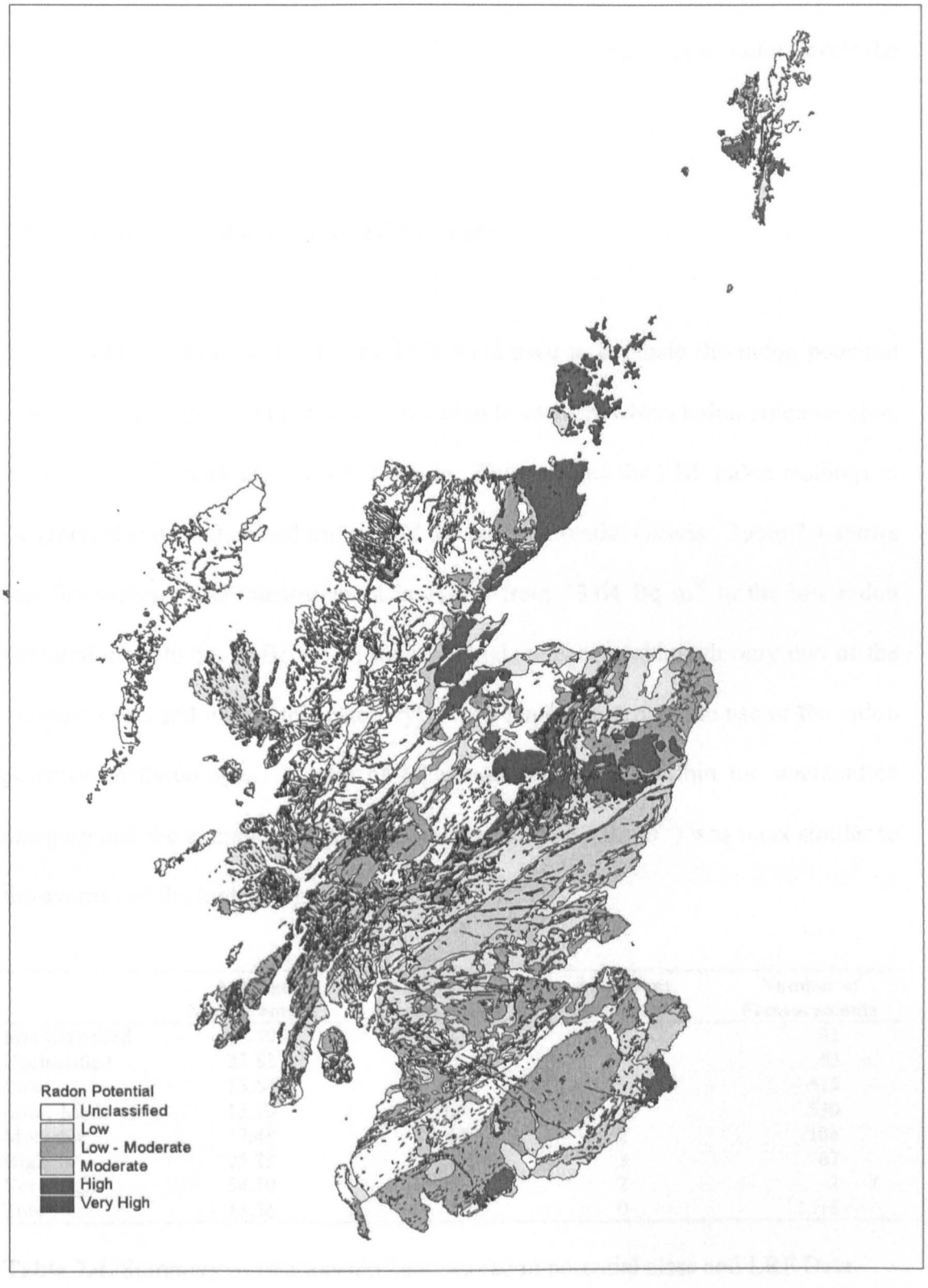


Figure 7.7. Radon potential classification in Scotland.

accuracy of the radon potential estimates. This was achieved using an independent data set of radon measurements to consider how good a predictor of radon levels the radon potential estimates are.

7.6.2 Evaluation of Radon Potential Estimates

The measurements collected by the LRF were used to evaluate the radon potential map. A point-in-polygon procedure was used to identify which radon potential class each of the LRF measurements fell within. This enabled the LRF radon readings to be grouped and summarised for each of the radon potential classes. Table 7.4 shows that the average LRF measurement increases from 13.64 Bq m^{-3} in the low radon potential class to 54.50 Bq m^{-3} in the very high category (although only two of the measurements fell within this category) which lends support to the use of the radon potential methodology. A total of 65 measurements fell within the unclassified category and the average of these measurements (23.82 Bq m^{-3}) was most similar to the average of the high category (25.25 Bq m^{-3}).

	Average Measurement	Maximum Measurement	Minimum Measurement	Number of Measurements
Mis-Classified	16.30	63	2	31
Unclassified	23.82	422	4	65
Low	13.64	234	1	513
Low / Moderate	15.79	763	0	530
Moderate	17.48	68	2	108
High	25.25	329	3	67
Very High	54.50	107	2	2
Total	16.26	763	0	1318

Table 7.4. Summary of relationship between radon potential class and LRF Data

A total of 31 of the measurements were misclassified which meant that the point did not fall within any of the polygons. They had an average radon reading of 16.30 which is similar to the average readings in the low-moderate and moderate

categories. Figure 7.8 demonstrates that this was an edge effect as a result of the greater degree of boundary generalisation in the BGS map compared to the boundaries from which the location of the LRF radon measurements were derived.

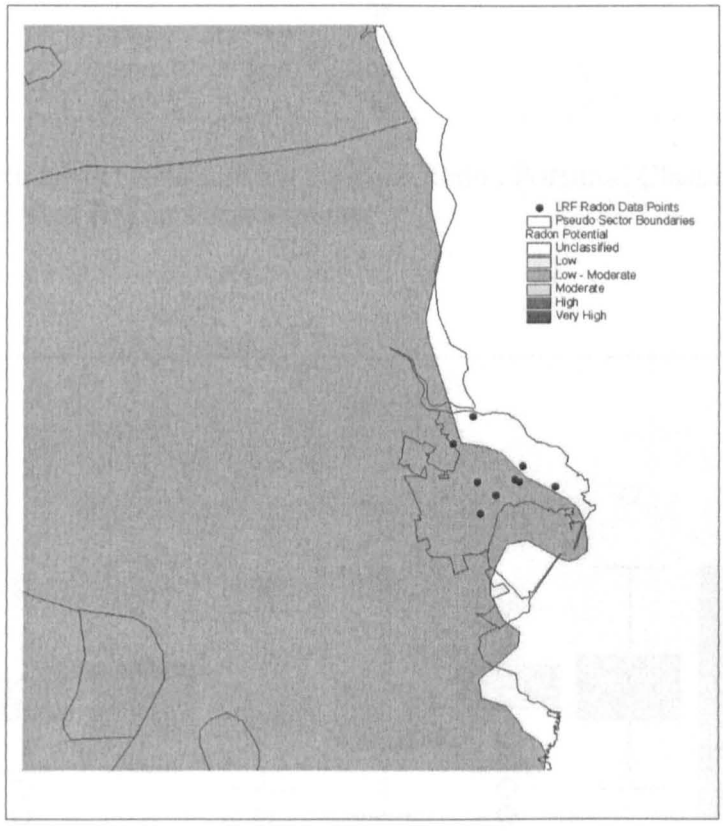


Figure 7.8. Edge effect between radon potential map and LRF data points

This resulted in some points falling just off the coast or in an unclassified feature such as a loch. Therefore, each of the points that did not fall within a radon potential class were allocated to the nearest polygon. The results from this are demonstrated in Table 7.5 which shows that each of the measurements that were previously misclassified has been associated with one of the other six categories. The average radon measurement still increased with a greater radon potential and this is demonstrated in Figure 7.9, which shows a boxplot of the radon potential class against the log of the seasonally adjusted radon measurements from the LRF dataset. The boxplot demonstrates a similar trend to Tables 7.4 and 7.5 with a generally

Radon Potential	Average Measurement	Maximum Measurement	Minimum Measurement	Number of Measurements
Mis-Classified	0.00	0	0	0
Unclassified	23.20	422	3	69
Low	13.70	234	1	526
Low / Moderate	15.87	763	0	537
Moderate	17.38	68	2	110
High	24.65	329	2	72
Very High	54.50	107	2	2
Total	16.05	763	0	1318

Table 7.5. Summary of Relationship between Radon Potential Class and LRF Data - Updated Radon Measurements

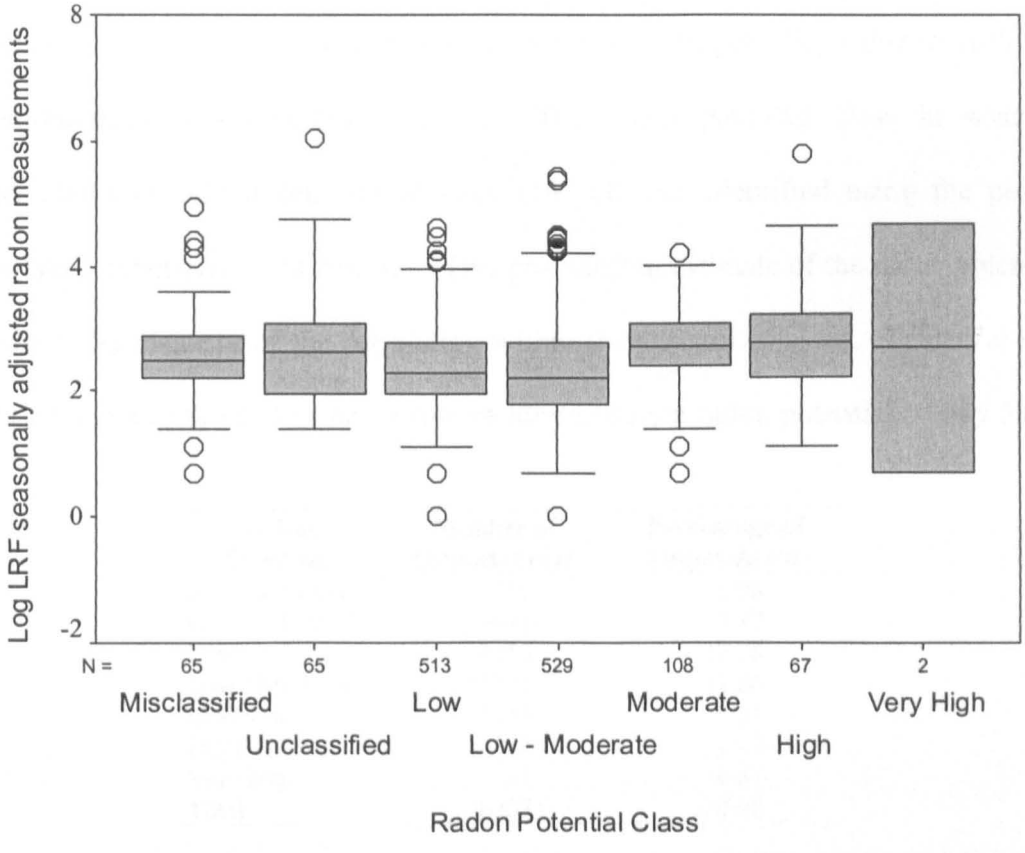


Figure 7.9. Boxplot of LRF radon measurements and radon potential class

increasing median radon value from the low radon potential class to the high. The exception was the low-moderate class which had a slightly lower median value than the low class. Similarly, the median of the very high class was slightly lower than the median of the high class, but the latter was based on only two measurements so cannot be interpreted with confidence. Most importantly, the boxplot shows the

overlap between the different radon potential classes. This demonstrates the large amount of variability that exists within each band and that radon levels are not exclusive to one particular class, reflecting the complex interaction of factors that determine radon levels.

7.6.3 Classification of Output-Areas into Radon Potential Classes

Having developed a method for estimating radon potential and techniques to justify this methodology, the next step was to assign a radon potential value to each of the output-areas (OAs) across Scotland. The radon potential class in which the population-weighted centroid of each OA fell was identified using the point-in-polygon functionality in ArcInfo. This provided an estimate of the radon potential to which the majority of the population within the OA are exposed. Table 7.6 shows that the majority of OAs had a low or low/moderate radon potential. Only 31 OAs

Radon Potential	Number of Output-Areas	Percentage of Ouput-Areas
Mis-Classified	759	1.98
Unclassified	1442	3.77
Low	14453	37.78
Low / Moderate	15783	41.26
Moderate	3713	9.71
High	2072	5.42
Very High	31	0.81
Total	38253	100.00

Table 7.6. Summary of relationship between radon potential class and OA lung cancer data

had a very high radon potential and 759 OAs had been misclassified (i.e. did not fall within any of the polygons or fall within a loch). This is demonstrated in Figure 7.10 which shows that a number of the population-weighted centroids in the centre of Aberdeen did not fall within any of the polygons that represent radon potential due to

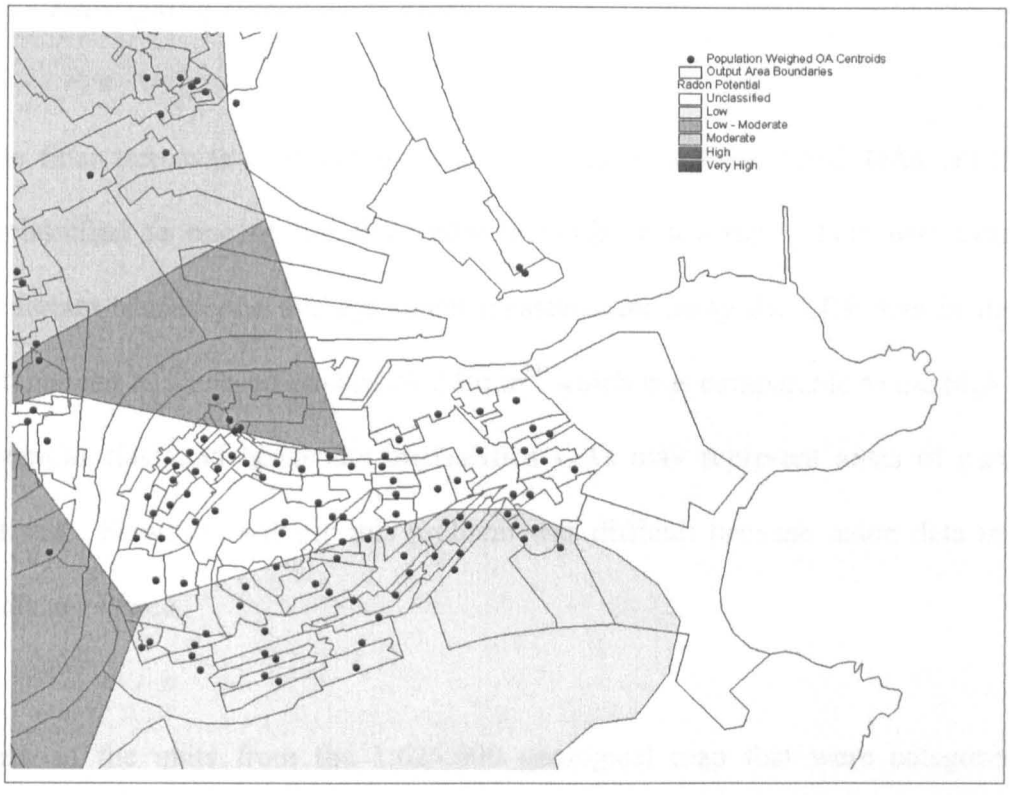


Figure 7.10. Edge effect between radon potential map and OA boundaries

the greater degree of generalisation of the radon potential boundaries. Therefore, each of these population weighted centroids was allocated to its nearest radon potential polygon. Table 7.7 shows how the 759 misclassified OAs were reallocated among the six radon potential classes.

Radon Potential	Number of Output-Areas	Percentage of Ouput-Areas
Mis-Classified	0	0.00
Unclassified	1562	4.08
Low	14723	38.49
Low / Moderate	16009	41.85
Moderate	3762	9.83
High	2166	5.66
Very High	31	0.81
Total	38253	100.00

Table 7.7. Summary of relationship between radon potential class and OA lung cancer data - updated radon measurements

7.6.4 Reassigning Unclassified Areas

The final step was to develop methods for assigning the 1,562 OAs labelled as unclassified to one of the five radon potential categories. This was especially important because the average radon measurement using the LRF data in the area categorised as unclassified was 23.2 Bq m^{-3} which was comparable to the high radon potential class and hence the unclassified OAs may represent areas of particular interest. However, solving this problem was difficult because radon data in these areas are scarce.

Each of the units from the 1:625,000 geological map that were categorised as unclassified were manually allocated to a radon potential class depending on the nature of the bedrock geology and overlaying drift deposits. Although the mean LRF measurement in the unclassified areas is relatively high, none of the unclassified areas can be considered to have a high likelihood of having high radon levels because none of the areas fall within a geological unit that is high in uranium. The unclassified areas in the central highlands and northern Shetlands are dominated by quartzite which is not associated with high levels of radon. Similarly, the Isle of Lewis has low radon potential because its geology is dominated by gneiss. The area in southern Scotland that is unclassified is dominated by sedimentary deposits and also has low radon potential. This anomaly probably arises because of the effect of two extremely high radon measurements taken in the unclassified areas. The presence of two high radon measurements reflects the localised variations in radon levels that can occur (even within seemingly similar geological areas) in response to variations in solid geology, drift deposits and housing characteristics. If the two

extreme values are not included in the calculation, then the mean LRF measurement in the unclassified areas is almost identical to that in the low/moderate category. Therefore, given the low radon potential of the unclassified areas, the areas that were previously labelled as unclassified were recoded to reflect their likely radon potentials which were all low or low-moderate.

An updated radon potential map of Scotland is shown in Figure 7.11 and it demonstrates that most of the areas that were previously unclassified were reclassified as low or low-moderate. Table 7.8 outlines the number of OA centroids that fell within each of the radon potential bands once the edge effect had been corrected for. All OAs in Scotland were classified into one of the six radon potential classes and the 1,562 OAs that were previously unclassified were then assigned to the low or low-moderate categories. Table 7.9 shows the classification of the LRF data into the radon potential bands once the edge effect was accounted for. The 69 previously unclassified values were distributed between the low and low-moderate classes. Importantly, the LRF radon reading still increased with each radon potential class.

Radon Potential	Number of Output-Areas	Percentage of Ouput-Areas
Mis-Classified	0	0.00
Unclassified	0	0.00
Low	15478	40.46
Low / Moderate	16815	43.96
Moderate	3762	9.83
High	2167	5.66
Very High	31	0.81
Total	38253	100.00

Table 7.8. Summary of relationship between radon potential class and OA lung cancer data - reclassified radon data

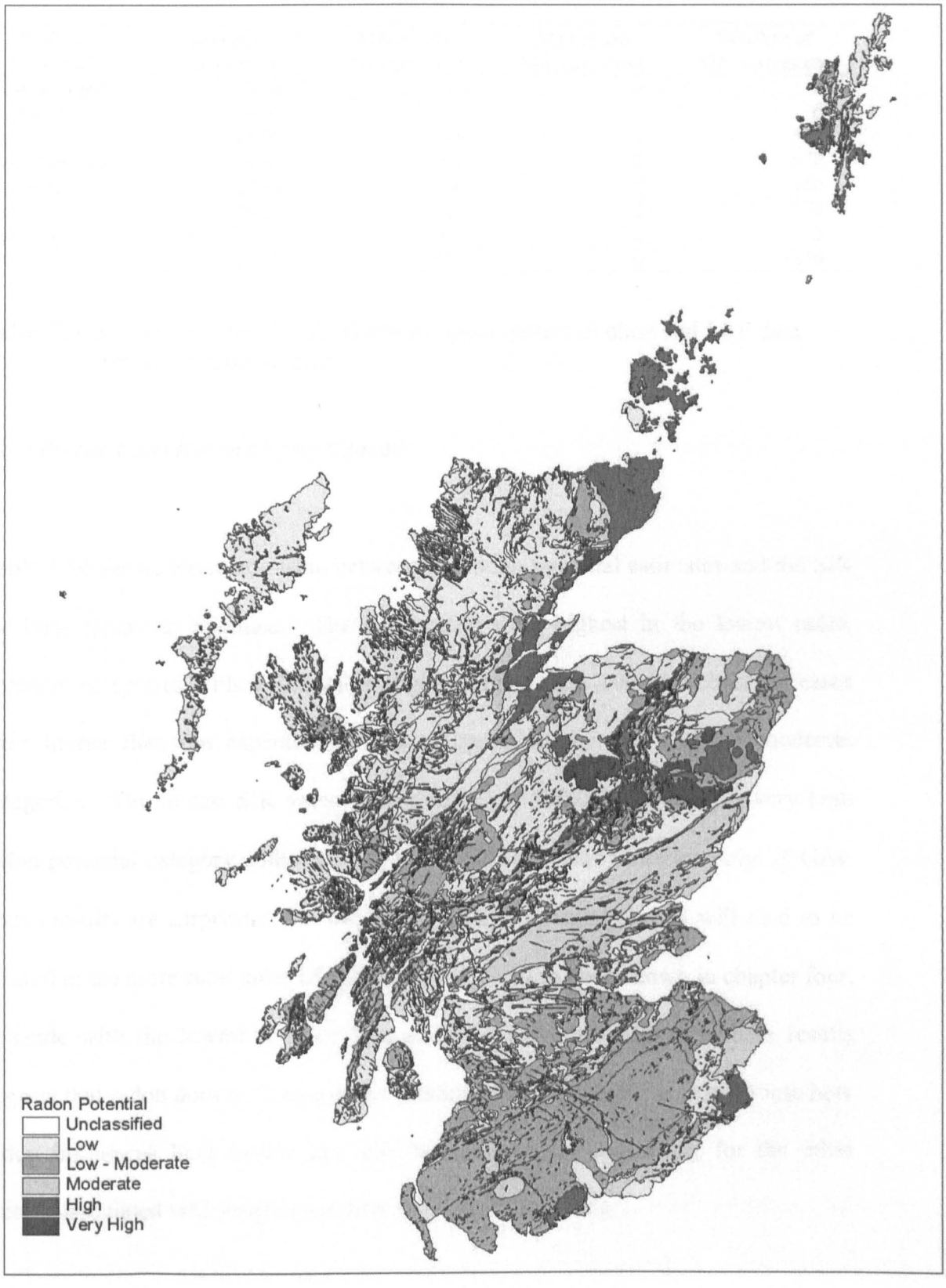


Figure 7.11. Radon potential map with recoding of unclassified areas

Radon Potential	Average Measurement	Maximum Measurement	Minimum Measurement	Number of Measurements
Mis-Classified	0.00	0	0	0
Unclassified	0.00	0	0	0
Low	13.99	234	1	562
Low / Moderate	16.60	763	0	570
Moderate	17.38	68	2	110
High	24.65	329	2	72
Very High	54.50	107	2	2
Total	16.05	763	0	1316

Table 7.9. Summary of relationship between radon potential class and LRF data – reclassified radon data

7.6.5 Radon Potential and Lung Cancer

Table 7.10 shows the relationship between the radon potential estimates and the SIR for lung cancer in Scotland. The SIR values were highest in the lowest radon potential categories with a SIR value slightly higher than 100 (the observed cases were higher than the expected number of cases) in the low and low-moderate categories. The lowest SIR value (65.14 cases per 100,000) was in the very high radon potential category, although this SIR calculation was based on only 31 OAs. These results are surprising, but the highest radon potential classes will tend to be located in the more rural areas of Scotland which, as has been shown in chapter four, coincide with the lowest rates of lung cancer. Therefore, although these results suggest that radon does not increase the incidence of lung cancer, an insight into how radon influences lung cancer can only be achieved by controlling for the other factors associated with lung cancer first, particularly smoking.

Radon Potential	Sum of Observed	Sum of Expected	SIR (per 100,000)
Low	7632	7519.51	101.50
Low / Moderate	8377	8111.36	103.27
Moderate	1666	1924.76	86.56
High	782	1059.34	73.82
Very High	13	19.96	65.14

Table 7.10. Relationship between radon potential class and lung cancer SIR.

7.7. Coal Mining

The second geologically related factor to be investigated as a potential cause of lung cancer was the impact of the coal mining industry. In chapter two it was suggested that there may be a causal link between coal mining and lung cancer, but this has not been considered in a Scottish context. This issue was addressed by examining how the coal mining industry influenced the incidence of lung cancer in 1991. This section outlines a method for estimating the proportion of people active in the coal mining industry in OAs across Scotland using data from the 1971 census. However, coal mining is not an occupation that is directly represented in the census and, therefore, a more aggregated mining occupational variable was used instead.

7.7.1 Methodology

Census data are a useful resource to identify the type of industry in which residents are employed. However, there has been a large reduction in coal mining activity during the post-war period to the extent that there were few remaining collieries at the time of the 1991 census (Halliday, 1990; Hollywood, 2002). Although using 1991 census data would have tied in with most of the other data sets that are used in this study, it would not have been sensible as many of the former mining

communities would have had very low levels of employment in the coal industry in 1991 due to the closures. It was more appropriate to use census data prior to 1991 in order to measure the level of employment in the coal industry at that time. In this analysis, the effect of the decline of coal mining during this period was so crucial that the difficulties of using lagged data were less relevant because contemporary data would not capture the distribution of the coal industry. For this reason data from the 1971 census were extracted instead.

The percentage of the population employed in mining was extracted for enumeration districts (EDs) from the 1971 census data and this included all people employed in any aspect of the mining industry. Estimates were calculated for 1991 PPSs using the data extracted for the 15,890 1971 EDs in Scotland. The 1971 census units have no relationship with the 1991 boundaries and so the number employed in mining in 1971 had to be recalculated for the 1991 census units in order that the data were consistent with other data used in this study. Digital boundaries of census units were never created in 1971 but a centroid was determined for each ED allowing the data to be mapped as points and then associated with the 1991 PPSs using the point-in-polygon procedure in ArcInfo. The population-weighted estimate of the percentage employed in the mining industry in 1971 was then calculated for 1991 PPS boundaries. It was found that 31 of the 1991 PPSs did not have a 1971 ED centroid within them and so the population weighted average of the contiguous PPSs was used instead.

7.7.2 Results

Figure 7.12 shows the distribution of the percentage of residents employed in the mining industry in 1971 for 1991 PPS boundaries. The map demonstrates that a large number of people were employed in mining in the coalfields of the Central Belt, southern Fife and south west Scotland. There were also high rates in some locations on the west coast, such as near to Tarbet and Oban, which probably reflect types of mining other than coal. The percentage employed in mining tended to be very low north of the central belt although there were a few high values around Pitlochry and Braemar. Therefore, although this measurement provides a sensible estimate of coal mining, this map shows that some areas have high values that are not within the coalfields of Scotland because the variable also incorporates other types of mining.

Table 7.11 shows the lung cancer SIR for OAs that have been divided into four categories depending on the percentage employed in coal mining in 1971. The table shows that there was not a clear relationship between the SIR and those employed in coal mining in 1971. The SIR was highest in the second lowest category (0.01 to 0.67 percent) and lowest in the second highest category (0.68 to 2.44). However, it is important to examine whether this relationship is maintained once other risk factors have been controlled for.

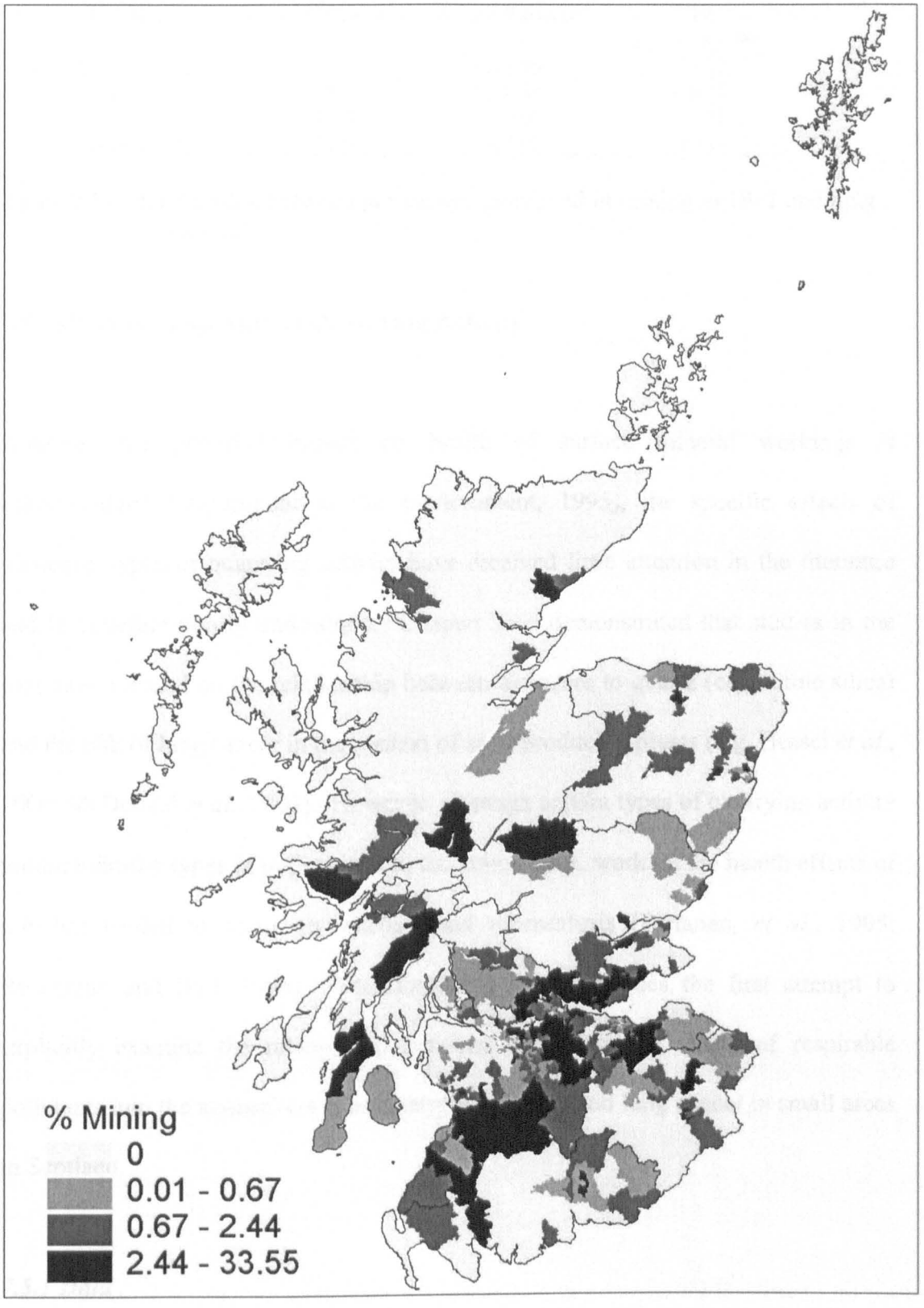


Figure 7.12. Percentage of residents employed in the mining industry in 1971.

% Mining in 1971	Sum of Observed	Sum of Expected	SIR (per 100,000)
0	8704	8897.59	97.82
0.01 - 0.67	4217	3878.88	108.72
0.68 - 2.44	2925	3183.09	91.89
2.45 - 33.55	2624	2674.67	98.11

Table 7.11. Relationship between percentage employed in mining in 1971 and lung cancer SIR.

7.8 Estimating Exposure to Quarrying Activity

Although the potential impact on health of surface mineral workings is acknowledged (Department of the Environment, 1995), the specific effects of particular types of quarrying activity have received little attention in the literature and is therefore poorly understood. Chapter Two demonstrated that studies in the past have focused on the relationship between exposure to quartz (crystalline silica) and the risk of lung cancer in the context of sand-producing plants (e.g. Hessel *et al.*, 2000; McDonald *et al.*, 2001). However, although certain types of quarrying activity produce similar types of pollutants into the atmosphere, work on the health effects of this has tended to focus on silicosis and tuberculosis (Partanen, *et al.*, 1995; Rosenman and Hall, 1996). Therefore, this study provides the first attempt to explicitly examine the relationship between the effect of release of respirable pollutants into the atmosphere from quarrying activity and lung cancer in small areas in Scotland.

7.8.1 Data

There are eight different types of quarries in Scotland: granite, basalt, dolerite, gabbro, limestone, sandstone, schist and greywacke. The most extreme dust

experiences are most likely to be experienced in close proximity to quarries (within 200 metres) but impacts could occur up to a distance of 500 metres (Department of the Environment, 1995). However, concerns about dust levels have been expressed up to a distance of one kilometre from the source (Department of the Environment, 1995a; Department of the Environment, 1995b). Information on the location and type of quarrying activity at all quarries in the UK is available in the national handbook of quarry information (Jones *et al.*, 1991). The postcode and type of activity for all the quarries in Scotland in 1991 were entered into a spreadsheet and supplemented with the quarries in northern Cumbria and Northumberland as these may influence the level of exposure in southern Scotland. Unfortunately, the amount of rock produced from each quarry is not available so it was not possible to develop estimates that were proportional to the size and output of the quarry.

Quarry data from 1991 were chosen to match up with the data on lung cancer which once again raises the issue of using contemporary measures of exposure to a carcinogen to predict levels of lung cancer rather than a measure that incorporates the lag effect between exposure and mortality. However, a geographically referenced and comprehensive database of quarries prior to 1991 is not available. Furthermore, it is likely that most quarries will have been active many years prior to 1991 (Jones *et al.*, 1991). For each quarry, the postcode was converted into an easting and northing using the postcode look-up table (cpd95) and in total, 147 quarries were included.

7.8.2 Methodology

In order to estimate the exposure to the different types of quarries, surfaces were produced that estimated the potential level of exposure for the whole of Scotland. Probably the most commonly used form of density estimation is the kernel estimate (Bailey and Gatrell, 1995), which is a statistical technique whereby a distribution of discrete points or 'events' are transformed into a continuous risk surface. A moving three-dimensional function (the kernel) of a defined radius or 'bandwidth' visits each of the events in turn, and weights the area surrounding the point proportionately to its distance to the event. The sum of these individual kernels is then calculated for the study region, and a smoothed surface produced (Figure 7.13). Increasing the bandwidth stretches the region around each point within which the observed events influence the kernel estimate (Bailey and Gatrell, 1995). The choice of bandwidth should therefore be conceptually driven and reflect the likely sphere of influence of the phenomenon under investigation.

Kernel estimation was used to provide an estimate of exposure for each of the quarry types. In addition, a surface of total exposure was constructed for all quarry types. Because quartz is considered to have the greatest carcinogenic effect (Holland, 1990; Bello *et al.*, 2002), a surface of the total exposure to quartz quarries (granite and sandstone) was also calculated separately. For each of the surfaces, bandwidths of one kilometre and two kilometres were selected so that the full extent of the potential quarrying exposure was incorporated. An output cell size of 250 metres was chosen because it provides a balance between allowing for the localised variations in the

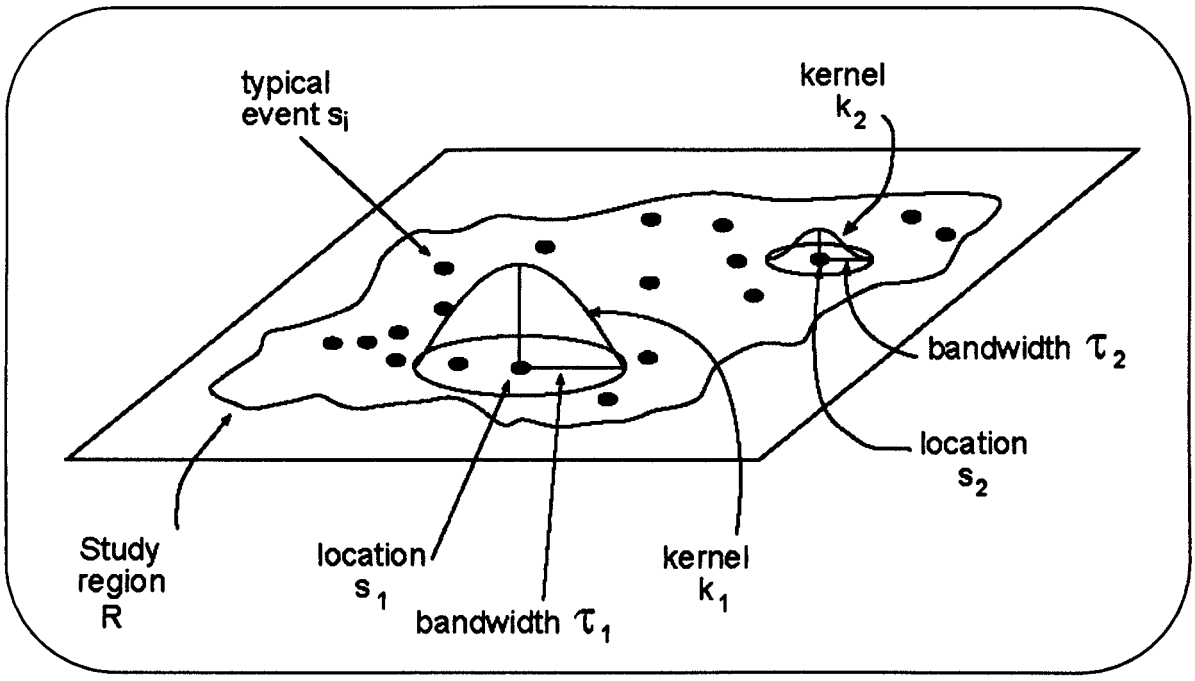


Figure 7.13. Principles of Kernel estimation (Sabel *et al.*, 2000)

levels of exposure to quarrying activity whilst also maintaining a relatively low number of cells and thus reducing the computational time that is required.

The final stage of the procedure was to associate the kernel estimate surfaces with the census geography so that all of the data sets were referenced to a consistent geography. Each OA in Scotland was associated with the exposure to each type of quarrying activity by using the point-in-polygon procedure to associate the population-weighted OA centroid with the exposure estimate for each surface. Therefore, each OA had a set of values attached that estimated the level of exposure to each type of quarry. In most cases the value was zero but if an OA was within two kilometres of a quarry then the value was greater than zero.

7.8.3 Results

Table 7.12 shows the number of OAs and the number of cases of lung cancer in OAs that are within two kilometres of each quarry type. A total of 2,418 OAs fell within two kilometres of any type of quarry and in those OAs there were 1,020 cases of lung cancer during the period 1988 to 1991. Table 7.12 also provides a summary of the expected number of cases and the Standardised Incidence Ratio (SIR). It was more sensible to calculate the SIRs for OAs within two kilometres of a quarry because there were so few OAs within one kilometre. The SIR provides a value that can be used to compare the cancer rate in the areas close to each type of quarry.

Quarry Type	No. of OAs	Observed Cases	Expected Cases	SIR (per 100,000)
Basalt	943	441	474.37	92.97
Dolerite	674	282	335.67	84.01
Gabbro	21	6	7.22	83.08
Granite	71	23	34.90	65.90
Greywacke	47	24	27.91	85.99
Limestone	186	95	96.64	98.30
Sandstone	354	125	170.20	73.44
Schist	122	25	55.01	45.45

Table 7.12. Standardised Incident Ratios (SIR) for OAs within two km of a quarry.

The table demonstrates that the SIR values tended to be low in the areas close to quarries because the SIR was below 100 for all cases which means the expected number of cases were more than the observed number of cases. The table also shows that the highest SIR values were in the regions close to limestone, basalt and greywacke quarries and the lowest values were found in the areas close to schist and granite quarries. This is perhaps surprising given that it is the quartz quarries that are thought to have the greatest potential for a carcinogenic effect as neither of these

show a higher than expected incidence of lung cancer. However, the location of quarries are similar to high levels of radon in that the areas where there is a large number of quarries tend to be the most rural, where lung cancer incidence is consistently low. Furthermore, these results assume that the areas are all homogenous in their exposure to all other risk factors such as smoking behaviour and air pollution, when infact this is unlikely to be the case. Therefore, it is more appropriate to use the estimates to examine the incidence of lung cancer once other factors have been controlled for (see chapter eight).

7.9 Conclusion

This section has developed methods for estimating the potential exposure to radon, coal mining and quarrying activity for all 38,253 OAs in Scotland. The calculation of the estimates is problematical due to the many factors that influence the level of exposure and the limit on data availability due to resources and confidentiality. For each estimate there are many different approaches that can be used to provide predictions and each one has some limitations in terms of methodology and/or the availability of data. Therefore, the most suitable technique has been used to provide estimates depending upon the nature of the process and the data that are available.

It has been demonstrated that there are large variations in indoor radon levels across Scotland. The radon level in a particular building depends upon a complex interaction of factors such as geology, drift and housing characteristics. Therefore, in the absence of comprehensive data on radon levels in buildings across Scotland it is difficult to make accurate predictions of radon levels for all areas of the country.

The radon level estimation procedure has incorporated the methodology developed by the NRPB to estimate radon potential and then the reliability of the estimates tested using an additional and independent data set that was supplied by the Leukaemia Research Fund. This work has produced two data sets that reflect the radon potential category for all OAs in Scotland. The first data set simply incorporates the radon potential values that have been calculated based upon a sufficient number of measurements to accurately reflect the combined effect of solid geology and drift deposits. The second data set is similar to the first but has recoded the areas that were previously unclassified into one of the five radon potential classes based on knowledge of the local geology and drift. These estimates provide a comprehensive estimate of radon potential for OAs in Scotland but the estimates are not based on a systematic collection of radon measurements for 1,562 of the 38,253 OAs. It is therefore useful to utilise both sets of estimates to understand the influence of radon on lung cancer bearing in mind the advantages and limitations of each method.

An estimate of coal mining activity has been calculated using 1971 census data to predict the proportion of people employed in the coal mining industry for 1991 PPSs. It was sensible to use 1971 data due to the closing of many coal mines between 1971 and 1991 but this created a geographical incompatibility between this data and all other data sets used in this study. Therefore, techniques were developed to estimate the level of employment in coal mining in 1971 for 1991 boundaries by associating 1971 ED centroids with the 1991 PPS boundaries

However, the use of 1971 census data to predict the proportion of people employed in the coal mining industry was further complicated by the dissemination of 1971 census data on occupation as a more aggregated variable that included all types of mining rather than coal mining specifically. Therefore, all people employed in any aspect of the mining industry were included. This aggregation is reflected in the resulting estimates of those employed in coal mining, which demonstrated that in addition to the high percentage values in coal mining communities there were high proportions of people employed in the mining industry in specific areas outside of these areas (Figure 7.12). For example, there was a high proportion of people employed in mining on the Mull of Kintyre which is probably a reflection of the presence of other mining industries, such as quarrying in this region.

The levels of exposure to different types and different combinations of quarries were calculated using kernel estimation. Kernel estimation has the advantage of allowing for the key parameter of distance to be incorporated but also allows for the interaction between quarries to be modelled. The approach was limited by the fact that information was not available on the size of the quarry. Furthermore, more complex models which allow other factors such as wind direction and topography to influence the spread of particles from each quarry were beyond the scope of this study. However, the approach does provide a simple but effective estimate of the exposure to different types of quarries.

The comparisons of the SIR values with each of the three estimates demonstrated that lung cancer incidence tends to be lower in areas where the exposure to each risk factor is higher. However, this is probably because high levels of geological risk

tend to be located in the more rural areas where the other risk factors associated with lung cancer (e.g. smoking) tend to be low. Therefore, to examine all three properly, lung cancer models have to be developed that incorporate the estimates of other risk factors, especially smoking.

There are important temporal differences between the three sets of estimates and so this once again raises the issue of using contemporary and non-contemporary measures of a risk factor to explore disease incidence. The radon potential estimates can be considered to be static over time as the category is a response to geology and drift which change only very slowly, the quarrying exposure estimates were provided for 1991 and the mining variables for 1971. The quarrying variables coincided temporally with the lung cancer data rather than incorporating a likely lag between the exposure to the carcinogen and lung cancer initiation. The use of 1971 data raises further issues about using explanatory variables explaining exposure prior to the disease data, as those that were employed in the coal industry in 1971 and subsequently diagnosed with lung cancer may have changed residence between 1971 and 1991. If this is common place then it raises methodological issues about using non-contemporary explanatory variables. These issues can be better addressed when the variables are used as explanatory variables in a model of lung cancer incidence.

8. MODELLING LUNG CANCER INCIDENCE IN SCOTLAND

8.1 Introduction

This section draws together the variables that may be associated with lung cancer which have been calculated in previous chapters, in order to examine whether they are important in explaining the incidence of lung cancer in Scotland. Earlier chapters have examined the relationship between lung cancer and factors such as age, sex, smoking, air pollution, radon, coal mining and quarrying. However, each variable was examined independently and so it is important to consider whether the relationships remain once other risk factors are controlled for. Here, the variables are examined in combination using Poisson regression, which is an appropriate method when the dependent variable is predominantly a count of low frequency. Each explanatory variable was incorporated into a set of Poisson models with the count of age and sex specific cases of lung cancer in each OA in Scotland set as the dependent variable.

The first part of this chapter introduces the principles of Poisson regression and provides some geographical examples of the use of this technique. A methodology for developing a Poisson model using the lung cancer data and the explanatory variables that represent the potential risk factors is then outlined. The resulting Poisson regression models are then presented. The age and sex structure of each area were controlled for and each of the potential causal factors were considered to see if they were significant. These factors were smoking behaviour, air pollution

levels, radon levels, quarrying activity, deprivation, population density, coal mining and occupation type.

It has been shown in earlier chapters that there are important variations in lung cancer incidence between different age groups, men and women, particular types of area and certain types of lung cancer. Therefore, the next section studies age and sex groups separately to see if the risk factors have different effects on certain population subgroups. The incidence of lung cancer in rural areas was then explored separately to examine whether different factors are more or less important in a rural context. Finally, the different types of lung cancer were examined individually to see if the risks associated with each type varied.

The final stage involved calculating and mapping the residuals from the model that provided the best prediction of the incidence of lung cancer in Scotland. This gave an insight into how successful the model is in explaining lung cancer in different areas and helped develop further hypotheses about how and why lung cancer incidence varies throughout Scotland. The residual values in Glasgow were given particular attention, due to the unusual rates of lung cancer found there. These results may have policy relevance because they can be used to identify and target areas where lung cancer is particularly bad.

8.2 Poisson Regression

Poisson regression is a form of generalised linear modelling that is suitable for modelling counts which are assumed to have a Poisson distribution and are

independent of each other. The Poisson distribution describes the probability that an event occurs k times in a fixed period given that each occurrence is independent and has a constant probability (Lovett and Flowerdew, 1989). Poisson regression is an appropriate method for examining relatively rare events such as the occurrence of cancer.

The natural logarithm of the estimate of the predicted value is equal to the linear combination of the corresponding values of the independent variable. If there is only one independent variable, then the predicted value of the dependent variable for case i is the maximum likelihood estimate $\hat{\lambda}_i$ of the mean of a Poisson distributed variable $\hat{Y}_i$:

$$\text{Ln}(\hat{\lambda}_i) = \beta_0 + \beta_1 x_i$$

or

$$\hat{\lambda}_i = \exp(\beta_0 + \beta_1 x_i)$$

In contrast to ordinary least squares regression, the Poisson model does not assume that the data are homoscedastic and the variance of each case is equal to the corresponding predicted value (O'Brien, 1992).

Poisson regression has often been used to examine ill health and disease. For example, Flowerdew and Geddes (1999) compared the distribution of long term

limiting illness with other variables for census wards in north-east England. They concluded that the geography of limiting long-term illness was related to that of coal mining in previous decades. Boyle *et al.,* (1996) used Poisson regression techniques to examine adult patients in Dyfed and Glamorgan who commenced chronic renal replacement therapy between 1985 and 1994. After controlling for age, population distribution, socio-economic variables and ethnic group, they found that there was a significant negative relationship between referral rates and distance of residence from the renal unit for patients aged over 60 years, but not for younger patients. Diamond *et al.,* (1999) examined the factors influencing the variation in census ward-level teenage conception rates using Poisson models. Deprivation indices were shown to be important in explaining geographical variations. Gatrell *et al.*, (2002) used Poisson regression to assess the use of cardiac services in parts of north-west England. They found that utilisation of the service was related to material deprivation and that areas with large Asian populations had far fewer angiograms and angioplasties than might have been predicted given the expected need of the population.

Poisson regression has also been used to examine the relationship between lung cancer and particular risk factors such as socio-economic status (Vacchino, 1999), occupation type (Boffetta *et al.*, 1992; Weston *et al.*, 2000), passive smoking (Jee *et al.*, 1999) and family history (Hemminki and Li, 2001). Vacchino (1999) used Poisson regression to examine lung cancer in administrative areas of Argentina and the research demonstrated that rates of lung cancer were higher in areas of high socio-economic status, although the study did not control for smoking behaviour. Boffetta *et al.* (1992) demonstrated that lung cancer incidence was raised among

workers involved in the production of man-made mineral fibres especially for the cohorts who worked in the rock-slag wool and glass wool industries. Similarly, Weston *et al.* (2000) found that occupational exposure to abrasive dusts was associated with lung cancer incidence in a large representative sample of the Canadian workforce between 1965 and 1971. Jee *et al.* (1999) examined the effects of husbands' smoking on the incidence of lung cancer in Korean women and found that wives of heavy smokers were more likely to develop lung cancer. Hemminki and Li (2001) used Poisson regression to demonstrate that a family history of carcinoids in first-degree relatives was one risk factor associated with carcinoid tumours in the lung. These provide just a handful of the geographical examples of the application of Poisson models in a health context. However, they highlight the widespread use of the methodology and demonstrate that it is a useful tool in the analysis of data sets that are distributed in this way.

8.3 Methods

8.3.1 Introduction

Poisson regression was used to predict the incidence of lung cancer for the period 1988 to 1991 in all OAs in Scotland using a set of explanatory variables that potentially influence lung cancer rates. These include data on the demographic profile of each output-area (OA) and the calculated estimates of potential causal factors such as smoking behaviour, air pollution levels and geological factors.

8.3.2 Data Set for Poisson Regression

The count of the number of cases of lung cancer in OAs was combined with population data and the potential explanatory variables (Table 8.1). For each OA the population count and the count of cases of lung cancer were calculated for 12 age-sex groups. Therefore, each OA was represented 12 times in the dataset, once for each age-sex group. The age and sex-specific population count was obtained from the 1991 census, which coincided temporally with the data on lung cancer incidence. The age and sex groups were chosen to coincide with the age and sex-specific smoking estimates (males and females aged 16 to 24, 25 to 34, 35 to 44, 45 to 54, 55 to 64 and over 65). Two categorical variables were included to define which age and sex group each record represented.

In addition, the explanatory variables outlined at the end of chapter four were associated with each record. These included smoking behaviour, air pollution levels, radon potential estimates, coal mining and quarrying activity. The smoking behaviour estimates were age and sex-specific and when the smoking probability was missing, the OA was excluded from the analysis[7]. All of the other explanatory variables were not age and sex-specific and hence the estimate was identical for the 12 times an OA was included in the dataset. A total of 37,907 of the 38,254 OAs in Scotland were included in the analysis and because each OA was represented 12 times there were 454,884 records in the entire data set.

[7] Smoking probabilities were absent when it was not possible to calculate estimates from the surrounding areas due to missing values in the contiguous OAs (see section 5.6)

Description	Label
Dependent Variable	
Count of lung cancer cases (age and sex specific)	Lung
Independent Variables	
Age and Sex-Specific variables	
Population count	Population
Smoking probability	Smoking
Age group (6 levels: 16-24, 35-34, 35-44, 45-54, 55-64, 65+)	Age
Sex group (2 levels: male/female)	Sex
Contextual Variables	
Carstairs index of deprivation	Carstairs
Population density	Pop. Density
Pollution	
Estimate of 1971 pollution levels in PPSs	Pollution 71
Estimate of 1991 pollution levels in PPSs	Pollution 91
Radon	
Radon potential classification using BGS classification (6 levels)	Rad A
Radon potential classification using BGS classification (5 levels - unclassified areas recoded)	Rad B
Radon potential classification using BGS classification (2 levels: i. very high and high. ii. low, low-moderate, moderate)	Rad C
Quarrying	
Estimate of exposure to basalt quarries for OAs within 1 km	Basalt 1km
Estimate of exposure to dolerite quarries for OAs within 1 km	Dolerite 1km
Estimate of exposure to gabbro quarries for OAs within 1 km	Gabbro 1 km
Estimate of exposure to granite quarries for OAs within 1 km	Granite 1km
Estimate of exposure to greywacke quarries for OAs within 1 km	Greywacke 1km
Estimate of exposure to limestone quarries for OAs within 1 km	Limestone 1 km
Estimate of exposure to schist quarries for OAs within 1 km	Schist 1km
Estimate of exposure to sandstone quarries for OAs within 1 km	Sandstone 1km
Estimate of exposure to granite and sandstone quarries for OAs within 1 km	GranSand 1km
Estimate of exposure to all quarries for OAs within 1 km	All Quarries 1km
Estimate of exposure to basalt quarries for OAs within 2 km	Basalt 2km
Estimate of exposure to dolerite quarries for OAs within 2 km	Dolerite 2km
Estimate of exposure to gabbro quarries for OAs within 2 km	Gabbro 2 km
Estimate of exposure to granite quarries for OAs within 2 km	Granite 2km
Estimate of exposure to greywacke quarries for OAs within 2 km	Greywacke 2km
Estimate of exposure to limestone quarries for OAs within 2 km	Limestone 2 km
Estimate of exposure to schist quarries for OAs within 2 km	Schist 2km
Estimate of exposure to sandstone quarries for OAs within 2 km	Sandstone 2km
Estimate of exposure to granite and sandstone quarries for OAs within 2 km	GranSand 2km
Estimate of exposure to all quarries for OAs within 2 km	All Quarries 2km
Coal Mining	
% of residents employed in mining in 1971	Mining 71
1991 Employment Variables	
% of residents aged over 16 who were in work and employed in the agriculture, forestry and fishing industries	Agriculture
% of residents aged over 16 who were in work and employed in the energy and water industries	Energy
% of residents aged over 16 who were in work and employed in the mining industries	Mining

% of residents aged over 16 who were in work and employed in the manufacturing metal etc. industries	Manufacture
% of residents aged over 16 who were in work and employed in the other manufacturing industries	Other Manuf.
% of residents aged over 16 who were in work and employed in the construction industries	Construction
% of residents aged over 16 who were in work and employed in the distribution and catering industries	Distribution
% of residents aged over 16 who were in work and employed in the transport industries	Transport
% of residents aged over 16 who were in work and employed in the banking and finance etc. industries	Banking
% of residents aged over 16 who were in work and employed in the other service industries	Other Service
Male-Specific Employment Variables	
% of males aged over 16 who were in work and employed in the agriculture, forestry and fishing industries	Male Agriculture
% of males aged over 16 who were in work and employed in the energy and water industries	Male Energy
% of males aged over 16 who were in work and employed in the mining industries	Male Mining
% of males aged over 16 who were in work and employed in the manufacturing metal etc. industries	Male Manufacture
% of males aged over 16 who were in work and employed in the other manufacturing industries	Male Other Manuf.
% of males aged over 16 who were in work and employed in the construction industries	Male Construction
% of males aged over 16 who were in work and employed in the distribution and catering industries	Male Distribution
% of males aged over 16 who were in work and employed in the transport industries	Male Transport
% of males aged over 16 who were in work and employed in the banking and finance etc. industries	Male Banking
% of males aged over 16 who were in work and employed in the other service industries	Male Other Service
Female-Specific Employment Variables	
% of females aged over 16 who were in work and employed in the agriculture, forestry and fishing industries	Female Agriculture
% of females aged over 16 who were in work and employed in the energy and water industries	Female Energy
% of females aged over 16 who were in work and employed in the mining industries	Female Mining
% of females aged over 16 who were in work and employed in the manufacturing metal etc. industries	Female Manufacture
% of females aged over 16 who were in work and employed in the other manufacturing industries	Female Other Manuf.
% of females aged over 16 who were in work and employed in the construction industries	Female Construction
% of females aged over 16 who were in work and employed in the distribution and catering industries	Female Distribution
% of females aged over 16 who were in work and employed in the transport industries	Female Transport
% of females aged over 16 who were in work and employed in the banking and finance etc. industries	Female Banking
% of females aged over 16 who were in work and employed in the other service industries	Female Other Service

Table 8.1. Variables used in Poisson regression

The pollution estimates were calculated for both 1971 and 1991. The first radon potential measurement was based on the BGS radon potential classification scheme, whilst the second estimate used this scheme but recoded each of the unknown radon potential values into one of the five classes based upon the underlying bedrock and drift deposits on the surface. Potential exposure to different types and combinations of quarrying activity for OAs within one kilometre and two kilometres of a quarry were also included in the analysis. These included exposure to basalt, dolerite, gabbro, granite, greywacke, limestone, schist and sandstone. In addition, variables that estimated the exposure to the quartz dominated quarries (granite and sandstone) and to all types of quarries were also added. An estimate of the percentage of people employed in coal mining in 1971 for each pseudo-postcode sector (PPS) was also included. Ten variables that represented the percentage employed in ten different industries in 1991 were also incorporated, as were the age and sex-specific equivalents.

Finally, the population density in 1991 and the 1991 Carstairs Index of Deprivation were also included. The population density provided an urban-rural proxy to test whether the urban-rural bias in lung cancer incidence identified in chapter four remained once the various explanatory factors had been controlled for. The Carstairs Index was chosen to capture the concept of material deprivation (Morris and Carstairs, 1991) in order to examine whether area deprivation was significant in explaining the incidence of lung cancer in Scotland, once other risk factors had been controlled for.

8.3.3. Modelling Procedure

There are a number of approaches that can be taken to build a Poisson regression model. This study follows the approach suggested by Lovett and Flowerdew (1989) who outlined a systematic strategy for developing a model. The first step is to fit the null model (or Grand Mean Model) where the predicted values ($\hat{Y}$) are estimated to be the mean of the observed values (hence a rather poor estimate). The deviance value from the null model provides a baseline measure of fit against which subsequent models can be compared. This is a measure of the overall difference between the observed values of Y and those predicted by the model. If the deviance value is high then it indicates that the correspondence between the observed and estimated values is poor, but if the deviance value is low then it indicates a closer correspondence. Other explanatory variables can be added and the reduction in the deviance is one measure of their significance.

Each of the potential explanatory variables were added one at a time to create a univariate model and any change in the model deviance was noted. Any reduction in deviance was compared with the critical value of chi-square. The introduction of a numerical variable results in the loss of one degree of freedom, therefore a decline in deviance of more than 3.84 indicates that the variable was significant at the 0.05 level. Each of the potential explanatory variables were tested individually, both in their raw and logged form, where applicable, and the more significant of the two was used in subsequent models. Once this had been done for each variable, they were then reintroduced one at a time, starting with the one that was the most successful in reducing the deviance, but this time in tandem with other variables. As each variable

was added, the change in the deviance was noted. Furthermore, the individual standard errors and estimates were compared in order that any insignificant variables could be removed from the model. An individual parameter is significant if the standard error is less than half the estimate value (or the T value is approximately greater than –2 but less than 2).

As new variables were added to the model, the significance of other variables could alter and so all of the parameter estimates and the corresponding standard errors had to be checked at each stage. The procedure was finished when the addition of each remaining explanatory variable failed to reduce the deviance significantly. At this stage interaction terms could be introduced to test whether they were significant. Finally, once this was completed, the parameter estimates from the parsimonious model were assessed.

Knudsen (1992) noted that *a priori* information can be incorporated into a Poisson model using an offset (a covariate with a known parameter value of one). Although the interpretation of the model will be altered by the use of an offset term, the statistical properties of the model will remain the same. Typically the *a priori* information is derived from established theory that identifies who the population at risk are (Diamond *et al.*, 1999). In the case of lung cancer, the number of cases of lung cancer increases with the size of the population. Therefore, the log of the population count in each age-sex group in an area was set as an offset which constrained the results to be proportional to the size of the cohort. In effect, this removed the effect of population size from the model and made the interpretation of the remaining parameters easier. However, the inclusion of an offset resulted in a

higher deviance value and so it was more difficult to interpret how much of the variation was explained by the model.

However, the inclusion of all cases of lung cancer in all OAs ignores the fact that some processes are more important for particular groups of people, or only operate in certain geographical areas or for particular types of lung cancer. Therefore, additional models were run that used subsets of the data to model lung cancer for males and females and the older and younger age groups separately. In addition, the explanatory variables were tested just in rural areas to examine whether there are certain factors that are more important in these areas compared to urban areas. Finally, the effect that the explanatory variables have on different types of lung cancer was also explored by changing the dependent variable to be the count of particular forms of lung cancer, as discussed in chapter four, rather than the total number of cases.

8.4 Results Using All Cases of Lung Cancer

The following discussion outlines the results of modelling each explanatory variable and different combinations of these variables to consider which factors were important in explaining lung cancer incidence in Scotland. A parsimonious model is suggested that best explains the variation in all cases of lung cancer incidence and identifies those significant variables that help to explain this variation. Each variable is discussed as it is added to the model and then a wider discussion of the parsimonious model follows.

8.4.1 Grand Mean Model

The Grand Mean Model, with the log of the population count defined as an offset, is shown in model one of Table 8.2. The scaled deviance was 113,522 and this provided a benchmark to which all subsequent models could be compared to assess the effect of adding independent variables. However, the deviance was unusually low compared to the number of degrees of freedom which suggests that the dataset was underdispersed due to the sparseness of the dataset (Boyle and Flowerdew, 1993; Boyle *et al.*, 1997). This makes the comparison of the deviance value with the critical value of chi-square unreliable and so it was hard to assess overall model fit and this was compounded by the inclusion of an offset term. However, although this makes the interpretation of the deviance values of this and subsequent models problematic (except for considering any reduction in the deviance), it was still valid to interpret the parameters in the model. Each of the potential explanatory variables were added to this model one at a time and then those variables that reduced the deviance by the most were added into the multivariate model.

8.4.2 Age and Sex

Models two and three of Table 8.2 show the effect of adding the categorical variables age and sex. These were the only individual-level variables in the dataset and the parameter estimates in this model were all significant and intuitively reasonable. Furthermore, the deviance fell from 113,522 to 84,745, demonstrating that much of the variation could be explained by the age and sex structure of the individuals. As expected, there were more cases of lung cancer in the older age

	Model 1			Model 2			Model 3		
Scaled Deviance		113,522			85,706			84,745	
Degrees of Freedom		454,883			454,878			454,877	
	Estimate	Std. Error	T-Value	Estimate	Std. Error	T-Value	Estimate	Std. Error	T-Value
Intercept	-5.442	0.007364	-739.00	-11.77	0.3466	-33.96	-11.560	0.3459	-33.42
Age (25-34)				1.592	0.3922	4.06	1.595	0.3916	4.07
Age (35-44)				3.845	0.3523	10.91	3.844	0.3516	10.93
Age (45-54)				5.589	0.3478	16.07	5.591	0.3471	16.11
Age (55-64)				6.956	0.3469	20.05	6.965	0.3462	20.12
Age (65+)				7.346	0.3467	21.19	7.315	0.3460	21.14
Sex (Female)							-0.470	0.0155	-30.32

Table 8.2. Poisson regression model with Grand Mean Model and demographic characteristics

	Model 4			Model 5			Model 6		
Scaled Deviance		83,280			82,582			82,561	
Degrees of Freedom		454,876			454,875			454,874	
	Estimate	Std. Error	T-Value	Estimate	Std. Error	T-Value	Estimate	Std. Error	T-Value
Intercept	-11.65	0.3447	-33.80	-10.89	0.3457	-31.50	-11.11	0.3491	-31.82
Age (25-34)	1.597	0.3905	4.09	1.537	0.3903	3.94	1.541	0.3903	3.95
Age (35-44)	3.884	0.3504	11.08	3.859	0.3502	11.02	3.863	0.3502	11.03
Age (45-54)	5.616	0.3459	16.24	5.579	0.3457	16.14	5.585	0.3457	16.16
Age (55-64)	6.969	0.3450	20.20	6.950	0.3448	20.16	6.953	0.3448	20.17
Age (65+)	7.333	0.3448	21.27	7.657	0.3448	22.21	7.636	0.3448	22.15
Sex (Female)	-0.473	0.0155	-30.52	-0.413	0.0156	-26.47	-0.419	0.0157	-26.69
Carstairs	0.085	0.0021	40.48	0.032	0.0030	10.67	0.033	0.0030	11.00
Log Smoking				0.677	0.0257	26.34	0.629	0.0278	22.63
Log Pop. Density							0.020	0.0046	4.35

Table 8.3. Poisson regression model with deprivation, smoking behaviour and population density

groups compared to the lowest age group (16 to 24) which was the base category. Lung cancer increased with each age group and the results were also more significant for the older age groups. The results were appropriate given that lung cancer was relatively rare in the younger age groups and that the number of cases among those under the age of 35 was close to zero. The sex variable also demonstrated that the incidence of lung cancer was lower for females compared to the males. These observations are in line with the discussion in chapters three and four.

8.4.3 Deprivation

Following age and sex, the variable that reduced the deviance most in a univariate model was the Carstairs Index of Deprivation. Model four in Table 8.3 shows that the deviance value of the model dropped to 83,280 when the Carstairs Index was added and that the variable was highly significant. The parameter estimate was positive which suggests that lung cancer rates were higher in more deprived areas, once age and sex were controlled for. This reflects the observations made in chapter four, which demonstrated a clear relationship between higher levels of deprivation and increased incidence of lung cancer.

There is a growing literature that suggests that residence in a disadvantaged area may have a negative effect on health that is independent of factors such as socio-economic status, income and education (e.g. Whitley *at al.*, 1999; Yen and Kaplan, 1999). Stead *et al.* (2001) consider there to be five factors that may explain this apparent effect. Firstly, deprived communities often have a more limited access to

and provision of facilities and services, which are necessary or beneficial to health. Secondly, these communities have poorer job prospects which directly affect physical and psychological health due to the stigmatisation of the neighbourhood which discourages employers from recruiting people from these areas. Thirdly, the higher levels of crime and violence increases the exposure to stress. Fourthly, there are fewer opportunities for social interaction and participation, which have been shown to have beneficial health effects. Finally, residence in a deprived area can result in feelings of exclusion, stigmatisation, segregation and abandonment, which can affect mental health. Although there is little research into how these specific factors influence the incidence of lung cancer, it seems likely that these circumstances create an environment that is conducive to the contraction of lung cancer. More specifically, local cultures in the most deprived areas such as the positive aspects of neighbourliness and attachment as well as factors such as coping with stress, feelings of exclusion and stigmatisation and the limited access to the triggers of smoking cessation have all been shown to influence smoking behaviour (Stead *et al.*, 2001). Furthermore, some of the deprived areas may have concentrations of people living in the most hazardous occupations, such as those exposed to asbestos (De Vos Irvine *et al.*, 1993), which may increase the incidence of lung cancer.

Although these factors have been considered in relation to other types of ill health, the direct effect of deprivation on lung cancer when the other key risk factors have been controlled for has not previously been examined. Model four suggests that there is an important relationship once age and sex are controlled for. However, at this stage it was not clear whether this observation was a reflection of the direct

influence on lung cancer or whether the Carstairs Index was acting as a proxy for a number of other risk factors. For example, areas with high levels of smoking and air pollution are likely to be more deprived than areas with low levels of smoking and air pollution and this may help to explain why the Carstairs Index was significant. If the Carstairs Index retains its significance when these factors are controlled for then this would suggest that living in a more deprived area independently increases lung cancer incidence.

8.4.4 Smoking

The effect of adding the log of the age and sex specific smoking probabilities to the model is shown in model five of Table 8.3. The deviance value of the model dropped to 82,582 and the variable was highly significant. It can be seen that having accounted for the age and sex structure of the population and deprivation, a higher probability of smoking increased the incidence of lung cancer as expected (Doll and Hill, 1950). Model five also shows that smoking was more significant than deprivation even though deprivation was more significant in a univariate model, as the smoking parameter had a T value that was more than twice the size of the deprivation parameter. The T value of the Carstairs parameter in model 5 had fallen to 25 percent of its value once smoking was added. Even so, it is interesting to note that deprivation had an independent effect.

It is worth remembering that the smoking estimates were calculated for 1991 to coincide with the lung cancer data rather than for an earlier date that would incorporate the lag effect between exposure and initiation of lung cancer. Earlier

chapters have suggested that it is often useful to use lagged variables in lung cancer studies instead of contemporary measures of risk factors, depending upon how the risk factor has changed over time. Although it would have been desirable to have produced estimates of smoking for an earlier date and then tested them within this model, spatially disaggregated smoking data are not available for an earlier date in Scotland. However, the 1991 estimates of smoking behaviour explained a significant amount of the variation in lung cancer incidence. Furthermore, as discussed in chapter five, the contemporary estimates of smoking are likely to reflect the historical patterns of smoking. Therefore, current smoking status is likely to be a reliable indicator of whether someone smoked in the past.

8.4.5 Population Density

Model six in Table 8.3 shows the effect of adding the log of population density to the model. The deviance value was reduced to 82,561 and the variable was just significant. The incidence of lung cancer increased with higher population densities, which shows that lung cancer was more common in densely populated areas than it was in sparsely populated areas. This suggests that there was an urban-rural effect that was not controlled for by the age structure of the population and the variations in smoking behaviour or deprivation. This was in line with the observations made in chapter four which consistently found that lung cancer was higher in urban areas of Scotland and that clusters of cases were located in the most urban areas, even when age and sex had been controlled for.

One explanation for this may be that the risk factors that are higher in urban areas, such as smoking, were inadequately controlled for in this model. An alternative explanation is that there are other features of urban areas that influence the incidence of lung cancer, particularly air pollution. In addition, a number of other factors have previously been identified that may, in combination, contribute to an urban effect (Haynes, 1988). These include passive smoking, particular types of employment that are common in urban areas, geographically biased mortality statistics and selective migration. Passive smoking may have a greater influence in urban areas than in rural areas because urban areas contain proportionally more smokers than rural populations (Matsukura *et al.*, 1984). Furthermore, people are more likely to be exposed to passive smoking in crowded, poorly ventilated public places in cities than they are in rural areas. There are also possible industrial explanations for the lung cancer urban effect. As was discussed in chapter two, a number of hazardous industries have been recognised as potential risk factors, particularly those industries that use materials such as asbestos, arsenic, chromium and chloromythyl. Occupational causes of lung cancer in Scotland are considered in subsequent sections.

Two other factors that are not connected with direct risk have also been suggested as helping to explain the urban effect on the incidence of lung cancer. Firstly, it has been suggested that there may be a geographical bias in mortality statistics, which has meant that cancers are more likely to be diagnosed and reported in urban areas than in rural areas (Bentham and Haynes, 1985). However, there is not any direct evidence of widespread under-diagnosis of lung cancer in rural Britain, and Scotland

is recognised as having some of the most accurate health data in the world (ISD, 2002).

Secondly, a stronger argument is that the urban effect may be created partially by selective migration (Haynes, 1988). Selective migration has been highlighted as one reason that may help to explain health differences between neighbourhoods, including between different areas on the urban-rural gradient (Verheij *et al.*, 1998; Brimblecombe *et al.*, 1999). For example, Brimblecombe *et al.*, (1999) used the 1991 British Household Panel Survey to investigate the extent to which selective migration influences current geographical variations in mortality. They found that local level selective migration since birth significantly altered the geographical pattern of mortality in Britain. Prior to the period of the lung cancer dataset (1988 to 1991), many rural communities grew at the expense of the urban populations, and studies have shown that adults who migrate relatively long distances tend to be healthier than the population that remains (Fox and Goldblatt, 1982). This trend would lead to an apparent improvement in health in rural areas but a worsening of health in urban areas. The effect may also be compounded by the net out-migration of the elderly and chronically sick to be closer to relatives or to health and service facilities (Bentham, 1988). Therefore, the effect of selective migration could obscure the geographical distribution of many diseases, especially those with a long latency period such as lung cancer (Haynes, 1988) and may help to explain why the population density variable was significant in the models of lung cancer. However, it is difficult to explicitly examine the effects of selective migration between urban and rural areas because the lung cancer data do not include a migration history and

other data on migrants do not tend to be available for OAs, nor do they identify individuals with lung cancer.

8.4.6 Air Pollution

The next most significant variables in a univariate model were the two estimates of air pollution levels. The estimates for 1971 and 1991 were added separately to examine whether it was more sensible to use contemporary measures of pollution (i.e. pollution levels that overlap temporally with the lung cancer data set in 1991) or estimates of pollution that attempt to incorporate the lag effect between exposure to the carcinogens and cancer initiation (i.e. pollution levels in 1971).

The results of adding the 1971 and 1991 measurements of pollution levels are shown in models seven and eight in Table 8.4. The table shows that the 1971 prediction was significant but that the 1991 prediction was not. The model that includes the 1971 prediction had a deviance value of 82,534 and the parameter estimate was positive, which demonstrates that the incidence of lung cancer was higher as pollution levels increased.

These models suggest that it is important to use 1971 estimates of pollution levels rather than contemporary measures that coincide temporally with the lung cancer cases. This is probably because the level of pollution was higher in 1971 due to the presence of heavy industry. Also, in 1991 the pollution estimates were relatively uniform across Scotland, with fewer areas of significant pollution as compared to 1971. This is different to other risk factors, such as smoking, where there is likely to

	Model 7			Model 8			Model 9			Model 10		
Scaled Deviance	82,534			82,559			82,490			82,497		
Degrees of Freedom	454,873			454,873			454,868			454,869		
	Estimate	Std. Error	T-Value	Estimate	Std. Error	T-Value	Estimate	Std. Error	T-Value	Estimate	Std. Error	T-Value
Intercept	-11.12	0.3491	-31.85	-11.12	0.3491	-31.85	-11.03	0.3495	-31.56	-11.06	0.3493	-31.66
Age (25-34)	1.540	0.3903	3.95	1.541	0.3903	3.95	1.538	0.3903	3.94	1.538	0.3903	3.94
Age (35-44)	3.864	0.3502	11.03	3.863	0.3502	11.03	3.862	0.3502	11.03	3.863	0.3502	11.03
Age (45-54)	5.585	0.3457	16.16	5.585	0.3457	16.16	5.583	0.3457	16.15	5.583	0.3457	16.15
Age (55-64)	6.953	0.3448	20.17	6.953	0.3448	20.17	6.952	0.3448	20.16	6.952	0.3448	20.16
Age (65+)	7.639	0.3448	22.15	7.636	0.3448	22.15	7.646	0.3448	22.18	7.645	0.3448	22.17
Sex (Female)	-0.419	0.0157	-26.69	-0.419	0.0157	-26.69	-0.417	0.0157	-26.56	-0.417	0.0157	-26.56
Carstairs	0.030	0.0030	10.00	0.032	0.0030	10.67	0.027	0.0030	9.00	0.027	0.0030	9.00
Log Smoking	0.635	0.0277	22.92	0.628	0.0277	22.67	0.650	0.0278	23.38	0.648	0.0278	23.31
Log Pop. Density	0.017	0.0046	3.70	0.020	0.0046	4.35	0.013	0.0047	2.77	0.015	0.0046	3.26
Pollution 71	0.001	0.0002	5.00				0.001	0.0002	5.00	0.001	0.0002	5.00
Pollution 91				0.001	0.0010	1.00						
Rad A (Low-Mod)							-0.027	0.0162	-1.67			
Rad A (Mod)							-0.093	0.0274	-3.39			
Rad A (High)							-0.216	0.0381	-5.67			
Rad A (Very High)							-0.121	0.3022	-0.40			
Rad A (Unclass)							-0.121	0.0454	-2.67			
Rad B (Low-Mod)										-0.018	0.0159	-1.13
Rad B (Mod)										-0.083	0.0273	-3.04
Rad B (High)										-0.205	0.0380	-5.39
Rad B (Very High)										-0.102	0.3021	-0.34

Table 8.4. Poisson Regression Models to explore the effects of pollution and radon.

have been a relatively uniform decrease across Scotland between 1971 and 1991. Air pollution levels have not decreased uniformly and this was reflected in the low correlation coefficient between the 1971 and 1991 predictions (section 6.5).

Although it appears to be more appropriate to use 1971 estimates, the use of this lagged variable is problematical because it does not account for any migration by the lung cancer patients in the period 1971 to 1991. If a lung cancer patient has migrated during this period then it is not appropriate to associate the 1971 level of pollution with them. If, however, they have remained within the same PPS between 1971 and 1991, then the 1971 estimate of pollution will provide a more appropriate measure of pollution during the crucial period of exposure. These results suggests that despite the methodological limitations of using 1991 SWS data to predict 1971 pollution levels (see chapter six), the 1971 value is a more appropriate measure of air pollution exposure. This highlights the importance of incorporating a lag effect into the air pollution estimates.

It is interesting to note that the log of population density retained its significance in model seven and this suggests that there was still an urban effect that was not explained by the larger populations, the age structure, higher smoking probabilities and higher pollution levels in urban areas (although its significance was lower having controlled for air pollution). Therefore, there were either other urban risk factors that were not controlled for in the model or there was an explicit urban effect that influenced lung cancer incidence.

8.4.7 Radon Levels

The next set of risk factors tested were the two categorical variables that provide estimates of radon potential in OAs in Scotland. As chapter seven demonstrated, radon levels are largely a response to a combination of the types of solid geology and drift at a particular location. These are factors that have remained relatively stable through time and hence contemporary measures of radon are suitable for this analysis. The first variable directly used the radon potential values calculated by the British Geological Survey for geological units in Scotland. Chapter seven showed, using data supplied by the Leukaemia Research Fund, that the missing category in this variable had a relatively high average radon level and therefore potentially represents areas of particular interest in an analysis of the relationship between radon and lung cancer. Therefore, OAs in this category were not excluded from the analysis because as it was a categorical variable, including this category did not interfere with the interpretation of the other results. The second estimate used the same categorisation but recoded the unclassified values into one of the other categories using a knowledge of the local geology (see chapter seven).

Models nine and ten in Table 8.4 show the results of adding the two radon variables. Each category was compared to the base category, which was the lowest radon potential class. When the variables were added to the model not all of the categories were significant and when they were significant the parameter estimates were negative. This indicates that lung cancer was lower in these areas compared to the base category, controlling for other variables. The very high radon potential estimate was not significant in either model but the high category was significantly negative.

There are a number of reasons why the radon categories were either insignificant or negatively related to lung cancer. The simplest explanation is that radon levels are not an important factor in controlling lung cancer incidence in Scotland. The insignificant results are supported by many studies that have found no link between radon and lung cancer in other areas (e.g. Blot *et al.*, 1990; Etherington *et al.*, 1996). On the other hand, it may be that due to the paucity of radon measurements in Scotland, the classification scheme is too generalised and cannot account for local variations that exist due to small-scale geological differences. The classification scheme used to represent the two variables was based on the 1:625,000 geological map of Scotland and this may be insufficient to capture these important local differences. Furthermore, chapter seven explained that large variations in radon readings can even occur between adjacent houses due to small-scale differences such as the degree of ventilation. Clearly, the localised variations were impossible to measure in the models due to these complex interaction of factors that affect the radon reading. The radon potential maps only represented the potential for having high levels of radon rather than the likely radon levels. Therefore, it is possible that there were high radon readings even in the low radon potential classes. Likewise, it is possible that there were low radon readings even in the higher radon potential classes. Both of these circumstances will have weakened any systematic effect. As chapter seven has demonstrated, high radon levels were found exclusively in the rural parts of Scotland where smoking and air pollution were relatively low. It is therefore useful to explore the effect of radon in rural locations alone and this is examined below.

In order to make a more confident assessment as to which of these factors is the more likely explanation, it would be necessary to undertake a comprehensive program of radon measurement collection. This would allow for a more accurate set of models to be developed that more satisfactorily assign the unclassified class and better reflect more localised variations.

8.4.8 1971 Coal Mining

The effect of adding the estimate of mining activity in 1971 is shown in model 11 (Table 8.5). The parameter estimate was negative and also significant which demonstrates that lung cancer rates tend to be lower in the coal mining communities, once other factors were controlled for. This was perhaps surprising because other studies have demonstrated that exposure to coal mining can have a strong detrimental effect on health outcomes (Boyle *et al.*, 1999; Flowerdew and Geddes, 1999; Frinckelman *et al.*, 2002). However, as chapter two demonstrated, very few studies have provided evidence of a positive relationship between lung cancer and coal mining.

It is possible that the estimate of mining activity in 1971 used here is unsatisfactory. First, the estimate was based on a census-based occupational definition, which included all people employed in the mining industry. Thus, both people who were employed in the non-manual aspects of coal mining and those employed in other types of mining were included in the definition. A second possibility is that the out-migration of former coal miners from the former coal mining communities due to economic restructuring that took place between 1971 and 1991 may have been so

	Model 11			Model 12			Model 13			Model 14		
Scaled Deviance	82,521			82,510			82,513			82,454		
Degrees of Freedom	454,872			454,871			454,871			454,869		
	Estimate	Std. Error	T-Value	Estimate	Std. Error	T-Value	Estimate	Std. Error	T-Value	Estimate	Std. Error	T-Value
Intercept	-11.09	0.3491	-31.77	-11.09	0.3491	-31.77	-11.10	0.3491	-31.80	-10.970	0.3496	-31.38
Age (25-34)	1.539	0.3903	3.94	1.539	0.3903	3.94	1.538	0.3903	3.94	1.539	0.3903	3.94
Age (35-44)	3.864	0.3502	11.03	3.864	0.3502	11.03	3.863	0.3502	11.03	3.862	0.3502	11.03
Age (45-54)	5.585	0.3457	16.16	5.585	0.3457	16.16	5.584	0.3457	16.15	5.583	0.3457	16.15
Age (55-64)	6.953	0.3448	20.17	6.952	0.3448	20.16	6.952	0.3448	20.16	6.953	0.3448	20.17
Age (65+)	7.641	0.3448	22.16	7.640	0.3448	22.16	7.647	0.3448	22.18	7.645	0.3448	22.17
Sex (Female)	-0.419	0.0157	-26.69	-0.419	0.0157	-26.69	-0.418	0.0157	-26.62	-0.736	0.0478	-15.40
Carstairs	0.029	0.0030	9.67	0.029	0.0030	9.67	0.029	0.0030	9.67	0.028	0.0030	9.34
Log Smoking	0.641	0.0278	23.06	0.640	0.0278	23.02	0.652	0.0280	23.29	0.741	0.0309	24.00
Log Pop. Density	0.017	0.0046	3.70	0.017	0.0046	3.70	0.016	0.0046	3.48	0.015	0.0046	3.31
Pollution 71	0.001	0.0002	5.00	0.001	0.0002	5.00	0.001	0.0002	5.00	0.001	0.0002	5.57
Mining 71	-0.006	0.0018	-3.33	-0.006	0.0018	-3.33	-0.006	0.0018	-3.33	-0.006	0.0018	-3.35
Energy				-0.003	0.0010	-3.00				-0.003	0.0010	-2.99
Other Service							0.001	0.0003	3.33	0.001	0.0003	2.61
Log Smok'g.Sex (F)										-0.235	0.0333	-7.08

Table 8.5. Exploring occupation type and developing a parsimonious model

high that the effect of coal mining has been dispersed. However, this seems unlikely given that out-migration from deprived areas is often relatively low (Kitching, 1990), and Hollywood (2002) used the UK Sample of Anonymised Records and the ONS Longitudinal Study to demonstrate that migration has not been a significant response to labour market adjustments among former miners. Furthermore, the study found that miners were less migratory than other groups in the population.

8.4.9 Quarrying

The estimates of exposure to different types of quarries for OAs within one kilometre and two kilometres were each added to model seven to examine the effect of quarrying activity on lung cancer incidence in Scotland. The resulting models found that none of the quarrying variables were significant in a univariate model or once other variables were controlled for. The parameter estimates were usually negative within all univariate and multivariate models which demonstrates that lung cancer was lower in OAs close to quarries. Importantly, none of the quartz quarries (sandstone and granite) or combination of quartz quarries were significant in explaining variations in lung cancer incidence.

The results suggest that living in close proximity of a quarry is not an important cause of lung cancer in Scotland. However, like radon, quarries tend to be located in more rural areas and it is important to consider the effect that quarrying has upon the rural population and this is explored in a subsequent section.

8.4.10 Employment Type

The ten variables that were used to represent the percentage employed in the different types of industry were each added to the model in turn (models 12 and 13). Two of the variables were significant once the age, sex, deprivation, smoking, population density, pollution and mining variables had been controlled for. These were the percentage of economically active residents who were employed in the energy and water industries, which was negatively related to lung cancer, and the percentage of economically active residents who were employed in the other service industries (service industries other than the ones specifically identified such as banking) which was positively related to lung cancer.

However, many of the industries that were considered here are gender biased because the workforce is dominated by males or by females (e.g. the mining industry was dominated by males). Therefore, it is important to also use male and female specific employment rates in models that consider each sex separately and this is addressed in subsequent sections.

8.4.11 Interactions

The final step was to test the interactions between the variables that were significant in models 12 and 13. The interaction between the two individual-level variables (age and sex) and all the other significant variables were tested. Also, the interaction between other combinations of variables were tested if there was a theoretical reason for doing so. For example, chapter three highlighted that the effect of deprivation

was stronger among females with lung cancer than it was for males and hence the interaction between deprivation and sex was tested. Chapter two highlighted that there was a multiplicative effect between radon and smoking (Pershagen *et al.*, 1994) and although radon was not significant in models 12 or 13, the multiplicative effect was examined by testing the interaction of smoking with all other variables in the model. Finally, the interaction between population density and the occupational variables were tested to examine whether particular types of employment are only important in explaining lung cancer in more urban or rural areas.

Model 14 demonstrates that the only significant interaction was between sex and the log of the probability of smoking. This interaction demonstrated that the effect of age and sex-specific smoking probabilities on lung cancer incidence was less for women than it was for men. The importance of exploring the relationship between the risk factors and gender is explored in more depth in the next section but it was interesting that the interaction between age and sex was not significant. This suggests that the effect of age was consistent for both males and females, which is in line with the graphs of the age distribution of lung cancer incidence in Scotland discussed in chapter four.

8.4.12 Developing a Parsimonious Model

The results from the previous sections have demonstrated that a number of the variables that were used to model lung cancer incidence had a significant influence upon the incidence rate. Model 14 in Table 8.5 was the optimum model that contained all of the significant variables and the significant interaction and which

had the lowest deviance value, given the number of degrees of freedom. The age and sex variables were especially significant in explaining the variation in lung cancer incidence, which demonstrates that it was particularly important to control for the demographic makeup of the area.

It was interesting to note that the Carstairs Index retained its significance in the parsimonious model, with lung cancer increasing with higher levels of deprivation. It was significant even when other key variables such as smoking, air pollution, population density and employment type were controlled for. This could be explained by the factors that were discussed previously which considered the link between deprivation and ill health, such as the provision of health care and services, job prospects and the stresses of crime and violence. Alternatively, it is possible that the Carstairs variable is acting as a proxy for other variables that have not been included in the models such as other lifestyle factors including diet and the amount of exercise taken by residents in more deprived urban areas. Similarly, the population density variable stayed significant even when other factors such as air pollution were controlled for. This adds weight to the previous discussion that the variable may be acting as a proxy for specific risks that are concentrated in urban areas such as passive smoking or particular occupational types, or it may reflect the confounding effects of selective migration. These alternative explanations are considered further in the analysis of the residual values.

The scaled deviance of the optimum model dropped from 113,522 in the Grand Mean Model to 82,454 in the optimum model (model 14). It is useful to examine the goodness of fit of Poisson models by comparing the deviance values of the null

model and optimum model to calculate the pseudo-R^2 value (the proportion of null deviance explained) (Lovett *et al.*, 1985). In model 14 the pseudo-R^2 was 0.274 which suggests that a large proportion of the variation in lung cancer remains unexplained by the variables in the model. However, although an examination of the amount of variance explained by the optimum model is a useful indicator of model fit, as has already been noted, the deviance value of the models is unusually low compared to the number of degrees of freedom. This suggests that the dataset is underdispersed which makes the interpretation of the deviance values and subsequent calculation of the pseudo-R^2 problematical (Boyle and Flowerdew, 1993).

8.5 Models for Age and Sex Subgroups

It has been demonstrated in chapters three and four that there were important differences in the incidence of lung cancer between particular age and sex groups. Most noticeably, there was a substantial difference in the rates of lung cancer between younger and older age groups and also between males and females. Payne (2001) explicitly called for research into lung cancer that 'reflects the complexity of biology and gender as influences on the risks of lung cancer'. For example, the construction industry is an industry traditionally dominated by males and so it was likely that the percentage employed in construction may be related to male lung cancer rates in a different way than it was to female lung cancer rates. Chapter five pointed out that the relationship between male and female smoking rates in Scotland has changed significantly over the past 30 years as the two rates of smoking have become more similar. In this section separate models are derived for different age

and sex groups using age and sex specific variables so that the effect of the risk factors may be better understood.

8.5.1 Sex

Models 15 to 18 in Table 8.6 show the results of examining the incidence of lung cancer for males and females separately. There were 12,073 lung cancer cases in the models of male incidence and 6,559 cases for female incidence. Although the models were similar, there were some differences between them, which demonstrated the gender differences in the factors that influence lung cancer. Model 15 shows the null model and Model 16 the parsimonious model for males. Age, the Carstairs Index, the log of smoking probability, the log of population density, 1971 pollution, 1971 mining and the percentage of males employed in energy and water industries were all significant. The parameter estimates were consistent with those in the parsimonious model that was developed for all cases in model 14.

The parsimonious model for females is shown in model 18. The same variables were significant as those for men in model 16 except that the percentage employed in the mining industry in 1971 and the sex-specific employment variables were not significant. This suggests that, despite some small differences in the parameters, the effects of age, deprivation, smoking, population density and pollution were consistent for both sexes. This suggests that the type of industry that females were employed in was not significant in explaining female lung cancer in Scotland. It was interesting to note that neither of the radon variables were significant because, as discussed in chapter two, previous work in Norway found that high levels of radon

	Model 15 (males)			Model 16 (males)			Model 17 (females)			Model 18 (females)		
Scaled Deviance	67,399			48,893			44,640			33,497		
Degrees of Freedom	227,441			227,429			227,441			227,432		
	Estimate	Std. Error	T-Value	Estimate	Std. Error	T-Value	Estimate	Std. Error	T-Value	Estimate	Std. Error	T-Value
Intercept	-5.197	0.0092	-571.10	-11.200	0.4928	-22.73	-5.778	0.0124	-465.97	-11.44	0.4989	-22.93
Age (25-34)				1.875	0.5363	3.50				1.049	0.5841	1.80
Age (35-44)				3.976	0.4951	8.03				3.743	0.4978	7.52
Age (45-54)				5.819	0.4892	11.89				5.284	0.4909	10.76
Age (55-64)				7.254	0.4882	14.86				6.542	0.4893	13.37
Age (65+)				7.830	0.4882	16.04				7.425	0.4894	15.17
Carstairs				0.025	0.0037	6.58				0.037	0.0051	7.25
Log Smoking				0.700	0.0349	20.09				0.544	0.0460	11.83
Log Pop. Density				0.013	0.0055	2.42				0.023	0.0081	2.84
Pollution 71				0.001	0.0003	5.29				0.001	0.0004	2.50
Mining 71				-0.006	0.0022	-2.73						
Male Energy				-0.002	0.0008	-2.42						

Table 8.6. Poisson regression models that consider males and females separately.

were only significant in explaining female lung cancer incidence (Magnus *et al.*, 1994).

Smoking was significant in both models but it was more significant in explaining male lung cancer than it was for female lung cancer, which is consistent with the interaction between the log of smoking and sex in model 14. This confirms that smoking behaviour is more important in explaining lung cancer incidence for males than it is for females. This is probably because male rates of smoking remain higher than female rates despite the reduction in the difference between male and female smoking rates over the past 30 years. Furthermore, as chapter five has shown, this has been consistent for all age groups. Therefore, given the long history of higher rates of smoking among males, it is unsurprising that the smoking variable was more significant in explaining male lung cancer incidence.

8.5.2 Age

Two sets of models were developed to examine the importance of age in understanding the risk factors associated with lung cancer. The first set of models examined lung cancer cases among the youngest age groups (16 to 24, 25 to 34, 35 to 44 and 45 to 54). These can be considered to be the premature cases of lung cancer because they occurred earlier in life than is usual. A total of only 1,740 lung cancer patients of the total of 18,633 cases during the period 1988 to 1991 were aged 16 to 54, which represented 9.3 percent of the cases. The lung cancer cases of patients in the age groups 55 to 64 and over 65 were modelled in a separate set of models. A total of 16,892 patients were aged over 54 at the time of diagnosis, which

represented 90.7 percent of cases. Models 19 to 22 in Tables 8.7 and 8.8 show the results of modelling the younger and older age groups.

Younger Age Groups

Models 19 and 20 in Table 8.7 represent the cases of lung cancer among younger people aged 16 to 54. Model 20 shows that age, sex, the Carstairs Index, the log of smoking probability, the log of population density, the pollution levels in 1971, the first measure of radon potential, the percentage of economically active residents who were employed in the energy and water industries, the percentage of economically active residents who were employed in other service industries and the interaction between the log of smoking probability and sex were significant. Most of these variables were consistent with the previous models that analysed all cases of lung cancer. In addition, the percentage of economically active residents who were employed in the energy and water industries, the percentage of economically active residents who were employed in other service industries and the interaction between the log of smoking probability and sex were significant. The parameter estimates were consistent with the optimum model developed for all cases of lung cancer so this supports the suggestion that the effect of living in areas with high levels of smoking was less for females than it was for males.

Of particular note in model 20 was the effect of the first measure of radon. The results suggested that lung cancer was less common in the low-moderate, moderate and high categories compared to the base category (low). However, the incidence of

	Model 19			Model 20		
Scaled Deviance	16,238			12,976		
Degrees of Freedom	303,255			303,240		
	Estimate	Std. Error	T-Value	Estimate	Std. Error	T-Value
Intercept	-7.502	0.0260	-288.54	-10.490	0.3640	-28.82
Age (25-34)				1.499	0.3895	3.85
Age (35-44)				3.843	0.3493	11.00
Age (45-54)				5.561	0.3447	16.13
Sex (Female)				-0.844	0.1370	-6.16
Carstairs				0.035	0.0101	3.46
Log Smoking				1.171	0.1067	10.97
Pollution 71				0.002	0.0006	3.25
Rad A (Low-Mod)				-0.040	0.0565	-0.71
Rad A (Mod)				-0.183	0.1049	-1.75
Rad A (High)				-0.146	0.1319	-1.11
Rad A (Very High)				1.560	0.5791	2.69
Rad A (Unclass)				-0.147	0.1645	-0.89
Energy				-0.013	0.0043	-2.92
Other Service				0.002	0.0010	2.40
Log Smok'g.Sex (F)				-0.420	0.1341	-3.13

Table 8.7. Poisson regression models including just the younger age groups (16-24, 25-34, 35-44 and 45-54).

lung cancer was greater in the very high radon potential category compared to the base category and this parameter was significant. This is an important finding because it suggests that exposure to very high levels of radon at a young age can lead to the premature development of lung cancer. It may only be possible to make this observation due to the lower net migration by the younger age groups between exposure and diagnosis, which meant that the lung cancer patients were more likely to be diagnosed in the area in which they were exposed to radon. Obviously, this result does not suggest that radon is not a cause of lung cancer in older age groups. Rather, exposure to high levels of radon could raise the risk of lung cancer among the wider population but because of the migration history of the older population and because the effects of smoking are more important in the older age groups, the effect may be obscured.

The log of population density variable was not significant, which demonstrates that there was not a significant independent urban effect on the incidence of lung cancer for the younger age groups. This is consistent with the earlier discussion of the urban effect on lung cancer, as it was suggested that one reason that may explain this effect is the selective migration of healthier people from large urban areas to more rural areas during the latency period. However, this is less likely to effect younger lung cancer patients as counterurbanisation has been shown to be lower for this group of people (Champion, 1989) which may help to explain why population density was not significant in this model.

None of the quarrying variables were found to be significant for the younger cases of lung cancer. This suggests that quarrying is not an important cause of lung cancer and does not even have an effect when the cohort are young and the levels of migration are less, which provides evidence for there not being an important effect from quarrying activity.

Older Age Groups

Model 22 in Table 8.8 shows the optimum model for the older age groups (55 to 64 and over 65). Age, sex, the Carstairs Index, the log of smoking probability, and the pollution levels in 1971 were again significant and the parameter estimates were consistent with model 14, which represented all cases of lung cancer. In addition, the log of population density variable was significant in this model when it was not for the younger age groups. This suggests that living in an urban area has a significant effect on the older age groups which is consistent with the discussion

above, as one of the potential reasons for the urban effect is the selective migration of healthier people from the most urban areas to more rural areas. This effect is likely to be more important in explaining lung cancer in the older age groups because net-migration is likely to be greater (Champion, 1989). Furthermore, other factors discussed earlier such as passive smoking and occupation type, may help account for the urban effect. These are factors that may only be important in causing lung cancer following long-term exposure to the risk factor and hence may only effect the older age groups.

	Model 21			Model 22		
Scaled Deviance	72,733			69,346		
Degrees of Freedom	151,627			151,617		
	Estimate	Std. Error	T-Value	Estimate	Std. Error	T-Value
Intercept	-4.538	0.007673	-591.42	-4.010	0.0646	-62.07
Age (65+)				0.533	0.0263	20.27
Sex (Female)				-0.700	0.0320	-21.88
Carstairs				0.028	0.0032	8.75
Log Smoking				0.617	0.0291	21.20
Log Pop. Density				0.013	0.0052	2.50
Pollution 71				0.001	0.0002	5.00
Agriculture				-0.002	0.0010	-2.00
Energy				-0.002	0.0010	-2.00
Mining 71				-0.006	0.0019	-3.16
Age (65+).Sex (F)				0.387	0.0369	10.49

Table 8.8. Poisson regression models including just the older age groups (55-64 and over 65).

The variables representing the percentage employed in mining in 1971, the percentage of economically active residents who were employed in the agriculture industry and the percentage of economically active residents who were employed in the energy and water industry were all significant and the parameter estimates were negative. In other words, after controlling for other variables in the model, the incidence of lung cancer among older people was lower in areas that had a high proportion of the population employed in the agricultural and fishing, the energy and

water industries in 1991 and the mining industry in 1971. The parameters were consistent with other models that have been discussed, such as models 14 and 16, although the agricultural variable was not significant before. Finally, the interaction between age and sex was significant and the parameter estimate was positive which shows that the effect of age is greater for older women than it is for older men. This could be a reflection of smoking patterns over the past 20 or 30 years. As chapter five has demonstrated, smoking rates have fallen during this period but the rates have traditionally been higher for males than for females. If smoking was taken up later by women then this would explain why the age effect is more important for females. However, this was not reflected in the analysis in chapter five of the General Household Survey data in 1973 and 1999, which demonstrated similar rates of smoking among males and females in the younger age groups. An alternative explanation is that if smoking is taken up at similar ages among men and women then the lag period between exposure to the risk factor and diagnosis could be longer for females than it is for males, although this has not been examined in previous studies.

Although this section has highlighted some interesting differences between males and females and between younger and older age groups, the results are broadly similar to those for the models of all cases of lung cancer. The models that considered males and females separately (Table 8.6) demonstrated that the significant variables were almost the same, except that specific occupations were significant in explaining male lung cancer incidence but not for females. Similarly, it was demonstrated that the variables in the models for the younger and older age groups were similar, except that radon was significant in explaining cancer in the

younger age groups but not for the older age groups (Tables 8.7 and 8.8). Furthermore, the population density variable was significant in explaining lung cancer incidence in the older age groups but not the younger age groups. However, despite these subtle differences, the age, sex, smoking, Carstairs Index and 1971 pollution variables were consistently the most significant variables in the models of all cases of lung cancer and in the models for particular age and sex groups.

8.6 Rural Areas

A number of studies have pointed to important differences in the rates of mortality and morbidity between urban and rural areas (e.g. Senior *et al.*, 2000) and, as chapter two has shown, other studies have demonstrated higher rates of lung cancer in urban areas compared to rural areas (Haynes, 1988; Barbone *et al.*, 1995). This has been reflected in this study which has shown higher age-standardised rates of lung cancer in urban areas. Chapter four demonstrated that the ASR of lung cancer increased with population density, which is largely a response to differential smoking rates but also to differences in other risk factors such as air pollution, deprivation and population density. However, model 14 demonstrated that the population density variable was significant in explaining lung cancer incidence in Scotland even when other risk factors such as smoking, air pollution, occupation type and deprivation were controlled for. There are potential risk factors that are likely to be more important in rural areas, particularly radon and quarrying activity because radon levels and the number of quarries are higher in rural areas but their effects may have been masked by higher rates of lung cancer in urban areas. Therefore, it is useful to

explore the effect that the potential risk factors have upon lung cancer in rural areas alone.

Rural areas can be defined in many ways, including measuring the settlement size, population density, nearest neighbour or access to services. Of these definitions, population density is probably the most widely used measure of rurality (Martin *et al.*, 2000). Although there is no consensus on the population density threshold to use, studies in the UK have tended to use a threshold of between 100 and 1,000 persons per km^2 (Shucksmith, 1990; Payne *et al.*, 1996). In this analysis, rural areas were defined as those OAs which had a population density of 200 people per km^2 or less and the OAs that were above this threshold were excluded from the analysis. Figure 8.1 shows the distribution of the 3,737 OAs that had a population density of less than 200 people per km^2. This represents 9.9 percent of the total number of OAs, but 98.1 percent of the total area of Scotland. A modelling procedure similar to the one used in the previous section was adopted to identify the variables that were significant in explaining lung cancer incidence in rural areas.

The Grand Mean Model for the rural areas is shown in model 23 of Table 8.9 and the scaled deviance was 7,967. Model 24 in Table 8.9 shows the optimum model for lung cancer incidence in rural areas and the model had a scaled deviance of 6,182. The age and sex variables show a relationship with lung cancer incidence that is consistent with the models for all cases of lung cancer (model 14), with lung cancer being higher in the older age groups compared to the youngest age group and higher for males than females. The sex parameter estimate was much larger in model 24 (-1.589) than it was in model 14 (-0.736), which suggests that lung cancer is relatively

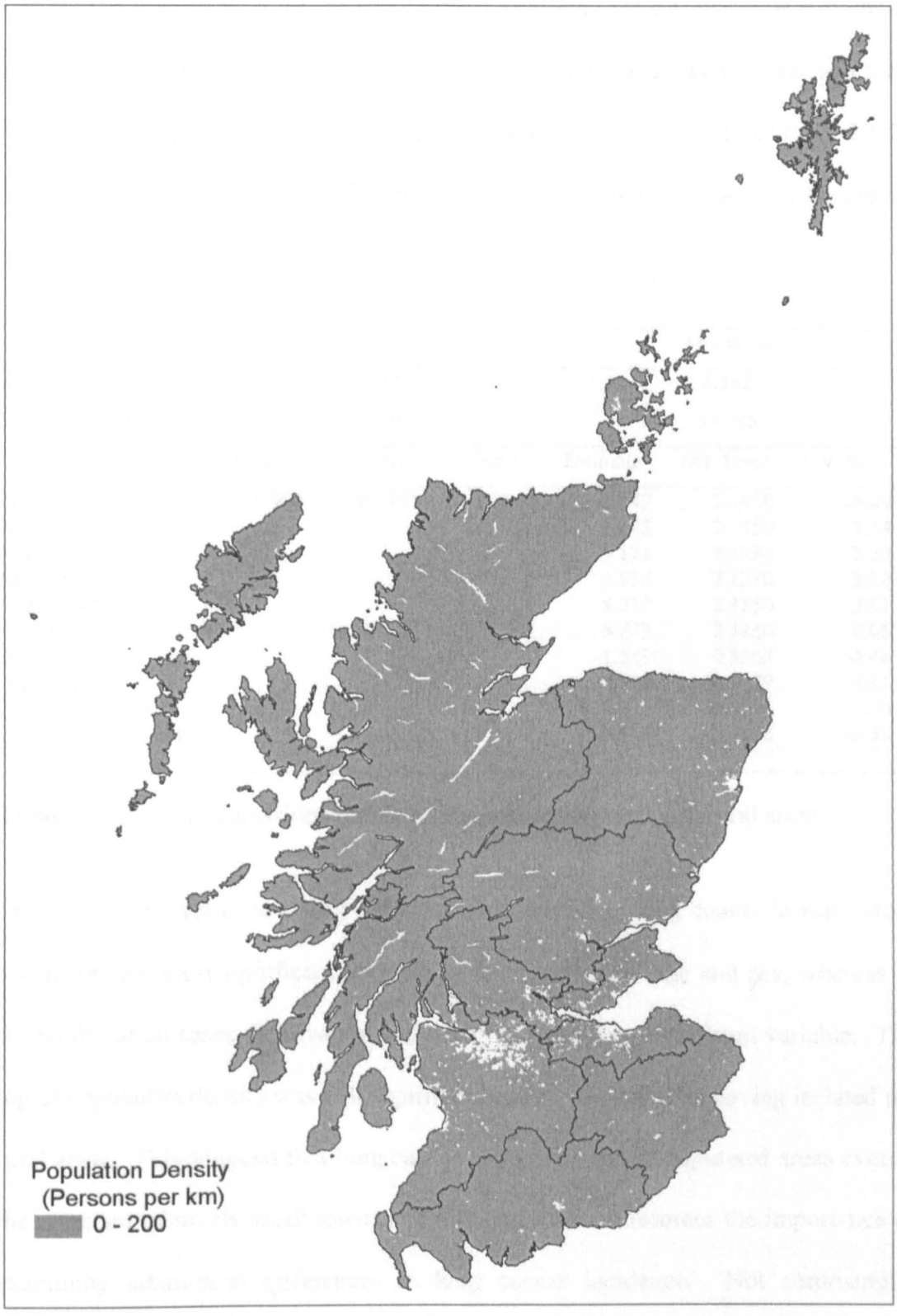


Figure 8.1. Output-areas in Scotland with a population density of 200 persons km^2 or less.

rare among women who live in rural areas. This may reflect historical patterns of smoking behaviour whereby female smoking rates are consistently lower than male rates but this difference is even greater in rural areas (Scottish Executive, 2002). Therefore, lung cancer rates are lower for females compared to males in all areas but especially so in rural areas.

	Model 23			Model 24		
Scaled Deviance	7,967			6,182		
Degrees of Freedom	44,795			44,786		
	Estimate	Std. Error	T-Value	Estimate	Std. Error	T-Value
Intercept	-5.900	0.0295	-200.00	-12.47	2.1930	-5.69
Age (25-34)				3.073	2.2950	1.34
Age (35-44)				5.121	2.1980	2.33
Age (45-54)				6.835	2.1870	3.13
Age (55-64)				8.317	2.1850	3.81
Age (65+)				8.873	2.1850	4.06
Sex (Female)				-1.589	0.3563	-4.46
Log Smoking				0.719	0.1329	5.41
Log Pop. Density				0.070	0.0206	3.40
Log Smok'g.Sex (F)				-0.679	0.1938	-3.50

Table 8.9. Poisson regression models of lung cancer incidence in rural areas

The estimated level of smoking was positively related to lung cancer in rural areas and it was the most significant variable in the model, after age and sex, whereas in the model of all cases, deprivation had been the next most significant variable. The log of population density was still significant and positive despite having isolated the rural areas. This suggests that lung cancer is higher in more populated areas even if the areas are relatively small towns and villages, which reinforces the importance of examining urban-rural differences in lung cancer incidence. Not surprisingly, neither of the pollution estimates were significant, probably because the predicted level of pollution is uniformly low throughout most of rural Scotland. The interaction between the log of smoking and sex was also significant, which was consistent with the parsimonious model of all cases of lung cancer (model 14).

The Carstairs Index was not significant once age and sex had been controlled for, which is a little surprising given that the variable had been significant in the models of all cases of lung cancer as well as the sex-specific and age-specific models. This may be because deprivation is not an important factor in explaining lung cancer incidence in rural areas. Alternatively, it has been suggested that the Carstairs Index is an inappropriate measure of deprivation in rural areas because the components of deprivation in rural areas are different from those in urban areas (Diamond *et al.*, 1999). For example, one of the components of the Carstairs Index is non-ownership of a car, which may be unsuitable as a measure of rural deprivation because a car is considered a necessity in many rural areas in the face of costly, or non-existent, public transport services. This results in higher than average levels of car ownership, often at the expense of low income households not being able to afford other types of household goods (Higgs and White, 2000). By ignoring accessibility to and provision of services, an important facet of deprivation in rural areas is neglected (Martin *et al.*, 2000). Consequently, it is difficult to understand the effects of deprivation on rural cases of lung cancer without developing a more appropriate measure of rural deprivation.

It was particularly interesting to examine the effects of radon and quarrying activity in rural areas where radon levels and exposure to quarrying activity tend to be higher than in urban areas. However, neither of the estimates of radon potential were significant in explaining rural lung cancer incidence, which suggests that radon is not an important cause of lung cancer in rural Scotland. Alternatively, and as discussed

above, it is possible that the estimates of radon are not sufficiently precise to isolate what is likely to be a small and geographically concentrated effect.

The results of incorporating each of the estimates of exposure to different types of quarrying activity for OAs within one kilometre and two kilometres of a quarry found that none of the variables were significant. This suggests that there is no evidence of a quarrying effect in rural areas, which supports the evidence from the models that incorporated all cases of lung cancer. If there is an effect then it may require an individual level study to confirm it, possibly because the numbers that were exposed to quarrying were so small.

None of the employment variables were significant in explaining lung cancer patterns in rural Scotland. This includes the percentage employed in coal mining in 1971, which was not particularly surprising given that coal mining tends to be concentrated in more industrial areas. These observations suggest that employment type is not an important influence on lung cancer incidence in rural areas and that the patterns are instead explained by the demographic characteristics of the area, smoking behaviour and population density. Therefore, there are a limited number of risk factors that cause lung cancer in rural areas.

Although there are some differences between the variables in the model of rural lung cancer and the variables in the optimum model of all cases (model 14), the models are broadly the same. It has been demonstrated that some of the variables in the model of rural lung cancer cases are not significant, such as the occupational

variables and the 1971 pollution variable, but the key explanatory variables such as age, sex, smoking and Carstairs are consistent in both models.

8.7 Lung Cancer Type

Most ecological studies have, in the past, treated lung cancer as a single disease. However, chapter two has demonstrated that there are a number of different forms of lung cancer. Although the causes of each type of lung cancer are similar, there is evidence to suggest that the risk factors may behave in subtly different ways for different forms of the disease. In this section, the different types of lung cancer are considered separately to examine whether the causal factors do affect each type of lung cancer differently.

In the data supplied by the ISD on the cases of lung cancer in Scotland, each lung cancer case was categorised according to the morphology of the cancer, which identifies the lung cancer type. As noted in chapter two, the most common clinical breakdown of lung cancer is into two main histological types: small cell lung cancer and non-small cell lung cancer. The two types not only differ in terms of the cell type but also in terms of the prognosis, as small cell lung cancer progresses very rapidly and is largely inoperable whilst non-small cell lung cancer can respond to radiotherapy (Simmonds, 1999). Therefore, small cell lung cancer and then non-small cell lung cancer were modelled separately. It should be noted that 7,754 of the cases did not have a specified type of lung cancer and so these were excluded from the analysis. However, there was no evidence to suggest that there was any particular bias in these cases towards either small cell or non-small cell lung cancer.

8.7.1 Small Cell Modelling Results

The count of cases of small cell and non-small cell lung cancer were calculated for each age-sex group in each OA. The models for all cases of small cell lung cancer are shown in Table 8.10. A total of 1,978 cases were classified as small cell lung cancer in Scotland during the period 1988 to 1991. Model 26 shows that there was much similarity between the parameters in this model and the parameters in many of the previous models. The age, sex, Carstairs Index, smoking probability and 1971 pollution variables were all significant and the parameter estimates were also similar to previous models. The model shows that smoking was the most significant variable in explaining small cell lung cancer.

	Model 25			Model 26		
Scaled Deviance	21,086			17,995		
Degrees of Freedom	454,883			454,871		
	Estimate	Std. Error	T-Value	Estimate	Std. Error	T-Value
Intercept	-7.667	0.02236	-342.89	-12.300	0.8328	-14.77
Age (25-34)				0.401	1.0600	0.38
Age (35-44)				3.516	0.8420	4.18
Age (45-54)				5.241	0.8295	6.32
Age (55-64)				6.518	0.8272	7.88
Age (65+)				6.866	0.8275	8.30
Sex (Female)				-0.538	0.1319	-4.08
Carstairs				0.030	0.0089	3.36
Log Smoking				0.873	0.0884	9.88
Pollution 71				0.002	0.0004	4.22
Manufacture				-0.005	0.0016	-2.98
Mining 71				-0.012	0.0059	-2.13
Log Smok'g.Sex (F)				-0.257	0.0963	-2.67

Table 8.10. Poisson regression modelling of all cases of small cell lung cancer

Other significant variables were the percentage of economically active residents employed in the manufacturing of metals etc. industries, the percentage of residents

employed in mining in 1971 and the interaction between sex and the log of smoking. These variables were consistent with other models, with the exception of the manufacturing variable which had not been significant before, and the parameter estimate was negative. The log of population density was not significant, which suggests that there is not a significant urban effect that independently explains the incidence of small cell lung cancer. However, the similarity between this model and many of the previous models suggests that the risk factors associated with small cell lung cancer were generally similar to those associated with lung cancer when it was treated as a single disease.

8.7.2 Non Small Cell Modelling Results

The models for cases of non-small cell lung cancer are shown in models 27 and 28 in Table 8.11. A total of 8,900 of the lung cancer cases were classified as non-small cell. The variables in the model were similar to those in the model of small cell cases. The main difference was that population density was significant which suggests that there was an urban effect that influences non-small cell lung cancer. Furthermore, smoking was less significant than in the models of small cell lung cancer which is surprising given that it is non-small lung cancer that is most strongly related to smoking, as discussed in chapter two. There were also different employment variables that were significant because the percentage of economically active residents employed in the energy and water industries and the percentage of economically active residents employed in the distribution and catering industries were significant and both parameter estimates were negative.

	Model 27			Model 28		
Scaled Deviance		67,607			53,416	
Degrees of Freedom		454,883			454,869	
	Estimate	Std. Error	T-Value	Estimate	Std. Error	T-Value
Intercept	-6.168	0.0106	-581.89	-11.110	0.3687	-30.13
Age (25-34)				1.436	0.4146	3.46
Age (35-44)				3.645	0.3668	9.94
Age (45-54)				5.263	0.3606	14.60
Age (55-64)				6.568	0.3593	18.28
Age (65+)				6.969	0.3595	19.39
Sex (Female)				-0.869	0.0675	-12.87
Carstairs				0.021	0.0043	4.88
Log Smoking				0.715	0.0431	16.58
Log Pop. Density				0.020	0.0066	3.11
Pollution 71				0.001	0.0003	5.09
Energy				-0.004	0.0014	-3.09
Distribution				-0.001	0.0005	-2.61
Mining 71				-0.005	0.0026	-2.05
Log Smok'g.Sex (F)				-0.229	0.0478	-4.78

Table 8.11. Poisson regression modelling of all cases of non-small cell lung cancer

8.7.3 Sex and Lung Cancer Type

Research has revealed that men and women differ in terms of their risk to the specific types of lung cancer. This was reiterated in chapter four, which demonstrated that the ratio of the number of cases of lung cancer between males and females depended upon the lung cancer type. For example, the ratio between males and females for the aggressive small cell lung cancer is much smaller than the ratio for non-small cell lung cancer (Stockwell *et al.*, 1998). Furthermore, within the non-small cell lung cancer group, women are more likely than men to suffer one form of the disease, adenocarcinoma, whereas among men the numbers are more evenly divided between adenocarcinoma and squamous cell lung cancer (Ferguson *et al.*, 1990). However, the information included in the lung cancer dataset did not enable the records to be reliably disaggregated into the subtypes of small cell and non small cell lung cancer (Brewster, private communication). Payne (2001) suggests that the

reasons for the variations in the form of lung cancer between males and females are a complex interaction of biological differences, such as hormonal factors and gendered factors such as smoking behaviour, as well as occupational differences.

In light of the differences between men and women in the type of lung cancer diagnosed and differences in trends in relation to histological type, it is sensible to study the different forms of lung cancer by sex. Therefore, models were run that considered males and females separately for the two major histological types of lung cancer: small cell and non-small cell, and the results are shown in models 29 to 36 (Tables 8.12 to 8.15). It is noticeable that there is much similarity between all of the optimum sex-specific models, for both types of lung cancer, because the age and smoking variables were significant in all of the models, whilst the Carstairs and 1971 air pollution variables were significant in three of the models. The main differences between the models was that the 1971 air pollution variable was not significant in the model of non small cell lung cancer for females and the Carstairs Index was not significant in the model of male small cell lung cancer. The population density variable was only significant in explaining female non small cell lung cancer. In addition there were some differences in the occupational variables that were significant in each model. Therefore, the analysis of the sex-specific models of small cell and non-small cell lung cancer incidence has demonstrated that although there are some differences between males and females, the key risk factors are broadly similar.

	Model 29			Model 30		
Scaled Deviance	12,063			10,354		
Degrees of Freedom	227,441			227,433		
	Estimate	Std. Error	T-Value	Estimate	Std. Error	T-Value
Intercept	-7.511	0.0291	-258.11	-11.610	0.7650	-15.18
Age (25-34)				0.355	0.9907	0.36
Age (35-44)				2.798	0.7929	3.53
Age (45-54)				4.707	0.7665	6.14
Age (55-64)				5.953	0.7626	7.81
Age (65+)				6.254	0.7625	8.20
Log Smoking				0.955	0.0764	12.49
Pollution 71				0.002	0.0005	4.71
Mining 71				-0.016	0.0077	-2.13

Table 8.12. Poisson regression modelling of males cases of small cell lung cancer.

	Model 31			Model 32		
Scaled Deviance	8,964			7,628		
Degrees of Freedom	227,441			227,434		
	Estimate	Std. Error	T-Value	Estimate	Std. Error	T-Value
Intercept	-7.858	0.0349	-225.09	-15.030	2.6210	-5.73
Age (25-34)				-0.243	3.7010	-0.07
Age (35-44)				5.771	2.6270	2.20
Age (45-54)				7.246	2.6200	2.77
Age (55-64)				8.587	2.6180	3.28
Age (65+)				9.042	2.6180	3.45
Carstairs				0.057	0.0137	4.14
Log Smoking				0.573	0.1197	4.78

Table 8.13. Poisson regression modelling of females cases of small cell lung cancer.

	Model 33			Model 34		
Scaled Deviance	67,607			33,286		
Degrees of Freedom	454,883			227,431		
	Estimate	Std. Error	T-Value	Estimate	Std. Error	T-Value
Intercept	-6.168	0.0106	-582.44	-11.460	0.5339	-21.46
Age (25-34)				1.885	0.5934	3.18
Age (35-44)				3.993	0.5409	7.38
Age (45-54)				5.699	0.5335	10.68
Age (55-64)				7.111	0.5320	13.37
Age (65+)				7.406	0.5321	13.92
Carstairs				0.020	0.0053	3.84
Log Smoking				0.689	0.0453	15.24
Pollution 71				0.002	0.0003	5.33
Male Agriculture				-0.003	0.0011	-2.78
Male Energy				-0.003	0.0012	-2.59

Table 8.14. Poisson regression modelling of males cases of non-small cell lung cancer.

	Model 35			Model 36		
Scaled Deviance	24,507			20,089		
Degrees of Freedom	227,441			227,431		
	Estimate	Std. Error	T-Value	Estimate	Std. Error	T-Value
Intercept	-6.585	0.01854	355.18	-11.470	0.5122	-22.39
Age (25-34)				0.932	0.5932	1.57
Age (35-44)				3.302	0.5034	6.56
Age (45-54)				4.782	0.4930	9.70
Age (55-64)				5.883	0.4905	11.99
Age (65+)				6.509	0.4909	13.26
Carstairs				0.024	0.0076	3.15
Log Smoking				0.554	0.0678	8.17
Log Pop. Density				0.033	0.0123	2.69
Pollution 71				0.001	0.0005	2.02
Female Distribution				-0.001	0.0006	-2.16

Table 8.15. Poisson regression modelling of females cases of non-small cell lung cancer.

8.8 Analysis of Residuals

In addition to examining the parameter estimates of the explanatory variables, the results of a regression model can be assessed by an examination of the residuals which represent the differences between values observed in each area and those predicted by the model (Flowerdew and Geddes, 1999). By calculating the residuals for each area and then mapping them, it was possible to visually identify which OAs had large discrepancies between the observed and expected values. The spatial pattern of the residuals can provide an indication of other important variables that may help to explain the variations in the incidence (Lovett and Flowerdew, 1989). Furthermore, the residual values can be used to identify anomalous areas where the incidence of lung cancer is higher or lower than expected, once the key risk factors have been controlled for. This has public health implications because the results provide information that could be used to spatially target neighbourhoods in Scotland where lung cancer is particularly high. Therefore, specific areas with a

particularly high or low incidence of lung cancer are identified and methods for comparing lung cancer between health authorities suggested.

8.8.1 Calculating Residual Values

The residual values were calculated in two ways. The first method can be termed the contribution to deviance method, which can be calculated by:

$$r_i = 2.(y_i.\ln(y_i / u_i) - y_i + u_i)$$

This calculates the amount that each observation contributes to the total deviance value in the model. Although this measurement is useful for identifying areas that contribute unusual amounts towards the total deviance, no indication is given as to whether a large value is indicative of an important observation or rather it only occurs because the original measured values are large.

An alternative method is the Pearson residuals (or standardised residuals). For the Poisson distribution, the residual, r, for each unit i, can be calculated as:

$$r_i = \frac{y_i - u_i}{\sqrt{u_i}}$$

where y is the observed value and u is the predicted value. This approach has the advantage of calculating a value that indicates whether the expected value is greater than the observed value or vice versa. If the residual value is positive then it indicates that the observed is greater than the expected but if the value is negative

then it suggests that the count of lung cancer cases is lower than the expected. This is useful because it discriminates between areas that have a higher or lower rate than was expected rather than just indicating how much the area contributed to the total deviance value. Without standardisation, large residual values may occur solely because the original measured values were large, but by standardising the residual values, this confusion is removed (O'Brien, 1992). Therefore, in tandem, the two approaches provide a useful way to interpret the spatial variation of the residuals.

8.8.2 Results

Model 14 in Table 8.5 provided the optimum model for explaining the incidence of all cases of lung cancer in all OAs in Scotland. Therefore, the Pearson residual values and deviance values were calculated for model 14 in order to investigate the spatial pattern of the residuals. This model was used because there were only subtle differences between this and the models for males, females, specific age groups, rural areas and different types of lung cancer. Furthermore, model 14 included enough cases of lung cancer to be confident that the results were reasonably robust.

Chapter four discussed the mean SIRs in each of the Scottish health boards and demonstrated that the highest mean SIRs for males and females tended to be in the most urban areas, particularly the Greater Glasgow and Lothian Health Boards (Table 4.4). Furthermore the health boards with the lowest mean SIRs were consistently the most rural ones (Shetland, Highland, Orkney and the Western Isles) for males and females. However, these results did not control for the key risk factors such as smoking, air pollution and deprivation. Once these factors are controlled for,

the lung cancer trend becomes less clear. Table 8.16 shows the total Pearson value of OAs in each of the Scottish health boards that were ranked according to the total Pearson value. The table demonstrates that the highest positive Pearson residuals were in the larger urban health boards such as Greater Glasgow (2936.64) and Lothian (1154.19), which suggests that there were more cases of lung cancer in these areas than expected even once the key risk factors had been controlled for. Beyond the two largest urban health boards, the pattern is less clear. Lanarkshire had a relatively high SIR for lung cancer, especially for females, but it has the second lowest Pearson residual value (-329.67), which demonstrates that there were far fewer cases of lung cancer than were expected. Similarly, some of the health boards that had a mean SIR of less than 100, such as Ayrshire and Arran, Argyll and Clyde and Dumfries and Galloway, have high positive Pearson residual values. Therefore, although the age-standardised lung cancer rates were relatively low in these health

Health Board	% Population	Total Observed	Total Expected	Total Deviance	% Deviance	Total Pearson
Grampian	9.85	1406	1533.40	6693.51	8.12	-381.14
Lanarkshire	11.06	1750	1924.26	8144.82	9.88	-329.67
Fife	6.85	1052	1154.76	5021.39	6.09	-272.42
Highland	4.04	435	592.28	2473.39	3.00	-227.27
Borders	2.08	349	365.93	1598.78	1.94	-99.25
Western Isles	0.58	69	96.20	378.59	0.46	-70.55
Shetland	0.44	45	57.17	240.93	0.29	-61.40
Orkney	0.39	43	55.85	236.05	0.29	-54.35
Tayside	7.68	1443	1517.93	6489.80	7.87	118.40
Dumfries and Galloway	2.97	495	512.55	2442.28	2.96	127.59
Forth Valley	5.40	910	909.96	4143.91	5.03	265.57
Ayrshire and Arran	7.51	1337	1334.94	6204.02	7.52	555.35
Argyll and Clyde	8.59	1646	1603.04	7272.58	8.82	597.60
Lothian	14.47	2740	2531.37	12347.15	14.97	1154.19
Greater Glasgow	18.09	4682	4213.36	18766.38	22.76	2936.64

Table 8.16. Contribution to deviance and Pearson values in Scottish Health Boards

boards there is a higher than expected incidence of lung cancer once the risk factors had been controlled for. Furthermore, the OAs in the Greater Glasgow and Lothian Health Boards contributed a large proportion (36.73 percent) of the total deviance in model 14. Similarly, four of the most rural health boards (Borders, Western Isles, Shetland and Orkney) contributed only a small proportion to the total deviance (2.98 percent). These results suggest that explanations of urban / rural differences in lung cancer may be more complicated than the results in chapter four suggested.

Table 8.17 divided all of the OAs into population density quintiles and summed the Pearson value in each category. The total deviance, the percentage deviance that each category contributed and the percentage of people in each group are also shown. Chapter four demonstrated that the SIR of lung cancer increased from the

Urban - Rural	% Population	Total Observed	Total Expected	Total Deviance	% Deviance	Total Pearson
Most Rural	19.64	3907	3154.34	19086.65	23.15	6409.31
	19.86	3556	3489.30	15994.08	19.40	641.46
Intermediate	20.07	3650	3588.07	15907.54	19.29	6.61
	20.15	3554	3808.61	15560.13	18.87	-889.86
Most Urban	20.27	3735	4362.67	15905.18	19.29	-1901.60

Table 8.17. Contribution to deviance and Pearson values in five urban/rural categories

most rural areas to the most urban (Table 4.3). However, Table 8.17 shows that it is the most rural areas which had the highest positive Pearson residuals and the most urban areas that had the highest negative residuals. This demonstrates a clear relationship between the total of the residual values and how urban or rural the OAs are. The lowest negative total Pearson value was in the most urban category (-1901.60) and the total Pearson value increased through the categories with the highest positive value in the most rural category (6409.31). Therefore, in the most

rural areas there were more cases of lung cancer than expected, having controlled for the variables in the optimum model. Furthermore, the total Pearson value in the most rural category was very high compared to all other categories. In the most urban areas the opposite was true because there were fewer cases of lung cancer than would be expected from the model. This is particularly interesting in light of the results in Table 8.16, which suggested that lung cancer was higher than expected in the most urban health boards. These results show that once the small areas are categorised according to how urban or rural they are, then the pattern is reversed. This reflects the variation in lung cancer incidence within the health boards and highlights the importance of considering lung cancer incidence in small geographical areas.

In addition, Table 8.17 shows that it is the most rural areas that provide the highest contribution to the deviance. Although the percentage of deviance was similar in the four most urban categories (approximately 19 percent each), the most rural category contributed more to the total deviance (23.15 percent). This suggests that the overall goodness of fit is worse in the most rural quintile compared to the other four groups.

The relationship between the residual values and the Carstairs Index of Deprivation is shown in Table 8.18. The PPSs in Scotland were divided into quintiles according to their Carstairs value and the total deviance value and total Pearson value were calculated for each category. Both the total deviance and the total Pearson value increased from the least deprived to the most deprived category. The least deprived category contributed 10.54 percent of the total deviance whereas 31.49 percent was accounted for in the most deprived category. This suggests that the model was better

at predicting lung cancer incidence in the least deprived areas compared to the most deprived areas. It was surprising to find that all of the Pearson values were positive in each of the deprivation categories and the total increased from 415.41 in the lowest category to 1902.92 in the highest deprivation category. This suggests that there are more cases of lung cancer than expected in each of the Carstairs categories but that this is particularly true in the most deprived category where there are far more cases of lung cancer than were predicted by the model. These results suggest that the incidence of lung cancer is higher than expected in the more deprived areas, even once the important risk factors, including the Carstairs Index, are controlled for.

Deprivation Quintile	% Population	Total Observed	Total Expected	Total Deviance	% Deviance	Total Pearson
Least Deprived	14.05	1697	1630.22	8691.56	10.54	415.41
	17.11	2493	2498.06	12249.79	14.86	504.51
Intermediate	20.23	3315	3334.21	15363.12	18.63	635.65
	23.73	4573	4593.92	20186.35	24.48	800.80
Most Deprived	24.88	6324	6346.59	25962.75	31.49	1902.92

Table 8.18. Contribution to deviance and Pearson values in five deprivation categories using the Carstairs Index

Although the incidence of lung cancer tends to be higher than expected in the rural areas, many of the very highest individual residual values are located in the urban areas. Figure 8.2 shows the distribution of the 50 highest and the 50 lowest Pearson residual values for all age-sex groups in Scotland and it shows that most of the highest and lowest values were located in urban areas. There was a cluster of extreme values in Glasgow which has many high positive and high negative values. There were also clusters of extreme cases in Falkirk, close to the Forth estuary, and in Dundee. Most of the extreme values in Edinburgh were positive but in Aberdeen they were negative. This may be because the risk factors, such as smoking, have exceptional values in these areas. Therefore, using the example of

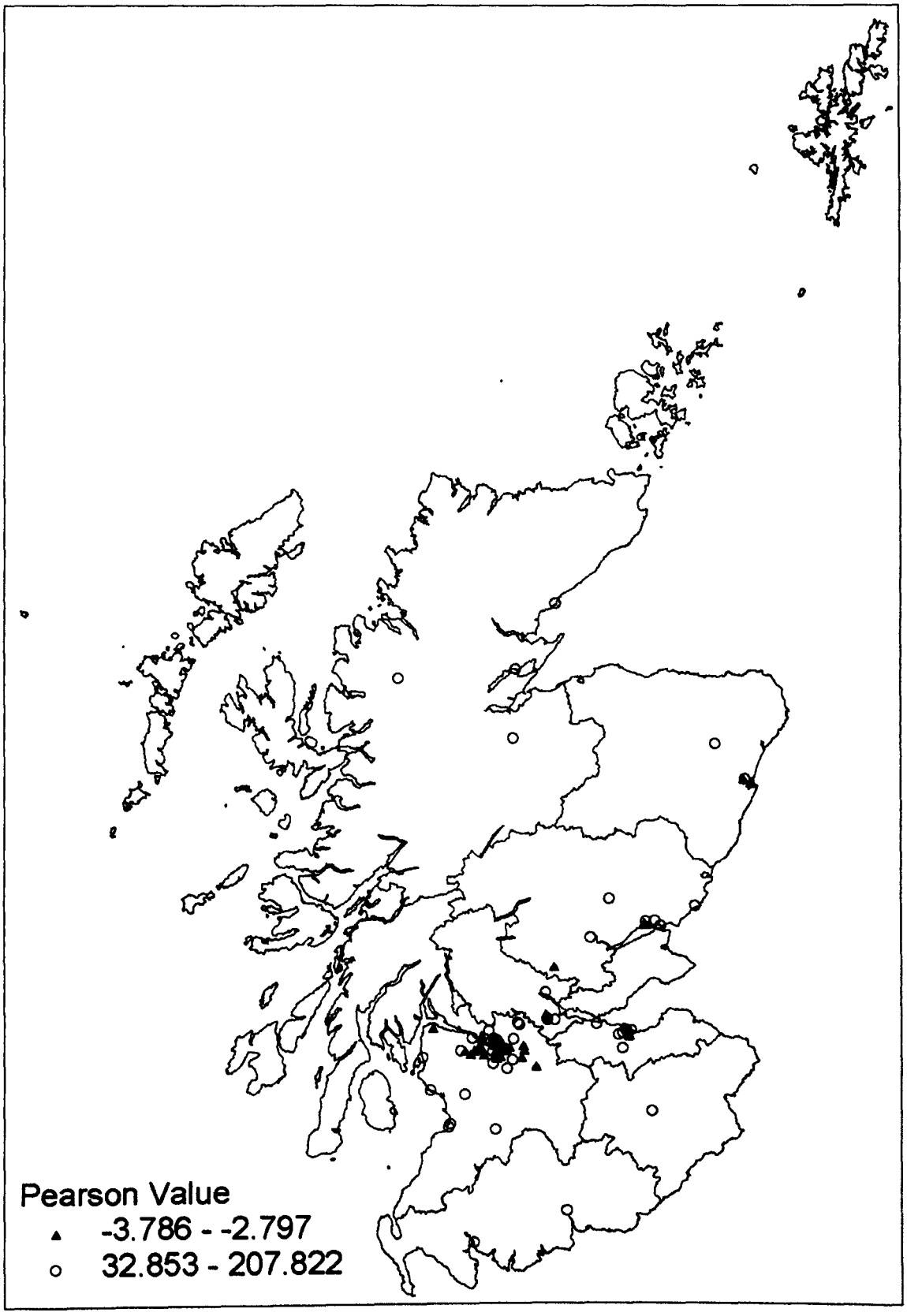


Figure 8.2. The 50 highest and 50 lowest summed Pearson residual values.

smoking, the probability of smoking was higher than expected in the parts of Edinburgh that had very high residual values but lower than was expected in parts of Aberdeen with extremely low residual values. There were also two highly positive values near to Ayr, which suggests that these areas had a higher incidence of lung cancer than may be expected. There were only a few extreme values in rural parts of Scotland with a small number in the northern Highlands and on the border with England, which suggests that the model is a good predictor of lung cancer in rural areas. However, the extreme values in rural areas were almost always positive and are often among the very highest values. Table 8.19 shows the OAs with the 50 highest Pearson values and their contribution to deviance value. The table shows that most of the OAs with the highest Pearson values were in large urban districts, particularly Glasgow and Edinburgh. However, it is noticeable that the majority of the ten highest values fell within more rural districts in the Highlands, Grampian, and Ayrshire and Arran Health Boards. The absolute values of the 50 highest residuals varied from 207.82 to 32.85 which are significantly higher than the majority of the OAs. For example, 6,275 of the 37,907 OAs included in the models had a Pearson residual greater than 1.99, which demonstrates that the OAs with the 50 highest residual values had unusually high values.

Figure 8.3 shows the distribution of the OAs with the highest values using the contribution to deviance method. The OAs were divided into those that had more cases of lung cancer than was predicted by the model (so have a positive Pearson residual value) and those that had fewer cases of lung cancer than predicted (a negative Pearson residual value) and the 50 highest residual values in each of these categories are shown. This approach confirms the observations made using the

Output Area	District	Health Board	Pearson Value	Deviance
27AF04A	Sutherland	Highland	207.82	47.01
40AF04C	Eastwood	Greater Glasgow	135.45	99.25
41BE13A	Glasgow City	Greater Glasgow	120.73	158.84
06AH15A	Falkirk	Forth Valley	101.96	140.71
36AL08B	Cumnock & Doon Valley	Ayrshire and Arran	92.79	132.94
53BJ04C	Perth & Kinross	Tayside	84.23	76.45
37BL06B	Cunninghame	Ayrshire and Arran	74.73	153.44
25AS03	Ross & Cromarty	Highland	74.01	27.07
49BG03A	Renfrew	Argyll and Clyde	68.83	112.93
17AU25C	Gordon	Grampian	59.42	84.47
45AP19C	Kyle & Carrick	Ayrshire and Arran	59.17	165.93
41BD15A	Glasgow City	Greater Glasgow	59.14	184.39
52AN23B	Dundee City	Tayside	57.78	170.82
11AD02B	Wigtown	Dumfrieas and Galloway	55.90	46.89
29BL23C	Edinburgh City	Lothian	53.66	155.32
29CQ13F	Edinburgh City	Lothian	52.10	154.36
06AH13A	Falkirk	Forth Valley	51.01	140.71
50AY06C	Strathkelvin	Greater Glasgow	50.81	75.04
09AE10B	Nithsdale	Dumfries and Galloway	50.26	91.42
49BJ07A	Renfrew	Argyll and Clyde	50.25	61.21
06AM02B	Falkirk	Forth Valley	46.78	179.30
45AQ02A	Kyle & Carrick	Ayrshire and Arran	45.82	172.84
35AF05B	Cumbernauld & Kilsyth	Lanarkshire	45.69	106.27
41CK04B	Glasgow City	Greater Glasgow	45.18	210.26
51AS18B	Angus	Tayside	43.95	142.69
41CC01B	Glasgow City	Greater Glasgow	43.20	158.74
41DD26B	Glasgow City	Greater Glasgow	42.35	128.05
52AX09C	Dundee City	Tayside	42.03	102.44
25AN06C	Ross & Cromarty	Highland	40.85	74.01
52AW06D	Dundee City	Tayside	40.43	140.27
53AP09B	Perth & Kinross	Tayside	40.30	215.49
29DH05A	Edinburgh City	Lothian	39.93	100.33
41AT02E	Glasgow City	Greater Glasgow	39.12	226.10
39AG01A	East Kilbride	Lanarkshire	38.74	138.69
37AX04A	Cunninghame	Ayrshire and Arran	38.34	138.16
44AL03B	Kilmarnock & Loudoun	Ayrshire and Arran	36.90	136.40
22AB06	Inverness	Highland	36.59	171.27
15AS21C	Aberdeen City	Grampian	36.56	185.73
29BH31B	Edinburgh City	Lothian	36.48	306.10
41CK20A	Glasgow City	Greater Glasgow	36.16	210.26
05AD26A	Clackmannan	Forth Valley	36.12	213.95
29BJ16A	Edinburgh City	Lothian	35.93	232.88
35AE10B	Cumbernauld & Kilsyth	Lanarkshire	35.81	98.97
29BU29B	Edinburgh City	Lothian	34.60	176.11
41BR14A	Glasgow City	Greater Glasgow	33.51	255.87
41CB07B	Glasgow City	Greater Glasgow	33.49	201.76
30AR02A	Midlothian	Lothian	33.39	52.18
02AR10A	Ettrick & Lauderdale	Borders	33.06	36.82
33AE07A	Bearsden & Milngavie	Greater Glasgow	32.92	69.15
42AD17B	Hamilton	Lanarkshire	32.85	197.27

Table 8.19. The 50 highest summed Pearson residual values

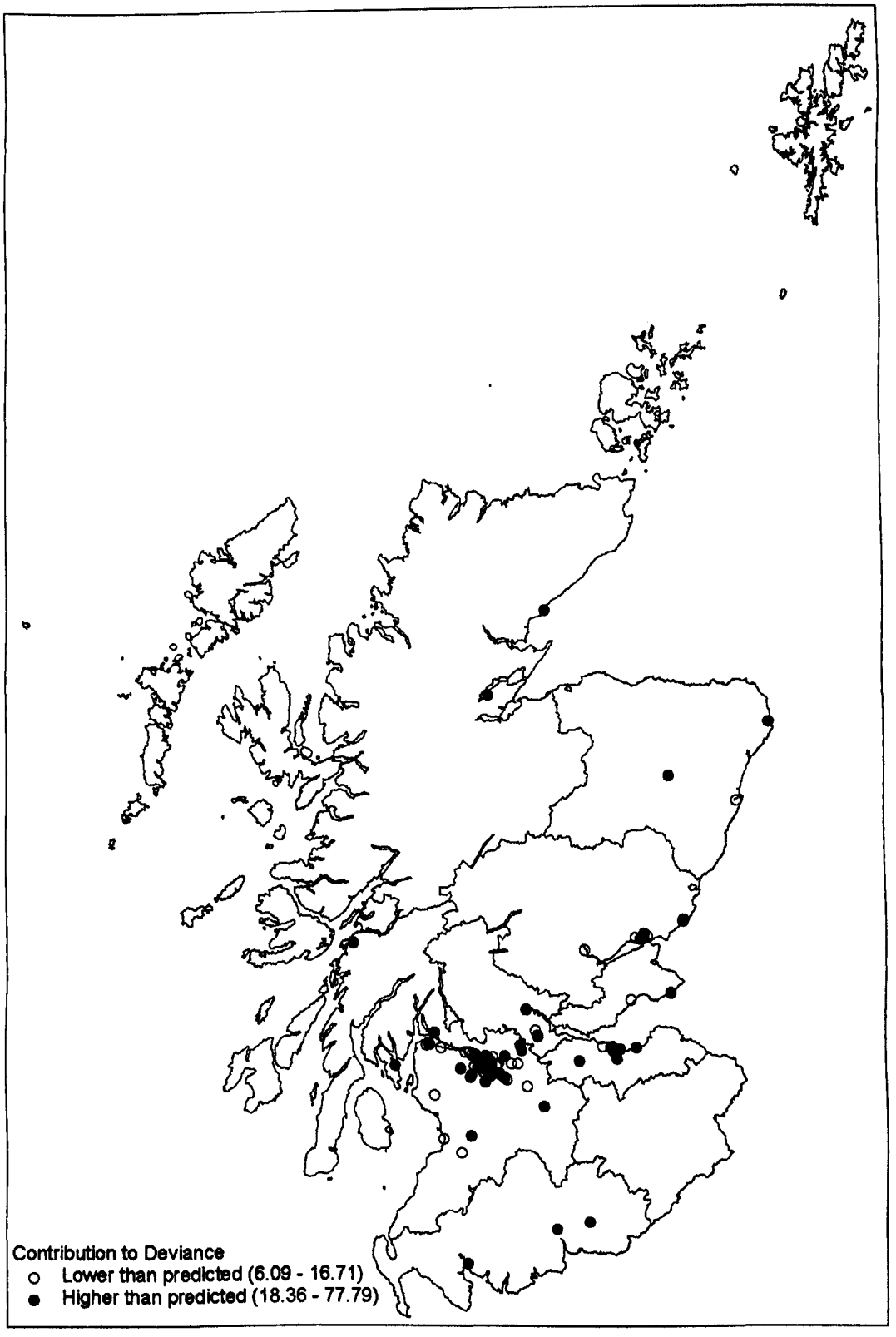


Figure 8.3. The 50 OAs with higher than expected cases of lung cancer and the 50 OAs with the lower than expected cases of lung cancer that contributed the most to the deviance value.

Pearson method because most of the highest values were in the urban areas (although this will also be related to the number of cases). There were a large number of high values in Glasgow and Edinburgh but relatively few in Dundee and Aberdeen. This is similar to the results for the Pearson residuals but suggests that much of the total deviation is explained by the residual values in parts of Glasgow and Edinburgh.

8.8.3 Discussion of Residuals in Rural Areas

Chapter four demonstrated a clear urban bias in lung cancer when the age structure of each area was controlled for, using mapped age-standardised rates and cluster detection methods. However, the results here have shown that this relationship is not as clear once factors such as smoking, air pollution, deprivation and population density are controlled for. When the important explanatory variables were incorporated into the model, rural areas appeared to have a higher incidence of lung cancer than expected while urban areas had a lower incidence than expected. This is despite the presence of the population density variable, which was specifically included to account for an urban effect. One possible reason why more cases were found in rural areas could be that there were localised raised pollution levels, which were not controlled for carefully enough. Although earlier results have demonstrated that air pollution was not significant in explaining the incidence of lung cancer in Scotland (section 8.6), it is possible that the measures of pollution do not fully capture some of the localised variations. Chapter six estimated the level of pollution in small areas across Scotland and the estimates were low and relatively uniform in the most rural areas. However, it is possible that there are localised 'hotspots' of high pollution in some areas that raise the incidence of lung cancer. These may also

explain some of the highest residual values in the rural areas. The 'hotspots' of pollution may not be incorporated into the 1991 estimates due to industrial changes during the lung cancer latency period. Furthermore, these unusual areas may not be incorporated into the 1971 estimates of pollution due to the inaccuracies of the predictions.

An alternative explanation for the high residual values in rural areas is that radon does increase the incidence of lung cancer in Scotland, especially in rural areas. As has already been highlighted in chapter seven, radon levels tend to be higher in the more rural areas of Scotland. This analysis has found some evidence that radon raises the incidence of lung cancer, especially in the younger age groups. However, in the optimum model of all cases of lung cancer the radon variables were not significant. As discussed earlier in this chapter, it is difficult to be confident that there is not a radon effect, given the limitations of the estimates of radon exposure. If radon does influence the incidence of lung cancer then it will not be controlled for in the optimum model and could be partially reflected in the pattern of the residual values.

8.8.4 Discussion of Residuals in Urban Areas

Although rural areas have more cases of lung cancer than expected, some of the most extreme residual values were located in the urban areas, particularly Glasgow. This is in line with previous studies which have also highlighted that parts of Glasgow have unusually high lung cancer rates (Gillis *et al.*, 1988; Lamont *et al.*, 1999). This is supported by the discussion in chapter four, which demonstrated consistently high

rates of lung cancer in small areas of Glasgow compared to the rest of Scotland. In addition, this chapter has demonstrated that having controlled for the key risk factors such as the population structure, smoking and air pollution, the Greater Glasgow Health Board had higher positive residual values than all other health boards in Scotland. Furthermore, the map of extreme residuals (Figure 8.2) has shown that Glasgow has both high positive and high negative values. In order to gain some further insights into the factors that explain lung cancer incidence in Scotland, it is useful to focus upon Glasgow to examine the socio-economic background of the neighbourhoods with extreme residual values. This may help to explain why Glasgow has many extreme residual values and provide further insights into the causes of lung cancer in Scotland.

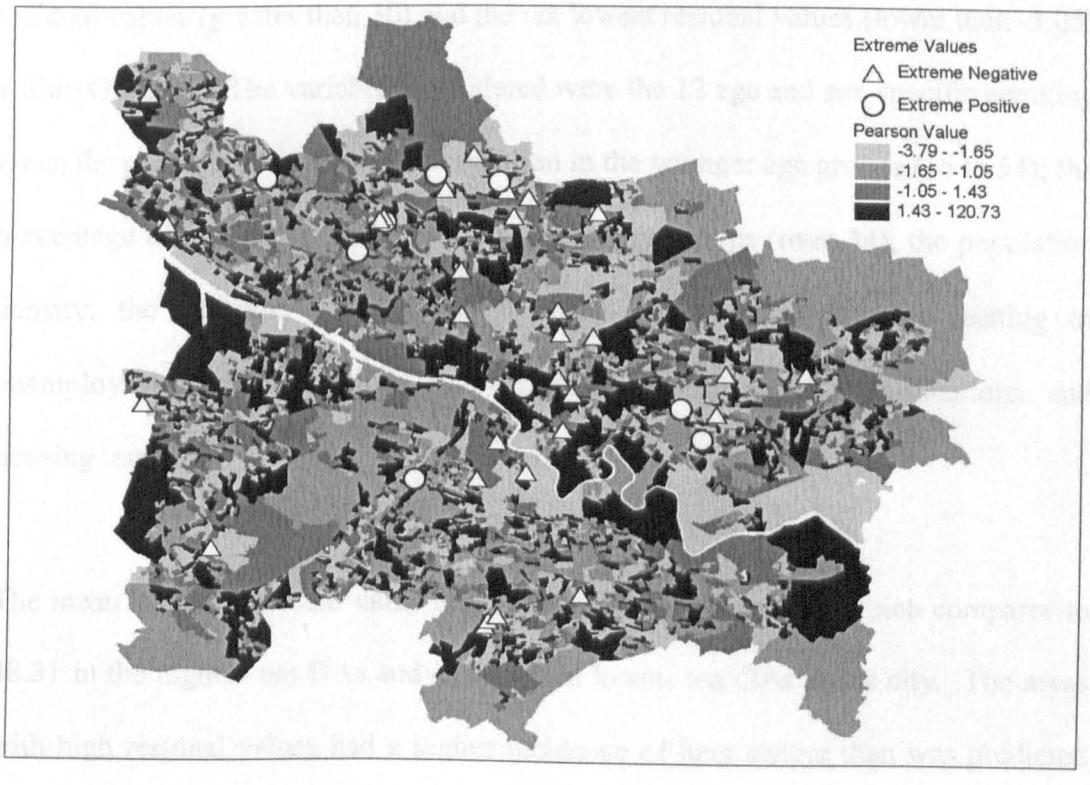


Figure 8.4. Pearson residual values in Glasgow.

Figure 8.4 maps the total Pearson residual values that were calculated from model 14 for all OAs in Glasgow and highlights the location of the extreme residual values in Glasgow identified in Figure 8.2. It can be seen that there were fewer extreme Pearson residual values south of the River Clyde and that most of these values were negative. North of the River Clyde, there were more extreme values but there were still a higher number of extreme negative values than there were positive values. However, it is not clear from this map whether the extreme residual values are located in similar types of areas and hence it is difficult to draw any conclusions about the location of these areas. Therefore, it is useful to examine the nature of the areas in which the extreme values were located. Table 8.20 shows the mean, minimum and maximum values for a selection of socio-economic variables for the whole of Scotland, the whole of Glasgow and for the OAs with the ten highest residual values (greater than 30) and the ten lowest residual values (lower than -3.05) within Glasgow. The variables considered were the 12 age and sex-specific smoking rates; the percentage of the adult population in the younger age groups (16 to 54); the percentage of the adult population in the older age groups (over 54); the population density; the Carstairs Index of Deprivation; and census variables relating to unemployment, overcrowding, car ownership, social class, qualifications and housing tenure.

The mean Pearson residual value in all OAs in Glasgow is 0.49, which compares to 48.31 in the highest ten OAs and -3.25 in the lowest ten OAs in the city. The areas with high residual values had a higher incidence of lung cancer than was predicted by the optimum model. This is because these OAs had high values for most of the variables in model 14, yet the incidence of lung cancer was even higher than was

	Scotland			Glasgow			High Residuals			Low Residuals		
	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
Pearson Value	-3.79	207.82	0.11	-3.79	120.73	0.49	30.24	120.73	48.31	-3.79	-3.06	-3.25
Prop. Smokers Males 16-24	0.170	0.792	0.355	0.000	0.791	0.433	0.263	0.574	0.454	0.416	0.791	0.554
Prop. Smokers Males 25-34	0.167	0.860	0.396	0.000	0.857	0.492	0.263	0.673	0.524	0.477	0.857	0.645
Prop. Smokers Males 35-44	0.146	0.876	0.390	0.000	0.873	0.495	0.236	0.697	0.532	0.479	0.873	0.665
Prop. Smokers Males 45-54	0.165	0.845	0.384	0.000	0.843	0.476	0.259	0.649	0.506	0.460	0.843	0.623
Prop. Smokers Males 55-64	0.142	0.848	0.364	0.145	0.845	0.463	0.228	0.649	0.494	0.442	0.845	0.620
Prop. Smokers Males > 65	0.092	0.652	0.220	0.000	0.651	0.283	0.150	0.399	0.296	0.260	0.651	0.388
Prop. Smokers Fem's 16-24	0.151	0.767	0.324	0.000	0.766	0.401	0.236	0.538	0.420	0.381	0.766	0.520
Prop. Smokers Fem's 25-34	0.148	0.841	0.365	0.000	0.838	0.460	0.236	0.640	0.490	0.441	0.838	0.612
Prop. Smokers Fem's 35-44	0.129	0.859	0.360	0.000	0.856	0.464	0.211	0.665	0.499	0.442	0.856	0.633
Prop. Smokers Fem's 45-54	0.146	0.825	0.353	0.000	0.822	0.444	0.232	0.615	0.472	0.424	0.822	0.590
Prop. Smokers Fem's 55-64	0.125	0.828	0.334	0.128	0.825	0.431	0.203	0.615	0.461	0.407	0.825	0.587
Prop. Smokers Fem's > 65	0.080	0.619	0.197	0.000	0.617	0.256	0.132	0.365	0.267	0.233	0.617	0.356
% Aged 16-54	0.00	100.00	62.21	0.00	100.00	61.15	43.09	94.51	69.20	29.06	78.89	49.22
% Aged 55+	0.00	100.00	37.79	0.00	100.00	38.85	5.49	56.91	30.80	21.11	70.94	50.79
Population Density	0.00	175000.0	7706.5	0.0	172000.0	12843.2	412.5	22200.0	8818.0	2990.9	172000.0	37937.8
Carstairs Index	-6.00	12.40	0.62	-3.57	12.40	4.71	-2.86	11.36	4.92	5.87	11.36	8.04
% Unemployed	0.00	100.00	13.70	0.00	100.00	25.67	0.00	52.63	30.44	8.11	51.11	32.05
% Overcrowded Households	0.00	60.78	7.11	0.00	60.78	12.77	0.00	38.10	16.07	0.00	17.92	10.13
% No Car Households	0.00	100.00	35.82	0.00	100.00	59.95	13.82	88.17	62.05	60.00	91.76	74.92
% Social Classes IV or V	0.00	100.00	21.76	0.00	100.00	24.50	0.00	100.00	34.89	0.00	57.14	24.55
% High Qualifications	0.00	100.00	12.19	0.00	100.00	9.95	0.00	62.50	13.87	0.00	11.76	3.22
% Limit Long Term Ill	0.00	98.18	14.47	0.00	98.18	19.34	2.54	27.83	18.06	15.43	58.36	26.64
% Owner Occupied H'holds	0.00	100.00	52.64	0.00	100.00	37.78	0.00	65.22	22.16	0.00	45.88	13.16
% House Assoc H'holds	0.00	100.00	40.07	0.00	100.00	55.91	0.00	100.00	62.58	54.12	100.00	85.19

Table 8.20. Summary data for OAs in Scotland and Glasgow, OAs with high Pearson residual values and OAs with low Pearson residual values.

expected. One explanation for this is the unusual effect of population density in Glasgow. The population density parameter estimate in the optimum model of lung cancer incidence in Scotland was positive and hence the predicted incidence of lung cancer was higher in areas of high population density. However, the areas with high residual values had a low population density (8,818) compared to the mean of all OAs in Glasgow (12,843), which could mean that lung cancer in Glasgow was under predicted by the optimum model. Therefore, although the optimum model demonstrates that higher population densities increase the incidence of lung cancer in Scotland, it would seem that this is not uniform and that in fact some areas of Glasgow have a low population density even though the levels of lung cancer incidence were in fact high. This could be related to the nature of the housing types in Glasgow as some areas have high density housing in relatively wealthy areas that would be expected to have low rates of lung cancer, particularly near to the city centre and, conversely, in some areas with a high proportion of local authority housing the density may be relatively low.

At the same time, the areas with high residual values had a high Carstairs Index value (4.92) compared to the mean of all areas of Scotland (0.62) but the value was similar to the mean for all OAs in Glasgow (4.71). The evidence for high levels of deprivation in the areas with large positive residual values is supported by other indicators of deprivation and low income such as high levels of unemployment, a low proportion of owner occupied housing and a high proportion of housing association households. Table 8.20 also demonstrates that the percentage of younger people (62.21) was greater than the percentage of older people (37.79) for the whole of Scotland, and very similar rates were found in Glasgow (61.15 and 38.85).

However, in the areas with high residual values there was a greater proportion of young people compared to older people (69.20 and 30.80). Earlier discussion has shown that lung cancer is predominantly a disease of older age and hence the relatively young population in these areas helps to explain why the predicted incidence of lung cancer rates are relatively low. Because there are high rates of lung cancer among the younger people, the model for Scotland underestimates incidence of the disease. Therefore, the areas with high residual values have an unusual combination of being relatively poor, not densely populated but also a relatively young population, which helps to explain the unusually high incidence of lung cancer.

The areas with low residual values represent OAs where lung cancer was over predicted by the optimum model (i.e. the incidence of lung cancer was lower than was expected). Perhaps surprisingly, these tend to be areas that were even more deprived than the areas with extreme high residual values. For example, the mean value of variables such as the Carstairs Index (8.04) and the proportion of housing association households (85.19) were higher than the corresponding values of areas with high positive residuals (4.92 and 62.58), whereas the mean of variables such as the proportion of owner occupied households were much lower (13.16 compared to 22.16). In addition, the mean smoking rates in all age-sex groups and the population density variable (37937.8) were relatively high in these areas. Furthermore, the population was relatively old, as the ratio of young to old persons was almost the same (49.22 and 50.79). The combination of high deprivation, high population density, high smoking and an old population would suggest that the predicted lung cancer rates were likely to be relatively high. However, it has been noted that the

population density may not be working in Glasgow the same way as it is for all of Scotland. If this is true then the particularly high population density values in these areas may cause lung cancer to be over predicted.

The unusual effects of urbanness and deprivation were explored by modelling lung cancer incidence in the OAs in Glasgow separately. Table 8.21 shows the Grand Mean Model (model 37) and the optimum regression model (model 38) for these OAs. The parameters in model 38 are similar to those in model 14 (the optimum model for all lung cancer cases in Scotland), except that the population density parameter estimate was negative. This suggests that the incidence of lung cancer tends to be lower in areas of high population density but higher in the less densely populated parts of Glasgow. Therefore, although the incidence of lung cancer is higher in the urban areas of Scotland, this effect is reversed within Glasgow. This reinforces the discussion of the high and low residual values in Glasgow, as it demonstrates that lung cancer may be over or under predicted in an area because the relationship between population density is inverted within Glasgow. As discussed above, this is likely to be because within the urban centre, the poorer, more deprived housing is not necessarily the most densely populated. A comparison of model 14 and model 38 also demonstrates that the Carstairs Index is less significant in explaining lung cancer in Glasgow, compared to all cases in Scotland. This may be because the Carstairs Index for PPSs does not capture the full extent of deprivation in Glasgow, especially as the analysis of lung cancer incidence has been undertaken for the smaller OAs. There may be important variations in deprivation within PPSs that are not captured in the models but which could be important in understanding the relationship between lung cancer and deprivation in Glasgow.

	Model 37			Model 38		
Scaled Deviance	21,200			15,155		
Degrees of Freedom	64,211			64,200		
	Estimate	Std. Error	T-Value	Estimate	Std. Error	T-Value
Intercept	-4.974	0.01592	-321.43	-10.620	0.8243	-12.88
Age (25-34)				1.282	0.9263	1.38
Age (35-44)				3.865	0.8257	4.68
Age (45-54)				5.756	0.8149	7.06
Age (55-64)				6.926	0.8133	8.52
Age (65+)				7.489	0.8133	9.21
Sex (Female)				-0.389	0.0334	-11.64
Carstairs				0.021	0.0062	3.42
Log Smoking				0.477	0.0617	7.73
Log Pop. Density				-0.025	0.0113	-2.25
Pollution 71				0.001	0.0003	3.21
Energy				-0.007	0.0030	-2.22

Table 8.21. Poisson regression models including just the OAs in Glasgow.

The examination of the extreme residual values in Glasgow also showed that the probability of smoking varied between the high and low values (Table 8.20). The summary of the smoking variables showed that the rates of smoking were higher in Glasgow compared to Scotland as a whole for all age-sex groups. Furthermore, the average smoking rates in the areas with extremely high residual values were higher than the rates for Glasgow but the rates for the extremely low residual values were higher again. It is possible that smoking may be under predicted in the areas with high residual values and over predicted in areas with the lowest residual values, which may be related to unusual cultures of smoking in Glasgow. The reasons for this are likely to be complex but Reece *et al.*, (2002) demonstrated that some areas of the city have extremely high rates of smoking and these areas tend to be the most disadvantaged. Furthermore, whilst the smoking rates have declined in the general population, they have not declined in the most disadvantaged communities to the same extent (Marsh and McKay, 1994). Qualitative work on smoking behaviour in Glasgow has demonstrated that there is a 'powerful association between smoking

and community disadvantage', which is compounded by the 'large-scale and pervasive disadvantage and exclusion' (Stead *et al.*, 2001). This study identified some of the specific features of disadvantaged communities that have an independent influence on smoking above the effects of individual disadvantage. Factors such as coping with the stress of personal circumstances, living in an inadequately resourced and unsafe environment, feelings of exclusion and stigmatisation, the lack of exposure to the triggers of smoking cessation and the positive aspects of neighbourliness and attachment which perpetuate smoking were all shown to reinforce people's decision to smoke (Stead *et al.*, 2001). These are influences that were not directly incorporated into the model that predicted smoking behaviour in Scotland, although some will be indirectly controlled for by the deprivation variable. Therefore, whilst none of these individual factors can be considered to be exclusive to Glasgow, it is apparent that some of the most disadvantaged communities in Scotland are found in Glasgow and the combination of the factors identified in this study are unlikely to be as extreme anywhere else.

In addition, the pattern of residuals in urban Scotland may be explained by risk factors that were not included in the models. As discussed in Chapter two, other factors influence the incidence of lung cancer but they have not been directly incorporated into the modelling procedure in this study. Passive smoking is likely to be strongly correlated with smoking behaviour, as it is possible that the areas with the highest smoking probabilities may have exposures to exceptionally high levels of passive smoking. For example, some of the high residual values in central Glasgow may be explained by high levels of passive smoking that are not directly incorporated into the optimum model of lung cancer. As was discussed above, levels

of passive smoking are often higher in urban areas and, although population density was significant in the optimum model of lung cancer, it is unlikely to capture the full extent of passive smoking. This is supported by the analysis of lung cancer in Glasgow, which demonstrated a negative relationship between lung cancer and population density.

Chapter two demonstrated that there is some research that suggests that diets that are high in cholesterol or saturated fat can significantly increase the incidence of lung cancer whereas diets that are high in fresh fruit and vegetables can have a preventative effect. Studies have, in the past, pointed to the relationship between poor diet and high levels of area deprivation (Ellaway and Macintyre, 1996) and in this study the Carstairs Index of Deprivation was significant in the optimum model of lung cancer in Scotland. This indicates that there is an area effect that influences lung cancer due to poor quality health provision, poor job prospects etc. Alternatively, deprivation is acting as a proxy for confounders such as poor diet and a lack of exercise that are features of disadvantaged communities (Ecob and Macintyre, 2000). This may be compounded by the poor access to food retail ('food deserts') that has been highlighted in deprived areas of British cities (Wrigley *et al.*, 2002). Furthermore, this could help to explain the relationship between deprivation and the residuals from the model of lung cancer discussed in Table 8.19 and also why the most extreme residual values in Glasgow were found in some of the more deprived parts of the city. If this factor has an effect then lung cancer will be higher in the more deprived areas than was predicted, resulting in the high positive Pearson values in these areas. This is partially supported by the results in Glasgow because the high positive Pearson residuals were located in deprived areas of the city as

measured by the Carstairs Index and other socio-economic variables. However, although this is true in relation to Scotland, it is confused by the observation that the extreme low residual values were located in even more deprived areas.

Similarly, as was highlighted in chapter two, asbestos has been shown to be a well established cause of lung cancer, especially in certain occupations. Although the overall level of exposure to asbestos is likely to be low in Scotland, studies have shown that there are areas where exposure will be higher, especially along the parts of the River Clyde in Scotland where shipbuilding was once a major industry (De Vos Irvine *et al.*, 1993). Although there has been a rapid decline in the shipbuilding industry in Glasgow since the 1960s (Dodds and Maguire, 1998), the industry may be important in explaining local lung cancer incidence in 1991. Again, given the 20 year lag between exposure and diagnosis, these factors could explain some of the extreme residual values that are observed in this area.

8.9 Conclusion

This chapter has explored the effects of a number of different risk factors in order to further our understanding of how they influence lung cancer incidence in Scotland. This has been achieved by examining all cases of lung cancer together, cases in rural areas, cases in particular age and sex groups and different types of lung cancer.

There are a number of consistencies between the models, particularly the effects of age and sex. Unsurprisingly, lung cancer incidence is consistently higher as age increases. The effect of sex is also consistent in all models because lung cancer

incidence is always higher for males than for females regardless of the cohort that was selected. Similarly, the smoking variable is consistently significant and shows that lung cancer increases with higher rates of smoking but, crucially, the relative importance of smoking varies according to the group of lung cancer cases being studied. In particular, it has been shown that smoking is more significant in explaining male lung cancer incidence, which is probably a response to the historical patterns of smoking behaviour. It has also been demonstrated that it is necessary to incorporate a measure of air pollution as this is consistently significant in explaining lung cancer incidence, except for in the rural areas. However, it is important to try and measure pollution levels at least 20 years prior to the lung cancer data because of the large scale reductions in pollution levels during the 1970s and 1980s, which resulted in the 1971 estimate of pollution being significant in almost all models of lung cancer and the 1991 estimate being insignificant.

However, other variables show a consistent relationship with lung cancer but are not always significant, depending on the group being studied. For example, in most models deprivation is an important variable in explaining lung cancer, with more cases of lung cancer in the more deprived areas even when other risk factors are controlled for. In these models, deprivation had an independent effect on lung cancer incidence. However, it has been demonstrated that deprivation is not a significant factor in rural areas or in models explaining male cases of small cell lung cancer. The reasons for this are probably multifaceted, depending on exactly what the deprivation index is measuring (for example provision of health care, quality of diet or amount of exercise taken), and how successful the deprivation index is at measuring deprivation in rural areas.

Similarly, population density tends to be a significant variable in many models, even in models for rural areas, with lung cancer cases being higher in the more densely populated areas, even when other risk factors are controlled for. An urban/rural effect has been consistently observed throughout this study and has been reiterated in the models and analysis of residuals in this chapter. The initial perusal of the lung cancer data demonstrated that lung cancer tends to be highest in urban areas, once the age and sex structure of each area has been controlled for. This has been supported by cluster detection work and the inclusion of the population density variable in most regression models of lung cancer incidence. However, the examination of the residual values has shown that although the very highest residual values are in the urban areas, once the major known risk factors associated with lung cancer have been controlled for, it is the more rural areas that have a higher lung cancer incidence than expected. This is the first study to demonstrate that lung cancer is higher in rural areas once the important risk factors are controlled for. Furthermore, the examination of the residual values in Glasgow demonstrated that the effect of population density was reversed within the city. This suggests that although lung cancer is higher in the more urban areas of Scotland, this relationship is not consistent within Glasgow, probably because higher income groups often occupy high density housing within the city.

Finally, there are also risk factors that are significant in some models but not others, demonstrating the importance of differentiating between the different cohorts of cases. For example, the 1971 coal mining variable was significant in many models, especially models of male lung cancer, but the parameter estimate was negative.

This is surprising, although very few studies have found an explicit link between coal mining and lung cancer. Similarly, it was found that the relationship between lung cancer and radon was not consistent nor was it significant in all models. Perhaps the most important observation, as far as radon is concerned, is the increased incidence of lung cancer among the youngest age groups in the highest radon potential category. This could be because the effect of radon in the older population is masked by the impact of smoking and hence the radon effect can only be detected when the younger age groups are examined independently. Furthermore, consideration of the residual values demonstrates that the total residual values were highly positive in rural areas, which suggest that lung cancer incidence is higher than expected in these areas, and radon is one possible explanation for this. Therefore, this is likely to have wider implications, as it is thought that there may be a significant relationship between radon and the larger population, but it is obscured by methodological limitations that arose from the paucity of radon measurements in Scotland that are available at the household level.

This chapter has demonstrated that there are a number of different risk factors that influence the incidence of lung cancer in Scotland but that these factors behave differently among particular groups of people and in specific geographical areas. The risk factors affect different age and gender groups, geographical areas and types of lung cancer in different ways. Therefore, studies of lung cancer should examine these different groups independently and not treat lung cancer as a homogeneous disease that affects the whole population equally (Payne, 2001). These results have given important insights into the incidence of lung cancer in Scotland, as it has been shown that:

- As expected, lung cancer is strongly dependent on the age and sex structure of an area.
- Lung cancer incidence is higher in urban areas than rural areas before the risk factors are controlled for.
- After controlling for the key risk factors, lung cancer is higher in areas with a higher population density, a proxy for how urban or rural an area is.
- Smoking probability is a powerful variable in all models of lung cancer incidence.
- Once the key risk factors are controlled for, the residual values are higher than expected in rural areas.
- Area deprivation has an independent effect on lung cancer incidence.
- Radon is significant in explaining lung cancer incidence in the younger age groups.
- The residual values suggest that air pollution and radon may be important causes of lung cancer in rural Scotland.
- In urban areas, the residuals demonstrate that there are probably other factors not considered here, which are important in explaining lung cancer in Scotland. Possible explanations include the effects of passive smoking and nutrition.

9. EXPLORING THE INCIDENCE OF LUNG CANCER IN SMALL AREAS ACROSS SCOTLAND: CONCLUSIONS AND FURTHER RESEARCH

9.1 Introduction

Scotland has one of the highest rates of lung cancer in the world. These rates vary between different regions of Scotland, between males and females, urban and rural areas and areas with high and low levels of deprivation. Prompted by these high rates and the geographical variations within Scotland, this study examines the importance of a number of risk factors in explaining lung cancer incidence in small areas across the country. Individual data on patients with lung cancer incidence are supplemented with ecological estimates of exposure to the key risk factors that are potentially associated with lung cancer, to consider the extent to which these factors explain the variations in the disease. This study is the first to carry out an ecological analysis of lung cancer for such small areas in order to reflect important localised variations in lung cancer and associated risk factors.

In this chapter consideration is given to how the risk factors relate to lung cancer and whether these relationships are consistent once other variables are controlled for. The first part of this chapter discusses the conclusions that can be drawn on the geography of lung cancer incidence in Scotland. More specifically, it discusses geographical differences between males and females, different age groups, urban-rural areas and categories of deprivation. The methods used to calculate the potential explanatory variables for small areas across Scotland are then reviewed. The chapter then considers the results of modelling lung cancer incidence in Scotland and

discusses the influence that risk factors such as smoking, air pollution, radon, coal mining and quarrying have upon the incidence of the disease. Conclusions are then drawn on the relationship that deprivation and the urbanness of an area have with the incidence of lung cancer as well as on the key findings from the examination of the residual values. The optimum model of lung cancer is then critically considered and the key findings from the pattern of the residual values that were calculated from the optimum model of all cases of lung cancer are then discussed.

The final section of this chapter will consider how this research can be advanced, firstly by suggesting methods for validating the smoking and pollution estimates and secondly, by proposing methods for calculating additional explanatory variables that have not been considered in this research, such as estimates of asbestos exposure and the quality of diet. Earlier chapters have highlighted the importance of the interrelated issues of the lag effect and migration. The lag effect is particularly relevant in studies of diseases with long latency periods such as lung cancer because changes in the risk factors during the latency period can undermine contemporary measures of exposure. Estimates of exposure to potential risk factors are further complicated if the lung cancer patient has moved between the time of exposure and diagnosis. Therefore, this section will also discuss the implications of these two factors for this study and suggest further work that would aid an understanding of, and ultimately incorporate, their effect. Despite these challenges, this analysis of lung cancer in small areas has provided some new insights into lung cancer in Scotland.

9.2 Geography of Lung Cancer in Scotland

Lung cancer rates vary between different places and at different geographical levels. For example, there are variations between different regions of the world, with more developed countries tending to have higher rates of lung cancer compared to developing countries. However, it has been shown that there are important variations in lung cancer incidence between seemingly similar countries in the same region and also within different parts of the same country. It has also been shown that certain features of lung cancer incidence are common across all scales such as higher rates for males, non professional social classes, more deprived areas and more urban areas.

The rates of lung cancer in Scotland are the highest in the UK and among the highest in the world. Using the world standard population, Scotland had an ASR for lung cancer of 60.09 for males in 1996, which compared to 44.22 in England and Wales and this was the tenth highest male lung cancer rate of all countries in the world. For females, the ASR was 30.86, which compared to 19.60 in England and Wales and globally was the highest rate of lung cancer. However, it has also been shown that there are important geographical differences within Scotland, with high rates in the Central Belt, particularly in Glasgow. In common with other studies, this investigation has pointed to the importance of disaggregating lung cancer incidence by sex (Payne, 2001). It has been demonstrated that males have a consistently higher incidence of lung cancer compared to women. However, chapter four demonstrated that there were also geographical differences in the incidence of lung cancer in Scotland between males and females using maps of counts of cases, age-standardised

rates and Poisson probabilities. For example, it was shown that, for males, the highest SIR values for lung cancer were consistently in the main urban centres of Scotland. However, for women this pattern was a little more dispersed, as there were a number of significant SIR values in the rural areas of the Borders, Dumfries, Angus and the Grampians.

In addition, it was shown that there were distinct differences in the rate of lung cancer between urban and rural areas. The pseudo-postcode sectors (PPSs) in Scotland were divided into quintiles according to the density of the population in each area and it was found that there were more cases of lung cancer in the urban areas in Scotland compared to rural areas, particularly in Glasgow and Edinburgh, for males and females. Furthermore, the maps of the age-standardised rates and Poisson probabilities demonstrated that the areas with significantly high lung cancer rates were mainly located in urban areas. These observations were supported by the work using cluster detection techniques, which showed that, having controlled for age and sex, all of the clusters of cases of lung cancer were located in the most urban areas of Scotland. Therefore, this examination of the geography of lung cancer in Scotland has shown that, regardless of the technique used for analysis, lung cancer incidence is consistently higher in urban areas compared to rural areas.

It was also found that lung cancer incidence varied with levels of area deprivation in Scotland. The PPSs in Scotland were divided into quintiles according to the Carstairs Index. It was found that the count of cases and the SIR of lung cancer were higher in the more deprived areas of Scotland, for both males and females and for all types of lung cancer. For example, the male ASR was 80.26 in the least deprived

category but 132.52 in the highest deprivation category. These results suggest that there is a strong relationship between lung cancer and deprivation, but this raises the question of whether deprivation has an independent effect on lung cancer or if these observations are explained by the key risk factors, such as smoking. The important next step, therefore, was to examine whether these factors are still important once the key factors are controlled for.

9.3 Calculation of Explanatory Variables

Chapter four introduced a dataset on lung cancer incidence that has been provided by the Information and Statistics Division of NHS Scotland for the period 1988 to 1991 (ISD, 2002). There were 18,632 new cases of lung cancer in Scotland during this period, of which 12,073 were male and 6,559 were female. Information on these individuals included their age, sex, tumour morphology, tumour site and the census output-area (OA) in which they lived when diagnosed with lung cancer. Although it has been shown that individual factors such as age and sex are important determinants of lung cancer incidence (Payne, 2001), chapter two highlighted that there are a number of other factors that have also been associated with lung cancer incidence. The key causes of lung cancer were outlined as: smoking; passive smoking; air pollution; radon gas; occupational carcinogens; nutrition; previous lung disease; and a genetic predisposition. Furthermore, chapters three and four noted a relationship between lung cancer and deprivation and also with urbanness that has been highlighted as having an independent effect on the incidence of lung cancer in other studies (Haynes, 1988). Data on these risk factors were not collected for the individuals in the lung cancer dataset and instead the key potential risk factors were

estimated for small geographical areas across Scotland and associated with the individuals using the OA identifier.

Some of the key variables such as the occupation, deprivation and urbanness could be directly calculated for small areas from the 1991 census, whereas others had to be derived using various sources of data. It has been noted that specific occupations, particularly the asbestos industry, are thought to have higher rates of lung cancer. However, comprehensive and geographically-specific data on particular occupation types are not available. Instead, the census disseminates information on occupation for more aggregated types of employment. Therefore, the 1991 census was used to directly calculate the percentage of people employed in the following industries: agriculture, forestry and fishing; energy and water; mining; manufacturing of metals etc.; other manufacturing; construction; distribution and catering; transport; banking and finance; and other services.

This study also considered the relationship between lung cancer and deprivation. Although it has been shown that lung cancer is higher in more deprived areas of Scotland, it is not clear whether this relationship is independent of the key risk factors, particularly smoking. Although previous research on lung cancer incidence has not tended to examine the relationship between lung cancer and deprivation, areal deprivation has been shown to have an independent effect on morbidity in general (Macintyre *et al.*, 1993; Boyle *et al.*, 1999). It is therefore particularly interesting to examine whether the relationship between deprivation and lung cancer remains once the key risk factors have been controlled for. Many measures of deprivation are available, but this study used the Carstairs Index of Deprivation as an

explanatory variable because it is thought to closely capture the concept of material deprivation (Morris and Carstairs, 1991) and it has been used in many previous studies of health and mortality (McLoone and Boddy, 1994).

It has also been noted that there is a clear relationship between lung cancer and measures of urbanness or ruralness in Scotland. Previous research has suggested that there is an urban-rural effect on lung cancer incidence, even once the key risk factors are controlled for (Haynes, 1988). This effect was tested using a measure of population density as a proxy for how urban or rural an area is (Martin *et al.*, 2000) which was calculated using 1991 census data.

The remaining potential explanatory variables could not be directly calculated from comprehensive datasets such as the census, so estimates had to be derived using sample data and various methods of estimation. First, comprehensive datasets on smoking rates are not available for small geographical areas and, although smoking estimates have been calculated for larger census units (Twigg *et al.*, 2000), this is the first time that probabilities have been calculated for such small areas. Furthermore, this has allowed smoking to be controlled for at a level of detail that has not been achieved before in an ecological study. The Scottish Household Survey was used to model smoking behaviour using multilevel modelling techniques and the parameters from this model were used to provide estimates of smoking probability in OAs in 1991 for 12 age/sex groups (Pearce *et al.*, 2003). In common with previous studies, it was found that smoking probabilities were higher in more urban and deprived areas and lower in more suburban and rural areas (Twigg *et al.*, 2000).

Second, the role that pollution plays in explaining lung cancer incidence in Scotland was explored by calculating pollution levels in 1971 and 1991. It was important to estimate pollution levels for these two periods because of the large reduction in pollution emissions between 1971 and 1991 (NETCEN, 2001). Estimates of air pollution were calculated using government collected data on 'black smoke', which is the pollutant that has been most strongly related to the incidence of lung cancer (Arden Pope and Dockery, 1999). Data on black smoke have been collected at monitoring stations across Scotland for a number of years. Interaction data on the workers in PPSs in Scotland, taken from the 1991 census, were used to predict the levels of black smoke pollution in these two years. The parameters from this model were then applied to all PPSs to predict pollution levels across Scotland. Although it was thought that the 1991 predictions were more accurate due to the temporal overlap between the explanatory and dependent variables, it was also important to incorporate the less accurate predictions of pollution in 1971 into the analysis, due to the wide-scale reduction in pollution levels during this period.

Third, the relationship between lung cancer and radon exposure has long been debated in the literature (Darby, 1999) but there has been little work that has examined this in Scotland, nor have ecological studies examined the relationship in such small geographical areas (Etherington *et al.*, 1996). This thesis has highlighted the difficulties of estimating exposure to radon because of the interaction of a multitude of factors that determine the radon levels (Appleton and Ball, 1995). This was exacerbated by the lack of radon measurements that have been taken in Scotland and the refusal of permission by the NRPB to use the radon data that have been collected. Therefore, this study has used an approach developed by the British

Geological Survey which predicted radon levels using lognormal models to provide estimates of the percentage of the housing stock above the UK Action Level (200 Bq m^{-3}) for each geological unit and drift deposit combination, in order to delimit areas of high and low radon potential (BGS, 1998). Therefore, each OA was assigned a radon potential category depending upon the combination of geology and drift found there. This is more sensible than other approaches that have been taken, such as using grid squares or administrative boundaries, because it reflects the natural differentiation in radon levels (Miles, 1998).

Fourth, chapter two demonstrated that the link between coal mining and lung cancer is poorly understood as some studies have found a relationship between the two (Une *et al.*, 1995), whilst many more have found no evidence for a link (Leigh *et al.*, 1994; Wang *et al.*, 1997). Estimates of the levels of coal mining activity were calculated for 1991 PPSs using the data on the percentage employed in mining for 1971 census enumeration districts as a proxy for coal mining. It was important to use the 1971 census rather than the census in 1991 because the numbers of people employed in coal mining in 1991 were a fraction of those that were employed in 1971, due to the rationalisation of the coal industry (Hollywood, 2002).

Fifth, it was reported in chapter two that there has been some research into the health effects of quartz, particularly in an occupational setting (e.g. McDonald *et al.*, 2001). This study examined the effects of quarrying activity on lung cancer by estimating the exposure to different types and combinations of quarries using kernel estimation (Bailey and Gatrell, 1995). This enabled estimates of the level of exposure to different types of quarries to be predicted for people living within one kilometre and

two kilometres of a quarry. This is the first time that the effect of quarrying activity on the incidence of lung cancer has been examined among the wider population in an ecological study of this type.

As noted above, the data on lung cancer incidence were geographically referenced by the 1991 census OA in which the patient was diagnosed with the disease. Therefore, in order to examine the extent to which the risk factors explain lung cancer incidence in Scotland the explanatory variables had to be calculated for a consistent geography. Therefore, each of the estimates of the risk factors that are potentially associated with lung cancer incidence in Scotland were either derived, or directly calculated, for 1991 OAs or PPSs.

9.4 Explaining Lung Cancer Incidence in Scotland

The data on lung cancer incidence in each OA were combined with population data and the derived and non-derived explanatory variables. The count of cases of lung cancer, the population data and smoking data were age and sex-specific and hence each OA was represented 12 times, once for each age and sex group. The data were incorporated into a set of Poisson regression models to examine the extent to which the explanatory variables explained the geographical variation in the incidence of lung cancer in these areas. Each of the variables were examined independently and in tandem with other variables in order to develop a parsimonious model and the key findings from this process are discussed here.

9.4.1 Population Structure

All of the models of lung cancer found a consistent relationship between sex and lung cancer with higher rates for males than females, which was consistent with the earlier discussion (Babb *et al.*, 2001). Furthermore, in line with previous work, the differences between males and females remains once the other risk factors are controlled for (Payne, 2001). Age also demonstrated a consistent relationship with lung cancer throughout this study. Lung cancer is extremely rare under the age of 50 and is most common between the ages of 61 and 80 and this is consistent for both males and females. Furthermore, this relationship remains once the risk factors have been controlled for.

9.4.2 Smoking

The comparison between the smoking probabilities, calculated using the records from the Scottish Household Survey, and the SIRs for lung cancer in small areas in Scotland found that lung cancer rates increased with higher probabilities of smoking. Furthermore, regression models of lung cancer incidence that controlled for the demographic profile of the area found that the age and sex-specific smoking rates were significant in explaining variations in the incidence, with higher rates of lung cancer in areas that had greater probabilities of smoking. The smoking variable was consistently significant in explaining lung cancer, regardless of the subgroup of people studied, how urban the area was or the type of lung cancer modelled. This is in line with many studies that have investigated the relationship between smoking

and lung cancer over many years (Doll and Hill, 1950; Wynder and Graham, 1954; Shairer and Schöniger, 2001).

9.4.3 Air Pollution

The relationship between lung cancer and air pollution continues to be debated in the literature, mainly due to of the difficulty in disentangling the effect from smoking behaviour (Doll, 1978) and due to the problems associated with estimating exposure (Dunn and Kingham, 1996). This study has shown that air pollution is a significant cause of lung cancer in Scotland, and this is consistent for both sexes and all age groups. The regression models of lung cancer incidence examined the effect of adding the 1971 and 1991 estimates of air pollution to different models and found that, despite the greater accuracy of the 1991 predictions, only the 1971 estimates of pollution were significant in explaining lung cancer incidence in 1991. In line with previous studies, this suggests that it is especially important to incorporate the lag between exposure and cancer initiation in an analysis that considers the effect that air pollution has upon lung cancer (Pike and Forman, 1991).

9.4.4 Radon

The literature examining the relationship between lung cancer incidence and radon gas is inconclusive (Darby, 1999), particularly in Scotland where the health effects of radon have received very little attention (Green *et al.*, 1996). This study found that radon was not significant in any of the models of lung cancer incidence, except when the youngest age groups were considered separately. This suggests that radon

may only be important in explaining 'premature' cases of lung cancer. This may be because it is a risk factor that people can be exposed to at an early stage in life, whereas exposure to some of the other risk factors, such as smoking, tend not to commence until later in life. Alternatively, it may reflect the lower levels of migration among younger people compared to older people, that have meant that they are more likely to reside in the same area in which they were exposed to the high radon levels as they were living in when they were diagnosed with lung cancer. Both suggestions are complicated by the difficulties of calculating individual or ecological estimates of the exposure to radon gas (Darby *et al.*, 1998) and the potentially synergistic relationship between radon and smoking (Lichtenstein *et al.*, 2000). However, if either of these suggestions are true then this research has provided some evidence of a link between radon and lung cancer incidence which, at the very least, calls for further research into the health effects of radon in Scotland among the young.

9.4.5 Coal Mining

A significant relationship between lung cancer and coal mining was found in this study, but it is counterintuitive because the incidence of lung cancer was lower in areas where the levels of mining were high. This suggests that coal mining is not an important cause of lung cancer, which is consistent with a number of previous studies (Miller and Jacobsen, 1985; Wang *et al.*, 1997). This result is difficult to interpret but it may arise because the measure of mining includes all people employed in the mining industry and not just those directly employed in mining coal. The rates of those employed in other types of mining tend to be highest in more rural

areas of Scotland where employment in coal mining and lung cancer incidence are relatively low.

9.4.6 Occupation

Chapter two highlighted that there are a number of occupations that have a higher incidence of lung cancer, due to the exposure to particular carcinogenic substances such as asbestos, arsenic, chloromethyl methyl ether and chromium compounds (Schottenfeld, 1996). However, many of the industries that use these substances, such as the rubber, iron and steel industries, are not major employers in Scotland, and those industries that are present are found in specific locations.

The effect of employment type on lung cancer incidence was tested using the occupational variables from the 1991 census. The results demonstrated that there was not a consistent occupational effect and the variables that were intended to represent the potentially important industries, such as those employed in manufacturing and construction, were not significant in any of the models of lung cancer incidence. However there have been a number of industrial changes and alterations to working practices during the lag period (1971 to 1991). Therefore, contemporary measures of employment may be inappropriate and it may require measures of occupation 20 years prior to the lung cancer data to examine the occupational risk factors that are potentially linked to lung cancer.

9.4.7 Quarrying

Although the relationship between quarrying activity and lung cancer in the wider population has not been examined before, previous work has suggested that particular occupational groups that are exposed to high levels of quartz, such as those working in industrial sand plants, may have higher rates of lung cancer (McDonald *et al.*, 2001). The results from this study found no relationship between any type of quarrying activity and the incidence of lung cancer in Scotland and this was consistent for all age and sex groups, in rural areas and for different types of lung cancer. The results from this analysis suggested that the release of respirable particles into the atmosphere from quarrying activity does not influence the incidence of lung cancer in the surrounding population and hence could be confined to the workers in these industries.

9.4.8 Deprivation

Discussion in early chapters of this thesis demonstrated a clear relationship between area deprivation and lung cancer incidence in Scotland. Chapter four demonstrated that the SIR values were higher in the more deprived areas compared to the less deprived areas for both males and females. However, it was not clear whether this reflected the higher rates of smoking and other risk factors that tended to be found in more deprived areas (Duncan *et al.*, 1999) or whether deprivation had an independent effect on lung cancer incidence. The regression analysis found that the Carstairs Index of Deprivation was consistently significant in the models of lung cancer incidence, even when other factors such as smoking and population density

were controlled for. The parameter estimate was consistently positive which suggests that lung cancer was higher in more deprived areas. This is consistent with other studies of lung cancer that have found deprivation to be significant (Hart *et al.*, 2001), but this is the first nationwide ecological study to explicitly examine this effect in small areas, controlling for other factors such as smoking.

There are general explanations of why deprivation might be related to morbidity and these include the limited access and provision of facilities and services, poorer job prospects, higher levels of crime and violence, lack of social interaction and participation, and feelings of exclusion, stigmatisation and abandonment which could each have a direct impact on health outcomes (Stead *et al.*, 2001). However, it is less clear why deprivation is related to lung cancer specifically, although deprivation could partly be acting as a surrogate for other risk factors that have not been identified. For example, it could be correlated with factors such as passive smoking and poor diet which are likely to be higher in more deprived areas.

9.4.9 Urban / Rural Effects

The preliminary analysis in this research suggested a consistent urban bias in lung cancer incidence as the SIRs were higher in more urban areas compared to rural areas both before and after controlling for the age and sex structure of an area. Furthermore, once other known risk factors such as age, sex, smoking behaviour and pollution had been controlled for in regression models, population density retains its significance. This suggests that there is still an important urban effect that is not explained by the known risk factors, which is consistent with other studies of urban

lung cancer (Haynes, 1988). Two key explanations were suggested for this observation. First, levels of passive smoking are likely to be higher in urban areas compared to rural areas (Matsukura *et al.*, 1984). Second, the urban effect may be exaggerated by the effect of selective migration because people who move from urban to rural areas are thought less likely to become ill than those who remain (Verheij *et al.*, 1998). Therefore, the lung cancer rates in urban areas may be raised because the people who remain in these areas are more likely to contract the disease (Haynes, 1988).

In addition, the separate analysis of the rural cases of lung cancer in Scotland showed that the risk factors are subtly different for urban and rural lung cancer. Most noticeably, there are fewer risk factors associated with rural lung cancer, and smoking is the most important risk factor that explains much of the variation in rural areas. For example, the Carstairs variable was not significant which suggests that deprivation is not an important factor in explaining rural lung cancer cases or that it is not an appropriate measure of rural deprivation. It was interesting to note that the potential risk factors that tended to have higher levels in rural areas of Scotland, such as radon and quarrying activity, did not have a significant effect on rural cases of lung cancer in this analysis. This study suggests that neither of these factors are important causes of lung cancer in Scotland.

9.4.10 Age and Sex

It was also found from the separate models of different age-sex groups that the risk factors were broadly similar to those for the models of all cases of lung cancer,

although there were some minor differences between males and females and between older and younger age groups. For example, it was found that the only variation between males and females was that different occupational variables were significant. Similarly, population density was significant in explaining lung cancer in the older age groups but not for the younger age groups which suggests that the urban effect may be confined to the older population. Perhaps most importantly, it was found that radon was only significant in explaining lung cancer in the younger age groups. However, variables such as smoking, deprivation and air pollution are consistently significant for all sub groups of the population. These results demonstrate that there are only minor variations in the risk factors that cause lung cancer for different age-sex groups in Scotland.

9.4.11 Lung Cancer Type

Lung cancer types can be broadly categorised into small cell and non-small cell, reflecting differences in their cell type and prognosis but also subtle differences in their aetiologies (Simmonds, 1999). This study has shown that models that considered the two types of lung cancer separately found that the variables in the models were broadly the same, which is in line with other studies that have suggested that the causes of the two types of lung cancer may be similar (Boyle *et al.*, 2000). There are some differences in the risk factors associated with them, such as population density being significant in explaining non-small cell lung cancer but not small cell lung cancer. There are also subtle differences between males and females for small cell and non-small cell lung cancer. For example, deprivation was significant in explaining female cases of small cell lung cancer but not for males.

9.4.12 Analysis of Residual Values

The residual values from the optimum model of all cases of lung cancer incidence in Scotland were calculated and then mapped. This identified areas that had a large discrepancy between the observed and expected values in the model and hence highlighted areas with a higher or lower incidence of lung cancer than was predicted by the model. Mapping the residual values helped to identify other factors that may be important in explaining the variations in lung cancer incidence in Scotland (Lovett and Flowerdew, 1989) and it also identified areas where lung cancer is particularly high, having controlled for the key risk factors. This is useful information as it allows areas with a particularly high incidence to be targeted in public health campaigns (HEBS, 2003).

It has been demonstrated that there is an urban effect on lung cancer incidence that is independent of the key risk factors, such as smoking and air pollution. The importance of the urban-rural effect was reiterated by the initial examination of the residual values, which suggested that the summed deviance values and Pearson residual values were highest in the most urban Health Boards in Scotland. Furthermore, mapping of the values showed that the highest deviance and most extreme Pearson residual values tended to be located in the urban centres. However, the large urban areas also contain the majority of the OAs in Scotland and once the total deviance value and Pearson residuals were calculated for different urban-rural categories it was found that there was a *rural* bias in the size of the values once the key risk factors had been controlled for. The sum of the Pearson residuals in the

most urban areas was negative, whereas in the most rural areas the sum was positive. This is an important finding because it suggests that once the key risk factors are controlled for, it is the rural areas that have a greater number of cases of lung cancer than would be expected, whilst the urban areas have a smaller number of cases. Although other studies have demonstrated urban and rural variations in health that are independent of differences in deprivation (e.g. Senior *et al.*, 2000), this research is the first to point towards higher lung cancer incidence in rural areas compared to urban areas once the key risk factors have been controlled for.

Although these broad trends in lung cancer incidence are useful, it is also important to examine the spatial distribution of individual residual values to provide insights into localised potential risk factors. The spatial distribution of the residual values suggested that other risk factors could be important in explaining the incidence of lung cancer in Scotland. The maps of residual values have shown that, despite the broad results for urban and rural areas, most of the highest values tend to be in the urban areas of Scotland, especially in Glasgow. Furthermore, the highest residual values in Glasgow tended to be in areas of high deprivation. One explanation for this could be differences in diet which, as discussed in chapter two, has both a causative and a preventative effect. For example, a diet that is high in cholesterol or saturated fat can significantly increase the incidence of lung cancer, whilst a diet that is high in fresh fruit and vegetable consumption can have a preventative effect (Schottenfeld, 1996). It is likely that the consumption of cholesterol and saturated fat is higher, and the consumption of fresh fruits and vegetables lower in the deprived urban areas (Ellaway and Macintyre, 1996; Ecob and Macintyre, 2000), which may help to explain why these areas had a higher incidence of lung cancer

than was predicted by the model. This was supported by the pattern of the residual values in Glasgow as it was found that the areas with high positive Pearson residual values tended to be deprived areas.

The distribution of residual values could also reflect the effect of passive smoking. This is a risk factor that is poorly understood and although it is likely to be strongly correlated with smoking, it may have a greater effect in the most urban areas (Jee *et al.*, 1999). Therefore, in some areas where the rates of smoking are high, there could be an additional effect from passive smoking that has not been controlled for but is reflected in the extreme residual values.

Asbestos is thought to have an effect on the incidence of lung cancer in specific geographical areas (Albin *et al.*, 1999). In Scotland, levels of asbestos exposure were thought to be relatively high along the River Clyde, which was a major shipbuilding centre (De Vos Irvine *et al.*, 1993). However, this is not explained by any of the 1991 occupational variables used in this analysis, such as the proportion of the workforce employed in the construction industries. This is probably due to the decline of the ship building industry in the 1960s, 1970s and 1980s, which meant that the level of employment in the construction industries was relatively low in this area by the 1991 census. This is reflected in the reduction in the number of people employed in the shipbuilding industry, which has declined from a post war high of 59,600 people in 1955, to 34,600 in 1975, and to just 10,100 in 1995 (Dodds and Maguire, 1998). Therefore, the influence of the shipbuilding industry may not be explained by the 1991 occupational variables and hence this may explain the location of some of the high residual values that were identified in Glasgow.

Although many of the extreme residual values are in the larger urban areas, it has been noted that the total residual values are highest in the most rural parts of Scotland, controlling for the established risk factors. Therefore, there may be important risk factors that are not controlled for within rural areas, which might be explained by factors such as radon levels and quarrying activity. Although this study has attempted to control for these effects, the challenge for future research is to develop methods to control for these risk factors with enough precision to allow their influence on lung cancer to be better understood.

9.5 Critical Assessment

This is the first study to carry out an ecological analysis of lung cancer incidence in areas that are small enough to reflect localised differences in the prevalence of the disease and the associated risk factors. As the previous section has highlighted, this approach has contributed to our understanding of lung cancer and the associated risk factors in a number of different ways. Section 8.4.12 presented an optimum Poisson regression model of lung cancer incidence in Scotland that included all of the significant variables and the significant interaction, and which had the lowest deviance value, given the number of degrees of freedom. The model demonstrated that age and sex were particularly significant in explaining variations in the disease but that the Carstairs Index, smoking, population density, the 1971 estimate of pollution, mining in 1971, the percentage employed in the energy sector, the percentage employed in other service industries and the interaction between smoking and sex were also significant. The estimates of radon potential and the quarrying

variables were not significant in the optimum model. Although the results from the optimum model have provided new insights into the causes of lung cancer in Scotland, there are a number of potential limitations that should be noted.

This study raises issues that are not only relevant to studies of lung cancer but also to other cross-sectional studies of health and ill health. The first set of issues is concerned with the problems presented by the lag period between exposure to a potential risk factor and the diagnosis of the disease, and the interrelated issue of migration. These issues are particularly relevant to studies of diseases with long lag periods such as lung cancer. Secondly, the research is complicated by the difficulties faced when calculating estimates of each of the potential risk factors associated with lung cancer. These challenges arise either because the data available included insufficient detail, or were not geographically specific, or else the access to the data was limited.

Individual-level lung cancer data, which includes historical information on the potential risk factors associated with the disease, are not collected in Scotland and hence ecological estimates of the risk factors are used instead. Earlier chapters have highlighted two complications that arise when employing ecological estimates of the risk factors. First, there is a latency period of approximately 20 years between exposure to potential risk factors and diagnosis (Williams, 1992). If there have been changes in the risk factors during the latency period then contemporary measures could be inappropriate. Second, the latency period is compounded by the effects of migration, as although the effect of the latency period may point to the use of lagged instead of contemporary variables, these may also be inappropriate if the lung cancer

patient has moved between the period of exposure and the time of diagnosis. This conundrum represents a challenge to all studies of the incidence of an illness that has a long lag period, such as lung cancer, and has bedevilled many studies in the past (Bentham *et al.*, 1995, Boyle *et al.*, 1999).

Although it is difficult to quantify the influence of the lag effect and migration on the results in this study, it is likely that the lack of individual data that includes information on personal exposure to the potential risk factors affects the confidence than can be placed in the optimum model. The optimum model of lung cancer had a low pseudo-R^2 value (0.274) which suggests that much of the variation in lung cancer incidence in Scotland remains unexplained (although, as discussed in section 8.4.1, the interpretation of the pseudo-R^2 value is problematical due to the underdispersion issue). This may be due to the use of ecological estimates of the risk factors rather than personal exposure measurements. Ecological estimates are inherently less reliable in the modelling procedure than information on individual characteristics. Therefore, it is difficult to be fully confident that factors such as smoking are fully controlled for, especially when contemporary smoking estimates are used rather than estimates for the 20 years prior to diagnosis with the disease. Furthermore, the use of ecological estimates makes if difficult to disentangle the effects of unconfirmed factors such as radon or quarrying exposure from the well established factors, particularly smoking behaviour.

The research is further complicated by the specific difficulties involved in estimating particular risk factors. Although the best estimate was calculated for each risk factor, a number of complications arose in response to imperfect datasets. The

estimates of smoking behaviour were calculated for small-areas across Scotland, using a multilevel model. The multilevel modelling approach allowed for the compositional and contextual determinants of smoking to be captured in a model and then the parameters from this model were used to estimate smoking behaviour in OAs across the country for 12 age-sex groups. Although the multilevel modelling approach is sensible, only a small number of individual-level variables were available (age, sex, smoking status) because of confidentiality restrictions, and hence OA variables were used as proxies for level one characteristics. However, previous work has noted the importance of other individual-level characteristics such as social class, marital status and ethnicity in explaining smoking behaviour (Twigg *et al.*, 2000). The use of OA variables instead of individual characteristics is likely to underrepresent the level one variation. Furthermore, it has been suggested that the higher level units used in multilevel models tend to utilise data collection areas (such as census areas) rather than reflect important sociological units that may explain differences in health behaviour (Mitchell, 2001). Nonetheless, administrative units such as census units are usually derived with some consideration of their sociological significance. These uncertainties are compounded by the lack of an alternative smoking dataset in Scotland that is geographically specific, which could be used to assess the accuracy of the estimates.

In addition, it would have been desirable to estimate smoking behaviour 20 years prior to the lung cancer data set in order to incorporate the lag between exposure and cancer initiation. However, there is not a dataset available for the earlier period that is as comprehensive and geographically specific as the one obtained from the results of the Scottish Household Survey. Furthermore, the 1991 estimates of smoking are

likely to be a good predictor of past smoking rates, because if a person is living in an area with a high proportion of smokers then there is a high probability that they smoke and that they smoked earlier in life too. This is reflected in how powerful this variable has been in explaining lung cancer incidence in 1991 throughout this study.

Similarly, a number of difficulties arose when estimating the exposure to other risk factors. It has been noted that the pollution estimates have been calculated for PPSs due to the paucity of pollution data that is collected across Scotland and because the explanatory data on workers in an area is only released at this level. However, it is probable that there is some variation in pollution levels within these areas and that the use of PPSs may be less appropriate than OAs as they do not capture these subtle but important variations. Furthermore, the 1971 estimates of pollution were thought to be less reliable than the 1991 estimates due to the use of 1991 data as explanatory variables. It is important to use a lagged pollution variable in an analysis of lung cancer in Scotland because there has been a wide-scale reduction in pollution levels across the country between 1971 and 1991, which has resulted in relatively low and uniform levels in 1991 (Lee, 1994). This reflects the shift away from heavy industry and steel and coal production to electronics and service industries, particularly tourism, during this period (Scottish Office, 1997). Furthermore, chapter six demonstrated that in 1971 there was a greater variation in the pollution levels between areas of Scotland, largely because some areas had much higher levels of pollution than in 1991. Consequently, past levels of air pollution are likely to be significant in explaining contemporary lung cancer rates. Given that the 1991 pollution variable was not significant at any stage of the analysis, it seems that if

black smoke levels remain at their present levels or continue to reduce, then pollution is unlikely to be a significant cause of future lung cancer cases in Scotland.

There were also difficulties in estimating radon exposure due to the localised variations in radon levels in response to small scale fluctuations in solid geology, drift deposits and housing characteristics. This is compounded by the unwillingness of the NRPB to release the radon data that has been collected. Therefore, although the results from this study suggest that radon is only a cause of lung cancer among younger people, the difficulties in estimating radon exposure suggest that an individual-level study of the health effects of radon may be necessary to ascertain that it is not a cause of lung cancer among the wider population.

The estimation of the coal mining and quarrying variables posed problems due to the lack of specific data relating to each. It was important to use data on coal mining from 1971, given the restructuring of the coal mining industry during the period prior to the 1991 (Hollywood, 2002). However, using 1971 data is problematical as it is incompatible with the 1991 data due to changes in the census boundaries during this period. Furthermore, the coal mining industry is aggregated with all other types of mining in the dissemination of the 1971 census. Given these issues, the coal mining variable is unlikely to fully capture the extent of the coal mining industry in 1971, which may help to explain why the parameter estimate was negative in the optimum model of lung cancer. Similarly, the estimation of the exposure to different types of quarrying activity was limited by the data on quarries that is available (Jones *et al.*, 1991). In the absence of quarry output data, the kernel estimation procedure assumed that the production level at each quarry was identical when in fact there is

likely to be considerable variation between quarries of different sizes. This may have implications for this analysis as it is possible that in certain locations the estimation procedure has not captured the full extent of exposure to quarrying activity and hence this hazard could be a cause of lung cancer only where the output is of sufficient magnitude.

9.6 Further Work

The previous section highlighted a number of limitations that are specific to a particular aspect of this research, such as the methods used to estimate some of the risk factors, and also more general issues relating to the lag effect and migration. The most appropriate method to address many of these issues would be to design a prospective case-control study for the whole of Scotland that contacts a large group of people and tracks them over a long period of time, collecting detailed information such as data on their smoking behaviour, place of residence, occupation, diet and exposure to pollution, radon and quarrying. This information could be used to examine precisely how the potential risks influence the incidence of lung cancer in Scotland. However, this would be very expensive to collect and would require many years to complete. Given these restrictions, there is, instead, potential for further work that could help to address a number of the key issues, without being prohibitively expensive and time consuming. Methods could be developed to improve the explanatory variables used here and also to calculate estimates for additional potential risk factors. Furthermore, future work could also focus on some issues that have posed consistent challenges to many features of this analysis, most noticeably the lag effect and the interrelated influence of migration.

9.6.1 Further Exploration of Risk Factors

Further research could usefully validate and improve upon the methods of estimating the level of exposure to some of the explanatory variables used in this research. In addition, research could focus upon estimating other potential risk factors that were highlighted by the examination of the pattern of residual values such as nutrition, passive smoking and asbestos exposure. The additional explanatory variables could be examined to see whether they are significant in explaining the variation in lung cancer incidence in Scotland.

Validation of Explanatory Variables

Given the absence of individual data on the exposure to the key risk factors associated with lung cancer, chapters 5 to 7 have suggested methods for estimating smoking behaviour, pollution levels (1971 and 1991), radon potential, employment in coal mining and quarrying activity. Although an appropriate method has been developed to estimate each risk factor, the limitations of each approach have been noted in the relevant chapter. Given the inherent inaccuracy in the estimates and the subsequent uncertainty in the interpretation of the variables in the models of lung cancer, it would be useful to develop methods to validate the estimates of the risk factors.

The absence of accurate and geographically disaggregated estimates of smoking behaviour in Scotland makes validation of the smoking estimates used in this study

difficult. However, the estimates can be compared to the age and sex-specific data collected General Household Survey (General Household Survey, 1999), although a comparison would only be possible at a cruder scale because the GHS data is only released for the larger standard regions rather than OAs. More usefully, future work could compare the estimates calculated here to the results of work currently being undertaken by Moon, Twigg and colleagues at the University of Portsmouth to calculate synthetic data on a number of health behaviour characteristics in Scotland, including smoking (Moon, personal communication).

This study has demonstrated that air pollution is a significant factor in explaining lung cancer. However, although the pollution estimates are intuitively reasonable, it is not clear how accurate they are, especially as this is the first time that pollution levels have been estimated using socioeconomic data. Ideally, it would be useful to validate the methods used to predict the pollution estimates used in this study. However, this is impossible due to the lack of alternative pollution datasets in Scotland. Instead, an assessment could be made of the methodology used by calculating estimates of pollution for 2001 and comparing them to data collected in the field. The same methodological approach could be taken to model air pollution levels in 2001, using data from the National Air Quality Information Archive (NETCEN, 2001) and the 2001 SWS data. The predictions could be validated by comparing them to pollution data that is independently collected in the field. If there is a close correspondence between the pollution values predicted by the model and the pollution measurements taken in the field then this would add a degree of confidence in the methodology used to calculate the estimates used in this study.

One of the initial driving forces for this piece of research was to investigate the geological causes of lung cancer, particularly the effect of radon and quarrying activity. This study has found some evidence of a link between radon and lung cancer incidence in the younger age groups. Furthermore, the total Pearson residual value was highly positive in rural areas, which suggested that lung cancer was higher than expected in these areas. One possible explanation for this is that radon does raise the incidence of lung cancer in rural Scotland but the variability in radon levels is not fully captured by the methods presented here. Predicting radon levels has been problematical due to the small-scale fluctuations in radon that make it difficult to precisely estimate exposure (Miles, 1998). Furthermore, the quantity of radon readings that have been taken in Scotland, and the difficulty in obtaining the existing data, have compounded this problem. This study has pointed towards a need for a systematic and comprehensive radon data collection strategy in Scotland and for these data to be made available to the research community. This will enable more precise estimates of radon to be calculated, particularly in areas where classification had not been possible due to a lack of radon data. In addition, past studies have shown that radon levels fluctuate within the same combination of geology and drift in response to variations in the housing type (Gunby *et al.*, 1993). Therefore, models of radon potential for census units could usefully incorporate data on the type of housing as an explanatory variable to differentiate between OAs within the same radon potential category. However, this would be complicated by fluctuations in radon levels between seemingly similar buildings in response to different levels of

insulation and ventilation as well as small-scale variations in the lithology and permeability of the local geology (Appleton and Ball, forthcoming).

The relationship between coal mining activity in 1971 and lung cancer incidence in Scotland during the period 1988 to 1991 was difficult to interpret because in the models where the coal mining variable was significant, the parameter estimate was negative. This suggested that lung cancer incidence was lower in areas where coal mining activity was high, which is hard to explain given that studies of lung cancer incidence and coal mining have either found that incidence of the disease is higher in coal mining communities (Une *et al.*, 1995) or that there is no relationship (Wang *et al.*, 1997). The estimate of coal mining used in this study is limited by the aggregated nature of the occupational variables in the 1971 census which meant that instead of identifying only those employed in the coal mining industry, all people employed in any type of mining were included. This resulted in high percentage values in areas outside of the former Scottish coal mining communities in places such as Mull of Kintyre, where other types of mining are located (section 7.7). Further work could usefully recode the mining variable in order that only the areas within or close to the coal measures have a non zero value. This may help to include only those employed in coal mining in the analysis rather than those employed in all types of mining industries. However, even this approach will not allow for those directly employed in mining operations to be identified rather than those working in other aspects of the coal mining industry (e.g. administration and management).

This study found no evidence that quarrying activity increases the incidence of lung cancer among the resident population, although the estimates of exposure to

quarrying were limited by the information available about quarries. Past studies have found that inhaling quartz in an occupational setting increases the incidence of lung cancer (McDonald *et al.*, 2001), which suggests that further research is necessary to be confident that quarrying does not increase lung cancer in the population who live near to quarries. Additional information on the quarries could be collected, particularly quarry output and historical data on when quarrying commenced and was completed. The kernel estimation procedure, used in this study, can then weight the quarries according to their total output and the length of time during which quarrying has taken place (Bailey and Gatrell, 1995), which will enable more accurate estimates of the total exposure to quarrying activity to be made. However, even if these data were collected, due to the small numbers of people involved, it may require an individual-level study to be confident that quarrying activity does not increase lung cancer incidence. An individual-level study would allow the residential history of each patient to be tracked, their exposure at each residential location to be calculated and hence their total exposure to quarrying activity could be more accurately estimated. However, even then it is difficult to be precise about the individual exposure to quarrying activity, given that quarrying activity is likely to vary over short periods of time due to daily patterns of movement among individuals (Gatrell, 2002) and the complicated interaction of meteorological and atmospheric factors that influence dispersal of quartz particles into the atmosphere.

The examination of the residual values has highlighted that there are particular areas in Scotland that have high or low residual values that may be explained by other risk factors that were not incorporated into the optimum model of lung cancer incidence. Therefore, this research could be extended by developing methods to estimate the exposure to additional suspected causes of lung cancer. More specifically, the spatial pattern of the residuals suggested that the exposure to nutrition, passive smoking and asbestos could usefully be measured. Estimates of nutrition could be made using a study such as the Scottish Health Survey (Shaw *et al.*, 2000) or the Scottish Household Survey (Dixon and Finlayson, 2000), which include questions about the consumption of different types of food such as fats, dairy, fruit and vegetables (Deepchand *et al.*, 2000). If specific nutritional data are obtained for small geographical areas, then different dietary factors could be modelled using comprehensive datasets such as the census. The parameters from this model could be used to make estimates for all areas of Scotland using a procedure similar to that employed to predict smoking behaviour and levels of air pollution. However, estimates of nutrition were not calculated in this study because the literature has shown that it likely to have only a minor effect.

Passive smoking is thought to be strongly correlated with smoking (Jee *et al.*, 1999) and chapter five has shown that levels of smoking are highest in the most urban and deprived areas in Scotland. Further work could explore whether the effect of passive smoking is greater in the areas where smoking is particularly high. Therefore, the areas with the highest smoking probabilities could be weighted to explore the effect

of passive smoking. However, measuring exposure to passive smoking is likely to be more complicated as the effect is likely to be confined to non-smokers (Stockwell *et al.*, 1992) and hence in order to fully explore the effect, an individual-level study would be needed to collect personal information on exposure levels to ambient smoke.

Previous studies have shown that asbestos has an important influence on lung cancer incidence among former workers in particular types of industries, such as shipbuilding (De Vos Irvine *et al.*, 1993). This may be an important cause of lung cancer in particular parts of Scotland, especially in Glasgow, where shipbuilding was a major employer (Dodds and Maguire, 1998). However, although the occupational variables from the 1991 census coincided with the data on lung cancer incidence, they may not incorporate this effect due to industrial changes and alterations in working practice that have taken place in recent years. Therefore, estimates of occupational exposure to asbestos could be better calculated from previous censuses, such as in 1971, and these estimates could be incorporated into the models of lung cancer incidence to see whether or not they are significant. However, although this approach would be useful, it is limited by the occupational classification used in 1971, which did not include a specific variable on those employed in shipbuilding. Instead a more aggregated measure would need to be used, such as the proportion of people employed in the construction or manufacturing industries.

9.6.2 Longitudinal Analysis

A recurring theme in this thesis has been whether to use contemporary measures of the risk factors that are associated with lung cancer, or instead to use estimates of the risk factors that incorporate the lag between exposure and diagnosis with the disease. Measures that coincide with the lung cancer data have the advantage of estimating the exposure that the population was subjected to at the time of diagnosis, but does not account for changes in the risk factor between the crucial period of exposure and the time of diagnosis. On the other hand, ecological measures that do incorporate the lag period have the advantage of reflecting the changes that have occurred in the risk factors but do not account for the fact that the patient may have moved during the lag period.

This issue has wider implications in the fields of medical geography and epidemiology due to data constraints that are placed upon researchers, which are designed to protect an individual's confidentiality. This means that data on the migration history of individuals with a disease and information on their personal exposure to the risk factors are rarely available. For example, various cross-sectional studies have explored the relationship between particular health outcomes, such as limiting long-term illness (Bentham *et al.*, 1995), and specific explanatory variables such as smoking or deprivation, without being able to consider the potential lag between the exposure to the risk and the diagnosis of the illness (e.g. Boyle *et al.*, 1999; Saul and Payne, 1999). Although the importance of this effect will vary by condition, it is relevant to many cross-sectional studies of health and ill-health.

This problem has been especially pertinent to this study because the approximated lag time of 20 years between exposure and the initiation of lung cancer has been well documented (Williams, 1992). Furthermore, the risk factors that have been associated with lung cancer have seen changes in their levels during the time between the key period of exposure and diagnosis. For example, levels of smoking have decreased over the past 30 years in response to cultural changes and a greater awareness of the consequences of cigarette consumption for health (Bridgwood *et al.*, 1998). Similarly, the average level of air pollution has decreased across Scotland during the same period (Lee, 1994).

If individual-level data on the people diagnosed with lung cancer, including full life histories, were available, then this issue could be fully addressed. However, in the absence of this information the issue could be examined through a longitudinal analysis of lung cancer patients. One data source that would allow such a study would be the England and Wales longitudinal study. The survey uses census records to track 800,000 individuals in 1971, 1981 and 1991 and links various census variables with data for various events such as births, deaths and cancer registrations for these individuals (CeLSIUS, 2003). The survey would allow the migration history of lung cancer patients to be tracked and hence their exposure to different risk factors estimated in each census year. For example, the residential history of people in the longitudinal study who contracted lung cancer could be traced and then estimates of exposure to smoking and air pollution could be calculated for 1971, 1981 and 1991 and associated with each patient. While this may be possible theoretically, even this data set would not allow many of the risk factors to be estimated reliably because geographically disaggregated data on risk factors such as

smoking behaviour are not available for these years. Furthermore, confidentiality constraints may prevent specific estimates of pollution or radon being associated with the records in the longitudinal study. In addition, a longitudinal survey does not yet exist in Scotland.

Therefore, in order to examine this issue longitudinally, it would be necessary to undertake a self-designed case-control study by interviewing patients with and without lung cancer over a set period of time and questioning them about their smoking behaviour, eating habits and occupational history. Furthermore, their residential history could be determined in order that their history of exposure to environmental variables such as radon, air pollution and quarrying activity could be estimated for small areas. The incidence of lung cancer could then be modelled using non-contemporary measures of exposure to the risk factors that are not undermined by the migration effect. Because the longitudinal approach would provide improved estimates of exposure to the risk factors, the results of this modelling procedure would provide a better insight into how the risk factors influence the incidence of lung cancer and how important it is to model the lag between exposure and diagnosis. Instead of assuming no migration between the time of exposure and diagnosis, the lung cancer patients could be tracked through time and estimates could be associated with them that reflected their movements. Furthermore, the results could be compared to the results from research that is constrained by the same data limitations as this study, to decide whether ignoring the lag effect has a significant effect on the outcome. However, although this would address a number of problems posed by the lag and migration effect, it would still suffer from the same problems of estimating past exposure to risk factors such as

pollution levels and radon. This type of information could only be estimated accurately by undertaking a large and expensive prospective study.

9.6.3 Lung Cancer and Deprivation

This research has demonstrated that area deprivation, as measured using the Carstairs Index, is positively significant in explaining the incidence of all cases of lung cancer in Scotland, independently of other risk factors, such as smoking. More specifically, the incidence of lung cancer in Scotland is higher in more deprived areas compared to less deprived areas, as measured using the Carstairs Index. This is the first time that the independent relationship between deprivation and lung cancer has been noted for such small areas and in such a large study. The results of this analysis raise two issues for further examination. First, why is area deprivation significant in explaining lung cancer incidence in Scotland? Second, why is area deprivation not significant in explaining rural lung cancer in Scotland?

Although this study has shown that deprivation has an independent effect upon lung cancer, the reasons for this are unclear. There are, however, a number of potential reasons to explain the relationship between deprivation and ill health more broadly. More particularly, deprived areas tend to have more limited access and provision of facilities and services; poorer job prospects; higher levels of crime and violence that increase exposure to stress; fewer opportunities for social interaction; and feelings of exclusion, stigmatisation, segregation and abandonment, which can affect mental health (Stead *et al.*, 2001). However, it is not clear precisely why these factors would relate to lung cancer in Scotland. Future research could usefully focus on

why deprivation is significant in explaining lung cancer in Scotland and if any of these factors are important in explaining the deprivation effect. For example, access to health care provision has been suggested as one cause that perpetuates high smoking rates in parts of Glasgow (Stead *et al.*, 2001). This could be measured using a GIS to calculate the distance by foot, bus and car to doctors' surgeries and hospitals (Lovett *et al.*, 2002). This would provide measures of accessibility to health care provision for all OAs in Scotland, and these could be tested to see if they are significant in explaining the incidence of lung cancer. Similarly, the deprived areas may represent a combination of particularly poor people who live in areas where localised pollution levels are particularly high and work in occupations that have been associated with lung cancer. Therefore, the multiplicative effect between these variables could be tested and a detailed nationwide survey of pollution levels could be carried out to enable estimates to be predicted for smaller geographical areas.

Chapter eight demonstrated that the Carstairs Index of Deprivation was not significant in explaining lung cancer in rural areas. It was suggested that this could be either because deprivation is not an important cause of lung cancer in rural areas or because the Carstairs Index is not an appropriate measure of rural deprivation. It was noted that one of the four components of the Carstairs Index is non car ownership which, it has been argued, is an inappropriate measure in many rural areas where ownership of a car is considered a necessity (Higgs and White, 2000). Therefore, in order to examine this effect it would be sensible to use alternative measures of deprivation that better capture rural deprivation and to incorporate them into models of lung cancer incidence instead. However, other indices of deprivation,

such as the Townsend Score or the Scottish Index of Deprivation, were not designed to explicitly capture rural deprivation (Scottish Executive, 2003). Instead, the influence of deprivation in rural areas could be explored by incorporating measures of deprivation that are more suitable in rural areas which include measures of factors such as rural transportation and access to health care (Martin *et al.*, 2000). The resulting models could be compared to examine how sensitive they are to the definition of deprivation. If the improved deprivation variable remains insignificant, then it would suggest that deprivation does not influence lung cancer in rural areas. However, if the alternative measure of deprivation is significant, then it would suggest that deprivation does increase rural lung cancer incidence in Scotland and that the Carstairs Index is an inappropriate measure of deprivation in rural areas.

9.6.4 Lung Cancer and Urbanness

This study also found an independent relationship between lung cancer incidence and how urban an area is in Scotland. This was in line with the discussion in chapter four that found that lung cancer was consistently higher in urban areas of Scotland and previous studies that have found an independent urban effect on the rates of lung cancer (Haynes, 1988). A number of variables that were not incorporated into the models have been suggested as potential explanations for this effect, particularly the high rates of passive smoking in urban centres and specific types of employment that use hazardous materials such as asbestos. Methods to estimate these were discussed above.

However, it has also been suggested that the urban-rural effect found in this study could, at least partially, be due to the effect of selective migration. This has been suggested in other studies of ill health as one explanation for health differences between urban and rural neighbourhoods (Verheij *et al.*, 1998; Brimblecombe *et al.*, 1999), especially for diseases with a long latency period such as lung cancer (Haynes, 1988). In this study, rates of migration away from urban areas were higher than those from the rural to urban areas in Scotland, prior to the period that the lung cancer data were collected (1988 to 1991) and, therefore, many rural communities grew at the expense of urban areas (Champion, 1989). However, there are not only differential rates of migration but it has also been shown that it is healthier adults rather then ill adults who are more likely to migrate long distances (Fox and Goldblatt, 1982). Therefore, those people who did not contract lung cancer between 1988 and 1991 are more likely to have moved than those with lung cancer and hence the rates of the disease may be exaggerated in urban areas.

Although it would be difficult to examine this effect in small areas across Scotland, a longitudinal analysis of lung cancer patients could examine the differential migration rates between those with and without lung cancer in 1991. For example, the Longitudinal Study in England and Wales could be used to trace the migration history of all individuals in 1991. By associating a measure or urbanness, such as population density, with each of the individuals with and without lung cancer in 1971, 1981 and 1991, their movement between urban and rural areas could be compared. If there is higher net migration of lung cancer patients from rural to urban areas than there is for people without lung cancer, then this would suggest that

selective migration may account for, at least part of, the urban effect found in this study.

9.7 Conclusion

This research into lung cancer in Scotland during the period 1988 to 1991 is the first to examine the incidence of the disease in small geographical areas and this approach has provided a number of new insights into its aetiology. By far the most important cause of lung cancer is smoking as it has been estimated that it may be responsible for as many as 85 percent of cases (Schottenfeld, 1996). Despite the importance of this factor, many previous ecological studies of lung cancer have not accurately controlled for smoking behaviour (e.g. Haynes 1988; Doll 1991), largely because of the absence of accurate and geographically specific data on smoking behaviour. This study has successfully developed a method for estimating smoking behaviour and is the first to do so for such small geographical areas. Multilevel modelling techniques have been employed to predict smoking behaviour for 12 age-sex groups at a level of detail that has not previously been attained (Pearce *et al.*, 2003). The multilevel approach has allowed the individual and contextual determinants of smoking to be controlled for and the parameters of the model were then applied to each age-sex group across Scotland (Twigg *et al.*, 2000). The smoking estimates were highly significant in explaining the incidence of lung cancer in Scotland which is unsurprising given the well established relationship between smoking and lung cancer (Doll and Hill, 1950). More importantly, the calculation of the smoking estimates has meant that for the first time it has been possible to examine the incidence of lung cancer in small geographical areas, whilst accurately controlling

for smoking behaviour. This has allowed new insights into the geographical distribution of lung cancer in Scotland as well as further understanding of the risk factors associated with the disease.

A novel approach was taken to predicting levels of air pollution in PPSs across Scotland using socioeconomic data from the 1991 census on the industries in which workers in each area are employed and data on the road network. This contrasted with more complicated air pollution models that incorporate factors such as meteorology and topography (e.g. Holland *et al.*, 2000) but that are limited to making estimates for large geographical areas. The approach taken in this study had the advantage of providing estimates for all areas across Scotland. It was found that the estimates of air pollution levels in 1971 were significant in explaining lung cancer incidence but that the 1991 estimates were not. This study is not only the first to consider air pollution and lung cancer in Scotland, but the results also add to the wider debate on the effect of air pollution on lung cancer incidence, concurring with the studies that have found a weak but significant relationship (e.g. Dockery *et al.*, 1993). Furthermore, the results suggest that due to the nationwide reduction in air pollution between the two periods, it was important to incorporate the lag effect for the air pollution variables, despite the fact that people may have moved between the period of exposure and the time at which they were diagnosed with the disease. This suggests that air pollution is unlikely to be a major cause of lung cancer in the future as the low estimates of pollution in 1991 were not significant. Furthermore, this is of wider significance because it suggests that cross-sectional studies of ill health should consider using lagged variables instead of contemporary measures to represent certain determinants of disease.

This study is the first to consider the health effects of radon gas in Scotland and despite the difficulty of estimating radon exposure, the results suggest that radon is a cause of premature cases of lung cancer in Scotland. This is an important finding because it is the first time that radon has been identified as a hazard to health in Scotland and lends support to previous work that has suggested that it is a significant cause of lung cancer in south-west England (Darby *et al.*, 1998). In particular, these results call for further research into the health effects of high radon exposure among the young. Furthermore, due to the complications in estimating personal exposure to radon, future work should also focus on the health effects of high exposure among the wider Scottish population in order to be confident that the health effects of the gas do not extend beyond the young.

The work has also added to the debate on whether there is an urban bias in lung cancer that is independent of demographic differences and smoking behaviour. The results of simple mapping, age-standardised rates, Poisson probabilities and methods of cluster detection have all shown that there were higher rates of the disease in the urban areas of Scotland, particularly Glasgow. Furthermore, the results of the Poisson regression have demonstrated that there is an independent urban effect that is consistent with previous studies in districts across England and Wales (Haynes, 1988) and has also demonstrated that this effect is also independent of air pollution. This research has improved upon the previous work by accurately controlling for smoking behaviour as well as examining the variations in lung cancer in much smaller geographical units. Similarly, this study is the first to demonstrate that area measures of deprivation are significant in explaining the incidence of lung cancer in

Scotland that is independent of demographic and smoking differences as well urbanness or ruralness.

An inspection of the residual values produced from the optimum model of lung cancer in Scotland found that having controlled for the key risk factors associated with the disease, there were more cases of lung cancer in rural compared to urban areas. This is an important finding because it suggests that there is a rural bias in lung cancer once the key risk factors have been controlled for. Although previous work has pointed to urban-rural differences in health outcomes (e.g. Senior *et al.*, 2000), this is the first study to identify a higher incidence of lung cancer in rural areas compared to urban areas after controlling for the key risk factors associated with the disease. However, it was also noted that most of the extreme residual values were located in the large urban areas of Scotland, particularly Glasgow. This may be related to poorer diets in some urban areas (Ecob and Macintyre, 2000), the urban concentration of passive smoking (Jee *et al.*, 1999) or exposure to asbestos used in the ship building industry (De Vos Irvine *et al.*, 1993). Each of these factors warrant further investigation.

On a final note, earlier chapters demonstrated that rates of lung cancer are consistently higher in developed countries compared to developing countries. This is attributed to differences in smoking rates between countries that reflect particular stages in what has been termed the smoking epidemic (Cavelaars, *et al.*, 2000). Many developed countries, such as Scotland, have rates of smoking that have fallen over a number of years and are considered to have reached or nearing the end of the epidemic. On the other hand, at present most developing countries are in the early

stages of the epidemic and have relatively low rates of smoking but it is anticipated that smoking prevalence will rapidly increase in the near future. This pattern of smoking behaviour has two important implications for lung cancer trends. First, lung cancer rates in most developed countries are expected to fall in response to historical patterns of smoking behaviour and therefore other factors, such as radon, will become relatively more important. Second, as the rates of smoking increase in developing countries, the prevalence of lung cancer will follow a lagged response. Therefore, lung cancer research is likely to shift away from developed countries that presently have high rates of the disease, such as Scotland, to the developing countries, maybe with a view to discouraging smoking and ultimately limiting the incidence of the disease. In addition to the focus on developing countries, research can usefully examine particular social and ethnic groups that continue to have high rates of smoking in developed countries despite the overall downward trend, such as young women in developed countries (WHO, 2000) and Maoris in New Zealand (Galgali *et al.*, 1998). Therefore, whilst this study has furthered our understanding of the aetiology of lung cancer and illuminated many of the social dimensions of the disease in Scotland, the future research agenda should be directed to considering patterns of the disease among particular social and ethnic groups and patterns in the developing world.

10. REFERENCES

Abercrombie, M., Hickman, M., Johnson, M.L. and Thain, M. (1992) *Dictionary of Biology*. Penguin, London.

Aggleton, P. (1990) *Health*. Routledge, London.

Acquavella, J., Leet, T. and Johnson, G. (1993) Occupational experience and mortality among a cohort of metal components manufacturing workers. *Epidemiology* 4, 428-434.

Adams, G.E. and Cox, R. (1991). Radiation carcinogenesis. In: Franks, L.M. and Teich, N.M. (eds.) *Introduction to the Cellular and Molecular Biology of Cancer*, 176-202. Oxford University Press, Oxford.

Aitkin, M. and Longford, N. (1986) Statistical modelling of data on teaching styles. *Journal of the Royal Statistical Society, Series A* 144, 419-461.

Åkerblom, G. (1987) Investigations and mapping of radon risk areas. In: *Proceedings of International Symposium on Geological Mapping, Trondheim, 1986: In the Service of Environmental Planning*, 96-106. Norg. Geol. Unders, Oslo

Alavanja, M.C., Brownson, R.C., Lubin, J.H., Berger, E., Chang, J. and Boice, J.D. (1994) Residential radon exposure and lung cancer among nonsmoking women. *Journal of the National Cancer Institute* 86, 1829-1837.

Alavanja, M.C., Brownson, C.C., Swanson, C. and Brownson, R.C. (1993) Saturated fat intake and lung-cancer risk among nonsmoking women in Missouri. *Journal of the National Cancer Institute* 85, 1906-1916.

Alavanja, M.C., Lubin, J.H., Mahaffey, J.A. and Brownson, R.C. (1999) Residential radon exposure and risk of lung cancer in Missouri. *American Journal of Public Health* 89, 1042-1048.

Albin, M., Magnani, C., Krstev, S., Rapiti, E. and Shefer, I. (1999) Asbestos and cancer: An overview of current trends in Europe. *Environmental Health Perspectives* 107, 289-298.

Alpha-Tocopherol, Beta Carotene Cancer Prevention Study Group (1994) The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. *New England Journal of Medicine* 330, 1029-1035.

American Cancer Society (2001) *American Cancer Society* www.cancer.org Site accessed April 2001.

Amos, A., Currie, C., Hunt, S. and Martin, C.J. (1990) Health related behaviour in a small Scottish community. *Public Health* 104, 131-140.

Anderson, S. and Hope, S. (2000) *Scotland's People: Results from the 1999 Scottish Household Survey. Volume 2: Technical Report.* Scottish Executive National Statistics Publication, Edinburgh.

Appleton, J.D. and Ball, T.K. (1995) Radon and background radioactivity from natural sources: characteristics, extent and the relevance to planning and development in Great Britain. *British Geological Survey Technical Report* WP/95/2.

Appleton, J.D. and Ball, T.K. (in press) Geological radon potential mapping.

Arden Pope, C. and Dockery, D.W. (1999) Epidemiology of particle effects. In: Holgate, S.T., Samet, J.M., Koren, H.S. and Maynard, R.L. (eds.). *Air Pollution and Health*, 673-705. Academic Press, London.

Arnold, D. (ed.) (1988) *Imperial Medicine and Indigenous Societies.* Manchester University Press, Manchester.

Australian National Health and Medical Research Council (1987) *Effects of Passive Smoking on Health.* Australian Government Publishing Service, Canberra.

Auvinen, A., Mäkeläinen, H., Hakama, M., Castrén, O., Pukkala, E., Reisbacka, H. and Rytömaa, T. (1996) Indoor radon exposure and risk of lung cancer: a nested case-control study in Finland. *Journal of the National Cancer Institute* 88, 966-972.

Babb, P., Brock, A., Jones, J. and Quinn, M. (2001) Geographic patterns in cancer incidence. In: Griffiths, C. and Fitzpatrick, J. (eds.) *Geographic Variations in Health.* Stationery Office Books, London.

Bailey, T.C. and Gatrell, A.C. (1995) *Interactive Spatial Data Analysis*, Longman, Harlow.

Balarajan, R. and Yuen, P. (1986) British smoking and drinking habits: regional variations. *Community Medicine* 8, 131-137.

Ball, T.K., Cameron, D.G. and Colman, T.B. (1992) Aspects of radon potential mapping in Britain. *Radiation Protection Dosimetry* 45, 211-214.

Ball, T.K., Cameron, D.G., Colman, T.B. and Roberts, P.D. (1991) Behaviour of radon in the geological environment. *Quarterly Journal of Engineering Geology* 24, 169-182.

Ball, T.K., Cameron, D.G., Colman T.B. and Roberts, P.D. (1995) The use of uranium exploration data for mapping radon potential in the UK - advantages and pitfalls. In: *Application of Uranium Exploration Data and Techniques in Environmental Studies. Proceedings of a Technical Committee meeting held in Vienna, 9-12 November 1993.* IAEA-TECDOC-827: 139-149. International Atomic Energy Agency, Vienna.

Ball, T.K. and Miles, J.C.H. (1993) Geological and geochemical factors affecting the radon concentration in homes in Cornwall and Devon, UK. *Environmental Geochemistry and Health* 15, 27-36.

Barbone, F., Bovenzi, M., Cavallieri, F. and Stanta, G. (1995) Air-pollution and lung-cancer in Trieste, Italy. *American Journal of Epidemiology* 141, 1161-1169.

Barnet, I. (1991) Radon risk mapping in the Czech Republic. In: Barnett, I. (ed.) *Radon Invesigations in Czechoslovakia II*, 13-19. Geological Survey, Prague.

Barnett, J.R. (2000) Does place of residence matter? Contextual effects and smoking in Christchurch. *New Zealand Journal of Medicine* 113, 433-435.

BEIR (Biological Effects of Ionizing Radiation) IV (1998) *Health Risks of Radon and Other Internally Deposited Alpha-Emitters*. National Academy Press, Washington DC.

BEIR (Biological Effects of Ionizing Radiation) VI (1998) *Effects of Exposure to Radon*. National Academy Press, Washington DC.

Bello, D., Virji, M.A., Kalil, A.J. and Woskie, S.R. (2002) Quantification of respirable, thoracic, and inhalable quartz exposures by FT-IR in personal impactor samples from construction sites. *Applied Occupational and Environmental Hygiene* 17, 580-590.

Bench, A. and Rabbitts, P. (1994) *Genetic changes in lung cancer. New perspectives in lung cancer.* BMJ Publishing, London.

Bentham G. (1988) Migration and morbidity: implications for geographical studies of disease. *Social Science and Medicine* 26, 49-54.

Bentham, G., Eimermann, J., Haynes, R., Lovett, A. and Brainard, J. (1995) Limiting long-term illness and its associations with mortality and indicators of social deprivation. *Journal of Epidemiology and Community Health* 49, s57-s64.

Bentham, G. and Haynes, R. (1985) Health, personal mobility and the use of health services in rural Norfolk. *Journal of Rural Studies* 1, 231-239.

BGS (British Geological Survey) (1998) *Radon Potential Based on Solid and Drift Geology (Sheet 53N 04W Liverpool Bay; 1:250,000 Scale Map: Natural Environmental Radioactivity Survey)* British Geological Survey, Keyworth.

Bhopal, R.S., Moffatt, S., Pless-Mulloli, T., Phillimore, P.R., Foy, C., Dunn, C.E. and Tate, J.A. (1998) Does living near a constellation of petrochemical, steel and other industries impair health? *Occupational and Environmental Medicine* 55, 812-822.

Birkett, N.J. (1992) Effect of nondifferential misclassification on estimates of odds ratios with multiple levels of exposure. *American Journal of Epidemiology* 136, 356-362.

Black, D. (1984) *Investigation of the Increased Incidence of Cancer in West Cumbria*. HMSO, London.

Blair, A., Rothman, N. and Zahm, S.H. (1999) Occupational cancer epidemiology in the coming decades. *Scandinavian Journal of Work Environment and Health* 25, 491-497.

Blaxter, M. (1990) *Health and Lifestyles*. Routledge, London.

Blot, W.J., Xu, Z.Y., Boice, J.D., Zhao, D.Z., Stone, B.J., Sun, J., Jing, L.B. and Fraumeni, J.F. (1990) Indoor radon and lung cancer in China. *Journal of the National Cancer Institute* 82, 1025-1030.

Boffeta, P., Jourenkova, N. and Gustavsson, P. (1997) Cancer risk from occupational and environmental exposure to polycyclic aromatic hydrocarbons. *Cancer Causes and Controls* 8, 444-472.

Boffetta, P. and Kogevinas, M. (1999) Introduction: Epidemiological research and prevention of occupational cancer in Europe. *Environmental Health Perspectives* 107, 229-231.

Boffetta, P., Saracci, R., Andersen, A., Bertazzi, P.A., Chang-Claude, J. and Westerholm, P. (1992) Lung cancer mortality among workers in the European production of man-made mineral fibers - a Poisson regression analysis. *Scandinavian Journal of Work, Environment and Health* 18, 279-286.

Bowling, A. (1997) *Research Methods in Health. Investigating Health and Health Services*. Oxford University Press, Oxford.

Boyle, P.J. and Flowerdew, R. (1993) Modelling sparse interaction matrices: interward migration in Hereford and Worcestershire, and the underdispersion problem. *Environment and Planning* A, 1201-1209.

Boyle, P.J., Flowerdew, R. and Williams, A. (1997) Evaluating the goodness of fit in models of sparse medical data: a simulation approach. *International Journal of Epidemiology* 26, 651-656.

Boyle, P.J., Kudlac, H. and Williams, A.J. (1996) Geographical variation in the referral of patients with chronic end stage renal failure for renal replacement therapy. *Quarterly Journal of Medicine* 89, 151-157.

Boyle, P.J., Gatrell, A.C. and Duke-Williams, O. (1999) The effect of morbidity on the variability in deprivation and population stability in England and Wales: an investigation at small-area level. *Social Science and Medicine* 49, 791-799.

Boyle, P., Norman, P. and Rees, P. (2002) Does migration exaggerate the relationship between deprivation and limiting long-term illness? A Scottish analysis. *Social Science and Medicine* 55, 21-31.

Boyle, P., Gandani, S. and Gray, N. (2000) Epidemiology of lung cancer: a century of great success and ignominious failure. In: Hansen, H.H. *Textbook of Lung Cancer*, 13-25. Dunitz, London.

Brennan, P., Butler, J., Agudo, A., Benhamou, S., Darby, S., Fortes, C., Jöckel, K.H., Kreuzer, M., Nyberg, F., Pohlabeln, H., Saracci, R., Wichman, H.E. and Boffetta, P. (2000) Joint effect of diet and environmental tobacco smoke on risk of lung cancer among nonsmokers. *Journal of the National Cancer Institute* 92, 426-427.

Breslow, R.A., Graubard, B.I., Sinha, R. and Subar, A.F. (2000) Diet and lung cancer mortality: a 1987 National Health Interview Survey cohort study. *Cancer Causes and Control* 11, 419-431.

Bridgwood, A., Lilly, R., Thomas, M., Bacon, J., Sykes, W. and Morris, S. (1998) *Living in Britain, Results from the 1998 General Household Survey*. HMSO, London.

Briggs, D.J., Collins, S., Elliot, P., Fischer, P., Kingham, S., Lebret, E., Pryl, K., Van Reeuwijk, H., Smallbone, K. and Van Der Veen, A. (1997) Mapping urban air pollution using GIS: a regression-based approach. *International Journal of Geographical Information Science* 11, 699-718.

Briggs, D.J., de Hoogh, C., Gulliver, J., Wills, J., Elliot, P., Kingham, S. and Smallbone, K. (2000) A regression-based method for mapping traffic-related air pollution: application and testing in four contrasting urban environments. *The Science of the Total Environment* 253, 151-167.

Brimblecombe, N., Dorling, D. and Shaw, M. (1999) Mortality and migration in Britain, first results from the British Household Panel Survey. *Social Science and Medicine* 49, 981-988.

Bromen, K., Pohlabelen, H., Jahn, I., Ahrens, W. and Jockel, K.H. (2000) Aggregation of lung cancer in families: Results from a population-based case-control study in Germany. *American Journal of Epidemiology* 152, 697-505.

Brooks, A. L., Khan, M.A., Duncan, A., Buschbom, R.L., Jostes, R.F. and Cross. F.T. (1997) Effectiveness of radon relative to acute 60Co gamma-rays for induction of micronuclei in vitro and in vivo. *International Journal of Radiation Biology* 66, 801-808.

Brown, T. and Duncan, C. (2000) London's burning: recovering other geographies of health. *Health and Place* 6, 363-375.

Brownson, R.C. and Alavanja, M.C.R. (2000) Previous lung disease and lung cancer risk among women (United States). *Cancer Causes and Control* 11, 853-858.

Brunsdon, C. (1991) Estimating probability surfaces in GIS: an adaptive technique. *Proceedings 2nd European Conference on GIS (EGIS), Brussels*, 155-164.

Burnley, I.H. (1997) Disadvantage and male cancer incidence and mortality in New South Wales 1985-1993. *Social Science and Medicine* 45, 465-476.

CancerBACUP (2001) CancerBACUP. www.cancerbacup.org.uk Site accessed April 2001.

Capelletto, F. and Merler, E. (forthcoming) Perceptions of health hazards in the narratives of Italian migrant workers at an Australian asbestos mine (1943-1966). *Social Science and Medicine.*

Carnes, B.A., Groer, P.G. and Kotek, T.J. (1997) Radium dial workers: issues concerning dose response and modeling. *Radiation Research* 147, 707-714.

Carstairs, V. and Morris, R. (1991) *Deprivation and Health in Scotland.* Aberdeen University Press, Aberdeen.

Cavelaars, A.E.J.M., Kunst. A.E., Geurts, J.J.M., Crialesi, R., Grötvedt, L., Helmert, U., Lahelma, E., Lundberg, O., Matheson, J., Mielck, A., Rasmussen, N.K.R., Regidor, E., do Rosário-Giraldes, M., Spuhler, T.H. and Mackenbach, J.P. (2000) Educational differences in smoking: international comparison. *British Medical Journal* 320, 1102-1107.

CeLSIUS (Centre for Longitudinal Study Information and User Support) (2003) www.celsius.lshtm.ac.uk/index.html Site accessed February 2003.

Chainey, S. (1997) *Crime mapping in the London Borough of Brent.* First workshop on GIS for crime and health data analysis, Sheffield Centre of Geographic Information and Spatial Analysis, University of Sheffield, 22-23 May, 1997.

Champion, A.G. (1989) United Kingdom, population deconcentration as a cyclic phenomenon. In: Champion, A.G. (ed.) *Counterurbanization, the Changing Pace and Nature of Population Deconcentration*, 83-102. Arnold, London.

Charlton, A. (1994) Tobacco and lung cancer. In: Thatcher, N. and Spiro, S. (eds.) *New Perspectives in Lung Cancer.* BMJ Publishing Group, London.

Checkoway, H., Mathew, R.M., Shy, C.M., Watson, J.E., Tankersley, W.G., Wolf, S.H., Smith, J.C. and Fry, S.A. (1985) Radiation work experience, and cause specific mortality among workers at an energy research laboratory. *British Journal of Industrial Medicine* 42, 525-533.

Chow, W.H., Schuman, L.M., Mclaughlin, J.K., Bjelke, E., Gridley, G., Wacholder, S., Chien, H.T.C. and Blot, W.J. (1992) A cohort study of tobacco use, diet, occupation, and lung-cancer mortality. *Cancer Causes and Control* 3, 247-254.

Cliff, A.D., Haggett, P. and Smallman-Raynor, M. (2000) *Island Epidemics*. Oxford University Press, Oxford.

Cocksedge, W., Rankin, W., Tostowaryk, W., Rahman, W., Charbonneau, B.W. and Grasty, R.L. (1993) *National native home radon survey – maximising resources through radon potential assessment.* Paper presented at the 26th Health Physics Society Meeting on Environmental Health Physics, January 24th-28th, 1993. Lake Coeur d'Alene, Idaho.

Cohen, A.J. (2000) Outdoor air pollution and lung cancer. *Environmental Health Perspectives* 108, 743-750.

Cohen, A.J. and Pope, C.A. (1995) Lung Cancer and Air Pollution. *Environmental Health Perspectives* 103, Sup 8, 219-224.

Cohen, B.L., Stone, C.A. and Schilken, C.A. (1994) Indoor radon maps of the United States. *Health Physics* 66, 201-205.

Colgan, P.A. and Gutierrez, J. (1996) National approaches to controlling exposure to radon. *Environment International* 22, S1083.

Conrad, R.A., Rall, J.E. and Sutow, W.W. (1966) Thyroid nodules as a late sequela of radioactive fallout, in a Marshall Island population exposed in 1954. *New England Journal of Medicine* 274, 1391-1399.

Copas, J.B. and Shi, J.Q. (2000) Reanalysis of epidemiological evidence on lung cancer and passive smoking. *British Medical Journal* 320, 417-418.

Cornwell, J. (1984) *Hard-Earned Lives: Accounts of Health and Illness from East London*. Tavistock Publications, London.

Craddock, S. (1995) Sewers and scapegoats: Spatial metaphors of smallpox in nineteenth century San Francisco. *Social Science and Medicine* 45, 957-968.

Cross, F.T. (1992) A review of experimental animal radon health effects data. In: Dewey, W.C., Edington, M., Fry, R.J., Hall, E.J. and Whitemore, G.F. (eds.) *Radiation Research, a Twentieth Century Perspective*, Volume 2, 333-339. Academic Press, New York.

Cross, F.T. (1994) Invited commentary: Residential risks of radon from the perspective of experimental animal studies. *American Journal of Epidemiology* 140, 333-339.

Cummins, S. and Milligan, C. (2000) Taking up the challenge: new directions in the geographies of health and impairment. *Area* 32, 7-9.

Curtis, S. and Taket, A. (1996) *Health and Societies. Changing Perspectives.* Arnold, London.

Dale, A. (1993) The content of the 1991 census: change and continuity. In: Dale, A. and Marsh, C. (eds.) *The 1991 Census Users Guide*, 16-51. HMSO, London.

Dale, A. and Marsh, C. (1993) *The 1991 Census User's Guide*. HMSO, London.

Darby, S. (1999) Radiation risks. *British Medical Journal* 319, 1019-1020.

Darby, S.C. and Doll, R. (1990) Radiation and exposure rate. *Nature* 344, 824.

Darby, S., Whitley, E., Silcocks, P., Thakrar, B., Green, M., Lomas, P., Miles, J., Reeves, G., Fearn, T. and Doll, R. (1998). Risk of lung cancer associated with residential radon exposure in south-west England: a case control study. *British Journal of Cancer* 78, 394-408.

Darnley, A. G., Richardson, K.A., Grasty, B.W., Carson, J.M., Holman, P.B. and Charbonneau, B.W. (1986) *Radioactivity Map of Canada*. Geological Survey of Canada, Map 1600A. 1:5,000,000 scale. 1st Edition.

Davis, J.M., Eder, B.K., Nychka, D. and Yang, Q. (1998) Modeling the effects of meteorology on ozone in Houston using cluster analysis and generalized additive models. *Atmospheric Environment* 32, 2505-2520.

Dean, G., Lee, P.N., Todd, G.F. and Wicken, A.J. (1978) *Report on a Second Retrospective Mortality Study in Northeast England: Parts 1 and 2*. Tobacco Research Council, London.

Dear, M. and Wolch, J. (1987) *Landscapes of Despair: From Deinstitutionalization to Homelessness*. Polity Press, London.

Deepchand, K., Shaw, A. and Field, J. (2000) Eating Habits. In: Shaw, A., McMunn, A. and Field, J. (eds.) *The Scottish Health Survey 1998*, 319-386. The Scottish Executive Department of Health, Edinburgh.

de Groh, M., Ellison, L., Gibbons, L., Luciana, S. and Schanzer, D. (2002) Lung cancer in Canada. www.hc-sc.gc.ca/hpd/lcdc/c/updates/lung_e.htm Site accessed April 2002.

deKlerk, N.H., Musk, A.W., Eccles, J.L., Hansen, J. and Hobbs, M.S. (1996) Exposure to crocidolite and the incidence of different histological types of lung cancer. *Occupational and Environmental Medicine* 53, 157-159.

Department of Health (1997) *Handbook on Air Pollution and Health*. HMSO, London.

Department of Health (1998a) *Saving Lives: Our Healthier Nation*. The Stationery Office, London.

Department of Health (1998b) *Statistics on Smoking: England, 1978 Onwards*. The Stationery Office, London.

Department of the Environment (1995a) *The Environmental Effects of Dust from Surface Mineral Workings. Volume 1, Summary Report and Best Practice Guidelines*. HMSO, London.

Department of the Environment (1995b) *The Environmental Effects of Dust from Surface Mineral Workings. Volume 2, Technical Report*. HMSO, London.

Devesa, S.S., Grauman, D.J., Blot, W.J. and Fraumeni, J.F. (1999) Cancer surveillance series: changing geographic patterns of lung cancer mortality in the United States, 1950 through 1994. *Journal of the National Cancer Institute* 91, 1040-1050.

De Vos Irvine, H., Lamont, D.W., Hole, D.J. and Gillis, C.R. (1993) Asbestos and lung cancer in Glasgow and the west of Scotland. *British Medical Journal (Clinical Research Edition)* 306, 1503-1506.

Diamond, I., Clements, S., Stone, N. and Ingham, R. (1999) Spatial variation in teenage conceptions in south and west England. *Journal of the Royal Statistical Society A* 162 Part 3, 273-289.

Diehr, P., Koespell, T., Cheadle, A., Psaty, B.M., Wagner, E. and Curry, S. (1993) Do communities differ in health behaviours? *Journal of Clinical Epidemiology* 46, 1141-1149.

Diem, J.E. and Comrie, A.C. (2002) Predictive mapping of air pollution involving sparse spatial observations. *Environmental Pollution* 119, 99-117.

Diggle, P.J. and Chetwynd, A.D. (1991) Second-order analysis of spatial clustering for inhomogenous populations. *Biometrics* 47, 1155-1163.

Dixon, F. and Finlayson, L. (2000) Research Note: the Scottish Household Survey. *Scottish Geographical Journal* 116, 143-148.

Dockery, D.W., Pope, C.A. III, Xu, X., Spengler, J.D., Ware, J.H., Fay, M.E., Ferris, B.G. and Speizer, F.E. (1993) An association between air pollution and mortality in six U.S. cities. *New England Journal of Medicine* 329, 1753-1759.

Dodds, C. and Maguire, B. (1998) The Scottish Shipbuilding Industry. *Scottish Economic Bulletin* 57.

Doll, R. (1955) Mortality from lung cancer in asbestos workers. *British Journal of Industrial Medicine* 12, 81-86.

Doll, R. (1978) Atmospheric pollution and lung cancer. *Environmental Health Perspectives* 22, 23-31.

Doll, R. (1991) Urban and rural factors in the aetiology of cancer. *International Journal of Cancer* 47, 803-810.

Doll R., and Hill, A.B. (1950) Smoking and carcinoma of the lung. *British Medical Journal* ii, 739-748.

Doll, R. and Hill, A.B. (1954) The mortality of doctors in relation to their smoking habits. *British Medical Journal* i, 1451-1455.

Doll, R. and Hill, A.B. (1964) Mortality in relation to smoking: 10 years of observations of British Doctors. *British Medical Journal* i, 1399-1414.

Doll, R. and Hill, A.B. (1966) Mortality of British doctors in relation to smoking. *National Cancer Institute Monographs* 19, 205-268.

Doll, R., Peto, R., Wheatley, K., Gray, R. and Sutherland, I. (1994) Mortality in relation to smoking: 40 years' observations on male British doctors. *British Medical Journal* 309, 901-911.

Doll, R., Vessey, M.B., Beasley, R.W.R., Buckley, A.R., Fear, C.C., Fisher, R.E., Gammon, E.J., Gunn, W., Hugher, G.O., Lee, K. and Norman Smith, B. (1972) Mortality of gas workers – final report of a prospective study. *British Journal of Industrial Medicine* 29, 394-406.

Doll, R. and Wakeford, R. (1997) Risk of childhood cancer from fetal irradiation. *The British Journal of Radiology* 70, 130-139.

Donaldson, K. and Borm, P.J. (1998) The quartz hazard: a variable entity. *The Annals of Occupational Hygiene* 42, 287-294.

Donaldson, K., Brown, G.M., Brown, D.M., Slight, J. and Li, X.Y. (1992) Epithelial and extracellular matrix injury in quartz-inflamed lung: role of the alveolar macrophage. *Environmental Health Perspectives* 97, 221-224.

Dorling, D. (1997) *Death in Britain*. Joseph Rowntree Foundation, York.

Dorn, M. and Laws, G. (1994) Social theory, body politics and medical geography: Extending Kearns's invitation. *Professional Geographer* 46, 106-110.

Droste, J.H.J., Weyler, J.J., Van Meerbeeck, J.P., Vermeire, P.A. and van Sprindel, M.P. (1999) Occupational risk factors of lung cancer: a hospital based case-control study. *Occupational and Environmental Medicine* 56, 322-327.

Duncan, C., Jones, K. and Moon, G. (1999) Smoking and deprivation: are there neighbourhood effects? *Social Science and Medicine* 48, 497-505.

Dunn, C. and Kingham, S. (1996) Establishing links between air quality and health: searching for the impossible? *Social Science and Medicine* 42, 831-841.

Dyck, I. and Kearns, R. (1995) Transforming the relations of research: towards culturally safe geographies of health and healing. *Health and Place* 1, 137-147.

Earickson, R. (2000) Geographic research at the end of the century: papers from the Eighth International Symposium on Medical Geography. *Social Science and Medicine* 50, 911-913.

Ecob, R. and Macintyre, S. (2000) Small area variations in health related behaviours; do these depend on the behaviour itself, its measurement, or on personal characteristics? *Health and Place* 6, 261-274.

EDINA (2002) www.edina.ac.uk Site accessed February 2002.

Ellaway, A. and Macintyre, S. (1996) Does where you live predict health related behaviours? A case study in Glasgow. *Health Bulletin* 54, 443-446.

Elliott, P., Shaddick, G., Kleinschmidt, I., Jolley, D., Walls, P., Beresford, J. and Grundy, C. (1996). Cancer incidence near municipal solid waste incinerators in Great Britain. *British Journal of Cancer* 73, 702-710.

Elwood, J.M. (2002) Developing areas in cancer in New Zealand. *Japanese Journal of Clinical Oncology* 32, S43-51.

Entrikin, J.N. (1991) *The Betweenness of Place. Towards a Geography of Modernity.* John Hopkins University Press, Baltimore.

Escobedo, L. and Pedicord, M. (1996) Smoking prevalence in US birth cohorts: the influence of gender and education. *American Journal of Public Health* 86, 231-236.

ESRI (2002) www.esri.com Site accessed February 2002.

Etherington, D.J., Pheby, D.F. and Bray, F.I. (1996) An ecological study of cancer incidence and radon levels in South West England. *European Journal of Cancer* 32A, 1189-1197.

Ettlinger, L.A., Sayala, D. and Smith, B.E. (1987) Predicting high radon levels: a study in North Virginia. In: *The Proceedings of the Second International APC Speciality Conference Indoor Radon 2*, Pittsburgh, USA.

Experian (2002) Scotland Mosaic. GB 2001 Profiler. www.experianintact.com/documents/Intactv4Docs/PDF/Scotland%20MOSAIC.pdf Site accessed December 2002.

Eyles, J. (1997) Environmental health research: setting an agenda by spinning our wheels or climbing the mountain? *Health and Place* 3, 1-13.

Eyles, J., Taylor, S.M., Johnson, N. and Baxter, J. (1993) Worrying about waste: Living close to solid waste disposal facilities in southern Ontario, *Social Science and Medicine* 37, 805-812.

Eyles, J. and Wood, K. (1983) *A Social Geography of Medicine and Health.* Croom Helm, London.

Ferguson, M.K., Skosey, C., Hoffman, P.C. and Golomb, H. (1990) Sex-associated differences in presentation and survival in patients with lung cancer. *Journal of Clinical Oncology* 8, 1402-1407.

Ferlay, J., Parkin, D.M. and Pisani, P. (1999) Estimates of the worldwide incidence of 25 major cancers in 1990. *International Journal of Cancer* 80, 827-841.

Finkelman, R.B., Tatu, C.A., Belkin, H.E., Zheng, B., Lerch, H.E., Maharaj, S.V., Bates, A.L., Orem, W. and Castranova, V. (2002) Health impacts of coal and coal use: Possible solutions. *International Journal of Coal Geology* 50, 425-443.

Flowerdew, R. and Geddes, A. (1999) Poisson regression analysis of limiting long-term illness data for parts of north-east England. *Geographical and Environmental Modelling* 3, 63-82.

Flowerdew, R. and Green, A. (1993) Migration, transport and workplace statistics from the 1991 census. In: Dale, A. and Marsh, C. (eds.) *The 1991 Census User's Guide*, 269-294. HMSO, London.

Foster, K. (1993) The use of standardisation in survey analysis. *Survey Methodology Bulletin* 33.

Fox, A.J. and Goldblatt, P.O. (1982) *Longitudinal Study: Socio-demographic Mortality Differentials*. OPCS Series LS No. 1. HMSO, London.

Francis, B., Green, M. and Payne, C. (1993) *The GLIM System. Release 4 Manual.* Oxford Science Publications, Oxford.

Franks, L.M. (1997) What is cancer? In: Franks, L.M. and Teich, N.M. (eds.) *Introduction to the Cellular and Molecular Biology of Cancer*, 1-30. Oxford University Press, Oxford.

Friberg, L. and Cederlof, R. (1978) Late effects of air pollution with special reference to lung cancer. *Environmental Health Perspectives* 22, 45-66.

Friis, S. and Storm, H.H. (1993) Urban rural variation in cancer incidence in Denmark 1943-1987. *European Journal of Cancer* 29A, 538-544.

Galgali, G., Beaglehole, R., Scragg, R. and Tobias, M. (1998) Potential for prevention of premature death and disease in New Zealand. *The New Zealand Medical Journal* 111, 7-10.

Gardner, M.J., Hall, A.J., Downes, S. and Terrell, J.D. (1987a) Follow up study of children born to mothers resident in Seascale, West Cumbria (birth cohort). *British Medical Journal* 295, 822-827.

Gardner, M.J., Hall, A.J., Downes, S. and Terrell, J.D. (1987b) Follow up study of children born elsewhere but attending schools in Seascale, West Cumbria (schools cohort). *British Medical Journal* 295, 819-822.

Garfinkel, L. and Stellman, S.D. (1986) Cigarette smoking among physicians, dentists and nurses. *World Smoking and Health* 11, 2-9.

Gates, A.E. and Gundersen, L.C.S. (eds.) (1992) Geologic Controls on Radon. *Special Papers of the Geological Society of America* 271.

Gatrell, A. (2002) *Geographies of Health*. Blackwell, Oxford.

General Household Survey (1999) The General Household Survey. www.data-archive.ac.uk/findingData/ghsAbstract.asp Site accessed May 2002.

GLOBOCAN (2000) Cancer Incidence, Mortality and Prevalence Worldwide. www-dep.iarc.fr/cgi-bin/exe/ Site accessed April 2002.

Godzic, B. (1997) Ground level ozone concentrations in the Krakow region, southern Poland. *Environmental Pollution* 98, 273-280.

Goldstein, H. (1984) The methodology of school comparisons. *Oxford Review of Education* 10, 69-74.

Goldstein, H. (1995) *Multi-level Statistical Models*. Edward Arnold, London.

Gould, M.I. and Jones, K (1996) Analysing perceived limiting long term illness using GBN census microdata. *Social Science and Medicine* 42, 857-869.

Gould, P. and Wallace, R. (1994) Spatial structures and scientific paradoxes in the AIDS pandemic, *Geografiska Annaler* 76B, 105-116.

Graham, E. (1997) Philospohies underlying human geography research. In: Flowerdew, R. and Martin, D. (eds.) *Methods in Human Geography: A Guide for Students Doing a Research Project*. Longman, Harlow.

Graham, H. and Der, G. (1999) Influences on women's smoking status – The contribution of socio-economic status in adolescence and adulthood. *European Journal of Public Health* 9, 137-141.

Green, B.M.R., Lomas, P.R. and Kendall, G.M. (1996) *Radon in Dwellings in Scotland: 1996 Review*. NRPB-M569.

Grimson, R.C. (1991) A versatile test for clustering and a proximity analysis of neurons. *Methods of Information in Medicine* 30, 299-303.

Gunby, J.A., Darby, S.C., Miles, J.C.H., Green, B.M.R. and Cox, D.R. (1993) Factors affecting indoor radon concentrations in the United Kingdom. *Health Physics* 64, 2-12.

Gundersen, L.C.S. and Schumann, E.R. (1996) Mapping the radon potential of the United States: examples from the Appalachians. *Environmental International* 22, S829-S844.

Gunderson, L.C.S., Schumann, R.R., Otton, J.K., Dubiel, R.F., Owen, D.E. and Dickinson, K.A. (1992) Geology of radon in the United States. In: Gates, A.E. and Gunderson, L.C.S. (eds.) *Geologic Controls on Radon.* Geological Society of America. Special paper 271.

Hackshaw, A.K., Law, M.R. and Wald, N.J. (1997) The accumulated evidence on lung cancer and environmental tobacco smoke. *British Medical Journal* 315, 980-988.

Halliday, R.S. (1990) *The disappearing Scottish colliery: a personal view of some aspects of Scotland's coal industry since nationalisation.* Scottish Academic Press, Edinburgh.

Hammond, E.C. and Horn, D. (1954) The relationship between human smoking habits and death rates. *Journal of the American Medical Association* 155, 1316-1328.

Hart, C.L., Watt, G.C.M., Hawthorne, V.M., Hole, D.J., Gillis, C.R. and Smith, G.D. (2001) Social class differences in lung cancer mortality: Risk factor explanations using two Scottish cohort. *International Journal of Epidemiology* 30, 268-274.

Härting, F.H. and Hesse, W. (1879) Lung cancer, mountain disease in Scheeberg mines (in German). *Vjsschr. Gerichtl. Med.* 31, 313-337.

Hatch, M. and Thomas, D. (1993) Measurement issues in environmental epidemiology. *Environmental Health Perspectives* 101, 49-57.

Haugen, A. (2000) Etiology of Lung Cancer. In: Hansen, H.H. (ed.) *Textbook of Lung Cancer*, 1-12. Martin Dunitz, Copenhagen.

Haynes, R. (1987) *The Geography of Health Services in Britain.* Croom Helm, London.

Haynes, R. (1988) The urban distribution of lung cancer mortality in England and Wales 1980-1983. *Urban Studies* 25, 497-506.

HEBS (Health Education Board for Scotland) (2003) www.hebs.scot.nhs.uk/ Site accessed January 2003.

Helman, C.G. (1990) *Culture, Health and Illness.* Wright, London.

Hemminki, K. and Li, X. (2001) Incidence trends and risk factors of carcinoid tumors: a nationwide epidemiologic study from Sweden. *Cancer* 92, 2204-2210.

Hessel, P.A., Gamble, J.F., Gee, J.B., Gibbs, G., Green, F.H., Morgan, W.K. and Mossman, B.T. (2000) Silica, silicosis, and lung cancer: a response to a recent

working group report. *Journal of Occupational and Environmental Medicine / American College of Occupational and Environmental Medicine* 42, 704-720.

Hewitt, N. (1991) Spatial variations in nitrogen dioxide concentrations in an urban area. *Atmospheric Environment* 25B, 429-434.

Higgs, G. and White, S. (2000) Alternatives to census-based indicators of social disadvantage in rural communities. *Progress in Planning* 53, 1-81.

Hirayama, T. (1981) Non-smoking wives of heavy smokers have a higher risk of lung cancer. *British Medical Journal* 282, 916-917.

Hjalmars, U., Kulldorf, M., Gustafsson, G. and Nagarwalla, N. (1996) Childhood leukemia in Sweden: Using GIS and a spatial scan statistic for cluster detection. *Statistics in Medicine* 15, 707-715.

Hoffman, D. and Hoffman, I. (1997) The changing cigarette, 1950-1995. *Journal of Toxicology and Environmental Health* 50, 307-364.

Holland, D.M., De Oliveira, V., Cox, L.H. and Smith, R.L. (2000) Estimation of regional trends in sulfur dioxide over the eastern United States. *Environmetrics* 11, 373-393.

Holland, L.M. (1990) Crystalline silica and lung cancer: a review of recent experimental evidence. *Regulatory Toxicology and Pharmacology* 12, 224-237.

Hollywood, E. (2002) Mining, migration and immobility: towards an understanding of the relationship between migration and occupation in the context of the UK mining industry. *International Journal of Population Geography* 8, 297-314.

Hornung, R.W. and Meinhardt, T.J. (1987) Quantitative risk assessment of lung cancer in U.S. uranium miners. *Health Physics* 52, 417-430.

Hughes, J.S. and O'Riordan, M.C. (1993) *Radiation exposure of the UK population – 1993 review.* Chilton, NRPB-R263. HMSO, London.

Hugod, C., Hawkins, L.H. and Astrup, P. Exposure of passive smokers to tobacco smoke constituents. *International Archives of Occupational and Environmental Health* 4, 21-29.

IARC (International Agency for Research on Cancer) (1997a) *Cancer in Five Continents.* Vol VII. IARC, Lyon.

IARC (International Agency for Research on Cancer) (1997b) Silica, some silicates, coal dust and para-aramid fibrils. In: *IARC Monographs on the Evaluation of the Carcinogenic Risks to Humans* 68. IARC, Lyon.

ISD (Information and Statistics Division, NHS Scotland) (2002) ISD Online www.show.scot.nhs.uk/isd/ Site accessed April 2002

Jacobson, B. (1981) *The Ladykillers: Why Smoking is a Feminist Issue.* Pluto Press, London.

Jarman, B. (1984) Underprivileged areas: validation and distribution of scores. *British Medical Journal* 289, 1587-1592.

Jarvis, M.J., Goddard. E., Higgins, V., Feyerabend, C., Bryant A. and Cook, D.G. (2000) Children's exposure to passive smoking in England since the 1980s: cotinine evidence from population surveys. *British Medical Journal (Clinical Research Edition)* 321, 343-345.

Jedrychowski, W., Becher, H., Wahhrendorf, J. and Basa-Cierpialek, Z. (1990) A case-control study of lung cancer with special reference to the effect of air pollution in Poland. *Journal of Epidemiology and Community Health* 44, 321-324.

Jee, S.H., Ohrr, H. and Kim, I.S. (1999) Effects of husband's smoking on the incidence of lung cancer in Korean women. *International Journal of Epidemiology* 28, 824-828.

Jones, D.I.E., Gill, H. and Watson, J.L. (1991) *Spon's Quarry Guide to the British Hard Rock Industry.* Chapman and Hall, London.

Jones, K. (1991) *Multi-Level Models for Geographical Research.* CATMOG 54. Environmental Publications, Norwich.

Jones, K. and Duncan, C. (1995) Individualies and their ecologies: analyzing the geography of chronic illness within a multilevel modelling framework. *Health and Place* 1, 27-40.

Jones, K. and Moon, G. (1987) *Health, Disease and Society: An Introduction to Medical Geography.* Routledge, London.

Jousilahti P., Patja K. and Salomaa V. (2002) Environmental tobacco smoke and the risk of cardiovascular disease. *Scandinavian Journal of Work, Environment and Health* 28, Supplement 2, 41-51.

Julious, S.A., Nicholl, J. and George, S. (2001) Why do we continue to use standardised mortality ratios for small area comparisons? *Journal of Public Health Medicine* 23, 40-46.

Kadafar, K., Freedman, L.S., Goodall, C.R. and Tukey, J.W. (1996) Urbanicity-related trends in lung cancer mortality in US counties: white females and white males, 1970-1987. *International Journal of Epidemiology* 25, 918-932.

Katsouyanni, K. and Pershagen, G. (1997) Ambient air pollution exposure and cancer. *Cancer Causes and Control* 8, 284-291.

Kearns, R.A. (1991) The place of health in the health of place: The case of Hokinanga Special Medical Area. *Social Science and Medicine* 33, 519-530.

Kearns, R.A. (1993) Place and health: Towards a reformed Medical Geography? *Professional Geographer* 45, 139-147.

Kearns, R.A. (1994) To reform is not to discard: A reply to Paul. *Professional Geography* 46, 505-507.

Kearns, R.A. (1995) Medical geography: making space for difference. *Progress in Human Geography* 19, 251-259.

Kearns, R.A. (1996) AIDS and medical geography: embracing the other. *Progress in Human Geography* 20, 123-131.

Kearns, R.A. (1997) Narrative and metaphor in health geographies. *Progress in Human Geography* 21, 269-277.

Kearns, R.A. and Gesler, W.M. (1998) *Putting Health into Place: Landscape Identity and Well-Being*. Syracuse University Press, New York.

Kearns, R.A. and Moon, G. (2002) From medical to health geography: novelty, place and theory after a decade of change. *Progress in Human Geography* 26, 605-625.

Kearns, R.A., Taylor, S.M. and Dear, M. (1987) Coping and satisfaction among the chronically mentally disabled. *Canadian Journal of Community Mental Health* 6, 13-22.

Kelsall, J.E. (1996) *Kernel Smoothing Methodology for Application in Environmental Epidemiology*. Unpublished PhD Thesis, Lancaster University.

Kendall, G.M., Miles, J.C.H., Cliff, K.D., Green, B.M.R., Muirhead, C.R., Dixon, D.W., Lomas, P.R. and Goodridge, S.M. (1994) *Exposure to Radon in UK Dwellings*. NRPB-R272. National Radiological Protection Board, Chilton.

King, A.C., Taylor, C.B. and Haskell, W. (1990) Smoking in older women: is being a female a risk factor for continued cigarette use? *Archives of Internal Medicine* 150, 1841-1846.

King, G. (1997) *A Solution to the Ecological Inference Problem: Reconstructing Individual Behaviour from Aggregate Data*. Princeton University Press, Princeton.

King, R.J.B. (1996) *Cancer Biology*. Longmans, Harlow.

Kitching, R. (1990) Migration behaviour among the unemployed and low-skilled. In: Johnson, J.H. and Salt, J. (eds.) *Labour Migration: The Internal Geographical Mobility of Labour in the Developed World*, 172-190. David Fulton, London.

Klaeboe R., Kolbenstvedt M., Clench-Aas J. and Bartonova A. (2000) Oslo traffic study - part 1: an integrated approach to assess the combined effects of noise and air pollution on annoyance. *Atmospheric Environment* 34, 4727-4736.

Kline, S.W. and Mose, D.G. (1990) Indoor radon – a prediction from aeroradioactivity generated by surficial materials. *Geoderma* 47, 243-260.

Knox, E.G., Stewart, A.M., Kneale, G.W. and Gilman, E.A. (1987) Prenatal irradiation and childhood cancer. *Journal of the Society of Radiological Protection* 7, 3-15.

Knudsen, D.C. (1992) Generalizing Poisson regression: Including a priori information using the method of offsets. *Professional Geographer* 44, 202-208.

Kogevinas, M., Sala, M., Boffetta, P., Kazerouni, N., Kromhout, H. and Hoar-Zahm, S. (1998) Cancer risk in the rubber industry: a review of the recent epidemiological evidence. *Occupational and Environmental Medicine* 55, 1-12.

Kreuzer, M., Kreienbrock, L., Gerken, M., Heinrich, J., Bruske-Hohlfeld, I., Muller, K.M. and Wichmann, H.E. (1998) Risk factors for lung cancer in young adults. *American Journal of Epidemiology* 147, 1028-1037.

Kuempel, E.D., Stayner, L.T., Attfield, M.D. and Buncher, C.R. (1995) Exposure-response analysis of mortality among coal miners in the United States. *American Journal of Industrial Medicine* 28, 167-184.

Kulldorff, M., Feuer, E.J., Miller, B.A. and Freedman, L.S. (1997) Breast cancer in the northeastern United States: A geographical analysis. *American Journal of Epidemiology* 146, 161-170.

Kulldorf, M., Rand, K., Gherman, G., Williams, G. and DeFrancesco, D. (1998) *SatScan v2.1: Software for the Spatial and Space-Time Scan Statistics*. National Cancer Institute, Bethesda, MD.

Kvåle, G., Bjelke, E. and Heuch, I. (1986) Occupational exposure and lung cancer risk. *International Journal of Cancer* 37, 185-193.

Lagarde, F. and Pershagen, G. (1999) Parallel analyses of individual and ecological data on residential radon, cofactors, and lung cancer in Sweden. *American Journal of Epidemiology* 149, 268-274.

Lagarde, F., Pershagen, G., Akerblom, G., Axelson, O., Bäverstam, U., Damber, L., Enflo, A., Svartengren, M. and Swedjemark, G.A. (1997) Residential radon and lung cancer in Sweden: risk analysis accounting for random error in the exposure assessment. *Health Physics* 72, 269-276.

Lam, T.H., Hedley, A.J., Chung, S.F., Betson, C.L. and Wong C.M. (1998) Respiratory symptoms due to active and passive smoking in junior secondary school students in Hong Kong. *International Journal of Epidemiology* 27, 41-48.

Law, M.R. and Morris, J.K. (1998) Why is mortality higher in poorer areas and in more northern areas of England and Wales? *Journal of Epidemiology and Community Health* 52, 344-352.

Lawther, P.J. and Waller, R.E. (1978) Trends in urban air pollution in the United Kingdom in relation to lung cancer mortality. *Environmental Health Perspectives* 22, 71-73.

Laxan, D. and Noordally, E. (1987) Nitrogen dioxide distribution in street canyons. *Atmospheric Environment* 9, 1899-1904.

Learmonth, A. (1988) *Disease Ecology*. Blackwell, Oxford.

Lee, D. (1994) Regional variations in long-term visibility trends in the UK, 1962-1990. *Geography* 79, 108-121.

Lee, M.E., Lichtenstein, E., Andrews, J.A., Glasgow, R.E. and Hampson, S.E. (1999) Radon-smoking synergy: A population-based behavioral risk reduction approach. *Preventive Medicine* 29, 222-227.

Lee, P.N. (1992) *Environmental Tobacco Smoke and Mortality*. Karger, Basle.

Lee, P.N., Chamberlain, J. and Alderson, M.R. (1986) Relationship of passive smoking to risk of lung cancer and other smoking-associated diseases. *British Journal of Cancer* 54, 97-105.

Legetic, B., Stanisavljevic, D., Jakovljevic, D., Marinkovic, J. and Niciforovic, O. (1996) Health care delivery and the status of the population's health in the current crisis in former Yugoslavia using EPI-design methodology. *International Journal of Epidemiology* 25, 341-348.

Leigh, J., Driscoll, T.R., Cole, B.D., Beck, R.W., Hull, B.P. and Yang, J. (1994) Quantitative relation between emphysema and lung mineral content in coalworkers. *Occupational and Environmental Medicine* 51, 400-407.

Létourneau, E.G., Krewski, D., Choi, N.W., Goddard, M.J., McGregor, R.G., Zielinski, J.M. and Du, J. (1994) Case-control study of residential radon and lung cancer in Winnipeg, Manitoba, Canada. *American Journal of Epidemiology* 140, 310-322.

Levi, F., Lucchini, F., Negri, E. and La Vecchia, C. (1999) Worldwide patterns of cancer mortality, 1990-1994. *European Journal of Cancer Prevention* 8, 381-400.

Lewis, S., Bennett, J., Richards, K. and Britton, J. (1996) A cross sectional study of the independent effect of occupation on lung function in British coal miners. *Occupational and Environmental Medicine* 53, 125-129.

Lichtenstein, E., Andrews, J.A., Lee, M.E., Glasgow, R.E. and Hampson, S.E. (2000) Using radon risk to motivate smoking reduction: evaluation of written materials and brief telephone counselling. *Tobacco Control* 9, 320-326.

Liu, B., Peto, R., Chen, Z., Boreham, J., Wu, Y., Li, J., Campbell, T.C. and Chen, J. (1998) Emerging tobacco hazards in China: 1. Retrospective proportional mortality study of one million deaths. *British Medical Journal* 317, 1411-1422.

Litva, A. and Eyles, J. (1995) Coming out: exposing social theory in medical geography. *Health and Place* 1, 5-14.

Lovett, A. and Flowerdew, R. (1989) Analysis of Count Data Using Poisson Regression. *Professional Geographer* 41, 190-198.

Lovett, A., Gale, S., Haynes, R. and Sünnenberg, G. (2002) Car travel time and accessibility by bus to general practitioner services: A study using patient registers GIS. *Social Science and Medicine* 55, 97-111.

Lovett, A., Whyte, I.D. and Whyte, K.A. (1985) Poisson regression analysis and migration fields: the example of the apprenticeship records of Edinburgh in the seventeenth and eighteenth centuries. *Transactions of the Institute of British Geographers* N.S. 10, 317-332.

Lubin, J.H. (1994) Invited commentary: Lung cancer and exposure to residential radon. *American Journal of Epidemiology* 140, 323-332.

Lubin, J.H., Boice, J.D., Edling, C., Hornung, R.W., Howe, G., Kunz, E. *et al.* (1994) *Radon and lung cancer risk: a joint analysis of 11 underground miners studies*. National Institutes of Health Publications, National Institutes of Health, Bethesda, MD.

Macintyre, S., Ellaway, A. and Cummins, S. (2002) Place effects on health: How can we conceptualise, operationalise and measure them? *Social Science and Medicine* 55, 125-139.

Macintyre, S., Maciver, S. and Sooman, A. (1993) Area, class and health: should we be focusing on places or people? *Journal of Social Policy* 22, 213-234.

Madge, C. (1997) Public parks and the geography of fear. *Tijdschrift voor Economische en Sociale Geografie* 88, 237-250.

Magadi, M.A., Madise, N.J. and Rodrigues, R.N. (2000) Frequency and timing of antenatal care in Kenya: explaining the variations between women of different communities. *Social Science and Medicine* 51, 551-561.

Magnus, K., Engeland, A., Green, B.M.R., Haldorsen, T., Muirhead, C.R. and Strand, T. (1994) Residential radon exposure and lung cancer – an epidemiological study of Norwegian municipalities. *International Journal of Cancer* 58, 1-7.

Mannino, D.M., Ford, E., Giovino, G.A. and Thun, M. (1998) Lung cancer deaths in the United States from 1979 to 1992: an analysis using multiple-cause mortality data. *International Journal of Epidemiology* 27, 159-166.

Mao, Y., Fincham, S., Hu, J., Ugnat, A. and Semenciw, R. (2001) Socioeconomic status and lung cancer risk in Canada. *International Journal of Epidemiology* 30, 809-817.

Marcinowski, F., Lucas, R.M. and Yeager, W.M. (1994) National and Regional Distributions of Airborne Radon Concentration in U.S. Homes. *Health Physics* 66, 699-706.

Markkanen, M. and Arvela, H. (1992) Radon emanation from soils. *Radiation Protection Dosimetry* 45, 269-272.

Martikainen, P., Lahelma, E., Ripatti, S., Albanes, D. and Virtamo, J. (2000) Educational differences in lung cancer mortality in male smokers. *International Journal of Epidemiology* 29, 264-267.

Martin, D., Diamond, I., Brigham, P., Roderick, P. and Barnett, S. (2000) The (mis)representation of rural deprivation. *Environment and Planning A* 32, 735-751.

Martland, H.S. (1929) Occupational poisoning in the manufacture in luminous watch dials. *Journal of the American Medical Association* 92, 466-473.

Martland, H.S. and Humphries, R.E. (1929) Osteogenic sarcoma in dial painters using luminous paint. *Archives of Pathology* 7, 406-417.

Matsukura, S., Taminato, T., Kitano, N., Seino, Y., Hamada, H., Uchihashi, M., Nakajima, H. and Hirata, Y. (1984) Effects of environmental tobacco smoke on urinary cotinine excretion in nonsmokers. Evidence for passive smoking. *The New England Journal of Medicine* 311, 828-832.

Mayne, S.T., Buenconsejo, J. and Janerich, D.T. (1999) Previous lung disease and risk of lung cancer among men and women nonsmokers. *American Journal of Epidemiology* 149, 13-20.

Mayne, S.T., Janerich, D.T., Greenwald, P., Chorost, S., Tucci, C., Zaman, M.B., Melamed, M.R., Kiely, M. and McKneally, M.F. (1994) Dietary beta carotene and lung cancer risk in U.S. nonsmokers. *Journal of the National Cancer Institute* 86, 33-38.

McClellan, R.O. (1996) Lung cancer in rats from prolonged exposure to carbonaceous particles: implications for human risk assessment. *Inhalant Toxicology* 8 (suppl), 193-226.

McDonald, A.D. and McDonald, J.C. (1980) Malignant Mesothelioma in North America. *Cancer* 46, 1650-1656.

McDonald, A.D., McDonald, J.C., Rando, R.J., Hughes, J.M. and Weill, H. (2001) Cohort mortality study of North American industrial sand workers. I. Mortality from lung cancer, silicosis and other causes. *The Annals of Occupational Hygiene* 45, 193-199.

McLoone, P. and Boddy, F.A. (1994) Deprivation and mortality in Scotland, 1981 and 1991. *British Medical Journal* 309, 1465-1470.

Mechanic, D. (1993) Social research in health and the American sociopolitical context: the changing fortunes of medical sociology. *Social Science and Medicine* 36, 95-102.

Merlo, J., Östergren, P.O., Hagberg, O., Lindström, M., Lindgren, A., Melander, A., Råstram, L. and Bergland, G. (2001) Diastolic blood pressure and area of residence: multilevel versus ecological analysis of social inequity. *Journal of Epidemiology and Community Health* 55, 791-798.

Miles, J. (1998) Mapping radon-prone areas by lognormal modeling of house radon data. *Health Physics* 74, 370-378.

Miles, J.C.H. and Ball, T.K. (1996) Mapping radon-prone areas using house radon data and geological boundaries. *Environmental International* 22 (Suppl. 1), 779-782.

Miles, J.C.H., Green, B.M.R. and Komas, P.R. (1993) Radon affected areas: Scotland *Documents of the NRPB* 4. National Radiological Protection Board, Chilton.

Miller, B.G., Buchanan, D., Hurley, J.F., Robertson, A., Hutchinson, P.A., Kidd, M.W., Pilkington, A.D. and Soutar, C.A. (1997) *The Effects of Exposure to Diesel Fumes, Low-Level Radiation, and Respirable Dust and Quartz, on Cancer Mortality in Coal Miners.* IOM Report TM/97/04. Institute of Occupational Medicine, Edinburgh.

Miller, B.G. and Jacobsen, M. (1985) Dust exposure, pneumoconiosis, and mortality of coal miners. *British Journal of Industrial Medicine* 43, 723-733.

Mitchell, R. (2001) Multilevel modelling might not be the answer. *Environment and Planning A* 33, 1257-1360.

Miyazaki, M. and Une, H. (2001) Risk of lung cancer among Japanese coal miners on hazard risk and interaction between smoking and coal mining. *Journal of Occupational Health* 43, 225-230.

Moon, G. (1995) (Re)placing research on health and health care. *Health and Place* 1, 1-4.

Moon, G. (1997) Applicability and theory in health-related research. *Health and Place* 3, iii-iv.

Moon, G., Gould, M. and Jones, K. (1998) Seven up – refreshing medical geography: an introduction to selected papers from the Seventh International Symposium in Medical Geography, Portsmouth, UK. *Social Science and Medicine* 46, 627-630.

Moon, G., Gould, M. *et al* (2000) *Epidemiology - An Introduction*. Open University Press, Buckingham.

Morris, R. and Carstairs, V. (1991) Which deprivation? A comparison of selected deprivation indexes. *Journal of Public Health Medicine* 13, 318-326.

Mossman, B.T. (1988) Carcinogenic potential of asbestos and non-asbestos fibres. *Journal of Environmental Science and Health* 6, 151-156.

Muir, D. and Laxen, D.P.H. (1995) Black Smoke as a surrogate for PM(10) in health studies. *Atmospheric Environment* 29, 959-962.

Murray, C.J. and Lopez, A.D. (1997) Mortality by cause for eight regions of the world: Global Burden of Disease Study. *Lancet* 349 1269-1276.

Muscat, J.E., Stellman, S.D. and Wynder, E.L. (1995) Insulation, asbestos, smoking-habits, and lung-cancer cell-types. *American Journal of Industrial Medicine* 27, 257-269.

National Research Council (1986). *Environmental Tobacco Smoke; Measuring Exposures and Assessing Health Effects.* National Academy Press, Washington DC.

National Statistics (2002) www.statistics.gov.uk/downloads/theme_health/DS16/ds16.asp Site accessed June 2002.

Natusch, D.F. (1978) Potentially carcinogenic species emitted to the atmosphere by fossil-fuelled power plants. *Environmental Health Perspectives* 22, 79-90.

Nero, A.V., Leiden, S.M., Nolan, D.A., Price, P.N., Rein, K.L., Revzan, H.R., Wollenberg, H.R. and Gadgil, A.J. (1994) Statistically based methodologies for mapping of radon 'actual' concentrations: the case of Minnesota. *Radiation Protection Dosimetry* 56, 215-219.

NETCEN (2001) UK National Air Quality Information Archive. www.aeat.co.uk/netcen/airqual/ Site accessed October 2001

Neuberger, J.S., Brownson, R.C., Morantz, R.A. and Chin, T.D. (1991) Association of brain cancer with dental X-rays and occupation in Missouri. *Cancer Detection and Prevention* 15, 31-34.

Ng, T.P., Chan, S.L. and Lee, J. (1990) Mortality of a cohort of men in a silicosis register: further evidence of an association with lung cancer. *American Journal of Industrial Medicine* 17, 163-171.

NRPB (National Radiological Protection Board) (1998) *Living with Radiation.* NRPB, Chilton.

NRPB (National Radiological Protection Board) (2000) *Health Risks from Radon.* NRPB, Chilton.

Nyberg, F., Agrenius, V., Svartengren, K., Svensson, C. and Pershagen, G. (1998) Dietary factors and risk of lung cancer in never-smokers. *International Journal of Cancer* 78, 430-436.

O'Brien, L. (1992) *Introducing Quantitative Geography. Measurements, Methods and Generalised Linear Models.* Routledge, London.

Oliver, M.A. and Badr, I. (1995) Determining the spatial scale of variation in soil radon concentration. *Mathematical Geology* 27, 893-922.

Openshaw, S., Charlton, M., Wymer, C. and Craft, A.W. (1987) A mark one analysis machine for the automated analysis of point data sets. *International Journal of Geographical Information Systems* 1, 335-358.

Openshaw, S., Craft, A.W., Charlton, M. and Birch, J.M. (1988) Investigation of leukaemia clusters by use of a geographical analysis machine. *Lancet*, 272-273.

Ordnance Survey (2002) Strategi. www.ordsvy.gov.uk/productpages/strategi/ Site accessed March 2002.

Osann, K.E. (1991) Lung-cancer in women – the importance of smoking, family history of cancer, and medical history of respiratory-disease. *Cancer Research* 51, 4893-4897.

Pandey, M., Mathew, A. and Nair, M.K. (1999) Global perspective of tobacco habits and lung cancer: a lesson for third world countries. *European Journal of Cancer Prevention* 8, 271-279.

Parkin, D.M., Pisani, P. and Ferlay, J. (1999) Estimates of the worldwide incidence of 25 major cancers in 1990. *International Journal of Cancer* 80, 827-841.

Partanaen, T. and Boffetta, P. (1994) Cancer risk in asphalt workers and roofers – review and metaanalysis of epidemiologic studies. *American Journal of Industrial Medicine* 26, 721-740.

Partanen, T., Jaakkola, J. and Tossavainen, A. (1995) Silica, silicosis and cancer in Finland. *Scandinavian Journal of Work, Environment and Health* 21, Supplement 2, 84-86.

Paul, B.K. (1994) Commentary on Kearns's 'Place and health: Towards a reformed medical geography'. *Professional Geographer* 46, 504-505.

Pawson, E. and Banks, G. (1993) Rape and fear in a New Zealand City. *Area* 25, 55-63.

Payne G., Payne, S. and Hyde, M. (1996) 'Refuse of All Classes'? Social Indicators and Social Deprivation. Sociological Research Online 1. http://www.socresonline.org.uk/1/1/3.html

Payne, S. (2001) 'Smoke like a man, die like a man'?: A review of the relationship between gender, sex and lung cancer. *Social Science and Medicine* 53, 1067-1080.

Pearce, J., Boyle, P. and Flowerdew, R. (2003) Predicting Smoking Behaviour in Census Output Areas across Scotland. *Health and Place* 9, 139-149.

Pershagen, G. (1990) Air pollution and cancer. In: Pershagen, G., Vainio, H., Sorsa, M. and McMichael, A.J. (eds.). *Complex Mixtures and Cancer Risk.* International Agency for Research on Cancer, Lyon.

Pershagen, G., Åkerblom, G., Axelson, O., Clavensjö, B., Damber, L., Desai, G., Enflo, A., Lagarde, F., Mellander, H., Svartengren, M., *et al.* (1994) Residential radon exposure and lung cancer in Sweden. *New England Journal of Medicine* 330, 159-164.

Pershagen, G., Elinder, C.G. and Bolander, A.M. (1977) Mortality in a region surrounding an arsenic smelting plant. *Environmental Health Perspectives* 19, 133-137.

Peto, R. (1994) Smoking and death: the past 40 years and the next 40. *British Medical Journal* 309, 937-939.

Peto, R., Lopez, A.D., Boreham, J., Thun, M. and Heath, C. Jr., (1992) Mortality from tobacco in developed countries: indirect estimation from national vital statistics. *Lancet* 339, 1268-1278.

Peto R., Lopez, A.D., Boreham, J., Thun, M. and Heath, C. Jr. (1994) *Mortality from smoking in developed countries 1950-2000: indirect estimates from national vital statistics*. Oxford University Press, Oxford.

Peto, R. and Darby, S. (1994) Radon risk reassessed. *Nature* 368, 97-98.

Pezzotto, S.M. and Poletto, L. (1999) Occupation and histopathology of lung cancer: A case-control study in Rosario, Argentina. *American Journal Of Industrial Medicine* 36, 437-443.

Phillimore, P. (1993) How do places shape health? Rethinking locality and lifestyle in northeast England. In: Platt, S., Thomas, H., Scott, S. and Williams, G. (eds.), *Locating Health: Sociological and Historical Explanations*. Avebury, Aldershot.

Phillips, P.H., Linet, M.S. and Harris, E.L. (1991) Assessment of family history information in case-control cancer studies. *American Journal of Epidemiology* 133, 757-765.

Philo, C. (1996) Staying in? Invited comments on 'Coming out: exposing social theory in medical geography'. *Health and Place* 2, 35-40.

Pike, M.C. and Forman, D. (1991). Epidemiology of Cancer. In: Franks, L.M. and Teich, N.M. (eds.) *Introduction to the Cellular and Molecular Biology of Cancer,* 49-97. Oxford University Press, Oxford.

Pinel, J., Fearn, T., Darby, S. C. and Miles, J.C. (1995) Seasonal correction factors for indoor radon measurements in the United Kingdom. *Radiation Protection Dosimetry* 58, 127-132.

Pirastu, R., Bartoli, D., Battista, G., De Santis, M., Iaia, T., Orsi, D., Tarchi, M. and Valiani, M. (1998) Cancer mortality of art glass workers in Tuscany, Italy. *Scandinavian Journal of Work Environment and Health* 24, 386-391.

Plane, D.A. and Rogerson, P.A. (1994) *The Geographical Analysis of Population with Applications to Planning and Business.* Wiley and Sons, New York.

Pless-Mulloli, T., Phillimore, P., Moffatt, S., Bhopal, R., Foy, C., Dunn, C. and Tate, J. (1998) Lung cancer, proximity to industry and poverty in Northeast England. *Environmental Health Perspectives* 106, 189-196.

Pobel, D. and Viel, J. (1997) Case-control study of leukaemia among young people near La Hague nuclear reprocessing plant: the environmental hypothesis revisited. *British Medical Journal* 314, 101-106.

Poland, D.B. (1998) Smoking, stigma and the purification of public places. In: Kearns, R.A. and Gesler, W.M. (eds.) *Putting Health into Place: Landscape, Identity and Wellbeing.* Syracuse University Press, New York.

Polgar, S. (1968) Health. In: Sills, D.L. (ed.) *International Encyclopaedia of the Social Sciences*, vi. 330-336. Macmill and Free Press, New York.

Powell, M. (1990) Need and provision in the NHS, an inverse care law. *Policy and Politics* 18, 31-37.

Price, B. (1997) Analysis of current trends in United States mesothelioma incidence. *American Journal of Epidemiology* 145, 211-218.

Redmond, C.K. (1983) Cancer mortality among coke oven workers. *Environmental Health Perspectives* 52, 67-73.

Reece, J., MacKintosh, A.M., MacAskill, S.M. and Stead, M. (2000) *An Investigation into Smoking Cessation in Disadvantaged Communities* CE1182/0101. Results from full survey (unpublished draft). Centre for Social Marketing, University of Strathclyde, Glasgow.

Rigby, J.E. and Gatrell, A.C. (2000) Spatial patterns in breast cancer incidence in north-west Lancashire. *Area* 32, 71-78.

Rinsky, R.A., Melius, J.M., Hornung, R.W., Zumwalde, R.D., Waxweiler, R.J., Landrigan, P.J., Bierbaum, P.J. and Murray, W.E. (1988) Case-control study of lung cancer in civilian employees at the Portsmouth Naval Shipyard, Kittery, Maine. *American Journal of Epidemiology* 127, 55-64.

Rodriguez, V., Tardon, A., Kogevinas, M., Prieto, C.S., Cueto, A., Garcia, M., Menendez, I.A. and Zaplana, J. (2000) Lung cancer risk in iron and steel foundary workers: A nested case control study in Asturias, Spain. *American Journal of Industrial Medicine* 38, 644-650.

Robinson, W.S. (1950) Ecological correlations and the behavior of individuals. *American Sociological Review* 15, 351-357.

Rodriguez, G. and Goldman, N. (1995) An assessment of estimation procedures for multilevel models with binary responses. *Journal of the Royal Statistical Society A* 158, 73-90.

Ron, E., Modan, B. and Boice, J.D. (1988) Mortality after radiotherapy for ringworm of the scalp. *American Journal of Epidemiology* 127, 713-725

Rosenman K.D. and Hall, N. (1996) Occupational risk factors for developing tuberculosis. *American Journal of Industrial Medicine* 30, 148-154.

RPII (Radiological Protection Institute of Ireland) (1996) *Radon in Irish Dwellings.* 1:750,000 scale. Radiological Protection Institute of Ireland, Dublin.

Ruosteenoja, E., Mäkeläinen, I., Rytömaa, T., Hakulinen, T. and Hakama, M. (1996) Radon and lung cancer in Finland. *Health Physics* 71, 185-189.

Rushton, L. (2001) Cancer and air pollution. In: Brimblecombe, P. and Maynard, R.L. *The Urban Atmosphere and its Effects. Air Pollution Reviews – Volume 1.* Imperial College Press, London.

Sabel, C.E., Gatrell. A.C., Löytönen, M., Masssilta, P. and Jokelainen, M. (2000) Modelling exposure opportunities: estimating relative risk for motor neurone disease in Finland. *Social Science and Medicine* 50, 1121-1137.

Samet, J.M., Humble, C.G. and Pathak, D.R. (1986) Personal and family history of respiratory disease and lung cancer risk. *American Review of Respiratory Disease* 134, 466-470.

Saracci, R. (1987) The interactions of tobacco smoking and other agents in cancer etiology. *Epidemiological Reviews* 9, 175-193.

Saul, C. and Payne, N. (1999) How does the prevalence of specific morbidities compare with measures of socio-economic status at small area level? *Journal of Public Health Medicine* 21, 340-347.

Schottenfeld, D. (1996) Epidemiology of lung cancer. In: Mitchell, J.B, Johnson, D.H. and Turrisi, A.T. *Lung Cancer: Principles and Practice*, 305-321. Lippincott-Raven, Philadelphia.

Schneider, D., Greenberg, M.R. and Choi, D. (1993) Screening for time-space cancer clusters at the minor civil division scale: a practical protocol. *Geographia Medica* 23, 177-187.

Schwartz, J., Ballester, F., Saez, M., Perez-Hoyos, S., Bellido, J., Cambra, K., Arribas, F., Canada, A., Perez-Boillos, M.J. and Sunyer, J. (2001) The concentration-response relation between air pollution and daily deaths. *Environmental Health Perspectives* 109, 1001-1006.

Scottish Executive (2002) *Social Justice a Scotland where everyone matters. Indicators of progress: Definitions, data, baseline and trends information.* Annex to the Social Justice Annual Report 2002. www.scotland.gov.uk/library5/social/emsjt-00.asp Site accessed December 2002.

Scottish Executive (2003) *Scottish Neighbourhood Statistics, Scottish Index of Deprivation.* www.scotland.gov.uk/stats/neighbours/tables/summary.pdf Site accessed February 2003.

Scottish Office (1997) *Keeping Scotland Moving. A Scottish Transport Green Paper.* The Stationery Office, London.

Selikoff, I.J., Hammond, E.C. and Churg, J. (1968) Asbestos exposure, smoking and neoplasia. *Journal of the American Medical Association* 204, 106-112.

Sellers, T.A., Bailey-Wilson, J.E., Elston, R.C., Wilson, A.F., Elston, G.Z., Ooi, W.L. and Rothschild, H. (1990) Evidence for mendelian inheritance in the pathogenesis of lung cancer. *Journal of the National Cancer Institute* 82, 1272-1279.

Senior, M., Williams, H. and Higgs, G. (2000) Urban-rural mortality differentials: controlling for material deprivation. *Social Science and Medicine* 51, 289-305.

Ševc, J., Kunz, E., Tomasášek, L., Placek, V. and Horácek, J. (1988) Cancer in man after exposure to Rn daughters. *Health Physics* 54, 27-46.

Shah, R., Uren, Z., Baker, A. and Majeed, A. (2001) Trends in deaths from drug overdose and poisoning in England and Wales 1993-1998. *Journal of Public Health Medicine* 23, 242-246.

Shairer, E. and Schöniger, E. (2001) Lung cancer and tobacco consumption. *International Journal of Epidemiology* 30, 24-27.

Shaw, A., McMunn, A. and Field, J. (2000) *The Scottish Health Survey 1998: Summary.* http://www.show.scot.nhs.uk/scottishhealthsurvey/

Shewry, M.C., Smith, W.C.S., Woodward, M. and Tunstall-Pedoe, H. (1992) Variations in coronary risk factors by social status: results from the Scottish heart health study. *British Journal of General Practice* 42, 406-410.

Shimizu, Y., Pierce, D.A., Preston, D.L. and Mabuchi, K. (1999) Studies of the mortality of atomic bomb survivors. Report 12, part II. Noncancer mortality: 1950-1990. *Radiation Research* 152, 374-389.

Shucksmith, M. (1990) *The Definitions of Rural areas and Rural Deprivation.* Report to Scottish Homes, Roseberry House, 9 Haymarket Terrace, Edinburgh EH12 5YA.

Simmonds, P. (1999) Managing patients with lung cancer: New guidelines should improve standards of care. *British Medical Journal* 319, 527-528.

Simon, S.L. and Graham, J.C. (1997) Findings of the first comprehensive radiological monitoring program of the Republic of the Marshall Islands. *Health Physics* 73, 66-85.

Sinha, R., Kulldorff, M., Curtin, J., Brown, C.C., Alavanja, M.C.R., and Swanson, C.A. (1998) Fried, well-done red meat and risk of lung cancer in women (United States). *Cancer Causes and Control* 9, 621-630.

Slunga, E. (1988) Radon classification of building ground. *Radiation Protection Dosimetry* 24, 39-42.

Smallman-Raynor, M. and Cliff, A. (1999) The spatial dynamics of epidemic diseases in war and peace: Cuba and the insurrection against Spain, 1895-98. *Transactions of the Institute of British Geographers* NS 24, 331-352.

Smallman-Raynor, M., Muir, K.R. and Smith, S.J. (1998) The geographical assignment of cancer units: patient accessibility as an optimal location problem. *Public Health* 112, 379-383.

Soutar, C.A., Robertson, A., Miller, B.G., Searl, A. and Bignon, J. (2000) Epidemiological evidence on the carcinogenicity of silica: factors in scientific judgement. *The Annals of Occupational Hygiene* 44, 3-14.

Speizer, F.E. and Samet, J.M. (1994) Air pollution and lung cancer. In: Samet, J.M. (ed.) *Epidemiology of Lung Cancer*, 131-146. Dekker, New York.

Stead, M., Reece, J., Eadie D., MacKintosh, A., Reece, J. and Eadie, D. (2001) "It's as if you're locked in": Qualitative explanations for area effects on smoking in disadvantaged communities and others. *Health and Place* 7, 333-343.

Steenland, K., Henley, J. and Thun, M. (2002) All-cause and cause-specific death rates by educational status for two million people in two American Cancer Society cohorts, 1959-1996. *American Journal of Epidemiology* 156, 11-21.

Steenland, K., Loomis, D., Shy, C. and Simonsen N. (1996) Review of occupational lung carcinogens. *American Journal of Industrial Medicine* 29, 474-490.

Steenland, K. and Sanderson, W. (2001) Lung cancer among industrial sand workers exposed to industrial silica. *American Journal of Epidemiology* 153, 695-703.

Steinmetz, K.A. and Potter, J.D. (1991) Vegetables, fruits, and cancer II. Mechanisms. *Cancer Causes and Control* 2, 427-442.

Stewart, A. (1997) A-bomb data: detection of bias in the Life Span Study cohort. *Environmental Health Perspectives* 105, 1519-1521.

Stewart, A., Webb, J., Giles, D. and Hewitt, D. (1956) Malignant disease in childhood and diagnostic irradiation in utero. *Lancet* 2, 447.

Stidley, C.A. and Samet, J.M. (1993) A review of ecologic studies of lung cancer and indoor radon. *Health Physics* 65, 234-251.

Stockwell, H.G., Armstrong, A.W. and Leaverton, P.E. (1990) Histopathology of lung cancers among smokers and nonsmokers in Florida. *International Journal of Epidemiology* 19 (Suppl. 1), s48-s52.

Stockwell, H.G., Goldman, A.L., Lyman, G.H., Noss, C.I., Armstrong, A.W., Pinkham, P.A., Candelora, E.C. and Brusa, M.R. (1992) Environmental tobacco smoke and lung cancer risk in non-smoking women. *Journal of the National Cancer Institute* 84, 1417-1422.

Takahashi, T., Trott, K.R., Fujimori, K., Simon, S.L., Ohtomo, H., Nakashima, N., Takaya, K., Kimura, N., Satomi, S. and Schoemaker, M.J. (1997) An investigation into the prevalence of thyroid disease on Kwajalein Atoll, Marshall Islands. *Health Physics* 73, 199-213.

Talbot, D.K., Appleton, J.D., Ball, T.K. and Strutt, M.H. (1998) A comparison of field and laboratory analytical methods for radon site investigation. *Journal of Geochemical Exploration* 65, 79-90.

Tango, T. (1994) Effects of air pollution on lung cancer: a Poisson regression model based on vital statistics. *Environmental Health Perspectives* 102 (Suppl 8) 41-45.

Tayanc, M. (2000) An assessment of spatial and temporal variation of sulfur dioxide levels over Istanbul, Turkey. *Environmental Pollution* 107, 61-69.

Tokuhata, G.K. and Lilienfield, A.M. (1963) Familial aggregation of lung cancer among hospital patients. *Public Health Report* 78, 277-312.

Townsend, P., Phillimore, P. and Beattie, A. (1988) *Health and Deprivation: Inequality and the North*. Croom Helm, London.

Travis, W.D., Ravis, L.B. and Devesa, S.S. (1995) Lung cancer, *Cancer* 75, 191-202.

Trichopoulos, D., Kalandidi, A. and Sparros, L. (1983) Lung cancer and passive smoking: conclusion of Greek study. *Lancet* 2, 667-678.

Trichopoulos, D. (1984) Passive smoking and lung cancer. *Lancet*, 684.

Turnbull, B.W., Iwano, E.J., Burnett, W.S., Howe, H.L. and Clark, L.C. (1990) Monitoring for clusters of disease; application to leukaemia incidence in upstate New York. *American Journal of Epidemiology* 132 (Supplement 1), s136-s143.

Twigg, L. (1999) Choosing a national survey to investigate smoking behaviour: making comparisons between the General Household Survey, the British Household Survey and the Health Survey for England. *Journal of Public Health Medicine* 21, 14-21.

Twigg, L., Moon, G. and Jones, K. (1998) Predicting small-area health related behaviour: a comparison of smoking and drinking indicators. In: Earickson, R. and Schneider, D. (eds.) *Proceedings of the Eighth International Medical Geography Symposium*, University of Maryland, Baltimore.

Twigg, L., Moon, G. and Jones, K. (2000) Predicting small-area health related behaviour: a comparison of smoking and drinking indicators. *Social Science and Medicine* 50, 1109-1120.

Uitenbroek, D.G. and McQueen, D.V. (1993) Trends in cigarette smoking by gender, age and occupational status. *Scottish Medical Journal* 38, 12-15.

UK Childhood Cancer Study Investigators (2000) The United Kingdom Childhood Cancer Study: objectives, materials and methods. *British Journal of Cancer* 82, 1073-1102.

UK Department of Health and Social Security (1988) *Fourth Report of the Independent Scientific Committee on Smoking and Health*. HMSO, London.

Une, H., Esaki, H., Osajima, K., Ikui, H. and Shigematsu, T. (1995) A prospective study on mortality among Japanese coal miners. *Industrial Health* 33, 67-76.

Uren, Z. and Fitzpatrick, J. (2001) Analysis of mortality by deprivation and cause of death. In: Griffiths, C. and Fitzpatrick, J. (eds.) *Geographic Variations in Health*. HMSO, London.

Uren, Z., Fitzpatrick, J., Reid, A. and Goldblatt, P. (2001) Geographic variation in mortality by Social Class and alternative social classifications. In: Griffiths, C. and Fitzpatrick, J. (eds.) *Geographic Variations in Health*. HMSO, London.

US Department of Health and Human Services (1986) *The Health Consequences of Involuntary Smoking. A Report of the Surgeon General.* DHHS, Washington, DC.

Vacchino, M.N. (1999) Poisson regression in mapping cancer mortality. *Environmental Research* 81, 1-17.

Vagero, D. and Persson, G. (1986) Occurrence of cancer in socioeconomic groups in Sweden. *Scandinavian Journal of Social Medicine* 14, 151-160.

Verheij, R.A., van de Mheen, H.D., de Bakker, D.H., Groenewegen, P.P. and Mackenbach, J.P. (1998) Urban-rural variations in health in The Netherlands: does selective migration play a part? *Journal of Epidemiology and Community Health* 52, 487-493.

Viel, J.F., Arveux, P., Baverel, J. and Cahn, J.Y. (2000) Soft-tissue sarcoma and non-hodgkin's lymphoma clusters around a municipal solid waste incinerator with high dioxin emission levels. *American Journal of Epidemiology* 152, 13-19.

Wald, N. and Nicolaides-Bouman, A. (1988) (eds.) *UK Smoking Statistics.* Oxford University Press, Oxford.

Wallace, M. and Denham, C. (1996) ONS Classification of local and health authorities of Great Britain, *Studies on Medical and Population Subjects* 59. HMSO, London.

Walter, S.D. (1994) A simple test for spatial pattern in regional health data. *Statistics in Medicine* 13, 1037-1044.

Wang, X., Yano, E., Nonaka, K., Wang, M. and Wang, Z. (1997) Respiratory impairments due to dust exposure: a comparative study among workers exposed to silica, asbestos, and coalmine dust. *American Journal of Industrial Medicine* 31, 495-503.

Washington, J.W. and Rose, A.W. (1992) Temporal variability of radon concentration in the interstitial gas of soils in Pennsylvania. *Journal of Geophysical Research* 97, 9145-9159.

Watts, S. (1997) *Epidemics and History: Disease, Power and Imperialism.* Yale University Press, New Haven.

Weston, T.L., Aronson, K.J., Siemiatycki, J., Howe, G.R. and Nadon, L. (2000) Cancer mortality among males in relation to exposures assessed through a job-exposure matrix. *International Journal of Occupational and Environmental Health* 6, 194-202.

Whitley, E., Gunnell, D., Dorling, D. and Davey Smith, G. (1999) Ecological study of social fragmentation, poverty, and suicide. *British Medical Journal* 319, 1034-1037.

Whittemore, A.S. and McMillan, A. (1993) Lung cancer mortality among U.S. uranium miners: a reappraisal. *Journal of the National Cancer Institute* 71, 489-499.

WHO (World Health Organization) (1988) *Emissions of Heavy Metal and PAH Compounds from Municipal Solid Waste Incinerators. Control Technology and Health Effects*. Report on a WHO meeting, Florence. WHO, Copenhagen.

WHO (World Health Organization) (2000) Report of the Second WHO European Meeting on Evidence-Based Treatment of Tobacco Dependence. WHO, Copenhagen.

WHO (World Health Organization) (2002) www.who.org Site Accessed February 2002.

White, S.B., Bergsten, J.W., Alexander, B.V., Rodman, N.F. and Phillips, J.L. (1992) Indoor ^{222}Rn concentrations in a probability sample of 43,000 houses across 30 states. *Health Physics* 62, 41-50.

Wiggs, L.D., Cox-DeVore, C.A., Voelz, G.L. and Reyes, M. (1991) Mortality among workers exposed to external ionizing radiation at a nuclear facility in Ohio. *Journal of Occupational Medicine* 33, 632-637.

Wilkinson, R.G., (1996) *Unhealthy Societies: The Afflictions of Inequality.* Routledge, London.

Willett, W.C. (2001) Diet and cancer: one view at the start of the millennium. *Cancer Epidemiology, Biomarkers and Prevention* 10, 3-8.

Williams, C. (1992*) Lung Cancer: The Facts*. Oxford University Press, Oxford.

Williams, F.L. and Lloyd, O. (1991) Trends in lung cancer mortality in Scotland and their relation to cigarette smoking and social class. *Scottish Medical Journal* 36, 175-178.

Williams, R.R. and Horm, J.W. (1997) Association of cancer sites with tobacco and alcohol consumption and socioeconomic status of patients: interview study from the Third National Cancer Survey. *Journal of the National Cancer Institute* 58, 525-547.

Woodward, M. (1996) Small area statistics as markers for personal social status in the Scottish heart health study. *Journal of Epidemiology and Community Health* 50, 570-576.

World Bank (1999) *Curbing the Epidemic: Governments and the Economics of Tobacco Control.* World Bank, Washington DC.

Wrigley, N., Warm, D., Margetts, B. and Whelan, A. (2002) Assessing the impact of improved retail access on diet in a 'food desert': A preliminary report. *Urban Studies* 39 11, 2061-2082.

Wrixon, A.D., Green, B.M.R., Lomas, P.R., Miles, J.C.H., Cliff, K.D., Francis, E.A., Driscoll, C.M.H., James, A.C. and O'Riordan, M.C. (1988) *Natural Radiation Exposure in UK Dwellings*. National Radiological Protection Board Report, NRPB-R190. NRPB, Chilton.

Wu, A.H., Fontham, E.T.H., Reynolds, P., Greenberg, R.S., Buffler, P., Liff, J., Boyd, P. and Correa, P. (1996) Family history of cancer and risk of lung cancer among lifetime nonsmoking women in the United States. *American Journal of Epidemiology* 143, 535-542.

Wu, A.H., Fontham, E.T.H., Reynolds, P., Greenberg, R.S., Buffler, P., Liff, J., Boyd, P., Henderson, B.E. and Correa, P. (1995) Previous lung-disease and risk of lung cancer among lifetime nonsmoking women in the United States. *American Journal of Epidemiology* 141, 1023-1032.

Wynder E.L. and Graham E.A. (1950) Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma. *Journal of the American Medical Association* 143, 329-336.

Xu, Z.Y., Brown, L., Pan, G.W., Li, G., Feng, Y.P., Guan, D.X., Liu, T.F., Liu, L.M., Chao, R.M., Sheng, J.H. and Gao, G.C. (1996) Lifestyle, environmental pollution and lung cancer in cities of Liaoning in northeastern China. *Lung Cancer* 14, Supplement 1, S149-S160.

Xuan, X.Z., Lubin, J.H., Li, J.Y., Yang, L.F., Luo, A.S., Lan, Y., Wang, J.Z. and Blot, W.J. (1993) A cohort study in southern China of workers exposed to radon and radon decay products. *Health Physics* 64, 120-131.

Yen, I.H. and Kaplan, G.A. (1999) Poverty area residence and changes in depression and perceived health status: Evidence from the Alameda County Study. *International Journal of Epidemiology* 28, 90-94.

Young, R. (1996) The household context for women's health care decisions: Impacts of UK policy changes. *Social Science and Medicine* 42, 949-963.

Zheng, W., Blot, W.J., Liao, M.L., Wang, Z.X., Levin, L.I., Zhao, J.J., Fraumeni, J.F. Jr. and Gao, Y.T. (1987) Lung cancer and prior tuberculosis infection in Shanghai. *British Journal of Cancer* 56, 501-504.