CLIMATE RELATED MORTALITY AND MORBIDITY IN SCOTLAND: MODELLING TIME SERIES OF COUNTS

by

ISLAY M. GEMMELL

A dissertation submitted to the
UNIVERSITY OF GLASGOW
for the degree of
DOCTOR OF PHILOSOPHY

Public Health Research Unit
Department of Statistics
April 2000

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

ACKNOWLEDGEMENTS

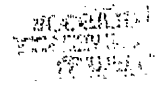
Firstly I would like to thank my supervisor Dr. Marian Scott for her support and advice throughout the period of my research.

I would also like to thank Dr. F. A. Boddy and Philip McLoone both formerly of the Public Health Research Unit, Glasgow University, for giving me the opportunity to conduct this research while working at the Unit.

Many thanks are also due to Noel Cressie who while visiting Glasgow took time to comment on the work described in Chapter 5.

My thanks are also due to my parents who insisted on paying my tuition fees for the first two years of the research and to my friends and family who have always been a source of support. I would also like to thank Adam Bolland for proof reading the final text.

I am grateful to the General Register Office for Scotland for the provision of the mortality data, to the Information and Statistics Division of the National Health Service in Scotland for the provision of the hospital discharge data and to the Meteorological Office for climatic data. During the course of this work the Public Health Research Unit was funded by the Chief Scientists Office of the Scottish Office Department for Health.



DECLARATION

The results in sections 2.3, 2.4, 3.4, 4.3 and 4.4 in this thesis have been published in 'Seasonal variation in Mortality and Morbidity in Scotland 1981-93' a Public Health Research Unit report based on this work. Some of the results in sections 2.3, 3.4, 4.2 and 4.3 are to be published in the April edition of 'The International Journal of Epidemiology'.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Summary

Earlier research has demonstrated that excess winter mortality is greater in the countries of the United Kingdom than in those on comparable latitudes elsewhere in Europe. The purpose of this thesis was to provide an up-to-date analysis of excess winter mortality in Scotland. This involved exploring the relationships between mortality, morbidity (as reflected in rates of emergency hospital admissions), climate, influenza epidemics, and socio-demographic variables. The majority of the analysis was concerned with temporal relationships between these variables, however, latterly spatial relationships were also considered.

Chapter 1 reviews the literature in support of seasonal patterns in health and assesses the merits of the various statistical techniques that have been used to demonstrate these patterns. Much of the previous analyses have used simple descriptive statistical methods with few acknowledging the Poisson time series nature of the data.

In chapter 2 the seasonal pattern of mortality and morbidity from three main disease groups was described using a generalised linear model with Poisson errors incorporating a cosine term. The method was used to analyse the seasonal pattern by sex, age group, social class, deprivation category and health board.

In chapter 3 the effect of climate on mortality and morbidity is explored. This chapter is chiefly concerned with the comparison of possible methods of analysis. Firstly the problems with summary methods are demonstrated before the principles of time series methodology are introduced. The final comparison involves three methods, ARIMA time series methods, Poisson regression and Zeger's method. Zeger's method is as a time series regression method for Poisson data. The methods are compared by assessing the effect of temperature on weekly deaths from respiratory disease. Examination of the residuals and the standard errors of the model coefficients reveal that Zeger's method is the most appropriate for this type of analysis.

Zeger's method is used in Chapter 4 to assess the relationship between temperature and mortality and morbidity in more detail, by considering the effects of age, socio-economic deprivation and city of residence. This chapter also includes a detailed examination of the effects on mortality of a variety of different temperature patterns.

In chapter 5 the spatial aspect of the data is included in the analysis. Space-time variations in emergency admissions for respiratory disease are assessed at various levels of aggregation. Overall there is no clear evidence of space-time patterns in emergency respiratory admissions over the time period, however spatial relationships are demonstrated. Finally, methods which account for spatial autocorrelation are used in an analysis of the relationship between emergency admissions and socio-economic deprivation in Glasgow. This analysis demonstrates, as with the previous temporal analysis, that if autocorrelation exists it is vital to account for this in any modelling procedure.

Chapter 6 provides a summary of the main findings of the analysis in terms of both the epidemiological results and the methodological concerns. The limitations of the study concerning problems associated with the use of routinely collected data are also recognised.

The thesis has demonstrated that seasonal patterns in mortality and morbidity are still a significant public health problem in Scotland and that Zeger's method is the most appropriate method to use when assessing the direct relationship between climate and ill health.

Chapter 1 - Introduction

1.1 Introduction

Ever since death records have been examined in an epidemiological manner, seasonal fluctuations in death rates have been recognised. Some of the earliest work in the UK was done by Quetelet[1] in 1842 and William Farr[2] in 1847. Recent studies in the UK and the rest of the world have demonstrated that seasonal variation in both mortality and morbidity exists in most countries. The extent of these seasonal fluctuations varies considerably around the globe but no consistent method of measurement has been used. Three studies[3-5] published between 1987 and 1991 identified that the British Isles experienced greater seasonal increases in mortality than most other European countries and that this seasonal variation was associated with outdoor temperature and influenza epidemics.

A study conducted by the General Register Office for Scotland[3] in 1987 computed the ratio of the death rate in the 1st quarter (Q1) of the year and the 3rd quarter (Q3) of the year between 1979 and 1984 in twelve European countries. This study found that although there was marked seasonal variation in every country studied, the greatest seasonal increases were in Scotland, England, Ireland, Wales and Italy. Death rates in Scotland in the 1st quarter of the year were 30% higher than death rates in the 3rd quarter of the year. In Germany, Netherlands, Denmark, Finland, Norway and Sweden this figure was between 15% and 20%. In Scotland the Q1/Q3 mortality ratio was greater for those over 65 years compared to those aged under 65.

A similar study conducted by the Office of Population Censuses and Surveys[4], computed an 'excess winter deaths index' (EWDI). This was defined as the percentage of excess deaths in the four winter months of highest mortality compared with the average of the numbers in the preceding and following four month periods. The authors compared the excess winter death index for several countries. It was found that countries in the British Isles experienced a greater seasonal increase in deaths than other European countries including the colder Scandinavian countries.

Curwen[5] described a regression model that used 3 variables to model the size of the EWDI in England and Wales from 1949 to 1985. These variables were the mean national winter temperature, the number of deaths registered as due to influenza and, to account for time trends, the year. From this model it was estimated that each degree Celsius by which the winter was colder than average was associated with around 8000 excess winter deaths. Each registered winter influenza death was associated with 3.5 excess winter deaths.

The three studies described above, using monthly aggregates, provided the impetus for an up-to-date and detailed investigation of excess winter deaths in Scotland. The broad aim of the study was to examine the seasonal variation of patterns of ill health in Scotland and to assess the direct effect of climate and influenza epidemics on these patterns. A further aim was to develop a more sensitive approach using appropriate statistical methods to fully establish the nature of the relationship between mortality, climate and influenza epidemics.

1.2 Evidence for Seasonality

The majority of the literature on climate related ill health is concerned with deaths or hospital admissions from respiratory disease, cardiovascular disease or all cause mortality. Epidemiological studies in this area have been carried out in many countries, several of which have a climate very different to that of Scotland. Due to the diversity of the work carried out in this field the relevant studies have been grouped into six semi-homogeneous groups based on disease group and country (UK or elsewhere). Within these groups, the statistical methods used range from simple measures of seasonal variation to more detailed Poisson regression and time series methodology.

1.2.1 UK Studies

1.2.1 (a) All causes

There have been several general studies of seasonal mortality in the UK. These studies have tended to use fairly crude but different measures of seasonality based on monthly data. The simplest measure of seasonal variation in death rates is the Q1/Q3 ratio. This measure provides an effective method of comparing several European countries using published data available from the UN Monthly Bulletin of Statistics. The Excess Winter

Deaths Index[4] provided another fairly simple method for comparing seasonal variation in death rates between several countries. Using this method, countries in which the number of deaths did not peak in the same months as European countries can be included in the analysis. The seasonal variation in deaths in Scotland was again shown to be considerably greater than most other European countries but not as great as that experienced in Australia, New Zealand and Israel. The EWDI was shown to vary by sex, age group, social class and cause of death. However there was little correlation between the regional EWDI and the regional average winter temperature in England & Wales. The regional EWDI was more closely related to the regional influenza rate than regional temperature.

In another OPCS publication[6] a different method of measuring seasonal variation in death rates was introduced. The seasonality ratio was computed as the death rate from January to March expressed as a ratio of the yearly rate. Using data as far back as 1840 the author showed how seasonal variation in death rates has changed over the centuries. The main focus of the paper was the apparent fall in seasonal variation since the 1960's. The author maintained that the fall in seasonality was not solely the product of fewer severe winters or less severe influenza epidemics but could be attributed in part to an increase in central heating systems and the fall in air pollution which has occurred since the 1960's.

The seasonality of disease in Scotland using both deaths and hospital admissions for different causes and age groups was assessed by Douglas[7] in 1991 using the method of 'cosinor analysis'. This method involved computing the amplitude of the seasonal curve from the coefficients of a cosine and sine term in a regression model and expressing this as a percentage of the mean yearly number of deaths. He found a mean to peak increase of 12.4% for all causes of death, 14.7% for ischaemic heart disease, 13.3% for cerebrovascular disease and 32.5% for respiratory disease. The degree of seasonality experienced increased with age and was greater for deaths than for hospital admissions.

The five papers discussed provide a general overview of the seasonal pattern of ill health in the UK. Each author has used a different method which makes a direct comparison of

the figures difficult. However, they all demonstrate a substantial seasonal increase in deaths.

One criticism of the first three methods is that the authors pre-determine the period of the year for which the peak in deaths is measured. If the peak does not fall exactly in the middle of the monthly grouping these methods will provide an underestimate of the true extent of seasonal variation. The method proposed by Douglas removes this obstacle but as with the other authors he uses monthly data. The data set is reduced to twelve monthly points, thus losing information.

Much of the work in the field of seasonality of disease has concentrated on specific diagnostic groups or causes of death. The disease groups which displayed the greatest seasonal variation, and on which most analysis has been conducted, are diseases of the respiratory system, cerebrovascular disease and ischaemic heart disease. A review of the findings and methods used in the different disease categories is now presented.

1.2.1 (b) Respiratory disease and influenza

Deaths from respiratory disease tend to display the greatest seasonal variation. Most papers concerned specifically with seasonal patterns in respiratory disease concentrated on the effect of influenza epidemics and of respiratory disease on all cause mortality. Specific respiratory conditions such as hypothermia and asthma have also attracted attention in studies of seasonality. Recently the effect of air pollution on deaths from respiratory disease has been the subject of much media interest.

Curwen[8] compared the number of deaths in a winter that contained an influenza epidemic to the number in the previous winter and estimated that the influenza epidemic in the winter of 1989-1990 was responsible for 25 000 deaths. Curwen used a fairly crude method to establish a 'hidden influenza factor'. He subtracted the number of deaths in the previous winter from the number in the epidemic winter to obtain the excess winter deaths. He then subtracted from the number of excess winter deaths those that were recorded as influenza deaths. The remaining excess winter deaths were deemed attributable to influenza but not recorded as influenza. To obtain a 'hidden influenza factor' he divided the number of deaths attributable to influenza by those recorded as

influenza. The hidden influenza factor is described as the number of deaths associated with the epidemic but ascribed to another cause for each death certified as due to influenza. The winters that contained influenza epidemics in England and Wales were the winters of 1969/70, 1972/73, 1975/76 and 1989/90 and for each of the epidemic winters respectively the hidden influenza factor was computed as 1.8, 4.6, 2.3 and 8.9. Curwen showed that the hidden influenza factor was more pronounced in the 65-74 age group rather than the very old or very young.

Fleming[9] considered not just the effect of influenza epidemics but the effect of the incidence of respiratory disease on deaths from all causes in England and Wales. Fleming proposed that the pattern occurring in all causes of death was driven by the underlying pattern of respiratory disease. Using four-weekly periods over the years 1986-1990 he removed the trend and seasonality from both the respiratory incidence series and from the mortality series. He then computed a correlation for the two residual series. This gave a strong positive correlation from which the authors concluded that there was a close association between respiratory disease incidence and total deaths. However, there may also have been serial autocorrelation present in both series and this autocorrelation may have contributed to the significant correlation.

Another study that associated influenza and respiratory syncytial virus (RSV) to total mortality in England & Wales was carried out by Nicholson[10]. Using four-weekly periods Nicholson demonstrated a relationship between influenza and RSV morbidity reports and all cause mortality. He concluded that mortality associated with RSV was considerably greater than that associated with influenza.

While there have been several studies in other countries concerning seasonal fluctuation in admissions for asthma there have been relatively few carried out in the UK. However, Khot[11] showed that childhood asthma admissions in southeast England peaked in early summer and again in September/October. The authors suggested that climatic factors might contribute to this rise in admissions in autumn. A recent editorial in the BMJ[12] reviewed the recent literature on asthma incidence and air pollution. The authors concluded that asthma is made substantially worse by current concentrations and combinations of air pollutants.

Intuitively respiratory disease is most likely to vary seasonally and be related to temperature fluctuations. However, a considerable amount of work has been done concerning other causes of death. The two other broad disease categories that display a strong seasonal pattern are cerebrovascular disease (CVD) and ischaemic heart disease (IHD).

1.2.1 (c) Ischaemic heart disease and cerebrovascular disease

Dunnigan[13] examined the seasonal variation in hospital admissions and mortality from ischaemic heart disease in Scotland between 1962-1966 by age and sex. The authors considered three case type categories; non-fatal admissions, fatal admissions and total deaths. The seasonal mean to peak increase in non-fatal hospital admissions was 7%, this increased to 17% for fatal hospital admissions and to 27% for total deaths. Dunnigan used simple correlation methods to assess the relationship between monthly temperature and monthly deaths. He took no account of the fact that both data series were seasonal and a correlation between any two seasonal curves is likely to produce a significant result. Anderson[14] in a letter in reply to Dunnigan's paper showed similar patterns in Ontario. He considered 'sudden' deaths separately from non-sudden deaths and showed a seasonal increase of around 15% in sudden deaths and a similar increase in non-sudden deaths. Both studies demonstrated a dicyclic pattern with peaks in deaths in winter and in spring.

Several studies into the seasonal pattern of cardiovascular disease in the UK have been published recently. Douglas[15] looked at the seasonality of coronary heart disease. Using monthly data from 1962 to 1971, he demonstrated a winter/summer pattern of seasonal variation in male and female deaths and female admissions. For male admissions he found a dominant spring peak up to the age of 55, followed by a bimodal pattern with a winter and a spring peak at older ages.

In a BMJ editorial Wilmshurst[16] reviewed the relationship between temperature and cardiovascular mortality. He commented that excess winter deaths from cardiovascular mortality are numerically far more important than any other cause of death and suggested that these excess deaths might be preventable. He discussed several possible mechanisms

in which environmental temperature would have an effect on cardiovascular mortality, such as raised blood pressure, peripheral vasoconstriction and blood clotting.

Several replies to this editorial were published in the following editions of the BMJ. Cooke[17] comments on how the seasonal pattern of pulmonary embolism is closely related to that of coronary disease and how the two diseases have similar causal paths. O.L. Lloyd[18] discussed the idea of a climate gradient of coronary mortality and commented on the 'Scottish Paradox'. He showed that within Scottish geographical regions the high cardiovascular and cerebrovascular standardised mortality rates (SMR's) occur in the milder west and southwest. In reply to this letter E.L. Lloyd[19] commented on the fact that it is not just absolute temperatures that have effects on health but also changes in temperature. E.L. Lloyd claimed there was no Scottish paradox and blamed the higher rates of CVD and IHD mortality in the west of Scotland on the more changeable climate which is experienced in the west.

A paper by Rothwell[20] suggested that previous studies of stroke deaths and hospital admissions were selective and potentially biased reflecting the seasonal nature of complications such as pneumonia. Using monthly data from 675 patients in Oxfordshire the authors computed a chi-square statistic to test for an equal number of events in each month. They found no significant difference between the monthly number of events in different diagnostic classifications for stroke, however they did find that primary intracerebral haemorrhage occurred more often than expected at low temperatures. Haberman[21] also used a chi-square test statistic to test for a seasonal pattern in the number of deaths from CVD occurring in England & Wales in 1975. He showed a clear winter peak that occurred in both males and females and at all ages.

Bull[22] used simple correlation and chi-square tests to determine the degree of association between climate and deaths from respiratory disease and cardiovascular disease. Using data from Belfast and London he showed that the incidence of cardiovascular disease was greater in winter and was related to temperature. He also showed that the relationship with temperature was stronger in older people. Bull commented that on his evidence it was unlikely that the increased incidence of respiratory and infective disease

in cold weather was causally related to the high incidence of cardiovascular disease in cold weather.

West[23] demonstrated that IHD mortality was correlated with temperature, rainfall and water calcium by computing correlation coefficients between IHD mortality and the three factors in 114 boroughs in England and Wales. Bainton[24] also demonstrated a relationship between mortality and temperature using data for Greater London between 1970 and 1974.

These three studies are among the studies in the UK which have assessed the direct effect of temperature on cardiovascular mortality, however they use simple correlations to infer causal associations. Most studies of climate related mortality and morbidity in the UK have used aggregate measures of temperature such as annual figures for each year of the analysis. Those studies that have used more detailed information on climate have used crude methods of relating the climate variables to the mortality data.

1.2.2 International studies

1.2.2 (a) All causes

There have been several studies on winter excess mortality in the Netherlands in recent years. In his paper on seasonal variation Mackenbach[25] assessed the seasonal fluctuation in deaths using a method similar to that used in this thesis, described in Chapter 2. He used Poisson regression and periodic functions of time to estimate the peak to trough ratio of the seasonal curve. Machenbach included higher order cosine terms to model more complicated patterns of variation. He considered all cause mortality and several specific causes of death. Machenbach estimated that in the Netherlands the trough to peak ratio for deaths from IHD was 34% and for deaths from cerebrovascular disease this figure was 25%. By using daily deaths standardised for age and sex he arrived at a single value for each diagnostic group and therefore an age-related increase or a sex difference was unidentifiable. For deaths from all causes Machenbach estimated a trough to peak percentage increase of 22%. An influenza epidemic occurred in the Netherlands in 1986, and the authors concluded that there were 12 deaths associated with each death registered as influenza during that year.

In his 1990 paper Kunst[26], using monthly deaths in the Netherlands from 1953-1988, described the changing pattern of excess winter deaths. The authors showed that excess winter mortality had decreased in the Netherlands throughout the time period. This decrease was not due to changes in the cause of death composition of total mortality nor had the effect of influenza declined over the years. The authors proposed that the decline in winter excess mortality experienced in the Netherlands was due to the decrease in winter smog and improved housing conditions. They also cite many social factors that may have affected this decrease, such as improvements in footwear, clothing, working conditions and transport.

The seasonal variation in mortality in Germany also fell between 1946 and 1995. Lerchl[27] demonstrated a seasonal increase of around 20-30% from the annual mean to the peak of the seasonal curve in the early years of the study. This fell to around 10% in the 1990's. The author proposes that increased use of central heating, better clothing and an improvement in the health care system have contributed to this decline in seasonal variation in Germany.

Laake[28] provided a direct comparison between the levels of excess winter mortality in England & Wales and Norway in a similar method to that used by Curwen in 1988. Excess winter mortality between 1970 and 1991 was shown to be 21% in England & Wales and only 11% in Norway. This paper used a different method of estimating the seasonal increase but confirmed the findings of the OPCS[4] study which showed that countries in the UK had a greater seasonal increase in mortality than those in the rest of Europe.

More recently McKee[29] demonstrated that the seasonal increase in mortality in Moscow between 1993 and 1995 was smaller than that in western countries. The authors suggested a fall in temperature of 1°C was associated with an increase in mortality of around 0.7 per cent. The Eurowinter[30] group, in a study that compared the change in mortality with temperature falls in several European countries, showed that the effect of falling temperature on mortality was greater in countries that had milder winters.

Saez[31] used a time series approach to assess the relationship between daily mortality and increases in temperature in Barcelona. Despite the fact that the main interest of this paper is increases in temperature rather than decreases in temperature it bears relevance to the work described in Chapter 4 and covers all cause mortality, cardiovascular mortality and respiratory mortality. Saez rejected the hypothesis that the data followed a Poisson distribution, he logarithmically transformed the data to normalise the distribution and then used ARIMA techniques to model the relationship between mortality and deaths. He checked the adequacy of his modelling technique using normality checks on the residuals. The models showed that, in winter, falls in temperature were negatively related to changes in mortality and unusual periods of temperature were associated with an increase in mortality. In summer, increases in temperature were associated with an increase in mortality and again unusual periods of temperature were associated with an increase in mortality.

1.2.2 (b) Respiratory disease and influenza

As in the case of the UK, there have been few international studies that have considered the seasonal variation in respiratory disease. The majority of studies involving the analysis of short-term fluctuations in respiratory disease are concerned with air pollution rather than climatic variables. These papers are discussed in section 4.5.1.

Dales[32] analysed the autumn increase in asthma admissions among pre-school children in Toronto. Using time series methods, incorporating measures of air pollution, allergens and meteorological variables, the authors concluded that respiratory infection was the major identifiable risk factor for the large autumnal increase in asthma admissions.

In a detailed study on the impact of influenza on mortality in the Netherlands, Sprenger[33] suggested that more than 2000 people died from influenza in the Netherlands each year but only a fraction were recognised as influenza deaths. He concluded that for each death recorded as influenza there were 2.6 other deaths attributable to influenza and that, of all deaths which were not registered as influenza, 47% were recorded as heart disease, 23% as lung disease and 30% as other causes of death.

1.2.2 (c) Ischaemic heart disease and cerebrovascular disease

In a study conducted in 1979 Anderson[34] compared the seasonal fluctuation of ischaemic heart disease mortality in England and Wales with that in Ontario. He showed a January peak in deaths existed in both areas, however the mean to peak increase in deaths in England and Wales was 31% compared with 9% in Ontario. Anderson discounted the fact that the difference in age structure of the two populations may account for the difference in mortality patterns by comparing England & Wales with selected counties in Ontario that had a similar age structure. He also considered increased use of central heating and an assumption of a lower underlying death rate in Ontario, however neither of these factors explained the difference in the seasonal fluctuation in mortality between the two areas. He concluded that the incidence of myocardial infarction was not increased by low environmental temperature but followed fluctuation in incidence of severe respiratory disease.

Douglas[35] used analysis of variance methods to assess seasonality of deaths from coronary heart disease and cerebrovascular disease in New Zealand between 1980 and 1984. He showed a clear seasonal pattern in both causes of death with a winter (June/July/August) peak and a summer trough. Previously Marshall[36] had demonstrated a seasonal increase in mortality from coronary heart disease in New Zealand of 35% between 1970 and 1983. He showed that this figure varied by age group with a greater seasonal increase occurring in older age groups. A seasonal increase in deaths from ischaemic heart disease of 33% was shown to occur in Los Angeles County[37]. This analysis was also based on monthly data but was extended to look at weekly patterns around the seasonal peak. The authors found that deaths peaked most markedly during the holiday season of Christmas and New Year. Another study[38] used monthly data for the whole of the United States from 1937 to 1991 to look at seasonal variation in mortality from coronary disease. Seasonal increases in mortality declined from around 38% in early years of the study to under 20% around 1970, the seasonal pattern then began to increase again to around 25% in the 1990's. The authors hypothesised that the initial decline was due to the increase in the use of central heating over the early years time period and the apparent reversal of this trend could be due to the uptake of air conditioning systems which began around 1970. The increased use of air conditioning systems would reduce the number of summer deaths from heat waves.

Green[39] carried out a study on excess winter mortality from ischaemic heart disease and cerebrovascular disease in Israel. He computed a measure of trough to peak percentage increase for both causes by sex and age group. The seasonal percentage increase in deaths from IHD in those aged over 45 was around 50% for males and 60% females, for deaths from CVD the increase was around 50% for males and 40% for females. For both causes of death there was an increase in seasonal variation with age group. Green also performed a cosinor analysis, splitting the data into warmer or colder years. He demonstrated that the amplitude of the cosine curve was greater in colder years than warmer years for both causes of death. He commented that the contribution of influenza and pneumonia to the seasonal pattern of deaths was relatively small.

Sheth[40] showed a 10% seasonal increase in mortality from AMI in Canada during 1980 to 1982 and 1990 to 1992 and seasonal increase in stroke mortality of 14% for the same time period. The authors also showed that these seasonal increases were significantly greater at older ages compared to younger ages.

The papers described above have concentrated on measuring the extent of the seasonal variation, however there have been several papers which have tried to establish a direct association between climatic conditions and deaths from ischaemic heart disease and cerebrovascular disease. Statistical methods such as Poisson regression and time series regression, have been used in some cases, however there have been many studies in which the distribution of the data and methodological issues have been ignored.

Pan[41] used odds ratios to establish a risk of dying from CVD at a particular temperature in Taiwan. The authors took no account of the seasonal pattern existing in both series and computed an odds ratio between the mean daily mortality at each degree Celsius and the mean daily mortality at a baseline temperature. They demonstrated that the odds of death were generally greater at temperatures lower than the baseline temperature. From the odds ratios they observed a U shaped relationship with temperature and computed a percentage increase in deaths due to each degree fall in temperature. While the methodology used in this study cannot establish a causal link it does support findings from other studies. Mannino[42] used linear regression to establish a link between mortality

rates from myocardial infarction and daily temperature in Wisconsin. The regression coefficients were significant for all age categories except those under 50 years. However the analysis did not adjust for the fact that both variables followed a seasonal pattern and could therefore produce significant results in a linear regression analysis when there was no causal relationship present. Baker-Blocker[43] used correlation coefficients to determine the relationship between minimum air temperature and cardiovascular mortality in Minneapolis-St Paul. She found a significant relationship between mortality and departures from normal minimum temperature on the days preceding death. She found that in two out of the five winters studied, snowfalls were significantly correlated to cardiovascular mortality several days after the snowfall. The limitations of this study are again failure to account for the seasonal pattern in both mortality and temperature

Using observed to expected ratios, Rogot and Padgett[44] demonstrated that both stroke and coronary heart disease mortality were related to snowfall for a period of up to 6 days after a snow fall in the USA. Anderson[45] assessed the effect of cold snaps and snow falls on deaths from IHD in Toronto over a 15 year period. He defined a cold snap as a day on which the temperature was at least 4.4°C lower than it was the day before. He considered the average number of deaths each day in the pair of days and found that deaths increased by 16% in males aged younger than 65. Similar analysis on snowfall days recorded an 88% increase in deaths, on pairs of days which experienced both a cold snap and a snowfall the increase in deaths was 113%. Heunis[46] considered the effect of short term changes in temperature on deaths from cardiovascular disease in people aged over 60 in Cape Town, South Africa for the period 1978-1985. Using t-tests he demonstrated that a significant increase in deaths occurred two days after a minimum temperature of below 4°C. He also considered deaths on very hot days associated with berg winds. These winds herald the approach of a cold front and are extremely hot. They are characterised by a sudden rise in temperature of at least 5°C and then drop in temperature of around 10°C. A significant increase in deaths was found two days after a berg wind.

The studies described have used various measures to determine the existence of an association between temperature and deaths. However, in most cases certain features of the data have been overlooked and simple summary statistics have been used to infer a causal relationship.

Ohlson[47] used Poisson regression to relate wind-chill, snowfall, atmospheric pressure and day of the week to daily admissions for IHD in a Swedish hospital during three winters. No relationship between the climatic features and IHD incidence was found, however this study did not consider lag effects and was concerned with climatic features other than temperature. The authors concluded that weather conditions were not a major triggering factor for myocardial infarction in Sweden. Enquselassie[48] studied fatal and non-fatal coronary events in New South Wales. Poisson regression analysis was used to determine the degree of association between the climatic variables and the coronary events. The main findings from the study were that fatal coronary events and non-fatal myocardial infarction were 20-40% more likely in winter (June-August) and spring (Sept-Nov) than at other times of the year. Fatal events were 40% more likely to occur on cold days than at moderate temperatures. The authors showed the change in deviance associated with the addition of each variable to the logistic regression model but did not give parameter estimates. Temperature and rainfall were both significant variables in the model.

Kunst[49] used data on all deaths in the Netherlands from 1979 to 1987 and daily meteorological data in his study of the relationship between temperature and mortality. He also had access to data on incidence of influenza-like diseases and air pollution measurements of sulphur dioxide from six stations. He used a Poisson regression model and demonstrated the presence of a lag effect. For all cause mortality he found that a 1°C drop in temperature was associated with a percentage increase in deaths of about 0.5% one week later. Falls in temperature had the greatest effect on deaths from respiratory disease and cardiovascular disease. He showed that the relationship between temperature and mortality was altered very little by the inclusion of sulphur dioxide levels and season into the model. Influenza incidence however was a significant term in the model. Kunst also considered hot temperatures and demonstrated a U shaped relationship between temperature and mortality.

Lanska[50] analysed seasonal patterns in stroke mortality in the USA from 1938 to 1988, concluding that both respiratory disease and temperature influenced the seasonal patterns in stroke mortality. The authors decomposed each time series into a trend, a seasonal effect and a residual effect and used cross correlation techniques to assess the

relationship between the residual time series. The cross correlations were greatest at a lag of zero indicating that the stroke and respiratory curves peaked at the same time, when the temperature curve was at its trough. However, the data were aggregated by month giving little opportunity to further investigate any time lags in the relationship.

Most of the studies described have used fairly simple statistical techniques such as correlation, chi-square tests and latterly Poisson regression or time series regression; few studies have recognised that weekly or daily mortality data is a Poisson time series.

1.2.3 The use of time series in epidemiology

Catalano[51] discussed the underuse of time series methods by epidemiologists stating that in the five years prior to the publication of his paper, the American Journal of Epidemiology had only published one article which used time series methods. However, Catalano failed to recognise that most epidemiological data do not satisfy the normality assumptions required in a classical time series analysis.

Giles[52], using a co-spectral time series analysis technique, found no relationship between asthma morbidity and daily meteorological data in Tasmania. Hoppenbrouwers[53] used a similar technique, spectral coherence methods, to demonstrate a relationship between environmental pollutants and cases of Sudden Infant Death Syndrome (SIDS). He showed that peak levels of pollutants preceded peak levels of SIDS by seven weeks. Bowie and Prothero[54], in a study of IHD mortality and temperature, used time series methods to filter their data and create residual series which consisted of independent identically distributed random variables before correlating the two series. This technique is similar to that described in section 3.3.3.

Alberdi[55] assessed the relationship between temperature and daily mortality in Madrid using a similar method. The authors removed any deterministic elements of the series such as trend and seasonal components and also removed autocorrelation using Box-Jenkins methodology. Several different weather variables and the mortality data were modelled using ARIMA models and the cross correlation function between the weather series and the mortality series was examined. The results showed that mortality was inversely related to cold temperature with a lag of 11 days and directly related to warm

temperature with a lag of 1 day. This was described as a J shaped relationship by the authors.

Albert[56] proposed two models for describing relapsing-remitting behaviour in a series of counts. The first was in the form of a Poisson time series with a periodic trend and terms for previous observations; the second model used a hidden Markov chain to describe relapsing-remitting fluctuations. He found that both were satisfactory models and their suitability depended on the vagaries of the data to which they were applied. Kuhn[57] compared Poisson regression and time series methods to detect changes over time in rates of child injury following a prevention program. He found that both methods provided similar parameter estimates and both provided a good fit to the data.

While these studies have applied time series techniques to epidemiological data, none have been able to deal with the fact that the data follow a Poisson distribution. A paper by Campbell[58] used a method developed by Zeger[59] to investigate the relationship between sudden infant death syndrome and environmental temperature. The authors concluded that the rate of SIDS in England and Wales from 1979 to 1983 increased by 4.3% for every 1°C drop in environmental temperature. In this paper the authors assess the validity of several methods of statistical analysis for time series data which follow a Poisson process. Zeger's method was developed to deal with this type of problem and the authors conclude that the method proposed by Zeger is the most appropriate one to use in these circumstances.

Zeger's method may provide the tool for time series methods to be applied to epidemiological problems and in this thesis three methods of analysis will be compared; classical time series methods, Poisson regression methods and Zeger's method when assessing the nature of the relationship between mortality and environmental temperature.

1.2.4 Physiological Hypotheses

Many studies have demonstrated a statistical relationship between seasonal increases in mortality and seasonal climatic change. However, to establish a causal relationship the mechanisms by which outdoor temperature affects the body's physiology need to be considered. Several studies have examined seasonal variation in three risk factors for

vascular disease; high blood pressure, high cholesterol levels and fibrinogen concentration in the blood.

Brennan,[60] using blood pressure measurements taken during a trial for mild hypertension, demonstrated that for each age, sex and treatment group, systolic and diastolic pressures were higher in winter than in summer months. Brennan also showed that seasonal variation in blood pressure was greater in older than in younger subjects and was related to daily air temperature measurements.

A study by Dobson[61] demonstrated a seasonal fluctuation in serum cholesterol with higher levels in winter than in summer. MacRury[62], in a similar trial, also showed some evidence of a seasonal variation in serum cholesterol levels, however both trials consisted of very few patients. A study conducted in Belfast by Stout[63] demonstrated a 23% increase in winter in plasma fibrinogen concentrations as well as significant seasonal variation in measurements of plasma viscosity and high density lipoprotein cholesterol. A similar study conducted in Cambridge by Woodhouse[64] also showed that fibrinogen levels were greater in winter than summer. Using results from the Northwick Park Heart Study[65] the authors estimated that the observed seasonal variation in plasma fibrinogen could account for 15% of the increase in ischaemic heart disease risk in winter. Keatinge[66] demonstrated significant changes in the levels of many cardiovascular risk factors by comparing healthy student volunteers in an experiment that involved the student being cooled over a period of 6 hours. He showed that blood viscosity increased by 21%, arterial pressure rose on average from 126/69 to 138/87 and both blood cholesterol and platelet count increased.

Donaldson[67], using data from men aged 50-69 attending BUPA health screening examinations in London, compared haematological and blood pressure data with outdoor temperatures. Regression analysis using data filtered for long term seasonal patterns showed that short-term falls in temperature were associated with significant increases in 9 out of the 13 variables recorded. The analyses also included daily mortality rates for IHD and CVD in London which also demonstrated a significant increase associated with a short term fall in temperature. The authors concluded that the relationship between

temperature changes and changes in arterial risk factors are sufficient to cause the observed increase in arterial disease mortality in winter.

The association between the increase in respiratory disease and outdoor temperature is more difficult to explain. Respiratory infections such as the common cold are so common that their importance is not recognised until such time as there is a major winter influenza epidemic. The respiratory tract is a major site for infections and identifying which infective agent is associated with a particular outcome is complex. Respiratory disease can be caused by viral infections and bacterial infections. Some viruses are temperature dependent and are only able to multiply freely at lower body temperatures of around 33°C[68]. Fleming[9] proposed that outdoor temperatures of a few degrees above freezing, together with relatively high humidity might encourage the spread of infection by the droplet method. Monto[69] examined the temporal patterns of respiratory syncytial virus and showed that the virus was prevalent during much of the cold season.

Influenza epidemics occur at cold temperatures and have been shown to be related to a significant increase in the number of deaths from all causes[8]. This phenomenon of hidden influenza deaths has been examined in several other studies[5,9] and it is generally recognised that an influenza epidemic can produce a significant increase in the number of deaths which are not recorded as influenza deaths.

1.3 Summary & Aims

Other studies[3,4] have established that seasonal variation in mortality in Scotland is greater than in most other European countries. The work described in this thesis looks at both seasonal variation in ill health and the direct effects of changing temperature on ill health in Scotland. Furthermore the analysis is also concerned with the role of socio-demographic variables and environmental pollution. The pattern of disease occurrence over both time and space is also examined by concentrating on hospital admissions for respiratory disease and examining how these vary over time throughout different geographical areas of Scotland.

Previous studies of seasonal variation in deaths in Scotland used fairly crude methods of measurement such as the Q1/Q3 ratio and the EWDI. The initial aim of this work was to

develop a more accurate method of measuring the seasonal variation in mortality and morbidity in Scotland. This analysis would then be examined in more detail to determine the effect of various socio-demographic variables on seasonal variation. The main effects to be studied were cause of death or hospital diagnosis, sex, age group, area of residence and social deprivation score. In Chapter 2 a method similar to that used by Douglas[7] is described, but using weekly data, for 1981 to 1993 and a time trend variable. This provides a more accurate but also generalisable estimate of seasonal variation in deaths and emergency hospital admissions in Scotland. Douglas[7] expressed the seasonal increase in deaths in Scotland by the percentage increase from the mean to the peak. Using the amplitude estimate from the sinusoidal curve it is possible to express the seasonal increase as the percentage increase from the trough to the peak thus providing a closer comparison to the previous work by OPCS[3,4].

The second main aim of the work was to describe in detail the relationship between climatic variables and ill health in Scotland, taking into account possible confounding variables such as influenza epidemics. Within this broad aim there were many potential areas of interest. These included determining which climatic features were most closely related to ill-health, determining the level of aggregation at which to assess the relationships, and incorporating the information obtained from the seasonality analysis described in Chapter 2. While these are principally matters of interest in terms of epidemiological analysis, there were also several statistical issues to be addressed. The data are a time series of events. The general approach to modelling time series data has in the past been Box-Jenkins methodology, which assumes normality. These data are counts and, after divisions by the various socio-demographic variables, the numbers may be small and therefore assumptions of normality may not be justified. When analysing data in the form of counts Poisson assumptions are usually valid, however in this case, because the data are a time series there is autocorrelation present in the data and Poisson regression methods require the counts to be serially independent. Chapter 3 provides a detailed discussion of the statistical issues and the results of the more detailed epidemiological analysis.

In Chapter 4 the relationship between temperature and ill health with regard to age group, socio-economic status and city of residence is assessed. An analysis of the effects

of changes in temperature, absolute temperatures and sudden cold snaps is described before concentrating on the effects of air pollution. Data on daily levels of Nitrogen Dioxide and Carbon Monoxide were available for Glasgow City and this information was incorporated into the analysis of the effects of climate on health in Glasgow City.

In Chapter 5 the concept of space-time analysis is introduced. The analysis described in this chapter acknowledges that while the main analysis is concerned with the temporal aspects of the data the spatial element is also important. The presence of a spatial pattern is established before the space-time interaction is examined. Finally Chapter 6 provides a summary of the main findings of the analysis, in terms of both the epidemiological results and the comparison of the statistical methods as well as a discussion of the limitations of the study.

Chapter 2 Seasonality of Deaths and Emergency Hospital Admissions

2.1 Introduction & Data

2.1.1 Aims of Chapter

The aim of the work of this chapter is to accurately describe the seasonal pattern of mortality and morbidity in Scotland during the period 1981 to 1993 and to determine how this seasonal pattern varies according to various demographic features. The method developed to describe these patterns was straightforward to use and understand, yet detailed enough to provide a reliable estimate of seasonal variation for which confidence intervals could be computed. Mortality data were obtained from the General Register Office, while emergency hospital admissions were used as a measure of population morbidity.

2.1.2 SMR1 and GRO data

The data used in this chapter were supplied by the Information and Statistics Division (ISD) of the NHS in Scotland and General Register Office (GRO) in Scotland. Summaries of hospital care have been collected on a 100% basis in Scotland for many years, these records are collectively known as the Scottish Morbidity Records (SMR). The database which records information on in-patient stays and day cases is known as the SMR1 database. ISD supplied computerised records of all continuous in-patient stays and day cases in general hospitals for the period 1981 to 1993 in Scotland while the GRO supplied computerised records of all deaths in Scotland for the same period. These records included demographic information as well as, in the case of the death records, 3 causes of death and in the case of the SMR1 records, 6 diagnoses, coded according to the International Classification of Diseases[70] (Ninth Revision) (ICD-9). A copy of the SMR1 form and a sample death record are shown in appendix I.

The further aim of this project was to describe in detail the relationship between climatic variables and ill health in Scotland (Chapters 3 & 4). The effect of short term climatic change on the pattern of ill health is likely to be fairly immediate, and for this reason only emergency admissions were selected for analysis from the SMR1 data base. It is

unlikely that a waiting list in-patient episode will be related to measurable climatic features. All death records were included in the analysis. The final data files for analysis consisted of over 4.3 million emergency hospital admissions and over 800 000 death records. The data were analysed separately for different causes of death and diagnostic groups. Weekly data were used in preference to daily data due to the amount of noise in the daily data. However, one problem with using weekly data was that there were 52 weeks and 1 day in most years and 52 weeks and 2 days in a leap year. In order to simplify the analysis, the data were adjusted such that each year contained exactly 52 weeks. This was done by making week number 52 contain 8 days in a normal year or 9 days in a leap year then adjusting the number of deaths in these weeks by either 7/8 or 7/9. Table 2.1 shows the distribution of deaths in the different causal groups.

Table 2.1 Total number of deaths in Scotland 1981-1993

Cause of Death	ICD-9 code	Sex	Average deaths per week	Total	Percent of all deaths
Cancer	140-239	M	145	98004	25
		F	137	92751	22
Ischaemic	410-414	M	183	123838	31
Heart Disease		F	155	104881	25
Cerebrovascular	430-438	M	60	40342	10
Disease		F	100	67861	16
Respiratory	460-519	M	68	46032	12
Disease		F	71	48061	12
Total	001-999	M	580	392298	100
		F	616	416075	100

Mortality from cancer and ischaemic heart disease (IHD) accounted for over half of all deaths in Scotland between 1981 and 1993, with deaths from cerebrovascular disease (CVD) and respiratory disease (RD) each accounting for around 10% of all deaths. Overall there were around 1200 deaths per week in Scotland. The patterns were quite different for emergency hospital admissions. Table 2.2 shows the distribution of emergency hospital admissions in different diagnostic groups.

Table 2.2 Total number of emergency hospital admissions in Scotland 1981-1993

Diagnostic Group	ICD-9 code	Sex	Ave emerg. admis. per week	Total	Percent of all emerg. admis.
Cancer	140-239	M	172	116476	5
		F	160	108103	5
Ischaemic	410-414	M	252	170235	8
Heart Disease		F	174	117528	6
Cerebrovascular	430-438	M	91	61496	3
Disease		F	109	73554	4
Respiratory	460-519	M	354	239418	11
Disease		F	289	195279	9
Total	001-999	M	3263	2205774	100
		F	3105	2099130	100

The majority of emergency hospital admissions were for accidents while respiratory admissions accounted for around 10% of all emergency admissions. Around 6-8% of emergency admissions were for IHD with 5% for cancer and around 3% for cerebrovascular disease.

2.1.3 Dividing the data into subgroups

2.1.3 (a) Cause of death/Hospital diagnosis

The data were split into the main disease groups from the ICD-9 chapter headings. The decision of which disease groups to concentrate on was made by considering two points; the proportion of all deaths which were attributable to that cause and the estimated size of seasonal variation in deaths from that cause. The expected seasonal variation from a cause was assessed by looking at a plot of the data and from information gained from the literature. The disease groups for which the seasonal analysis was most relevant were ischaemic heart disease, cerebrovascular disease, and respiratory disease. These categories accounted for over half of all deaths and all had demonstrated a seasonal pattern in a simple plot of the weekly number of deaths. The other disease group that was considered was cancer, cancer deaths accounted for a quarter of all deaths. Of these four groups respiratory disease showed the greatest seasonal variation, followed by IHD and

CVD and there was virtually no seasonal variation in deaths from cancer. From this preliminary analysis and results from the literature it was decided to concentrate the detailed analysis on respiratory disease, IHD and CVD.

2.1.3 (b) Age group

The data were split into 5 age groups. The age groups were determined by considering age groups in which the population is at varying risks of deaths from certain diseases and with reference to previous studies which showed that seasonal variation is greatest at older age groups. The age groups were 0-9, 10-59, 60-69, 70-79 and 80+.

2.1.3 (c) Socio-economic status

Deprivation scores were assigned to each case using the Carstairs scoring system. Carstairs[71] attached a score to each postcode sector in Scotland based on (i) the proportion of male unemployment in the postcode sector, (ii) the proportion of overcrowding (iii) the proportion of car ownership and (iv) the proportion of the population classed as social class I & II. The Carstairs scores were divided into 5 groups ranging from deprived to affluent, with each category containing approximately 20% of the population. These were then further grouped into 3 categories. The 'affluent' category contained the top 20% of the population, the 'average' group contained the middle 60% of the population and the 'deprived' group contained the bottom 20% of the population.

Information about an individual's occupational social class was provided on death records but not on SMR1 forms, therefore emergency hospital admissions could only be assessed by the area-based deprivation score, whereas the death records could also be analysed according to social class. Occupational social class for females is unreliable as it frequently is based on the occupation of the spouse, therefore a social class analysis was only conducted for males aged 16 and over.

The distribution of deprivation and social class amongst the population when compared by age at death is not even, as shown in table 2.3. For each cause of death, in the younger age groups the proportion of deaths is greatest in the deprived areas, whereas in the older age groups the proportion of deaths is greatest in the affluent areas. This is due to the fact that the population of deprived areas in general die younger than those who live in affluent areas. The relationship between social class and age group, while not as strong

as that for deprivation categories, also indicates that the population of social classes III, IV & V die younger than the populations in social classes I & II. From table 2.3 it appears that there is a clear trend across deprivation categories, whereas with the social class data there are similar age distributions in social classes III, IV & V which are different to those in social classes I & II. There has been much work on the inequalities in health related to social class and socio-economic deprivation[72,71] which is beyond the scope of this thesis. However, these observations suggest that when assessing seasonal variation according to deprivation categories and social class, the data should be age standardised to avoid the problems of confounding. Seasonal patterns have consistantly shown to be greater at older age groups and not standardising for age would mean that, because affluent areas have higher proportions of older people they may be shown to experience greater seasonality. This could be wrongly attributed to socio-economic status when, in fact, it is purely an age association.

Table 2.3 Distribution of age group, social class and deprivation

Cause of Death	Age Group	Deprivation Category			Social Class		
		Aff(%)	Ave(%)	Dep(%)	I & II(%)	III(%)	IV & V(%)
IHD	<65	23.6	29.1	36.3	26.7	30.8	30.1
	65-74	32.3	34.2	34.6	32.4	34.1	35.0
	75+	44.0	36.7	29.0	40.9	35.0	34.8
CVD	<65	11.4	14.3	20.5	11.8	15.3	15.7
	65-74	24.4	29.5	32.0	26.1	29.8	30.3
	75+	64.2	56.2	47.4	62.1	54.8	54.0
RD	<65	9.9	14.0	19.8	10.4	14.0	14.6
	65-74	21.3	26.4	29.9	22.1	27.2	28.0
	75+	68.8	59.7	50.3	67.5	58.9	57.4
All Cause	<65	24.2	28.6	35.6	25.6	28.6	28.5
	65-74	27.6	30.2	31.3	28.8	30.9	31.4
	75+	48.2	41.2	33.2	45.6	40.5	40.2

Age standardisation was carried out using data from both the 1981 and 1991 censuses. The population were grouped into 18 quinenial year age groups and three deprivation

groups 'affluent', 'average' and 'deprived' for both 1981 and 1991. The data were then interpolated and extrapolated to provide the weekly population figures for the time period Jan 1981 to Dec 1993 in each of the 54 (18*3) age/deprivation categories. The rate for each age/deprivation category was computed by dividing the weekly number of deaths by the weekly population estimate and this was standardised to the 1991 Scottish population figures using direct standardisation techniques.

Population figures by social class were not available in the 1981 census, however they were available in the 1991 census. In order to interpolate and extrapolate the social class data the percentage change in the overall population for each of the 18 age groups from 1981 to 1991 was applied to each social class group within the age group. This method assumed that the change in population patterns within a 5-year age group was spread evenly throughout the social class groups. This may not be the case, but it was the only method in which the weekly population by age group and social class from 1981 to 1993 could be estimated. Again, the mortality data was divided by the population data and direct standardisation techniques were used.

2.1.3 (d) Health Board of residence

The data were also analysed according to Health Board of residence. As with the analysis by deprivation category and social class, age effects may confound a geographical effect, so the data were age standardised in a similar method. Population data by five year age group and Health Board were available from both the 1981 and 1991 census. The three Island Health Boards; Orkney, Shetland and the Western Isles were grouped together as 'Islands' to avoid the problems of very small numbers.

2.1.4 Exploratory plots

The weekly number of deaths in Scotland from 1981 to 1993 showed a clear seasonal pattern (figures 2.1 & 2.2). On average, there were around 1200 deaths per week in Scotland. This figure increased to over 1400 deaths per week in most winters and, in summer, fell to around 1000 deaths per week. The pattern is less clear for emergency hospital admissions (figures 2.3 & 2.4). The weekly number of admissions for all diagnostic groups showed a clear upward trend. The number of admissions per week was

around 2600 for females and 2800 for males in 1981, this had increased to around 4000 admissions per week for both males and females by 1993.

Figure 2.1 Weekly deaths from all causes Scotland 1981-1993

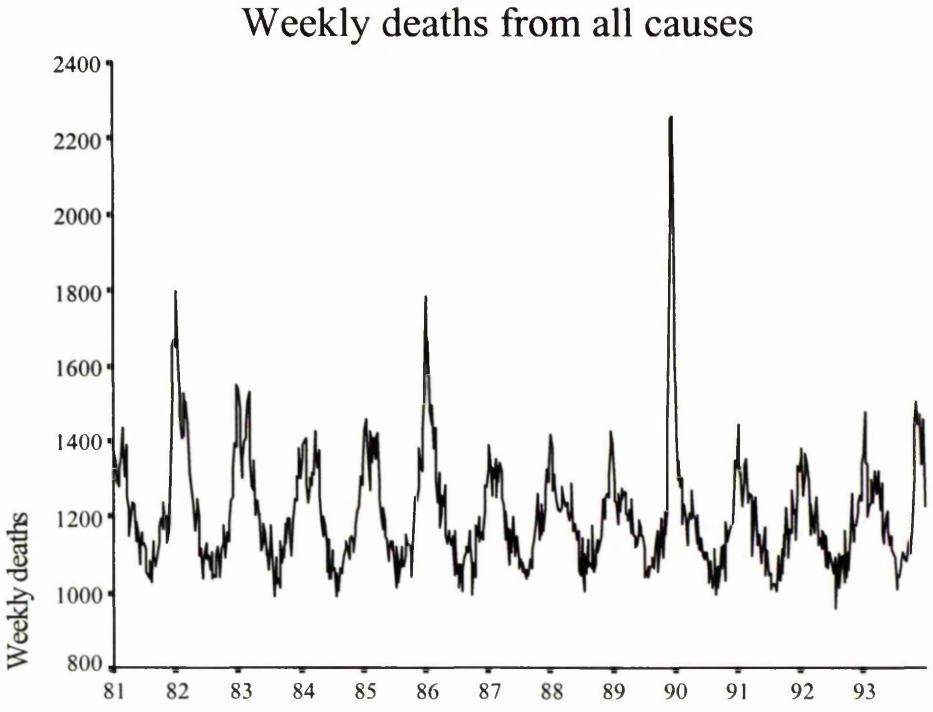


Figure 2.2 Average weekly deaths from all causes, males and females

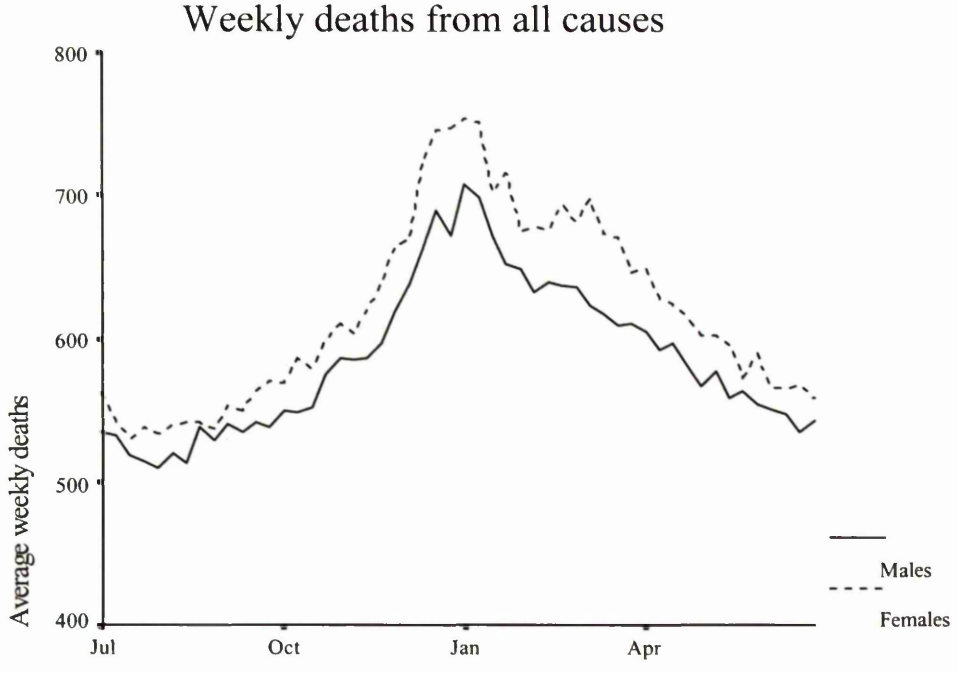


Figure 2.3 Weekly emergency admissions in Scotland 1981-1993

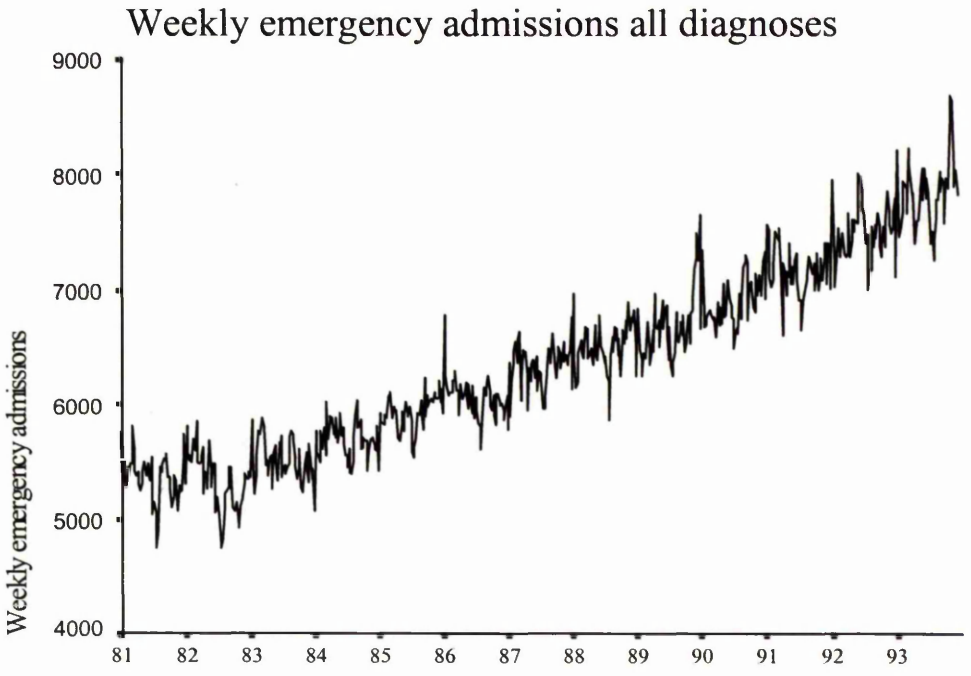
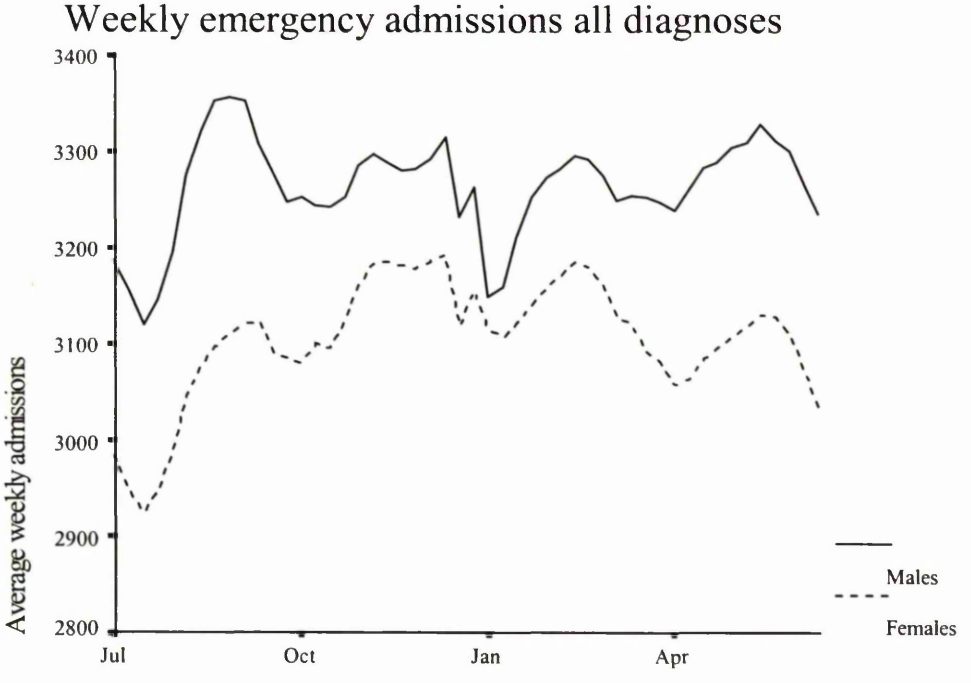


Figure 2.4 Average weekly emergency admissions in Scotland 1981-1993



When the data were averaged over a year, several peaks and troughs throughout the year were evident (figure 2.4). There was indication of a trough in the summer months but little evidence of any peak in winter.

2.2 Development of Methods

2.2.1 Trend Analysis

Before estimating the seasonal component in weekly deaths and hospital admissions a simple linear regression was fitted to the data shown in figures 2.1 and 2.3. There was found to be a decreasing trend in the number of deaths and an increasing trend in weekly admissions. In order to study the existence of seasonal patterns the linear trend in the series was removed by fitting a straight line to the data. The residual series from this fit contained the seasonal and the random variation present in the data.

2.2.2 Spectral Analysis

Spectral analysis of time series describes the fluctuations in the series in terms of sinusoidal curves. The spectrum of a series is defined by the autocovariance of the series and when based on the sample autocovariances is known as the periodogram. The periodogram is a plot of the amount of variation in the series that could be associated with a cycle at a particular frequency or period. A sharp peak in the periodogram may indicate the presence of a cycle at the corresponding frequency. It can be used to identify cyclical patterns in data which may not have been evident in a plot of the data and which may not have been predictable before the data were examined. Smoothed estimates of the spectrum which gain precision as n increases can also be used to identify cycles in data. These smoothed estimates are known as 'windowed' estimators for the spectrum. Once cyclical behavior has been establised harmonic modelling can be used to model these observed patterns.

If we imagine that our series consists of several sinusoidal components, then the data series can be written as

$$y_t = \sum_{k=1}^{m} \left\{ \alpha_k \cos(\omega_k t) + \beta_k \sin(\omega_k t) \right\} + u_t \qquad t = 1,.....,n \qquad (2.1)$$

where u_t is a white noise sequence and each ω_k is one of the Fourier frequencies. If we further specify $\omega_k = 2\pi k / n$, for some positive integer $k < n/2$ then the regression sum

of squares associated with a particular frequency ω can be computed and the periodogram ordinates $I(\omega)$ defined

$$I(\varpi) = \left[\left\{ \sum_{t=1}^{n} y_t \cos(\omega t) \right\}^2 + \left\{ \sum_{t=1}^{n} y_t \sin(\omega t) \right\}^2 \right] / n \qquad (2.2)$$

The periodogram is a plot of the periodogram ordinates $I(\omega)$ against ω. The periodogram will show a peak in the value of $I(\omega)$ if a cycle of frequency ω results in a large regression sum of squares. This plot will demonstrate where cycles may be found in the data.

For example in our data set of 13 years giving a series of 676 weeks an annual cycle would be represented by a peak in the periodogram at the frequency $\omega = \frac{2\pi k}{n}$, or

$$\omega = \frac{2\pi 13}{676} = \frac{2\pi}{52} = 0.12083.$$

Cycles at other frequencies that were not predictable before using spectral analysis would also be identified by a peak in the periodogram. To make identification of these cycles easier, the periodogram ordinate can be plotted against the periodicity of the curve rather than the frequency. In this case a yearly cycle would show as a peak at a periodicity of 52 weeks rather than a frequency of 0.1208 ($2\pi/52$) and a 6 month cycle would show as a peak at 26 weeks rather than a frequency of 0.24166 ($2\pi/26$) or 2ω.

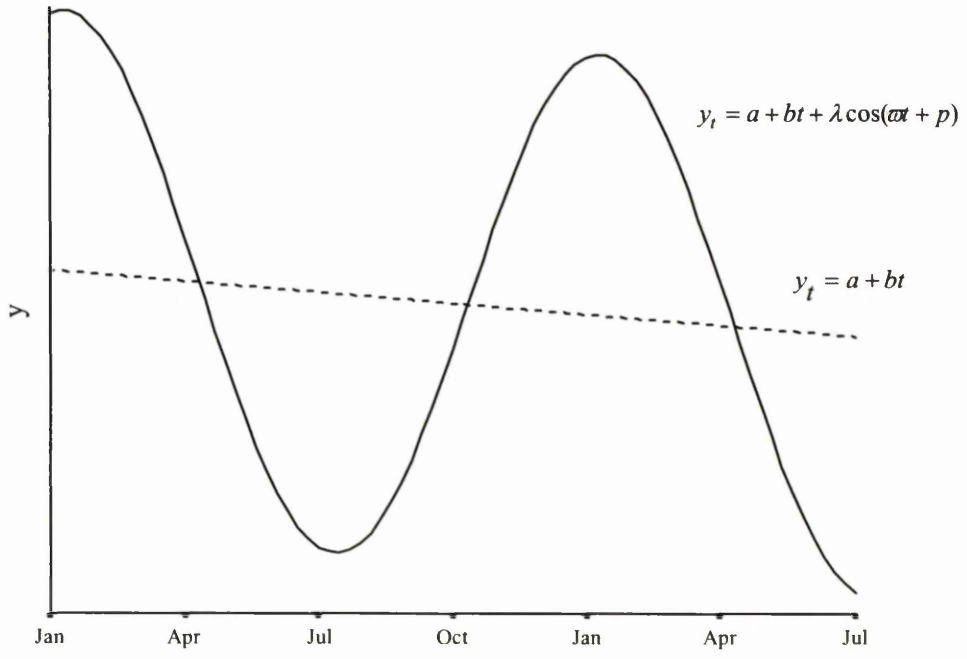
2.2.3 Harmonic modelling

The all cause male mortality data y_t showed a clear yearly cycle and a negative trend. These features were evident simply from looking at the data (figure 2.1). The yearly cycle in the data was the cycle of most interest and spectral analysis, to identify any other patterns in the data, was not necessary at this stage. The model fitted was of the form

$$y_t = a + bt + \lambda \cos(\omega t + p) \qquad t = 0,1,.....675 \qquad (2.3)$$

A plot of this curve is shown in figure 2.5

Figure 2.5 Sinusoidal curve



In this model a represents the constant, b the trend, λ the amplitude of the seasonal curve and p the phase of the curve. The phase of the curve influences where in the year the peak occurs. This model gave a basic pattern for the data, incorporating a simple sinusoidal curve peaking once every 52 weeks and a linear term. The model accounted for the average number of deaths, the trend in the number of deaths and the yearly cycle. Spectral methods were then used to look at the residuals from this fitted model. A scatter plot of the residuals seemed to show a random scatter with a few outliers (figure 2.6). In order to use spectral methods with time series data the series must be stationary. A time series can be described as stationary if there is no trend in the data and if the covariance between two points depends only on the absolute difference between the points and not their position in the series. The residual series was found to be second order stationary (see section 3.3.3 for a more detailed description of stationarity). From a spectral analysis of the residual series other cyclical patterns which had not been accounted for by the simple yearly sinusoidal model could be determined. Spectral analysis of the residual series in the form of a periodogram showed that there were still other cycles occurring within the residuals (figure 2.7).

Figure 2.6 Residuals from linear regression on all cause male mortality

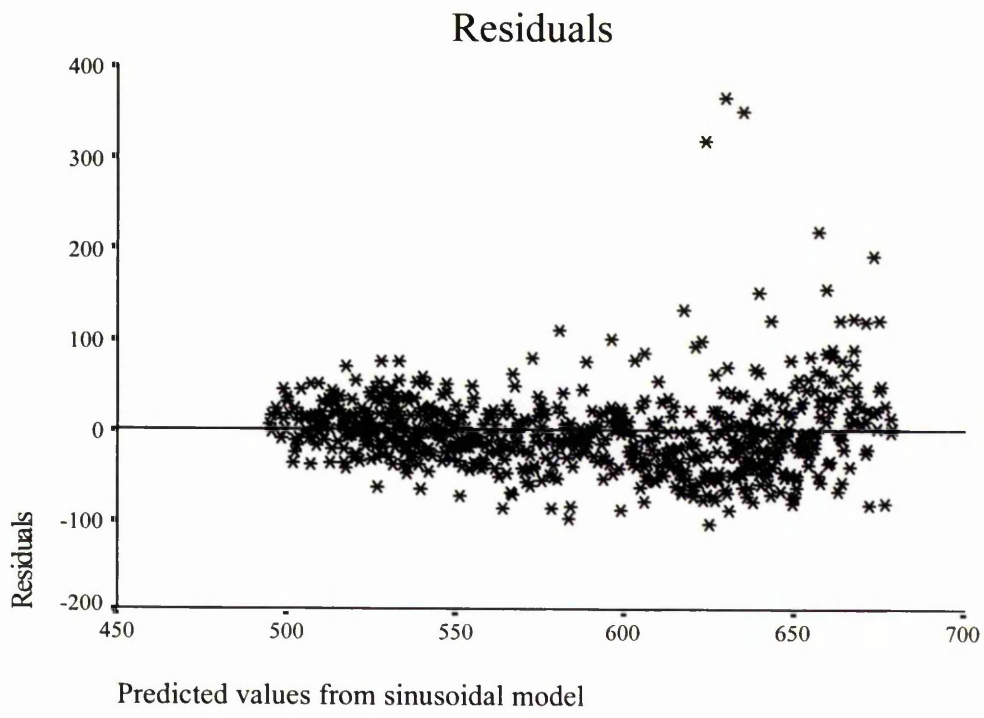
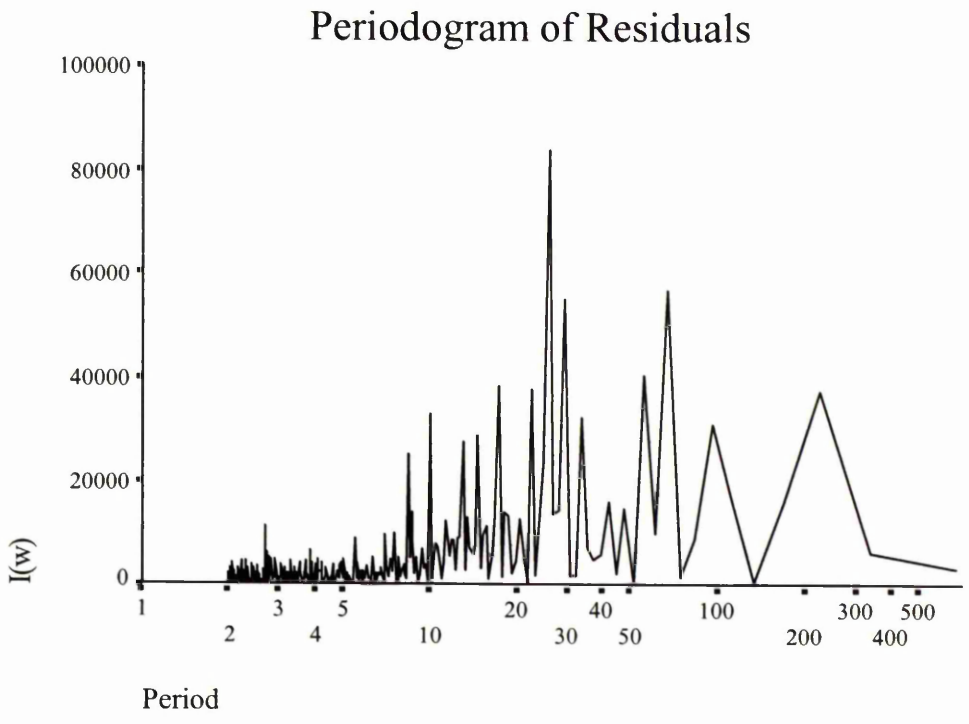


Figure 2.7 Periodogram of residuals for all cause male mortality



The plot of the periodogram using period rather than frequency on the x-axis showed a peak at a period of 26 weeks, indicating an additional 6 monthly cycle.

This apparent cycle appearing at a frequency of 2ω was however not a real cycle, this peak in the periodogram occurs due to the fact that the data follow a non-sinusoidal curve. The model that had been fitted contained one cosine term to represent the cyclical nature of the data. The data however did not follow a purely sinusoidal curve, therefore at a certain time each year the model does not fit well, thus examination of the residuals from the fit using spectral methods displays evidence of other cycles in the data. The weekly number of deaths from all causes does not follow a 6 month cycle but the presence of this other possible cycle in the periodogram indicates that the data deviates from a straightforward sinusoidal curve. These other 'pseudo-cycles' which often appear in spectral analysis are known as harmonics.

Harmonics occur in a cycle when the cyclical variation in the original series is non-sinusoidal. If a series has a clear cycle at frequency 'ω' then if this is a non-sinusoidal cycle there will be peaks in a periodogram at frequencies '2ω','3ω','4ω' etc. If there are peaks at these frequencies in the periodogram they should not be interpreted as separate cycles in the series but as indicators that the cycle is not sinusoidal.

The harmonic terms can be represented in a model in the same way as the yearly cycle was represented. The inclusion of these terms in the model leads to the model being a better fit of the data. By including a term for all the harmonics found in the data the residual series will show no cyclical features and no trend.

These full harmonic models were of the form

$$y = a + bt + \lambda_1 \cos(\omega t + p_1) + \lambda_2 \cos(2\omega t + p_2) + \lambda_3 \cos(3\omega t + p_3) + \ldots\ldots \qquad (2.4)$$

This method of modelling the data in such detail was done for both male and female deaths from all causes, IHD, CVD and respiratory disease. It provided a good insight into the form of the data and the fact that the patterns in the data were not straightforward. However, the information gained from the individual parameters from this type of model is not easy to interpret. The main problem with the full model containing several cosine terms is that no simple summary measure of amplitude could be obtained using the parameter estimates in the model.

2.2.4 Simple model

The second method of modelling provided a simple summary measure of the amplitude of the seasonal curve expressed as the average percentage increase in deaths from summer to winter. The method developed by Halberg, Tong and Johnson[73] involves fitting a sinusoidal curve to time series data using least squares regression. Lentz[74] provides a straightforward description of the technique, while Teicher[75] and Faure[76] discuss the technique in more detail. Bloomfield[77] provides a good general introduction to the sinusoidal analysis of time series.

From (2.1) assuming the presence of only 1 cyclical term, our series can be written as

$$y_t = \alpha\cos(\omega t) + \beta\sin(\omega t) \text{ where } \omega = 0.1203\ (2\pi/52) \qquad (2.5)$$

using trigonometric expansions this equation can be written as

$$y_t = \lambda\cos(\omega t + p) \qquad (2.6)$$

where λ is the amplitude of the fitted curve and p is the phase of the curve. In this model $\lambda = \sqrt{\alpha^2 + \beta^2}$ and $p = \arctan(-\beta/\alpha)$.

In the previous section concerned with harmonic modelling the data was modelled using the non-linear model function in SPSS. However, in this case, when the aim is not to get the most accurate fit to the data but to develop a method that can provide a summary measure of seasonal variation with confidence limits, the linear model (2.5) was more appropriate. Equation (2.5), the trigonometric expansion of (2.6), is equivalent to (2.3) but without the trend.

The clear trend in the data was modelled using

$$y_t = a + bt \qquad (2.7)$$

this gave an expected number of deaths per week assuming no seasonal variation was present. No other structured form was considered for the data, anything greater than linear terms would introduce potential extrapolation problems. Next a generalised linear model with a poisson error and a log link of the form

$$\ln(y_t) = \alpha\cos(\omega t) + \beta\cos(\omega t) \qquad (2.8)$$

was fitted using the expected values $a+bt$ as an offset.

The inclusion of the offset term meant that when $\lambda=\sqrt{\alpha^2+\beta^2}$ was computed λ represented the increase from the mean to the peak of the fitted curve expressed as a ratio and $p=\arctan(-\beta/\alpha)$.

The variance of λ was given directly in the modelling procedure using equation (2.3), but when using the expansion necessary for the log-linear analysis (2.8), the variance of λ had to be computed separately. This was done using the result that for a function $Z=H(x,y)$, where x is $N(\mu_x,\sigma_x^2)$ and y is $N(\mu_y,\sigma_y^2)$ then

$$Var(Z)=\left[\frac{\partial H}{\partial x}\right]^2\sigma_x^2+\left[\frac{\partial H}{\partial y}\right]^2\sigma_y^2+\left[\frac{\partial H}{\partial x}\right]\left[\frac{\partial H}{\partial y}\right]Cov(x,y)$$

this gives

$$Var(\lambda)=\left[\frac{\alpha}{\sqrt{(\alpha^2+\beta^2)}}\right]^2\sigma_\alpha^2+\left[\frac{\beta}{\sqrt{(\alpha^2+\beta^2)}}\right]^2\sigma_\beta^2+\left[\frac{\alpha\beta}{\alpha^2+\beta^2}\right]\sigma_{\alpha\beta}^2$$

An S-Plus function was written to estimate λ which was then expressed as the percentage increase from the trough to the peak of the fitted seasonal curve using $\lambda^*=\frac{2\lambda}{1-|\lambda|}$, 95% confidence intervals for this estimate were also provided.

This method provided a means of comparing the seasonal increase in mortality between different populations or different subgroups of the same population. The method was used to analyse the seasonal pattern of mortality and morbidity by sex, age group, social class, deprivation category and health board.

A copy of the S-PLUS function is provided in appendix II.

2.3 Mortality Results

As mentioned in section 2.2, two methods of modelling the data were explored. Firstly the data were modelled using a fairly complex harmonic model, which was later found to be of little use in the descriptive analysis as it did not provide a simple measure of amplitude for which a confidence interval could be computed. The second method used a

simple model, which could be used to provide an estimate of the amplitude expressed as the percentage increase in mortality from the summer trough to the winter peak. This method also accounted for the fact that the data were Poisson, which while not of great concern when considering deaths from all causes, would be important when dealing with smaller numbers. When describing the seasonal patterns for each cause of death the data have firstly been examined using simple summaries such as the average pattern over the 13 years and have then been described according to the results from the sinusoidal model. The results should be looked at bearing in mind that the simple model averages the data and, as shown by the more detailed modelling in section 2.2.3, the data in fact follow a non-sinusoidal pattern.

2.3.1 All Causes

As described in section 2.1.4 there was a clear seasonal pattern apparent in the plots of mortality from all causes over the 13 year period (figure 2.1) and in the average number of deaths per week averaged over one year (figure 2.2). This section of the thesis reports on the results from modelling these data using a simple sinusoidal curve.

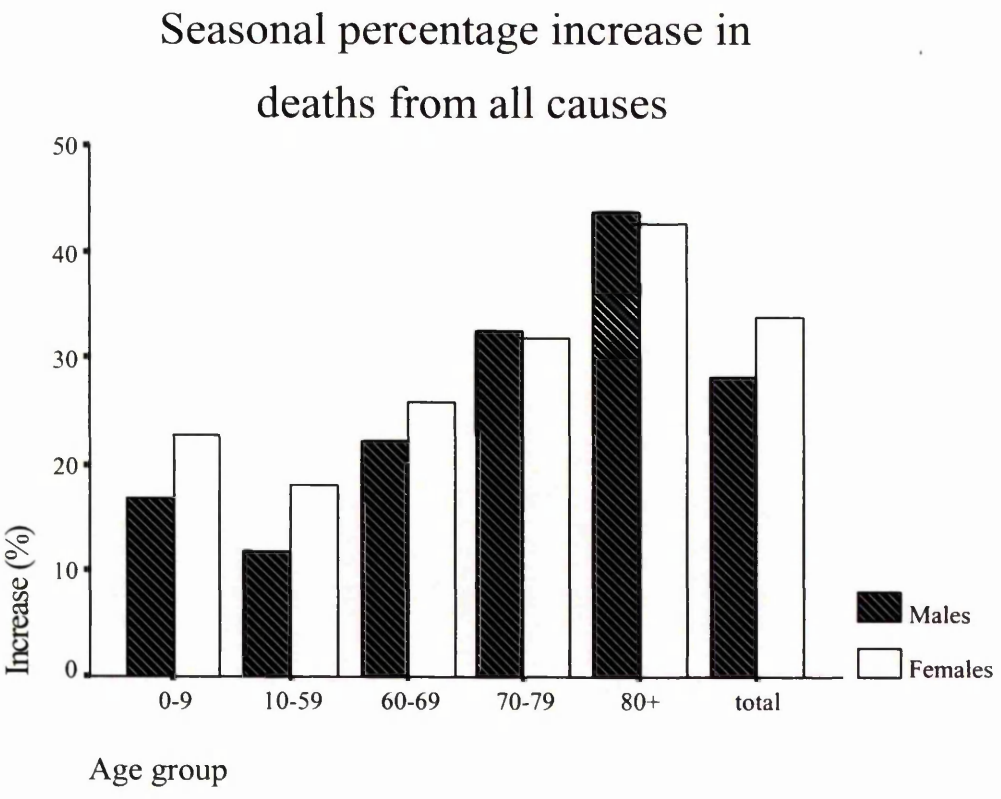
2.3.1 (a) Results from sinusoidal model for all cause mortality

When modelled using a simple sinusoidal curve with just one cosine term the percentage increase in deaths from all causes from trough to peak for males was 28.3% with a 95% confidence interval (27.5,29.0) and 33.8% (33.1,34.6) for females. The confidence intervals for male and female total deaths from all causes did not overlap, implying that the difference in amplitudes was significant, the seasonal increase in female deaths from all causes was significantly greater than the seasonal increase for males. The seasonal increase in mortality from all causes was greatest at the very young age groups and the older age groups. These age groups are the more vulnerable members of society, those who may struggle to keep warm in the winter or who are more vulnerable to infection. The seasonal increase when the data are split by age and sex is given in table 2.4 and figure 2.8

Table 2.4 Seasonal percentage increase in deaths by sex & age group, all causes

Age group	Males		Females	
	Percent Increase	CI	Percent Increase	CI
0-9	16.8	(11.2,22.4)	22.6	(15.8,29.5)
10-59	11.7	(10.2,13.3)	17.9	(15.8,20.0)
60-69	22.1	(20.7,23.6)	25.8	(24.0,27.5)
70-79	32.5	(31.3,33.8)	31.8	(30.5,33.1)
80+	43.7	(42.2,45.3)	42.7	(41.6,43.8)
All Ages	28.3	(27.5,29.0)	33.8	(33.1,34.6)

Figure 2.8 Seasonal percentage increase in deaths by sex & age group, all causes



The assessment of the seasonal increase in mortality from all causes for different deprivation categories and different social classes was done using age standardised data to take account of the fact that the age structure of the population varied in different deprivation categories and social classes. There appeared to be no difference in the

degree of seasonal variation experienced in each of the deprivation categories. However, when the data were analysed according to occupational social class the seasonal increase in male mortality in social classes I & II was considerably smaller than that experienced in social classes III, IV & V. People in the lower social classes experienced a greater seasonal increase in mortality than those in the higher social classes. This may be related to a person's ability, financial or otherwise, to protect themselves from the effects of cold winter conditions. The results of this analysis are shown in table 2.5.

Table 2.5 Seasonal percentage increase in mortality by deprivation category and social class, all causes (age standardised)

	Males		Females	
Deprivation Category	%	CI	%	CI
Affluent	30.1	(26.4,33.8)	35.1	(31.4,38.8)
Average	30.4	(27.0,33.8)	34.0	(30.5,37.5)
Deprived	31.7	(28.5,34.9)	36.6	(33.1,40.0)
All categories	30.6	(27.2,34.1)	34.7	(31.2,38.2)
Social class				
I & II	23.2	(18.8,27.7)	-	-
III	30.2	(27.2,33.3)	-	-
IV & V	30.9	(27.7,34.1)	-	-
All categories	29.3	(25.9,32.6)	-	-

In order to assess whether seasonal increases in mortality varied according to geographical area of residence the data were split into separate Health Board areas. In general there was little difference in the degree of seasonal variation in mortality between the health boards however the seasonal increase in male mortality in the Islands was significantly lower than in all other health boards. The seasonal increase in mortality was greater in females than in males in each health board. The results of this analysis can be seen in table 2.6.

Table 2.6 Seasonal percentage increase in mortality by Health Board of residence, all causes, (age standardised)

Health Board	Percent Increase	Males 95%CI	Percent Increase	Females 95%CI
Argyll & Clyde	31.1	(27.7,34.5)	33.5	(30.0,36.9)
Ayrshire & Arran	25.7	(22.4,29.1)	29.0	(25.6,32.4)
Borders	38.2	(34.3,42.0)	36.6	(32.9,40.3)
Dumfries & Galloway	27.4	(23.8,30.9)	32.4	(28.8,35.9)
Fife	27.1	(23.7,30.6)	33.1	(29.5,36.6)
Forth Valley	29.2	(25.8,32.6)	36.0	(32.4,39.5)
Grampian	30.2	(26.6,33.8)	32.0	(28.3,35.7)
Greater Glasgow	30.2	(26.9,33.4)	36.8	(33.4,40.3)
Highlands	25.2	(21.8,28.7)	33.2	(29.6,36.9)
Lanarkshire	28.8	(25.4,32.1)	30.6	(27.2,34.0)
Lothian	29.7	(26.2,33.2)	38.0	(34.3,41.7)
Tayside	29.7	(26.2,33.3)	39.6	(35.9,43.2)
Islands	17.2	(13.9,20.6)	29.8	(26.1,33.4)

2.3.1 (b) Summary of seasonal analysis for all cause mortality

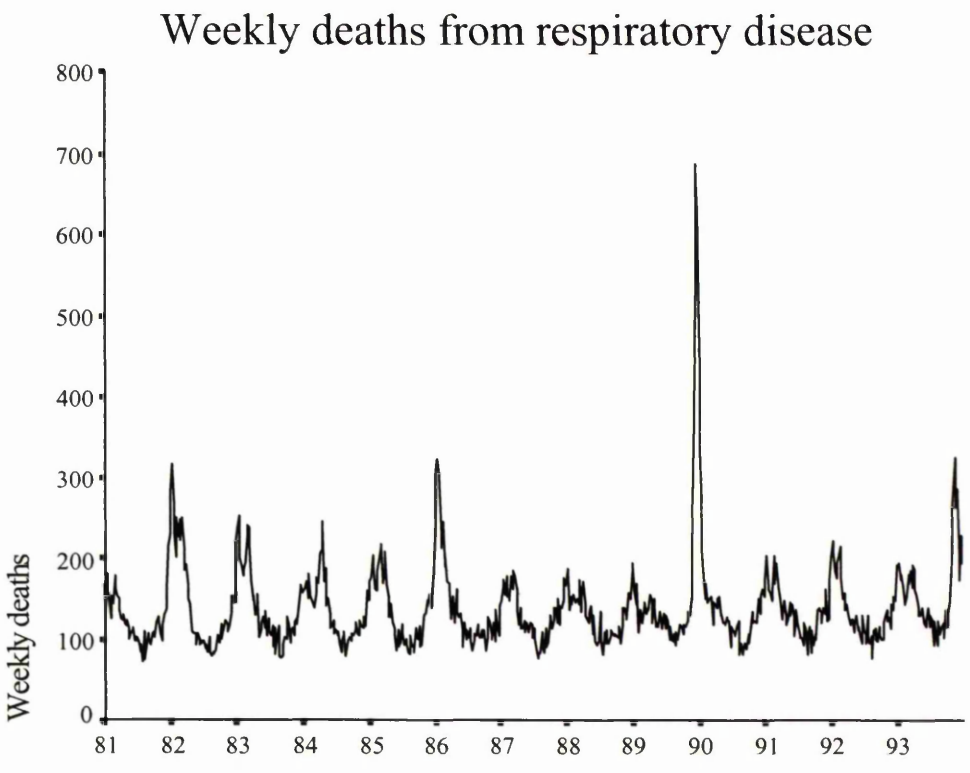
The seasonal increase in mortality in Scotland from 1981 to 1993 was 28% for males and 34% for females, these figures were greater in the very young and the older age groups. There was little variation in seasonal mortality according to area based deprivation measures, however when the data were grouped according to occupational social class, social classes III, IV and V experienced a greater seasonal increase in mortality than social classes I & II. Seasonal increases in mortality were similar in each health board apart from the Islands which experienced lower male seasonal variation than all the other health boards.

2.3.2 Respiratory disease

2.3.2 (a) Exploratory analysis for respiratory disease

There were over 94 000 deaths from respiratory disease (ICD 460-519) during the time period 1981-1993 and deaths from this cause showed a greater degree of seasonal variation than any other cause of death. Figure 2.9 shows that the seasonal effect was not the same each year, there were larger than normal peaks in years 82,83,84,86,90 and 92. For males the average number of deaths over the full time period was 68 per week, this decreased to 35 at the minimum and increased to 285 at the peak. For females the average was 71 deaths per week, the minimum number of deaths in any week was 31 and this increased to 404 at the peak.

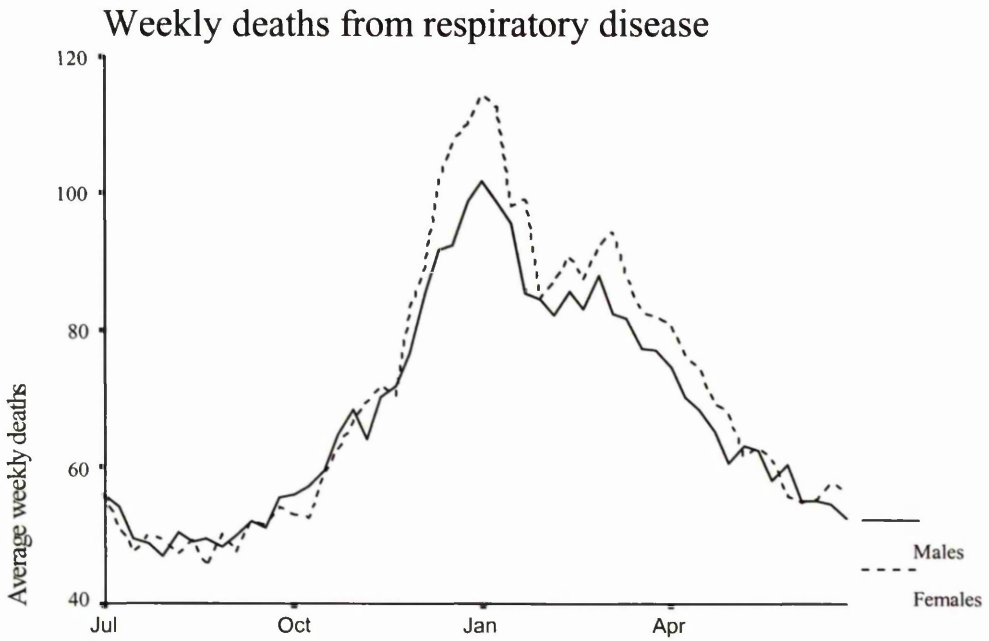
Figure 2.9 Weekly deaths from respiratory disease, Scotland 1981-1993



When the data were averaged over the 13 years (figure 2.10) the average value of the trough was 47 deaths per week for males and 46 deaths per week for females. The average value for the peak was 102 deaths per week for males and 115 deaths per week for females. It can be seen from these figures that there is a steep peak which is not

mirrored in the depth of the trough and that this peak was more than double its average size in the largest peak during Dec 1989 and Jan 1990. From the average figures we can see that there is approximately a 100% increase in the number of weekly deaths from trough to peak in males and that this figure is greater in females.

Figure 2.10 Average weekly deaths from respiratory disease, males and females



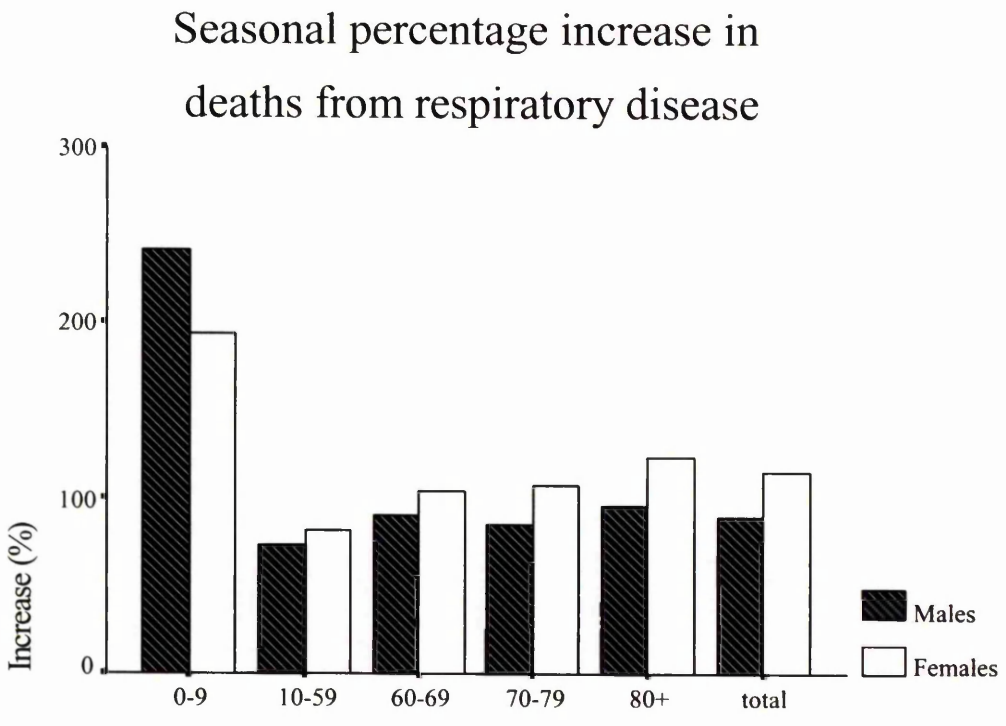
2.3.2 (b) Results from sinusoidal model for respiratory disease

When modelled using a simple sinusoidal curve with just one cosine term, the percentage increase in deaths from respiratory disease from trough to peak was 88.7% (86.1,91.3) for males and 114.2% (111.5,117.0) for females. The model gives a 95% confidence interval for each of the parameter estimates. The confidence intervals for male and female total deaths from respiratory disease did not overlap, implying that the difference in amplitudes was significant, the seasonal increase in female deaths from respiratory disease is significantly greater than the seasonal increase for males. The results for different age groups are shown in table 2.7 and figure 2.11

Table 2.7 Seasonal percentage increase in deaths by sex & age group, respiratory disease.

Age group	Males		Females	
	Percent Increase	CI	Percent Increase	CI
0-9	240.6	(189.5,291.9)	193.1	(137.2,249.1)
10-59	72.7	(63.8,81.5)	80.6	(70.0,91.2)
60-69	89.7	(83.4,96.0)	104.0	(96.3,111.7)
70-79	84.5	(80.3,88.7)	107.1	(101.8,112.5)
80+	95.6	(91.3,99.9)	123.6	(119.9,127.3)
All Ages	88.7	(86.1,91.3)	114.2	(111.5,117.0)

Figure 2.11 Seasonal percentage increase in deaths by sex & age group, respiratory disease



The seasonal increase in mortality from respiratory disease increased with age group after ages 0-9. Seasonal increases were greater in females at each of these age groups except for age group 0-9. The greatest seasonal increase, 240% for males and 193% for

females, occurred in the very young age group. This again is likely to be related to the fact that the elderly and the very young are those who require more protection from the effects of cold weather than other members of the population. The figures for the 0-9 age group are based on small numbers.

Again, age standardised data was used to assess the seasonal increase in mortality from respiratory disease for different deprivation categories and different social classes. The results of this analysis are shown in table 2.8. The analysis by deprivation category appeared to indicate a pattern of greater seasonal variation in the more affluent areas. However, the estimates have large confidence intervals and the suggestion is that there is no difference in the size of the seasonal variation in respiratory mortality according to socio-economic deprivation for both males and females. When the data were analysed by social class there was again no evidence of a difference in the size of the seasonal increase in mortality experienced by the different social classes.

Table 2.8 Seasonal percentage increase in mortality by deprivation category and social class, respiratory disease, (age standardised)

	Males		Females	
Deprivation Category	%	CI	%	CI
Affluent	96.1	(82.1,110.1)	129.6	(115.3,144.0)
Average	93.2	(80.5,105.7)	117.0	(103.8,130.3)
Deprived	87.1	(76.5,97.7)	114.0	(102.3,125.7)
All categories	92.5	(80.2,1.05)	119.1	(105.4,132.8)
Social class				
I & II	92.3	(72.7,112.0)	-	-
III	95.3	(84.3,106.3)	-	-
IV & V	88.5	(77.3,99.7)	-	-
All classes	92.8	(80.3,105.2)	-	-

The results from the analysis by health board of residence are supplied in table 2.9. The seasonal increase in mortality for respiratory disease was lower in the Borders than in the other health boards for both males and females. In the Islands the seasonal increase was lower for males only. The greatest seasonal increase occurred in Dumfries and Galloway.

No clear conclusions can be drawn from these results as Borders HB and Dumfries and Galloway HB border each other, are both rural health boards and yet one experiences lower than average seasonal increases in respiratory mortality and one experiences greater than average seasonal increases in mortality.

Table 2.9 Seasonal percentage increase in mortality by health board of residence, respiratory disease, (age standardised)

Health Board	Percent Increase	Males 95%CI	Percent Increase	Females 95%CI
Argyll & Clyde	84.9	(72.8,96.9)	113.0	(99.0,127.0)
Ayrshire & Arran	78.7	(66.6,90.7)	93.3	(80.2,106.5)
Borders	45.2	(38.1,52.2)	34.3	(26.8,41.8)
Dumfries & Galloway	127.5	(112.1,143.0)	138.6	(123.0,154.2)
Fife	68.6	(56.8,80.4)	101.1	(87.4,114.8)
Forth Valley	110.3	(97.7,122.9)	106.0	(93.1,119.0)
Grampian	103.8	(89.4,118.2)	116.9	(102.1,131.6)
Greater Glasgow	87.5	(76.7,98.3)	119.0	(106.6,131.2)
Highlands	106.0	(91.2,121.0)	100.4	(85.7,115.1)
Lanarkshire	73.4	(62.5,84.4)	102.5	(90.3,114.8)
Lothian	95.3	(82.3,108.3)	132.5	(117.7,147.4)
Tayside	92.3	(79.3,105.3)	125.1	(111.0,139.2)
Islands	43.4	(31.0,55.8)	117.5	(102.5,132.6)

2.3.2 (c) Summary of seasonal analysis for respiratory disease

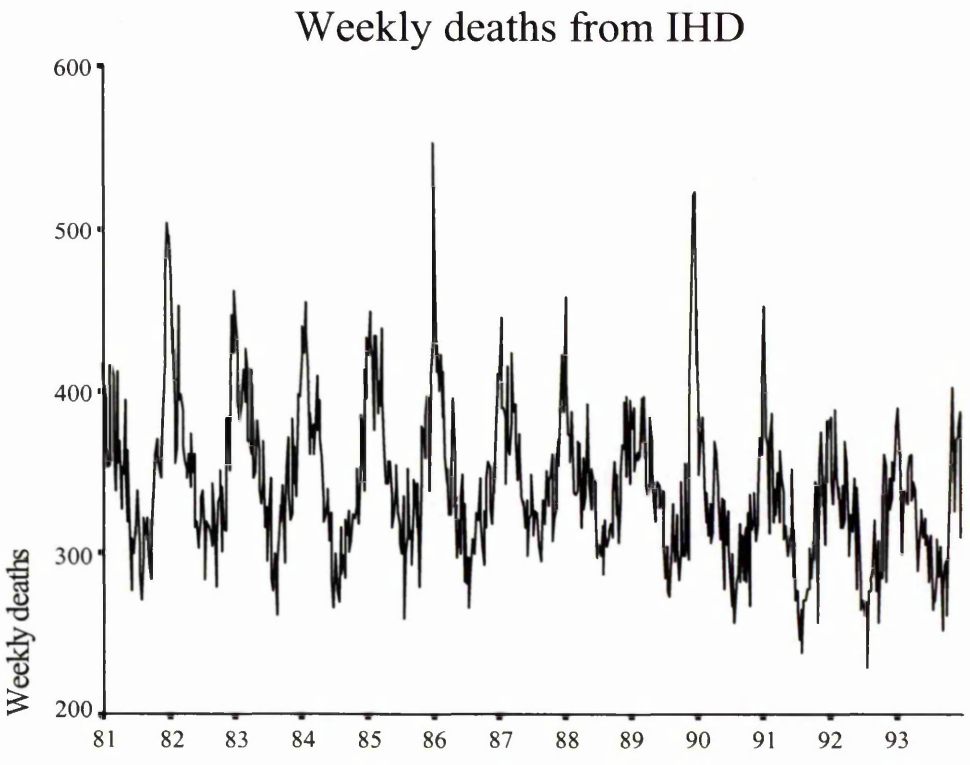
Mortality from respiratory disease displayed the greatest amount of seasonal variation, with male mortality increasing by 89% from summer to winter and female mortality increasing by 114% from summer to winter. The size of the seasonal increase was greater at older ages and was very large in the 0-9 age group. There was no variation in the seasonal mortality from respiratory disease when the data were analysed according to deprivation categories and occupational social class. Borders Health Board and the Island Health Boards appeared to experience a lower seasonal increase in mortality from respiratory disease than the other health board areas.

2.3.3 Ischaemic Heart disease

2.3.3 (a) Exploratory analysis for ischaemic heart disease

Ischaemic heart disease accounted for 31% of all deaths in males over the time period and 25% of all deaths for females, the largest single cause of death. There was a total of 228 719 deaths from IHD from Jan 1st 1981 to Dec 31st 1993. Figure 2.12 shows a similar pattern to the plot for deaths from respiratory disease. There is a clear seasonal pattern peaking in the winter weeks, however the irregular large peaks every 4 years were not as prominent for deaths from IHD as they were for deaths from respiratory disease.

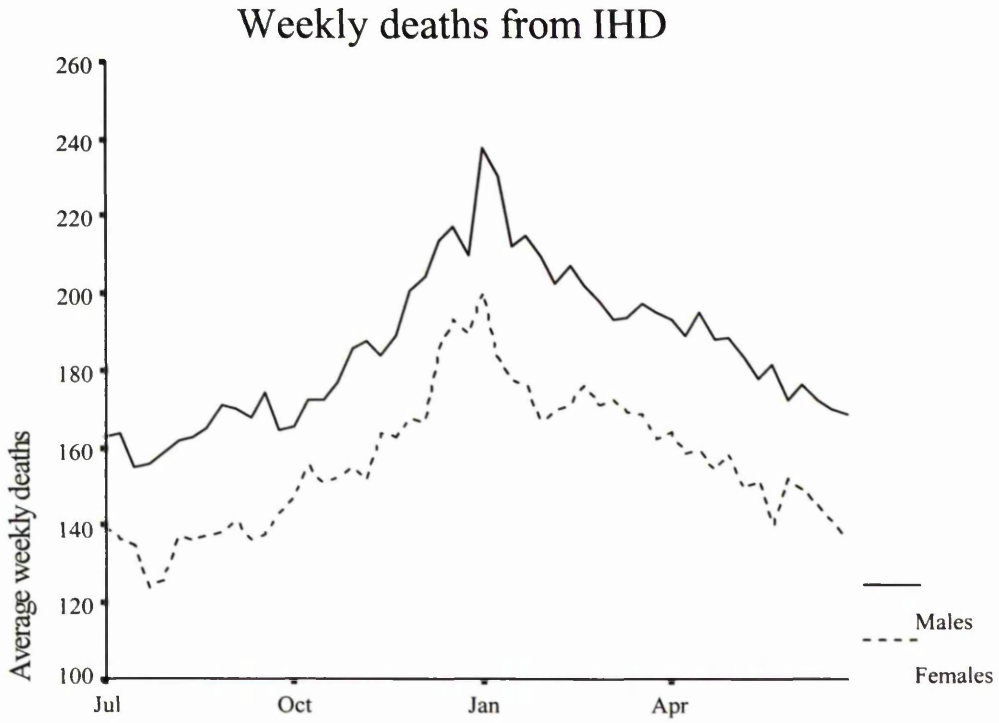
Figure 2.12 Weekly deaths from IHD, Scotland 1981-1993



The average number of male deaths from IHD was 186 per week, the minimum number of deaths in any week was 123 and the maximum was 321. For females the average weekly number of deaths was 156, the minimum was 92 deaths in a week and the maximum was 265 deaths per week. Again it can be seen from the raw data that the difference from peak to average is considerably larger than the difference from average

to trough. When the data were averaged over all years (figure 2.13) the peak for males was 238 and for females was 200 and the trough was 155 for males and 124 for females. From this averaged data the estimated values of the percentage increase from trough to peak was 54% for males and 61% for females.

Figure 2.13 Average weekly deaths from IHD, males and females



2.2.3 (b) Results from sinusoidal model for ischaemic heart disease

When modelled using the simple sinusoidal curve the percentage increase from trough to peak was 34.1% (32.8,35.4) for males and 36.1% (34.7,37.6) for females. The seasonal variation in mortality from IHD was greater in females than for males. Again one has to bear in mind that the estimate from the averaged data is effected by outliers and that the estimate from the fitted curve smoothes over the extreme values in each year. As with respiratory disease this method results in an underestimation of the true value of the size of the curve due to the fact that it averages out the data and imposes symmetry on the estimate. This method of modelling, however, provides a good method of comparing subgroups of the population. Table 2.10 shows the seasonal increase in deaths from IHD

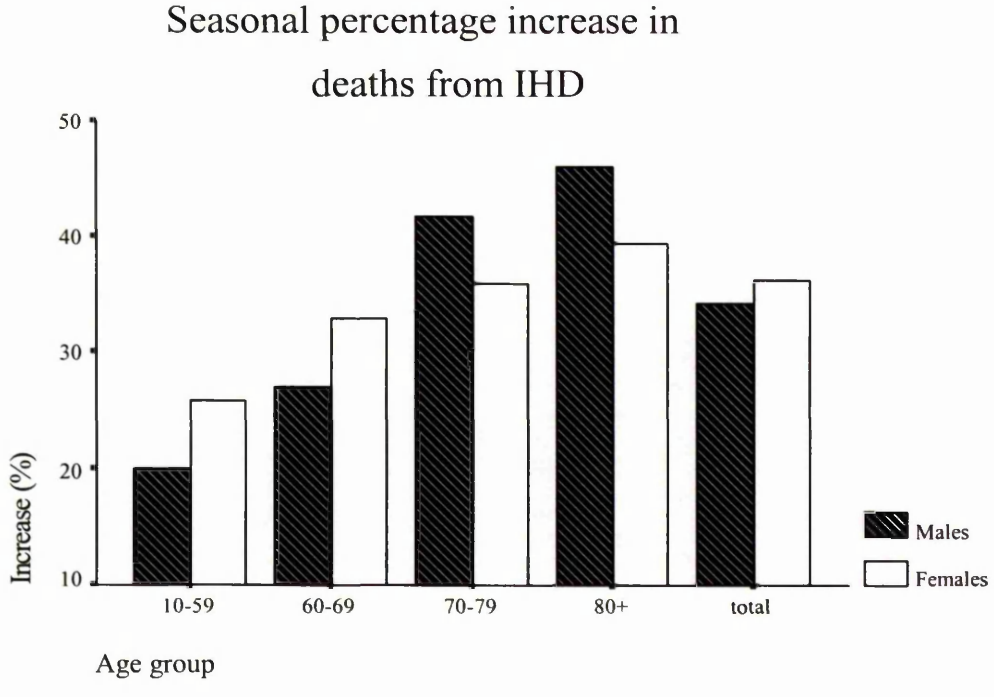
when the data were split by age group and sex. There were very few deaths from IHD in the youngest age group 0-9 so this group was excluded from the seasonal analysis.

Table 2.10 Seasonal percentage increase in deaths by sex & age group, IHD.

Age group	Males		Females	
	Percent Increase	CI	Percent Increase	CI
0-9	-	-	-	-
10-59	20.0	(17.0,22.9)	25.7	(20.2,31.3)
60-69	26.9	(24.5,29.3)	32.7	(29.4,36.2)
70-79	41.7	(39.5,44.0)	35.9	(33.5,38.3)
80+	46.0	(42.9,49.1)	39.3	(37.1,41.5)
All Ages	34.1	(32.8,35.4)	36.1	(34.7,37.6)

A clear pattern of increasing seasonality with increasing age was seen for both males and females in deaths from IHD. Again those in the oldest age group suffer the greatest seasonal increase in mortality from IHD. These results are displayed in figure 2.14.

Figure 2.14 Seasonal percentage increase in mortality by sex & age group, IHD



The results of the age standardised analysis of seasonal variation by deprivation category and social class are shown in table 2.11.

Table 2.11 Seasonal percentage increase in mortality by deprivation category and social class, IHD, (age standardised)

	Males		Females	
Deprivation Category	%	CI	%	CI
Affluent	36.7	(29.9,43.5)	33.4	(26.2,40.5)
Average	35.9	(29.7,42.1)	38.8	(32.0,45.6)
Deprived	37.2	(31.2,43.3)	42.9	(36.2,49.6)
All categories	36.3	(30.0,42.6)	38.5	(31.3,45.6)
Social class				
I & II	25.6	(17.0,34.2)	-	-
III	36.0	(30.0,42.1)	-	-
IV & V	34.5	(28.1,40.8)	-	-
All classes	33.7	(27.1,40.4)		

There was no difference between deprivation categories for male deaths from IHD. However, for female mortality it appeared that those people living in affluent areas may experience a lower seasonal increase in mortality from IHD than those living in deprived areas. There appeared to be considerably greater seasonal variation in mortality from IHD in social classes III, IV & V as compared to social classes I & II. While there are no significant differences observed in this analysis, the patterns suggest that there may be a social class/deprivation effect on seasonal increases in mortality from IHD.

The seasonal increase in mortality from IHD by health board is shown in table 2.12. In general there are no clear patterns between the health boards however for both males and females the Island health boards display significantly less seasonal variation than several of the other health board areas.

Table 2.12 Seasonal percentage increase in mortality by health board of residence, IHD, (age standardised)

Health Board	Percent Increase	Males 95%CI	Percent Increase	Females 95%CI
Argyll & Clyde	35.1	(29.0,41.1)	36.3	(29.5,43.2)
Ayrshire & Arran	35.0	(28.9,41.1)	33.7	(26.8,40.5)
Borders	45.2	(38.1,52.2)	34.3	(26.8,41.9)
Dumfries & Galloway	27.7	(21.7,33.8)	38.8	(31.8,45.9)
Fife	34.1	(27.9,40.3)	43.0	(35.7,50.2)
Forth Valley	24.4	(18.4,30.4)	37.8	(30.7,44.9)
Grampian	37.1	(30.4,43.9)	33.9	(26.5,41.4)
Greater Glasgow	38.3	(32.1,44.5)	35.4	(28.4,42.4)
Highlands	34.4	(28.0,40.8)	34.9	(27.4,42.4)
Lanarkshire	31.6	(25.8,37.5)	33.3	(26.7,39.9)
Lothian	36.8	(30.2,43.4)	40.5	(32.9,48.1)
Tayside	35.2	(28.7,41.6)	45.8	(38.4,53.3)
Islands	23.5	(17.6,29.3)	26.6	(19.4,33.7)

2.3.3 (c) Summary of seasonal analysis for ischaemic heart disease

The ratio from peak to trough in weekly deaths from IHD was around 34% for males and 39% for females and this value increased with increasing age. There was little difference between deprivation categories for males, however for females those people living in more affluent areas may have experienced less seasonal variation than those who lived in deprived areas. Males in social classes III, IV and V experienced greater seasonal variation in mortality from IHD than those in social classes I & II. There was little variation by health board although people who lived in the Island Health Boards experienced lower seasonal increases from IHD than the other health board areas.

2.3.4 Cerebrovascular disease

2.3.4 (a) Exploratory analysis for cerebrovascular disease

There was a total of over 100 000 deaths from CVD from 1st Jan 1981 to 31st Dec 1993, these deaths accounted for 10% of all male deaths over the time period and 16% of all female deaths over the time period. From a plot of the raw data (figure 2.15) a clear seasonal pattern can be seen, and there is also evidence of increased peaks in years 82, 86 and 92 (as seen in the deaths from respiratory disease). The average number of deaths per week for males was 60, the minimum number of deaths in a week was 33, rising to 116 as the maximum number of deaths in any week. For females the average number of weekly deaths was 100, minimum number of deaths in any week was 64 and the maximum was 189. Again, these figures show that the increase to the peak is greater than the decrease to the trough and that the data do not follow a symmetric curve. When the data is averaged over all years (figure 2.16) the trough was 46 deaths per week for males and 83 deaths per week for females with a peak of 76 male deaths per week and 128 female deaths per week from CVD.

Figure 2.15 Weekly deaths from CVD, Scotland 1981-1993

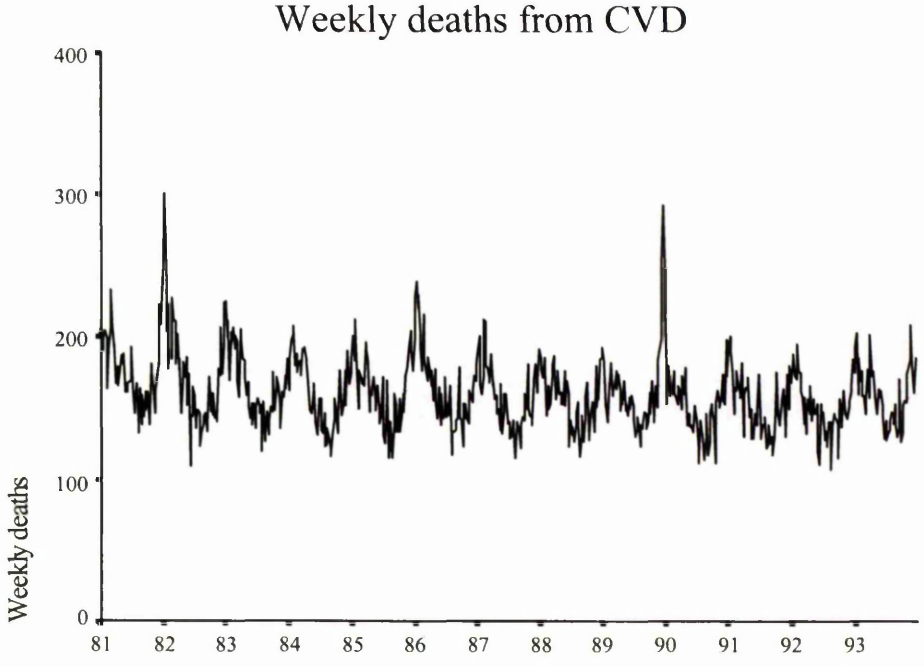
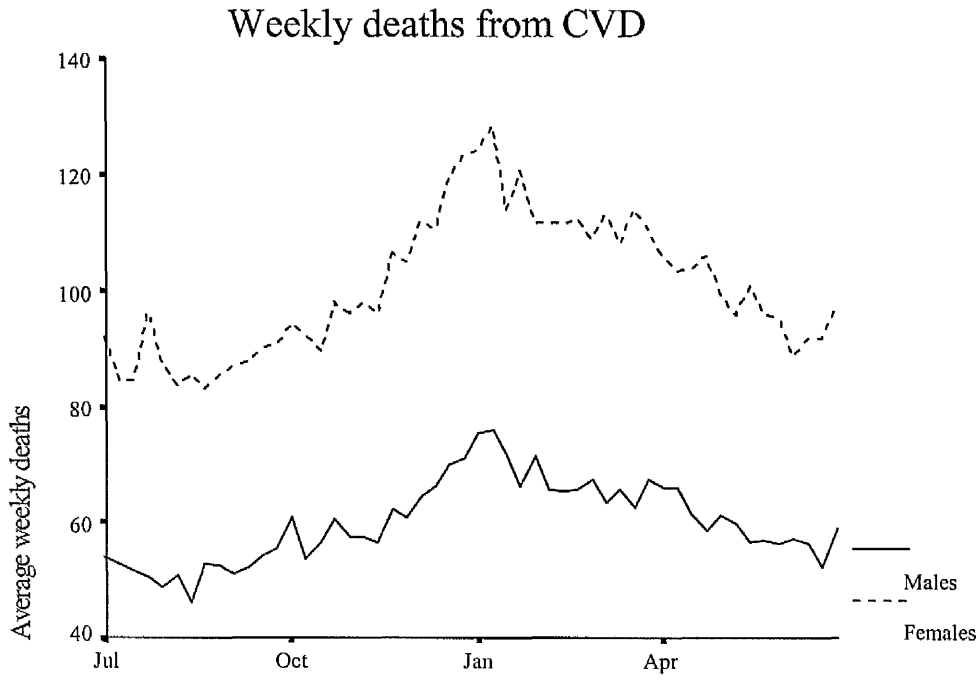


Figure 2.16 Average weekly deaths from CVD, males and females



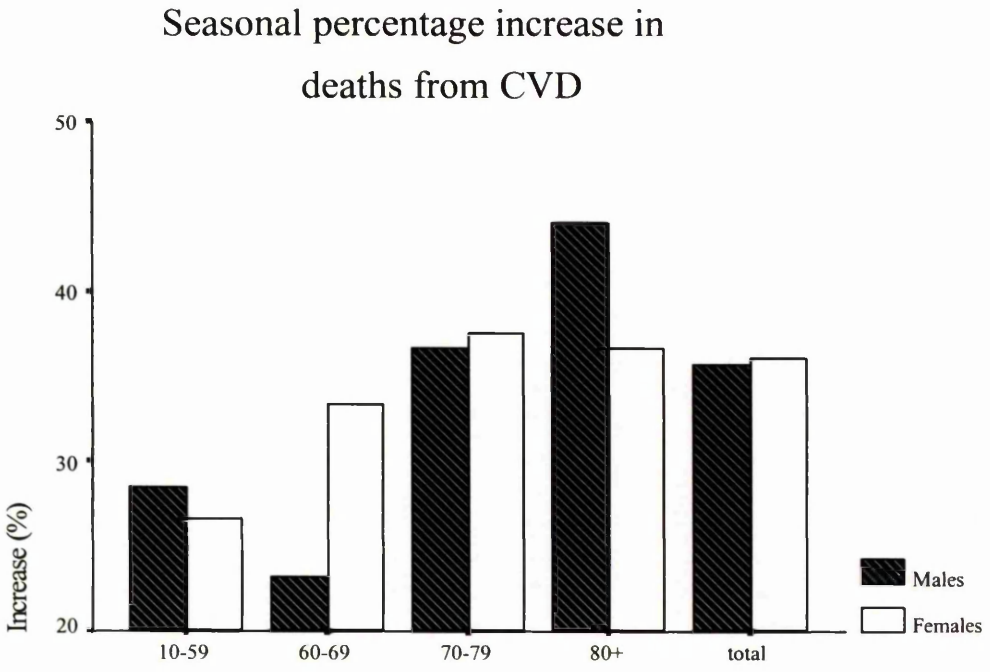
2.3.4 (b) Results from sinusoidal model for cerebrovascular disease

Analysis using the simple sinusoidal curve gives an estimate of the ratio of average to trough or peak as 36% for both males (33.4,38.0) and females (34.3,37.9). As with IHD there are very few deaths from CVD per week in the 0-9 age groups and for this reason the analysis was not performed for this age group. The results from the analysis by age group are shown in table 2.13 and figure 2.17.

Table 2.13 Seasonal percentage increase in deaths by sex & age group, CVD.

Age group	Males		Females	
	Percent Increase	CI	Percent Increase	CI
0-9	-	-	-	-
10-59	28.4	(20.6,36.1)	26.5	(18.7,34.4)
60-69	23.2	(18.1,28.3)	33.4	(27.8,39.0)
70-79	36.6	(33.0,40.3)	37.5	(34.3,40.8)
80+	44.0	(40.0,48.1)	36.6	(34.3,39.0)
All Ages	35.7	(33.4,38.0)	36.1	(34.3,37.9)

Figure 2.17 Seasonal percentage increase in deaths by sex & age group, CVD



Again, as with respiratory disease and IHD, the seasonal increase in mortality from CVD increased with increasing age. The results of the age standardised analysis of seasonal variation by deprivation category and social class are shown in table 2.14.

Table 2.14 Seasonal percentage increase in mortality by deprivation category and social class, CVD, (age standardised)

	Males		Females	
Deprivation Category	%	CI	%	CI
Affluent	35.2	(23.8,46.6)	38.7	(30.0,47.4)
Average	39.5	(28.7,50.4)	38.9	(30.6,47.3)
Deprived	35.6	(23.9,45.3)	38.1	(29.6,46.6)
All categories	38.1	(27.1,50.0)	38.5	(29.7,47.3)
Social class				
I & II	36.1	(21.5,50.7)	-	-
III	34.0	(24.6,43.5)	-	-
IV & V	40.7	(30.3,51.0)	-	-
All classes	36.2	(25.5,46.9)		

It appears from the social class analysis that the seasonal increase in mortality from CVD is slightly lower in social class III compared to social classes I & II and IV & V. However, the confidence intervals in this analysis are too large to assume that any real differences between the social classes exists. There was no effect of deprivation on the size of the percentage seasonal increase in mortality from CVD.

There was considerable variability in the seasonal increase in mortality from CVD within health boards. For males the seasonal variation in the Highlands and Islands was very low while in the Borders the seasonal increase was very large. The variation between health boards was not as great for females as males and no real patterns could be established. The results of this analysis are shown in table 2.15.

Table 2.15 Seasonal percentage increase in mortality by health board of residence, CVD, (age standardised)

Health Board	Percent Increase	Males 95%CI	Percent Increase	Females 95%CI
Argyll & Clyde	37.5	(27.1,48.0)	40.9	(32.5,49.3)
Ayrshire & Arran	32.9	(22.6,43.2)	27.5	(19.5,35.5)
Borders	70.6	(57.5,83.7)	37.3	(28.1,46.5)
Dumfries & Galloway	42.6	(31.5,53.8)	29.0	(20.7,37.2)
Fife	33.0	(22.7,43.4)	47.8	(39.1,56.5)
Forth Valley	39.7	(28.7,50.6)	38.5	(29.9,47.1)
Grampian	33.9	(22.5,45.3)	31.4	(22.2,40.5)
Greater Glasgow	33.0	(22.2,43.8)	39.7	(30.7,48.7)
Highlands	15.2	(5.4,25.0)	33.8	(25.3,42.4)
Lanarkshire	32.6	(21.9,43.4)	20.8	(12.6,29.1)
Lothian	43.3	(32.0,54.6)	40.1	(31.1,49.1)
Tayside	49.1	(37.1,61.2)	45.9	(36.3,55.6)
Islands	28.1	(17.5,38.6)	52.1	(42.7,61.4)

3.3.4 (c) Summary of seasonal analysis for cerebrovascular disease

The size of the seasonal increase in mortality from CVD from trough to peak of the seasonal curve is around 36% for both males and females. This figure varied by age

group and increased with increasing age, there was no variation by social class or deprivation category and considerable variation by health board area.

2.3.5 Cancer

Other studies[3,25] have shown that there is very little seasonal variation in the pattern of deaths from cancer. This was confirmed in this analysis. Cancer deaths accounted for one in four deaths and so, although they did not appear to show much variation, they are an important cause of death and for that reason they were included in the descriptive part of the analysis. There was an average of 145 cancer deaths in males per week over the time period and an average of 137 deaths from cancer per week for females (figure 2.18). There was a significant increasing trend in the number of deaths from cancer in both males and females. When the data were averaged over one year a slight seasonal pattern was visible but the data were very noisy (figure 2.19). No clear pattern could confidently be seen in the averaged data. When the data were modelled using a sinusoidal curve the percentage increase from the trough to the peak of the fitted seasonal curve was 4% (3.0,5.5) for males and 5% (3.2,5.9) for females. While this seasonal variation is significantly greater than zero, seasonal patterns in cancer mortality were negligible.

Figure 2.18 Weekly deaths from cancer, Scotland 1981-1993

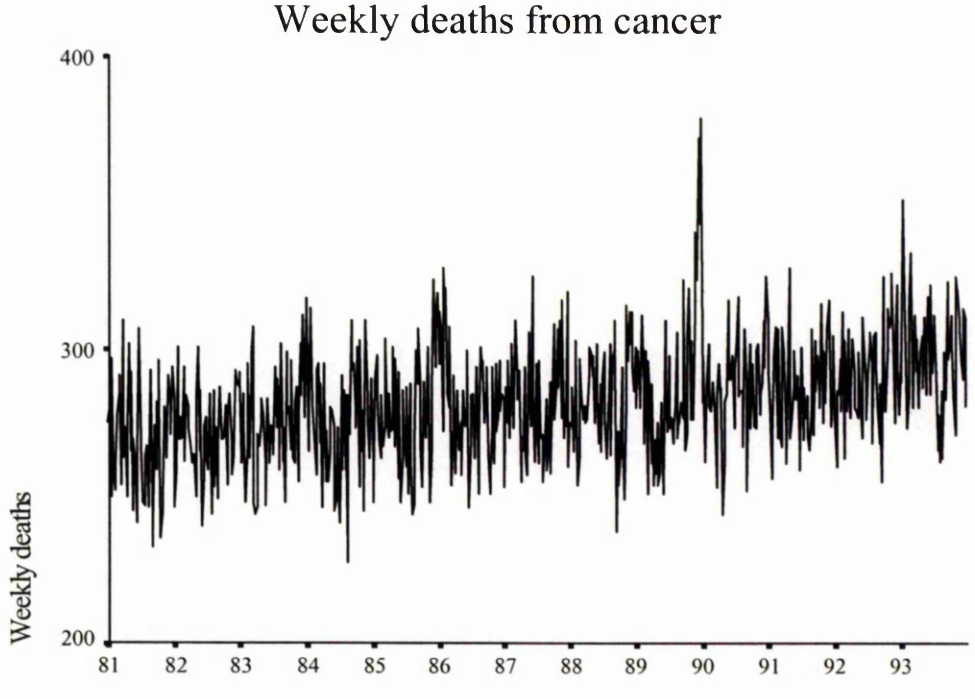
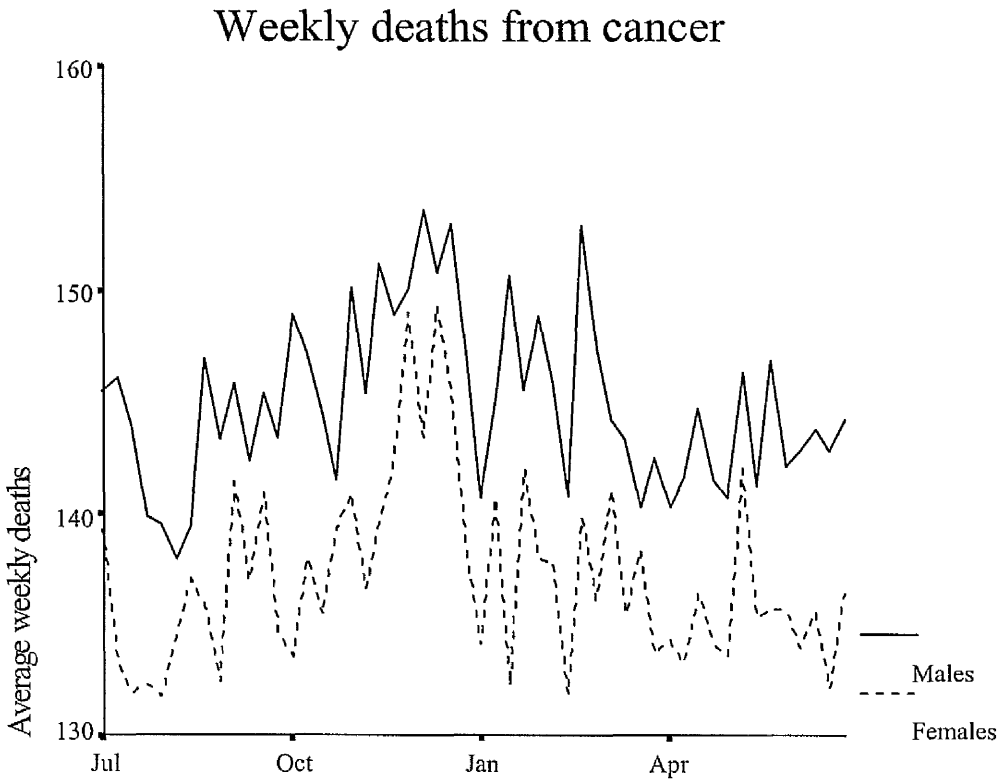


Figure 2.19 Average weekly deaths from cancer, males and females



An analysis by age group showed a pattern of increasing seasonality with increasing age for males while no pattern was visible for females. These results are shown in table 2.16

Table 2.16 Seasonal percentage increase in deaths from cancer by sex & age group

Age group	Males		Females	
	Percent Increase	CI	Percent Increase	CI
0-9	-	-	-	-
10-59	0.9	(-2.1,3.9)	6.2	(3.2,9.2)
60-69	3.5	(1.1,5.8)	5.6	(2.9,8.3)
70-79	5.0	(2.9,7.2)	3.8	(1.5,6.2)
80+	9.3	(6.1,124)	4.3	(1.7,7.0)
All Ages	4.2	(3.0,5.5)	4.6	(3.2,5.9)

The overall 5% increase from summer to winter in cancer mortality and the fact that it was difficult to see even an age effect resulted in the analysis of the seasonal variation in cancer mortality not being taken any further. It was considered unlikely that there would be any practical advantage in studying the variation between deprivation categories, social classes and health boards from cancer mortality as the overall variation was so small.

2.4 Emergency Hospital Admission Results

The modelling procedure for the emergency admissions data was the same as that for the deaths data, however on first examination the emergency admissions data appeared to be considerably more complex than the deaths data (see figures 2.3 and 2.4). In the case of the mortality data there was a clear seasonal pattern with the number of deaths reaching a peak in the winter whereas this was not the case for all emergency admissions. It appeared that there were several significant cycles in the data resulting in about four cycles per year. When the data were split by age group and by diagnostic group the patterns observed by simple curves were even more complex. Within each age group certain disease groups would form the dominant diagnoses. The seasonal pattern appeared not only to be greater in magnitude in certain age groups and disease groups as was the case for deaths, but the pattern also peaked at different parts of the year according to age group and diagnosis. For this reason any analysis of the total number of emergency hospital admissions using sinusoidal curves was uninformative.

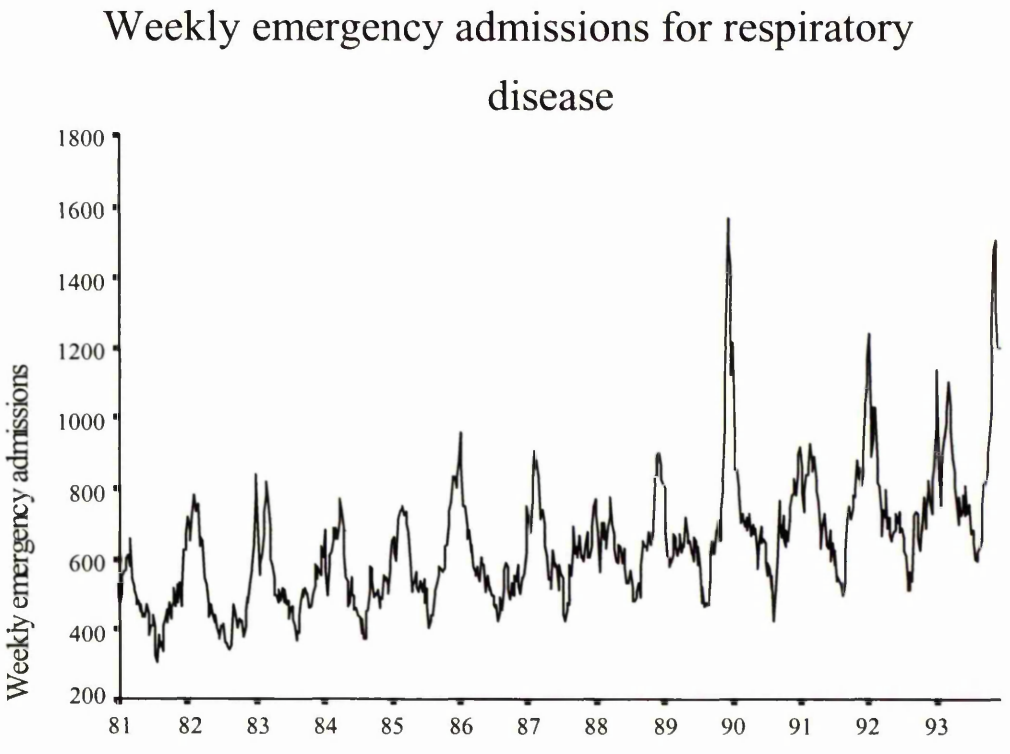
Having already developed the methodology for the deaths data the same techniques were applied to cause specific emergency admissions data. The diagnostic groups used in the analysis of admissions data were respiratory disease, ischaemic heart disease and cerebrovascular disease. Initial results from the analysis of the emergency admissions data and the previous results concerning deaths meant that the analysis for emergency admissions and future work reported on in this thesis concentrated on respiratory disease, IHD and CVD only. In order to reduce the number of tables shown in this chapter, details of seasonal variation in emergency admissions has been restricted to an analysis

by age and sex only. The results for the analyses by deprivation category and health board are given in appendix III.

2.4.1 Respiratory disease

Unlike the deaths data where the number of deaths showed only a slight decline over the time period the number of emergency admissions for respiratory disease increased considerably from 1981 to 1993. In the first week of January 1981 there were 299 male emergency admissions and 232 female emergency admissions from respiratory disease, in the same week in 1993 these figures were 575 for males and 564 for females (figure 2.20). Thus if one were to look simply at the average value over the time period, the maximum value and the minimum value as was the case with deaths, then the trend would be represented rather than any seasonal variation.

Figure 2.20 Weekly emergency admissions for respiratory disease, Scotland 1981-1993



Even if the data were averaged over the years the trend would still have an effect in the estimate of seasonal variation. An increasing trend would result in an underestimate of

the seasonal effect and for this reason no estimates of the seasonal effect using the raw data have been computed.

The simple sinusoidal model accounted for the trend and computed the percentage increase from trough to peak by considering it relative to the mean number of deaths at that time. From this model the estimate of the seasonal variation accounted for the trend component in the data. Therefore this model, which was used in the deaths data where a slight downward trend was evident, was equally applicable in the admissions data where a much stronger increasing trend was evident.

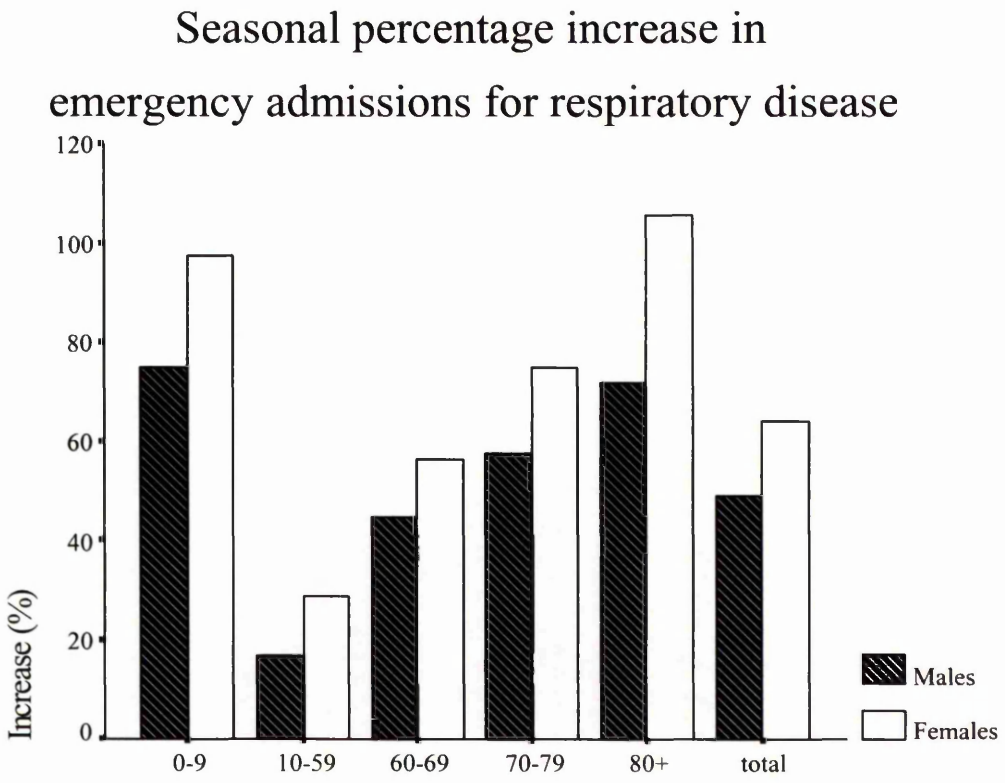
The percentage seasonal increase from summer trough to winter peak in emergency admissions for respiratory disease was 49%(47.7,49.7) for males and 64%(62.6,64.9) for females. When the data were split by age group (table 2.17) there was a similar pattern of increasing seasonality with increasing age as was the case for the deaths data.

Table 2.17 Seasonal percentage increase in emergency admissions by sex & age group, respiratory disease

Age group	Males		Females	
	Percent Increase	CI	Percent Increase	CI
0-9	74.9	(73.1,76.8)	97.1	(94.5,99.7)
10-59	16.4	(14.6,18.1)	28.5	(26.6,30.4)
60-69	44.5	(42.0,47.2)	56.0	(53.0,59.0)
70-79	57.3	(54.8,59.7)	74.7	(71.8,77.5)
80+	71.8	(68.5,75.1)	105.3	(101.8,108.8)
All Ages	48.7	(47.7,49.7)	63.8	(62.6,64.9)

As with the mortality data, the size of the seasonal increase in the number of emergency hospital admissions for respiratory disease was greater at older age groups and in the very young age group. Overall the seasonal increase for emergency hospital admissions for respiratory disease is just over half the size of the seasonal increase in deaths from respiratory disease.

Figure 2.21 Seasonal percentage increase in emergency admissions by sex & age group, respiratory disease



Due to the evidence of an age effect and the assessments of variation in mortality by deprivation category and social class it was decided to age standardise the emergency admissions data for further analysis. Social class information, however, was only available for the death records so the socio-economic analysis for emergency hospital admissions could only be carried out using deprivation categories. Again, as with the mortality data, the analysis by health board was also done using age standardised data. The results from the analysis of seasonal variation in emergency hospital admissions by deprivation category and health board are given in appendix III.

2.4.1 (a) Summary of seasonal analysis for respiratory disease

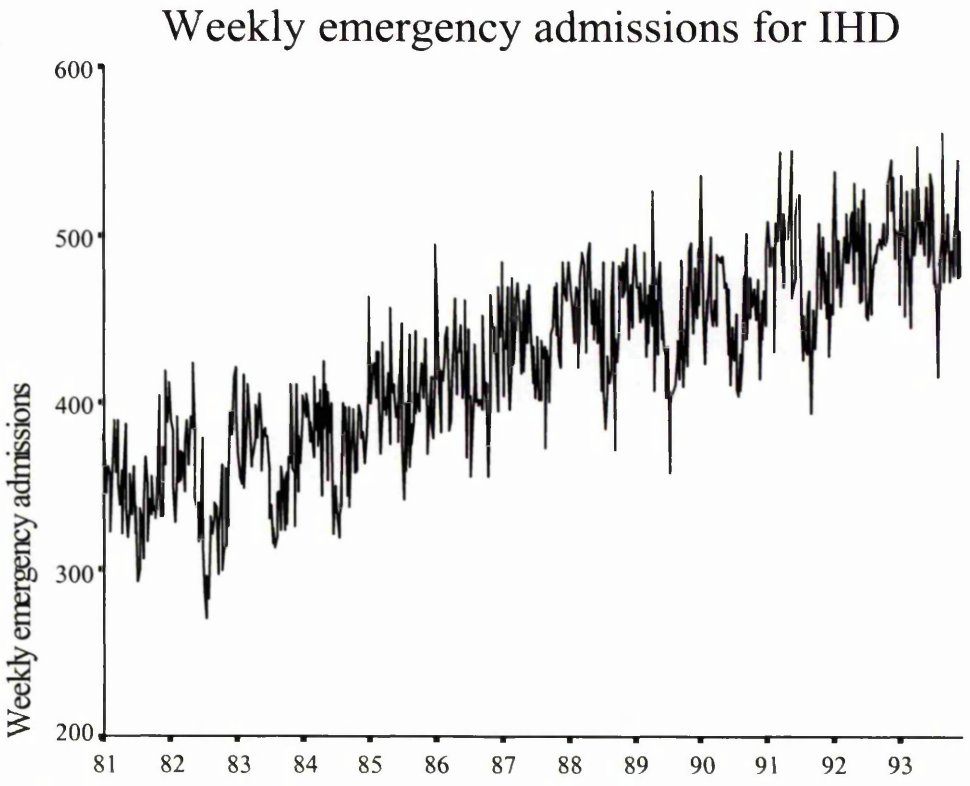
The seasonal increase in emergency admissions for respiratory disease is 49% for males and 64% for females. This is around half the size of the seasonal increase in mortality from respiratory disease. These figures increase for older age groups and for the very

young age group but do not vary according to socio-economic deprivation categories. The seasonal increase in emergency admissions for respiratory disease varies significantly between health board areas. For both males and females the increase is significantly higher in the Borders and in Dumfries & Galloway, both rural areas, but it is considerably smaller in the Island health boards. Overall there is no distinct spatial pattern.

2.4.2 Ischaemic Heart Disease

As with emergency hospital admissions for respiratory disease, there was a strong increasing trend in emergency admissions for ischaemic heart disease and the trend dwarfed any seasonal pattern when the data were inspected graphically (figure 2.22).

Figure 2.22 Weekly emergency admissions for IHD, Scotland 1981-1993



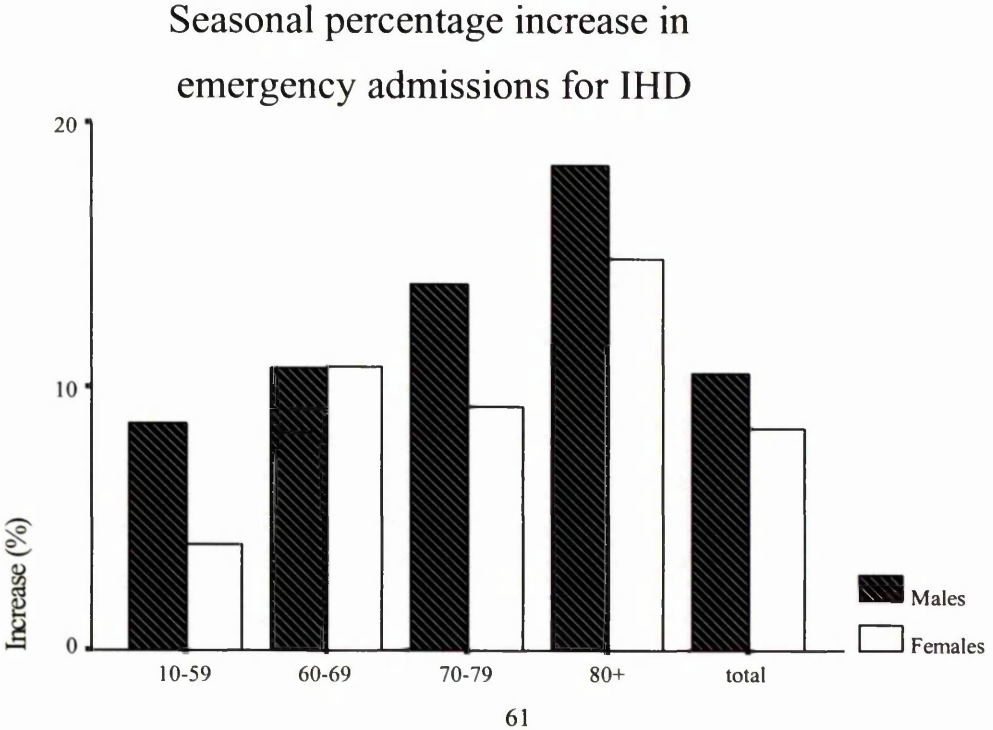
In week 1 of 1981 the number of admissions was 247 for males and 131 for females, by week 1 of 1993 this number had risen to 625 for males and 333 for females. The total number of emergency admissions for IHD over the time period was 287 763.

Using the model which accounted for the trend, the percentage increase in emergency admissions for IHD from trough to peak was 10%(9.5,11.5) for males and 8%(7.2,9.6) for females. Again, as with the deaths data, the degree of seasonal variation increased with age group.

Table 2.18 Seasonal percentage increase in emergency hospital admissions by sex & age group, IHD

Age group	Males		Females	
	Percent Increase	CI	Percent Increase	CI
0-9	-	-	-	-
10-59	8.6	(7.0,10.2)	4.0	(1.4,6.5)
60-69	10.6	(8.8,12.4)	10.7	(8.4,13.0)
70-79	13.8	(11.7,15.9)	9.2	(7.1,11.3)
80+	18.4	(14.6,22.2)	14.8	(12.0,17.6)
All Ages	10.5	(9.5,11.5)	8.4	(7.2,9.6)

Figure 2.23 Seasonal percentage increase in emergency hospital admissions by sex & age group, IHD



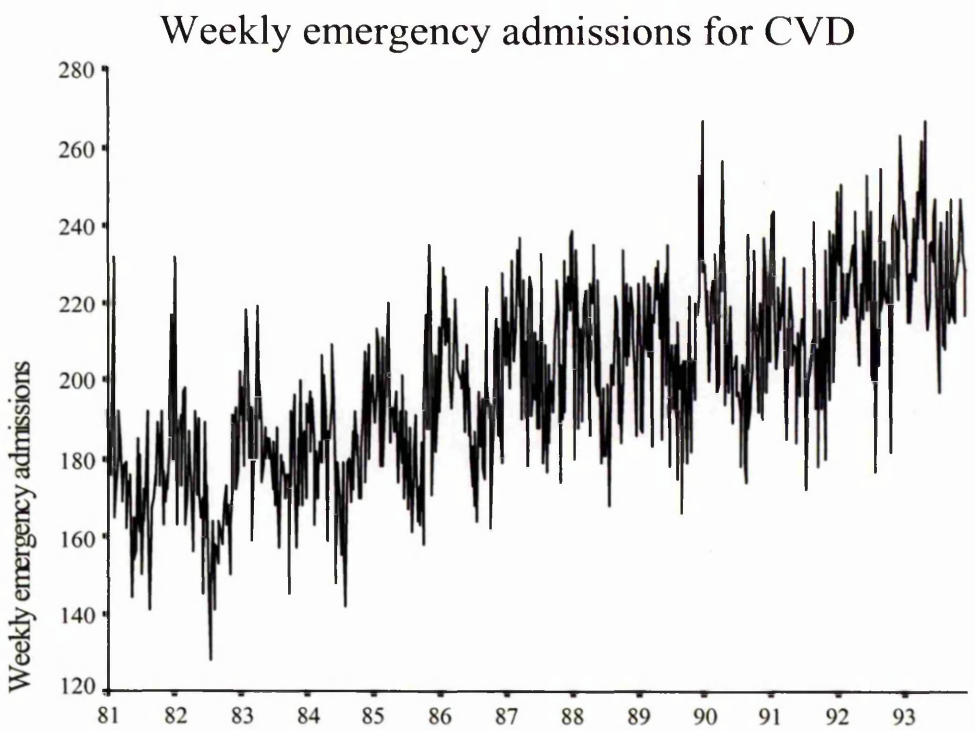
2.4.2 (a) Summary of seasonal analysis for ischaemic heart disease

The seasonal increase in emergency admissions for ischaemic heart disease is 10% for males and 8% for females. These figures are less than a third of the size of the seasonal increase in mortality from IHD. The size of the seasonal increase is greater at older age groups but there is little variation according to deprivation categories. Considerable variation exists among health boards but this is not significant.

2.4.3 Cerebrovascular Disease

The total number of emergency admissions for cerebrovascular disease was just over 135 000. Again, admissions for this diagnosis showed an increase over the time period. In the first week of 1981 the number of admissions for CVD was 102 for males and 101 for females, by 1993 these figures had increased to 131 and 119 a 20% increase (figure 2.24).

Figure 2.24 Weekly emergency admissions for CVD, Scotland 1981-1993



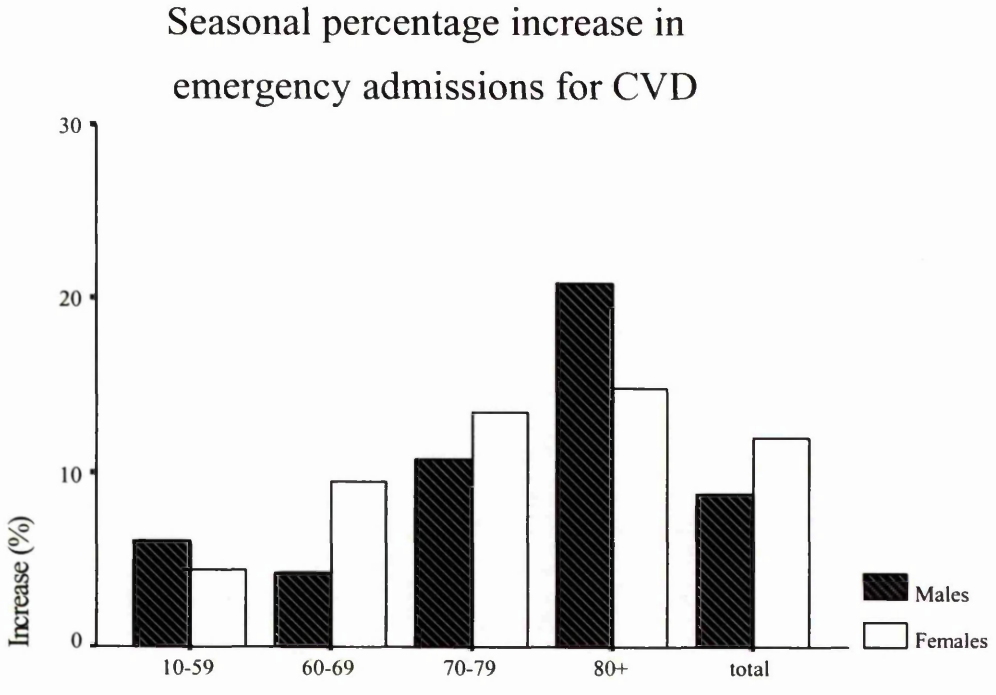
Although the increase over time for emergency admissions for CVD was not as great as for IHD and respiratory disease it was still the most dominant feature in the plot of the

data, although the seasonal pattern was still evident. The simple sinusoidal model gave a percentage increase from trough to peak of 9%(7.1,10.4) for males and 12%(10.4,13.5) for females. Again, this increase varied by age group and the general pattern was an increase in seasonality with increasing age (table 2.24 & figure 2.25). This pattern was not as clear as for IHD and respiratory disease.

Table 2.24 Seasonal percentage increase in emergency hospital admissions by sex & age group, CVD

Age group	Males		Females	
	Percent Increase	CI	Percent Increase	CI
0-9	-	-	-	-
10-59	6.1	(2.2,9.9)	4.4	(0.1,8.8)
60-69	4.2	(1.0,7.4)	9.5	(5.8,13.2)
70-79	10.7	(7.9,13.4)	13.4	(10.9,16.0)
80+	20.8	(16.9,24.8)	14.8	(12.2,17.4)
All Ages	8.7	(7.1,10.4)	12.0	(10.4,13.5)

Figure 2.25 Seasonal percentage increase in emergency hospital admissions by sex & age group, CVD



2.4.3 (a) Summary of seasonal analysis for cerebrovascular disease

As with emergency admissions for IHD, the seasonal increase in emergency admissions for CVD is around a third of that experienced for deaths from CVD. There is an increase in the degree of seasonal variation at older age groups but no association with deprivation categories or with health board of residence.

2.4.4 Cancer

Cancer admissions were included in this analysis simply for completeness. They were included in the analysis of mortality because, although they do not show much seasonal variation, they are an important cause of death. While emergency admissions provided a reasonable estimate of the incidence of CVD and IHD and an estimate of the incidence of more serious respiratory events there were relatively few emergency admissions for cancer due to the nature of the disease. Cancer accounts for 25% of all deaths but only about 5% of all emergency admissions.

As with deaths from cancer, emergency admissions for cancer did not show a great deal of seasonal variation. The percentage increase from trough to peak was 2.1%(0.9,3.2) for males and 3.5% (2.3,4.7) for females. The size of the increase for all emergency admissions for cancer is barely significant. Since emergency admissions for cancer account for only 5% of all admissions the analysis of seasonal variation of emergency admissions for cancer was limited to all ages and not considered by age group, deprivation category or health board of residence.

2.5 The Effect of Influenza

Influenza epidemics have been shown to affect the seasonal pattern of deaths from causes of death not recorded as influenza[4; 5; 8; 10]. There were a total of 1524 deaths from influenza (ICD-9 487) over the 13-year time period, 501 male deaths and 1023 female deaths. The seasonal pattern in flu deaths is clearly evident from a plot of the deaths from flu over time (figure 2.26). Virtually all deaths from flu occur in the winter. There were 3081 emergency admissions for influenza, 1451 male admissions and 1630 female admissions. A plot of emergency admissions for flu displays a very similar pattern to that for deaths from flu (figure 2.27). There were several flu epidemics apparent from the plots of the data, these occurred in winters 81/82, 82/83, 85/86, 89/90 and 93/94. During

the most prominent epidemic in the winter of 1989/90 there were 150 deaths recorded as flu deaths and around 90 emergency admissions for flu in Scotland. In each of the smaller epidemics there were between 20 and 30 emergency admissions and deaths per week at the peak of the epidemic. The emergency admission data shows that in the years in which the epidemic was less strong there was still a significant seasonal pattern in emergency admissions for respiratory disease.

The effect of these flu epidemics is reflected in the pattern of deaths from all causes and in deaths and emergency admissions from IHD, CVD and respiratory disease. There was evidence of an increase in the number of deaths and emergency admissions from these causes in the years that there was a flu epidemic. The effect of flu epidemics could be that the influenza virus acts as a trigger and although the person died from a different underlying cause of death, contracting influenza may have been a contributory factor to the patient's subsequent death from the underlying cause. The presence of flu epidemics that occur in winter increases the percentage increase in deaths from other causes from summer to winter. The extent of this effect however is difficult to measure and will be discussed in more detail in chapter 3.

Figure 2.26 Weekly deaths from influenza, Scotland 1981-1993

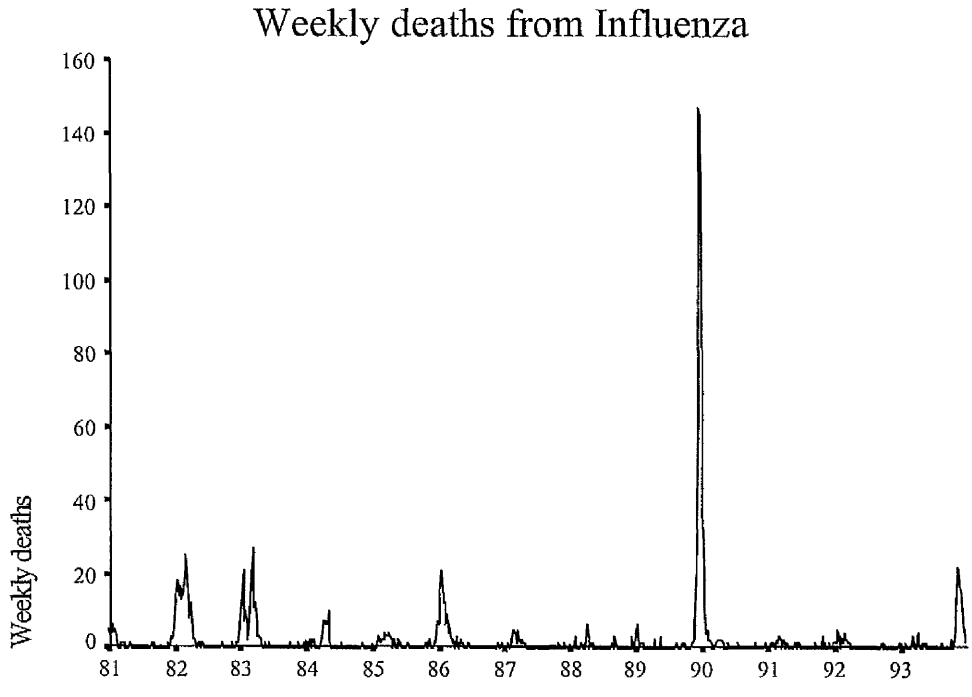
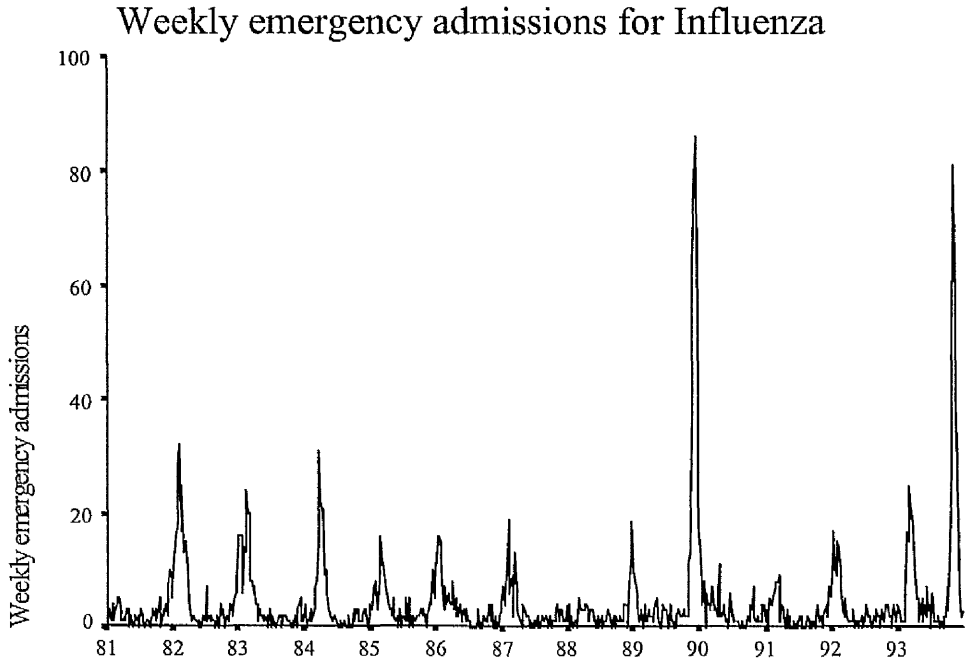


Figure 2.27 Weekly emergency admissions for influenza, Scotland 1981-1993



2.6 Conclusions & Discussion

2.6.1 Summary of results

There is seasonal pattern in deaths from all causes for both males and females. The extent of this seasonality results in a 28% increase from summer to winter in male mortality and a 34% increase in female mortality. The seasonal increase was greatest in deaths from respiratory disease with an 89% increase from summer to winter for men and a 114% increase for women. Deaths from CVD and IHD showed a 35% increase from trough to peak for both sexes, while cancer deaths showed very little seasonal pattern. The size of the seasonal increase in mortality increased with increasing age group and particularly for deaths from respiratory disease was greater in the very young age group. Several other studies have demonstrated an increase in seasonal mortality in the older age groups[40,4,3,15]. There was little difference in the degree of seasonal variation experienced when the data were grouped by deprivation category, this finding is supported by a recent study in London[78] which also showed no evidence of an effect of deprivation on excess winter mortality. Analysis by social class appeared to indicate that people in social classes I & II experienced a lower seasonal increase in mortality from IHD and CVD than people in social classes IV & V. The amount of seasonal variation

experienced by the population varied according to health board of residence. Generally those people living in the Island health boards experienced less seasonal variation than those resident in other health boards.

The analysis of seasonal patterns in mortality in Scotland from 1981 to 1993 has shown that seasonal variation in Scotland was considerable (around 30%) and that this figure was greater for the more vulnerable members of the population such as the elderly, the very young and those in lower social classes.

Seasonal patterns in emergency admissions were smaller than those for deaths. The seasonal increase in emergency admissions for respiratory disease was 49% for males and 64% for females. Emergency admissions from IHD and CVD showed a seasonal increase of between 8% and 12%, whereas emergency admissions for cancer showed a seasonal increase of around 3%. As with the mortality data, seasonal variation in emergency hospital admissions was greater in the older age groups and in the very young. There was no difference in the degree of seasonal variation experienced by the population when areas were grouped according to socio-economic deprivation scores. Social class data was not available for the SMR1 data and the analysis by health board did not produce any clear spatial pattern. The issue of how the seasonal pattern in emergency hospital admissions varies spatially is described in more depth in Chapter 5.

2.6.2 Implications of results

The results described in this chapter further strengthen findings from other work on the seasonality of disease in Scotland. Using Q1/Q3 ratios for the period 1977-83 GRO(S)[3] showed that variation in the degree of seasonal increase in mortality existed by cause of deaths, age group, and social class. They also commented that with a Q1/Q3 ratio of around 30% the seasonal increase in mortality in Scotland was greater than in most other European countries. The work described in this chapter has demonstrated that a seasonal percentage increase in mortality of around 30% still exists in Scotland and that it is greater in older age groups, lower social classes and in females. Other work from the Netherlands[25] indicates a seasonal variation (percentage increase from trough to peak) of around 22% implying that at 30%, Scotland is still experiencing greater seasonal increases than some other European countries.

The seasonal increases in mortality are of concern as it is clear from the demographic analysis that it is the most vulnerable and the poorest members of society who

experience the most severe effects of winter. The increase in emergency admissions during winter puts more demand on the NHS, a system that is running at capacity throughout the whole year. Almost every year in the UK there are media reports of the sudden increase in demand for hospital beds in winter and the strain on the health service that this causes. The work in this chapter, combined with other earlier work carried out by the GRO in Scotland, indicates that this situation is long standing and should perhaps be tackled in a preventative manner. If resources were supplied to those who may be more susceptible to ill health in the winter it may help to reduce the size of the increased demand on the NHS in winter and reduce the high number of deaths which occur each winter in Scotland.

Chapter 3 - The effect of Climate on Mortality and Morbidity

3.1 Introduction & Aims

In Chapter 2 the seasonal pattern in deaths and emergency hospital admissions was described. Generally when the word 'seasonal' is used it refers to weather patterns at a particular time of year i.e. winter, spring, summer, and autumn. The fact that mortality follows a seasonal pattern may lead to the hypothesis that there is an association between climate and mortality, this supposition being strengthened by an established experience of ill health in winter. Several authors[23; 44; 45] have used correlation methods to assess the relationship between climate and mortality. In this analysis, however methods which establish a causal relationship will be developed. Poisson regression methods[30; 49] and time series methods[31] have also been used to assess the strength of the relationship between climate and mortality. The work described in this chapter builds on these results and assesses the merits of a method that combines both time series and Poisson regression methods.

In this chapter, Meteorological Office data from January 1981 to December 1993 were analysed in conjunction with the mortality data and SMR1 data to investigate the existence of a causal relationship between climate, mortality and morbidity. Section 3.2 describes the data obtained from the Meteorological Office in Edinburgh, while the development of various methods is described in section 3.3. The methodology used in section 3.3 includes simple statistics, time series methods, Poisson regression and finally a method that is a combination of both time series methods and Poisson regression. A comparison of the methods is given at the end of the section. The majority of the development work used daily data as opposed to the weekly data used in Chapter 2. The methods were developed using mortality from respiratory disease and applied to mortality from all causes, ischaemic heart disease and cerebrovascular disease; the results are given in section 3.4. Section 3.5 describes the relationship between emergency hospital admissions and climate and a summary of the chapter is given in section 3.6.

3.2 Scotland's Climate

3.2.1 Data

Climate data for Scotland from 1st January 1981 to 31st December 1993 were obtained from the Meteorological Office in Edinburgh. The data consisted of daily maximum and daily minimum temperatures, average daily wind speed and total daily rainfall for three weather stations located at Aberdeen Airport (Dyce), Edinburgh Airport (Turnhouse) and Glasgow Airport (Abbotsinch). From this data an average daily temperature for Scotland was computed by calculating the average daily temperature at each station as $(max + min)/2$ and then computing a Scottish average $(Glas + Edin + Aber)/3$. The Scottish average daily wind speed and total daily rainfall were also calculated as above.
The average daily temperature in Scotland over the time period 1981 to 1993 was 8.5°C, the daily maximum temperature was on average 12.17°C and the minimum was on average 4.77°C. The maximum reached on any day over the time period was 28.07°C and the minimum temperature reached on any day was -16.5°C. On average there was 2.5mm of rain per day and the wind speed was 8.7 knots.

3.2.1 (a) Temperature

The average daily temperature over the time period was 8.69°C in Glasgow, 8.58°C in Edinburgh and 8.14°C in Aberdeen. Glasgow experienced a milder climate than Edinburgh, which in turn experienced a milder climate than Aberdeen, however, these differences were slight. Temperature summaries for each of the three cities are given in Table 3.1

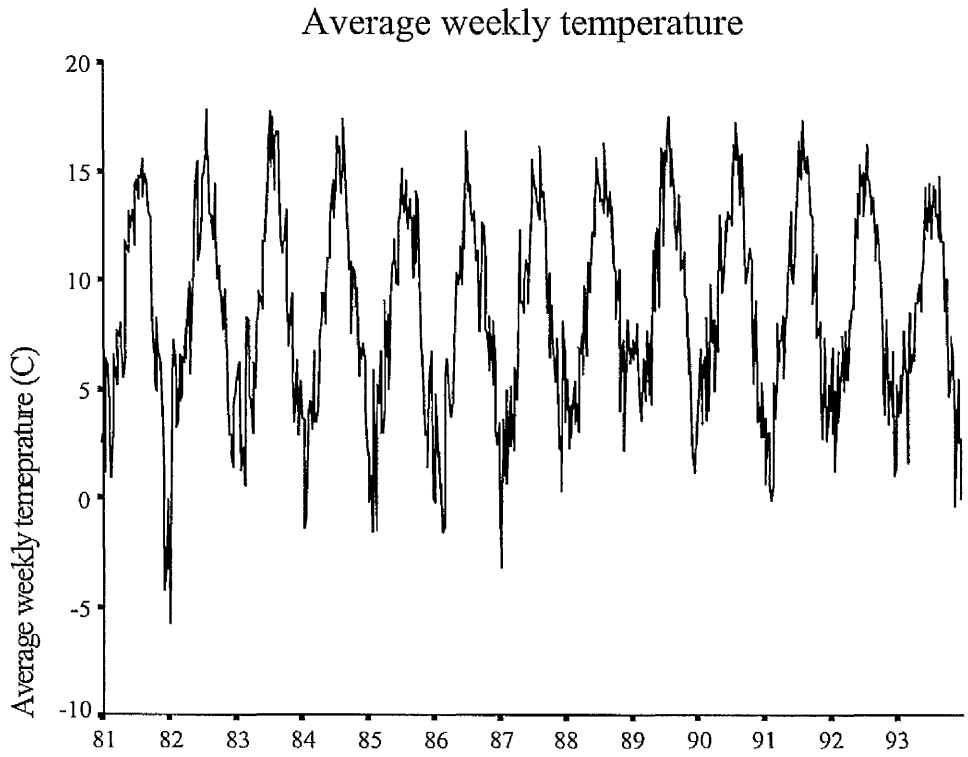
Table 3.1 Summaries of daily temperature(°C) in Glasgow, Edinburgh & Aberdeen

	Average	Average maximum	Average minimum	5%ile of average	95%ile of average
Glasgow	8.69	12.46	4.91	0.35	16.45
Edinburgh	8.58	12.36	4.80	0.60	16.35
Aberdeen	8.14	11.70	4.59	0.70	15.75

There were five occasions when the temperature in Aberdeen was less than or equal to 0°C for 7 consecutive days, there were three such occasions in Edinburgh and four in Glasgow. For all days when temperature was 0°C or below, the average temperature was around -2.2°C. The average temperature over the three sites (Scottish average) was less than or equal to 0°C for 7 consecutive days on three occasions. The temperature was below or equal to zero on 3.5% of days in Aberdeen, 3.9% of days in Edinburgh and 4.3% of days in Glasgow.

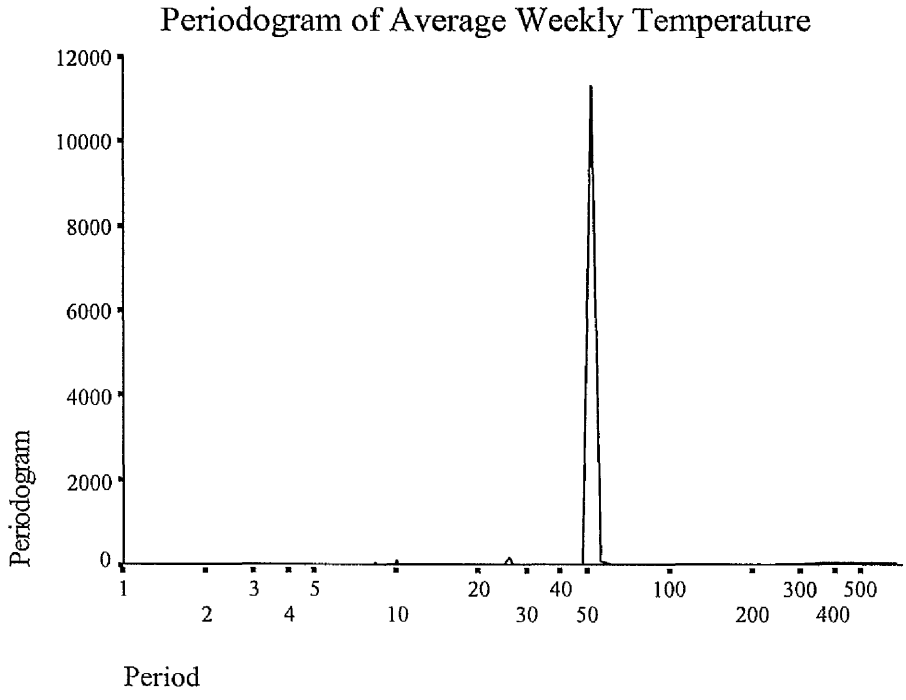
So although average temperature, over the time period, was higher in Glasgow than in the other two sites there were more days in Glasgow when the temperature was below or equal to zero. However, in Aberdeen the temperature was below or equal to 0°C for 7 consecutive days more often than in Glasgow. Overall Aberdeen was colder for longer periods while Glasgow's cold spells were less cold, more frequent and for shorter periods of time. Edinburgh fell somewhere in the middle of Glasgow and Aberdeen. A plot of daily temperature over 13 years contained a lot of noise therefore to demonstrate the seasonal pattern effectively average weekly temperature is used in Figure 3.1.

Figure 3.1 Average weekly temperature(°C) in Scotland from 1981 to 1993.



The seasonal pattern in the temperature data is clear from figure 3.1 and when these data were examined using spectral methods (as described in section 2.2.2) the periodogram identified the presence of a yearly cycle (figure 3.2)

Figure 3.2 Periodogram of average weekly temperature



3.2.1 (b) Wind Speed

Table 3.2 displays the average wind speed in the three locations and gives the 5% and 95% percentiles of the distribution. Aberdeen experienced the highest average wind speed and Edinburgh the lowest. A plot of the Scottish average wind speed over the time period (figure 3.3) did not clearly identify a seasonal pattern, however spectral analysis determined that a yearly pattern did exist. (figure 3.4).

Table 3.2 Summaries of daily wind speed (knots) in Glasgow, Edinburgh & Aberdeen

	Average	5%ile	95%ile
Glasgow	8.70	2.60	17.30
Edinburgh	8.58	2.70	16.80
Aberdeen	8.85	3.30	16.60

Figure 3.3 Average weekly wind speed (knots) Scotland 1981-1993

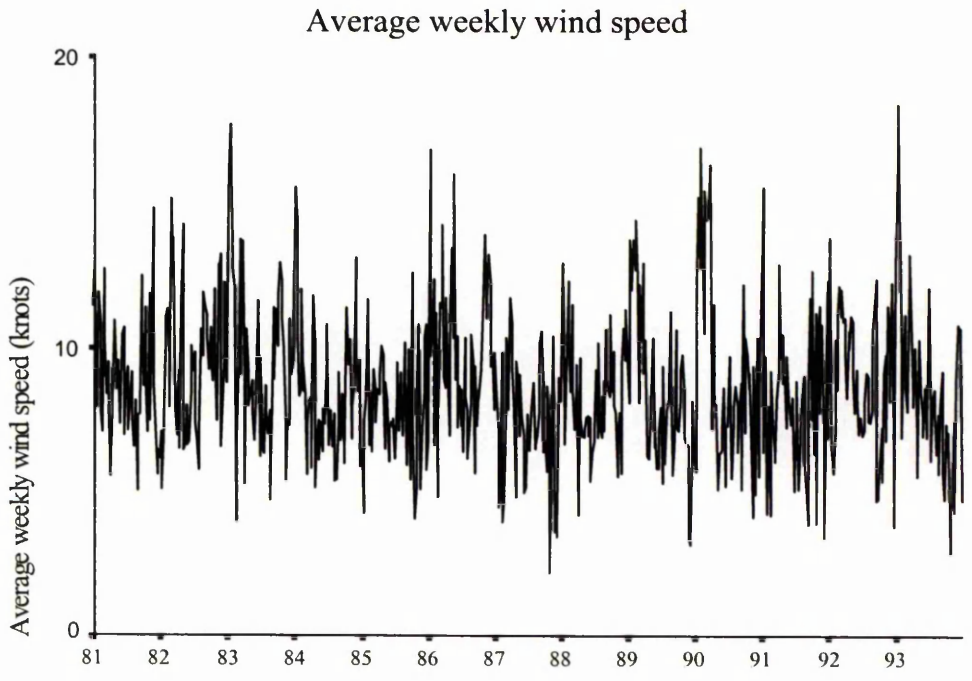
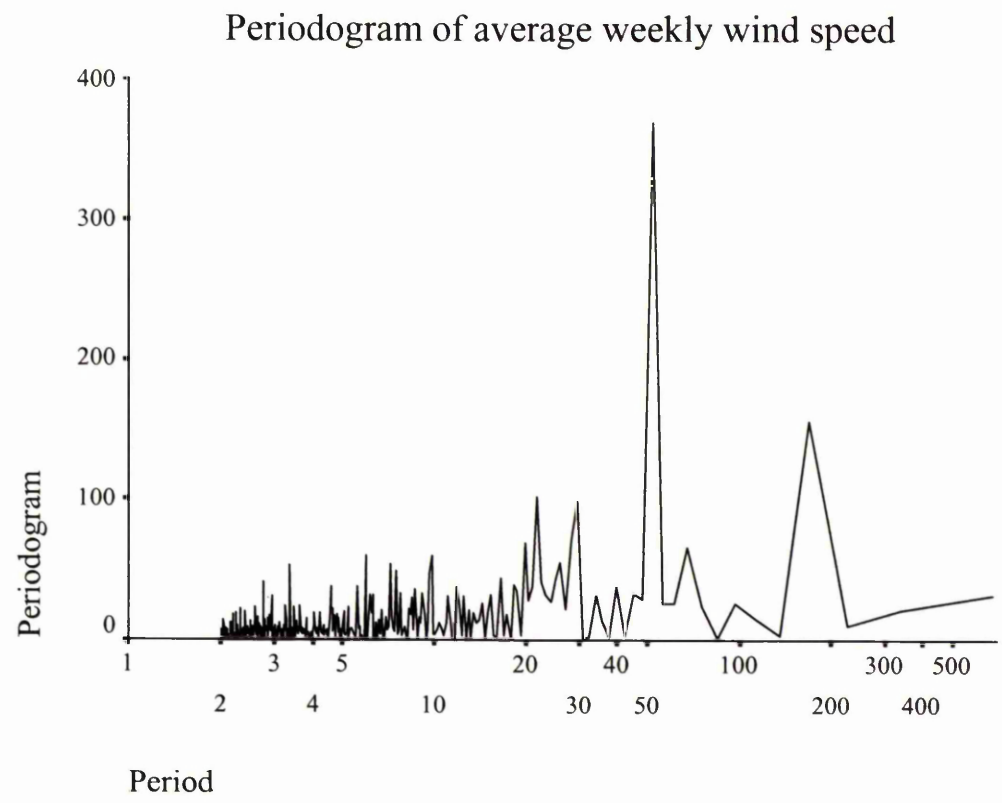


Figure 3.4 Periodogram of average weekly wind speed



3.2.1 (c) Rainfall

Glasgow experienced much more rain than either Aberdeen or Edinburgh. The average daily rainfall in each site with the 5% and 95% percentiles is shown in Table 3.3.

Table 3.3 Summaries of daily rainfall (mm) in Glasgow, Edinburgh & Aberdeen

	Average	5%ile	95%ile
Glasgow	3.14	0.00	14.40
Edinburgh	2.00	0.00	10.00
Aberdeen	2.18	0.00	10.85

As with the plot of wind speed a plot of the average Scottish rainfall did not display a clear seasonal pattern, however a closer inspection of figure 3.5 indicates that there is increased rainfall in autumn. Again, a spectral analysis of the data confirmed the presence of a yearly cycle (figure 3.6).

Figure 3.5 Average weekly rainfall (mm) Scotland 1981-1993

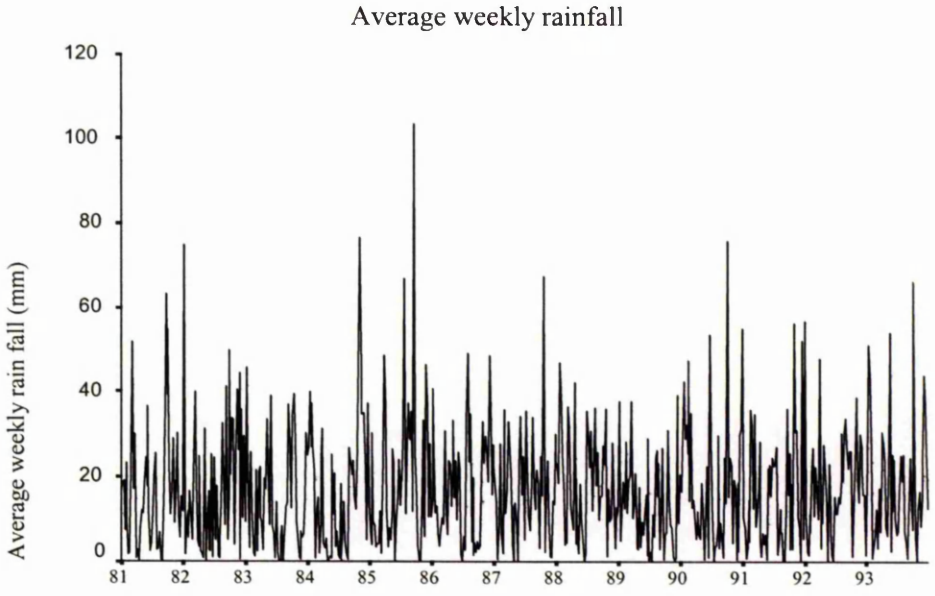
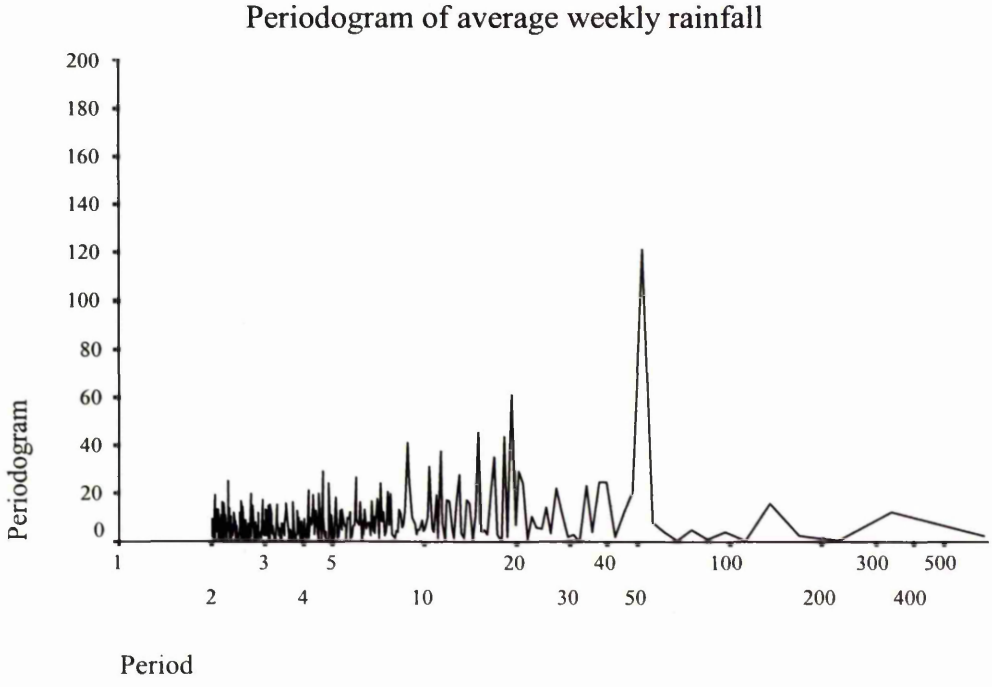


Figure 3.6 Periodogram of average weekly rainfall



3.2.2. Seasonal modelling of variables

All three of the weather variables displayed a yearly seasonal cycle, and so a simple sinusoidal curve of the form $y = (a + bt)(1 + \lambda \cos(\omega t + p))$ was fitted to the data. As with the models described in section 2.2.4 this model expressed the amplitude of the seasonal curve as the percentage increase from the mean, the parameter p determines the time of the year at which the curve peaks and the model adjusts for the trend in the data. Using $\lambda^* = \frac{2\lambda}{1-|\lambda|}$ it is possible to express λ^* as the percentage increase from the trough to the peak of the seasonal curve. The results from applying this modelling procedure to the climate data were that the percentage increase from the trough to the peak of the seasonal curve for average temperature was 431%, for wind speed this was 27% and for rainfall this was 66%. These three fitted curves did not peak at the same time of the year, from the values of p it was computed that the temperature reached its peak around mid July, wind speeds reached their peak in late January and rainfall was greatest around mid November.

There was considerable unexplained variation in the sinusoidal models for wind speed and rainfall. The model explained only 3.8% of the variation in wind speed and for rainfall this was 1.2%. For average temperature the simple sinusoidal curve explained over 70% of the variation. These results are summarised in table 3.4.

Table 3.4 Results of initial sinusoidal models

	λ	Seasonal increase (%)	Peak day in year	Variation explained (%)
Average temperature	-0.68	431	203	70.7
Wind speed	0.12	27	28	3.8
Rainfall	0.25	66	320	1.2

There was a decreasing trend in size of the seasonal variation in temperature even though the daily average temperature showed an increasing trend. Having now determined that the climate data, in particular daily average temperature, and the mortality and morbidity data follow a yearly seasonal pattern an investigation into the degree of association between climate, mortality and morbidity was undertaken.

3.3 Development of methods

3.3.1 Introduction

Computation of a simple correlation coefficient between two variables which both follow a yearly seasonal pattern would result in a correlation coefficient which was statistically significant. The level of significance could then be used to infer that the two series are associated with each other when in fact this may not be the case. A paper by Campbell[58] refers to Yule who found a correlation of 0.95 between the standardised annual mortality rate and the yearly proportion of marriages celebrated in the Church of England for the years 1866 to 1911. He pointed out it would be incorrect to infer causality from this high correlation.

In order to determine an association between two seasonal time series the seasonal components present in the data have to be removed. There is also the problem in these data of serial auto-correlation. Auto-correlation refers to the fact that observations close in time to each other are similar in value i.e. the temperature today is likely to be closely related to the temperature yesterday. With deaths or hospital admissions the autocorrelation in the data is more difficult to comprehend, while one death does not necessarily cause another death there are time periods where many deaths from one cause occur and periods where very few deaths occur. Because temperatures are influenced by global weather systems it is easy to believe that today's temperature has a direct bearing on tomorrow's temperature. For deaths and hospital admissions it is likely that the relationship between today's death rate and tomorrow's death rate involves some other processes which display serial autocorrelation and which affect the daily number of deaths.

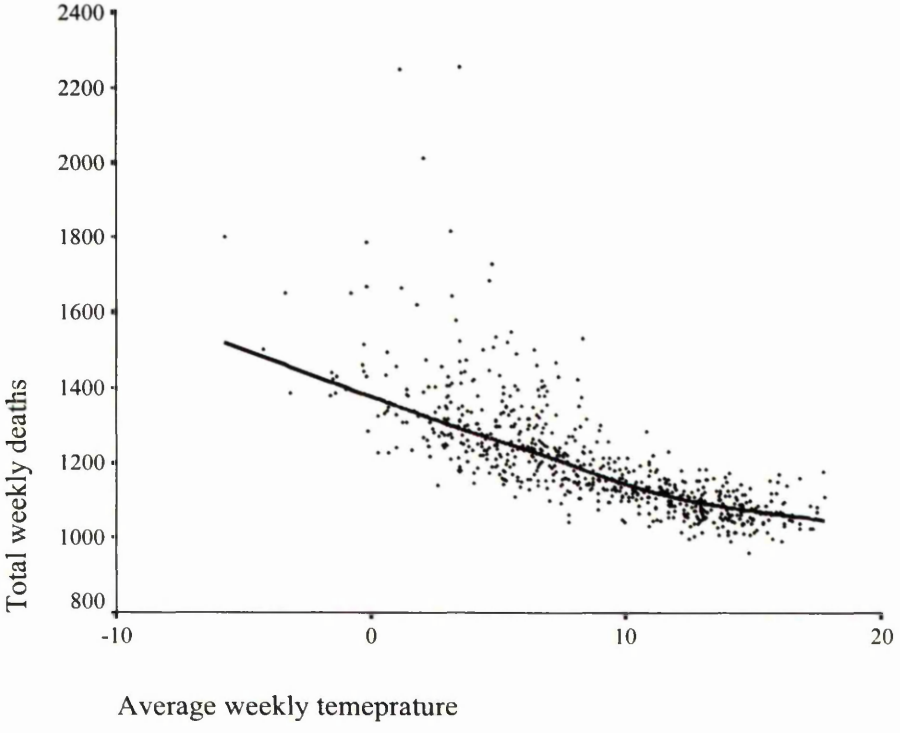
3.3.2 Summary statistics

Results from Chapter 2 indicated that mortality from respiratory disease, ischaemic heart disease and cerebrovascular disease displayed the greatest amount of seasonal variation. The more detailed analysis described in this chapter concentrates on these three causes of death, as it is likely that these three causes will display the strongest relationship with climate. Scatter plots and simple summary statistics are used in this section with two objectives, firstly to summarise the data and secondly to highlight the ways in which the relationship between seasonally autocorrelated time series could be misinterpreted if inappropriate statistical methods were used. In section 3.3.2(a) the weekly number of deaths and hospital admissions are plotted against average weekly temperature to give an overall sense of what the data look like. The other forms of analysis concentrate on the mortality data and three methods of summarising the relationship between temepratrue and mortality analysis are described. However, these summary methods summary are found to be innapropriate for accurately describing the relationships and more appropriate time series methods are introduced in section 3.3.3.

3.3.2 (a) Scatter plots of weekly deaths and admissions against temperature

The relationship between mortality and temperature was assessed visually before considering any more detailed summary statistics. Figure 3.7 shows a scatter plot of weekly deaths and average weekly temperature

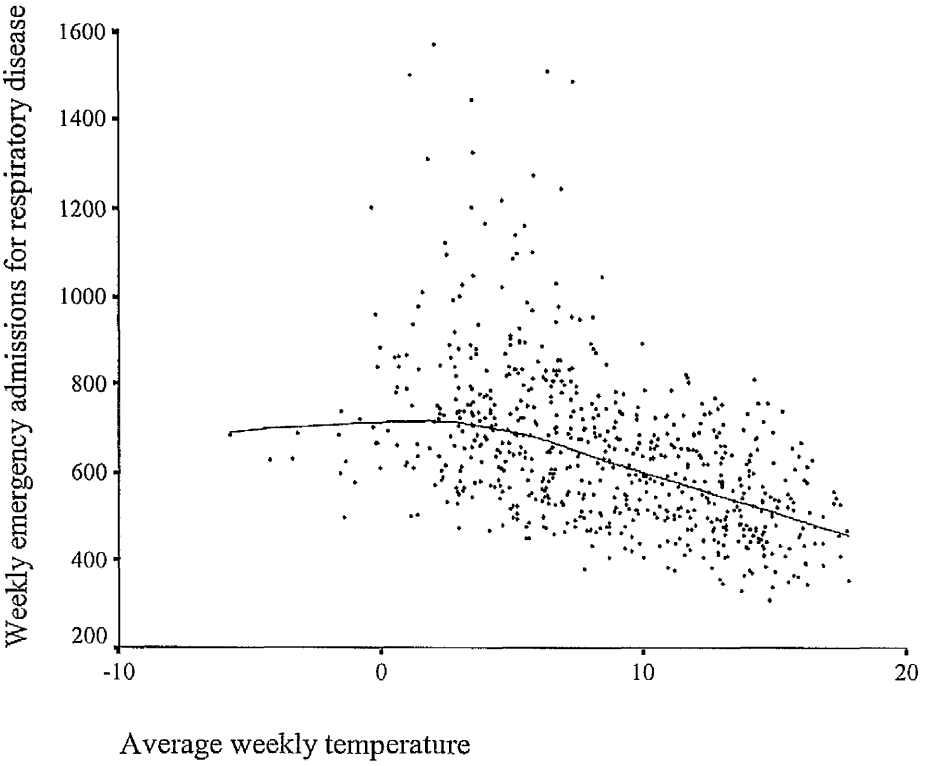
Figure 3.7 Scatter plot of weekly deaths and average weekly temperature



The plot shows increased weekly mortality at lower temperatures and a curvilinear relationship between mortality and temperature. The loess curve levels out as average temperature increases. Several outliers can be seen and these relate to the flu epidemic of 1989/90 where there was a greater than usual winter increase in deaths. This chart also shows a greater spread in the numbers of deaths when the temperature is lower compared to when the temperature is between 10°C and 20°C.

It was demonstrated in the exploratory plots of emergency admissions described in section 2.1.4 that there was little seasonal pattern in emergency admissions for all diagnostic groups. A similar plot for weekly emergency admissions for respiratory disease and average weekly temperature is shown in figure 3.8.

Figure 3.8 Scatter plot of weekly emergency admissions and weekly temperature



The scatter plot of emergency admissions for respiratory disease against temperature demonstrates an almost linear relationship for temperatures above 3°C and virtually no relationship at temperatures <3°C. Again the scatter is greater at temperatures below 10°C compared to at temperatures above 10°C and there are several outliers. The results from the two plots above combined with the evidence of a smaller seasonal variations in emergency admissions would lead us to believe that the relationship between temperature and emergency admissions may not be as strong as that for temeprature and mortality. Further development of summary statistics is based on mortality data only.

3.3.2 (b) Correlation's using Q1/Q3 ratios

Before the complex relationship between deaths or emergency hospital admissions and temperature was analysed some simple relationships were explored. In Chapter 1 the Q1/Q3 ratio was defined as the ratio between the number of deaths occurring in the first quarter of the year and the number of deaths occurring in the third quarter of the year.

This gives a crude measure of the degree of seasonal variation present in the data. It was felt that these values might be of use in an initial analysis of the relationship between mortality and temperature. From a Q1/Q3 ratio a rough estimate of the degree of seasonal variation present for any year can be determined. Q1/Q3 ratios were used to establish whether years with large temperature swings were associated with years that displayed large seasonal variations in mortality. Assuming a constant summer temperature over the years, then in a cold winter the ratio Q1/Q3 would be lower than when a mild winter occurred. In the case of deaths, assuming a constant summer death rate over the years, then in a winter where many deaths occurred the Q1/Q3 ratio would be greater than in a winter where fewer deaths occurred. Therefore a relationship between increased seasonal variation in temperature and increased seasonal swing in deaths could be shown by a negative correlation between the respective Q1/Q3 ratios. The correlation between the yearly Q1/Q3 ratios for deaths from different causes and the corresponding yearly Q1/Q3 ratios for temperature is shown in Table 3.5.

Table 3.5 Correlation (r) between Q1/Q3 ratios for mortality and temperature

	Males		Females	
	r	p-value	r	p-value
All Causes	-0.58	0.037	-0.59	0.035
Respiratory disease	-0.41	0.166	-0.46	0.116
Ischaemic heart disease	-0.67	0.012	-0.52	0.068
Cerebrovascular disease	-0.55	0.051	-0.61	0.027

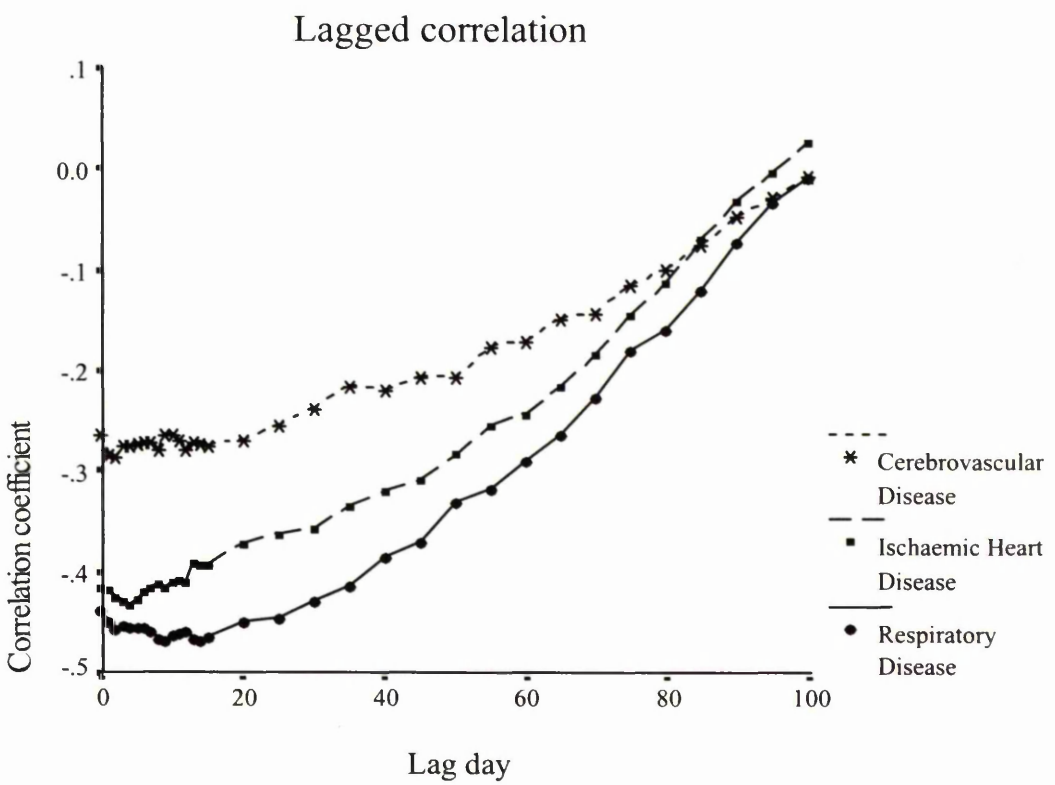
As can be seen from the table above the correlation is negative in each case and is significant for deaths from all causes, ischaemic heart disease and cerebrovascular disease but not for deaths from respiratory disease.

These results indicate that in years when a large seasonal change in temperature occurred a large seasonal change in the number of deaths also occurred. This may indicate the presence of a relationship between temperature and deaths. Using Q1/Q3 ratios tells us about the effect of seasonal swings in temperature on mortality, however a more direct measure of the effect of climate on deaths can be obtained by considering daily lagged correlations.

3.3.2 (c) Lagged correlations

As described previously a simple correlation between two seasonal data series that displayed auto-correlation will give a statistically significant result. However, lagged daily correlations between temperature, wind speed, rainfall and mortality provided a simple method of summarising the relationship between mortality and the three climatic variables before attempting more detailed appropriate methods. Spearman's correlation coefficient was used to estimate the correlation between deaths and the weather variables at lag days. From this simple measure the effect of the three climate variables on the three causes of death can be compared. Figure 3.9 shows the lagged correlation of temperature and the three causes of death while Figure 3.11 shows the lagged correlation between deaths from all causes and temperature, wind speed and rainfall.

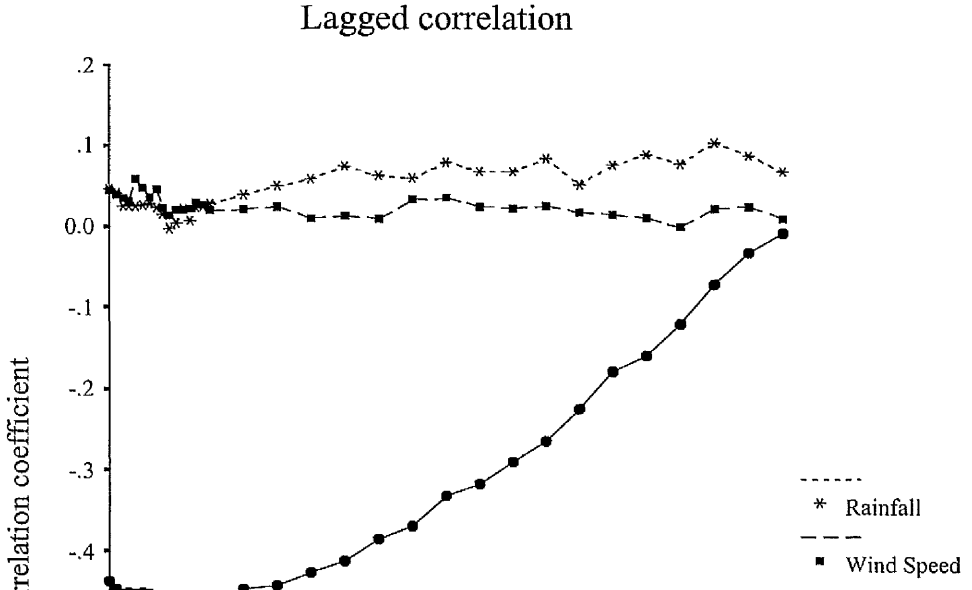
Figure 3.9 Lagged correlation between temperature and deaths



It can be seen from Figure 3.9 that the relationship between mortality and temperature is stronger for deaths from respiratory disease than for deaths from IHD and CVD and that the relationship is strongest for each cause at a lag of a few days. For deaths from CVD the relationship with temperature is strongest at a lag of 3 days, for deaths from IHD this

lag period was around 5 days and for deaths from respiratory disease the relationship between the two variables is greatest up to 15 days apart. This lagged relationship can be explained by the fact that a change in temperature may not have an immediate effect on mortality. For example, if the temperature suddenly fell it would take 3 days to notice an increase in deaths from CVD, 5 days before the greatest increase in deaths from IHD was observed and up to 15 days for the number of associated deaths from respiratory disease to peak.

Figure 3.10 Lagged correlation between deaths from all causes and climate variables



From Figure 3.10 it is clear that the climatic variable which is most closely associated with mortality is temperature. The correlation between wind speed and rainfall and deaths was around 0.05 whereas the correlation with temperature was around –0.4. At this stage of the analysis it was decided to concentrate on the relationship between mortality and morbidity and temperature only.

3.3.2 (d) Linear Regression of deaths and temperature

A regression model that related the number of deaths per day to the average daily temperature may provide a clearer measure of the relationship between the two variables. A simple linear regression of the form $y_t = \alpha + \beta(temp_t) + \varepsilon_t$ was fitted to the data. The results of this analysis are shown in Table 3.6.

Table 3.6 Results from simple linear regression of deaths and temperature

	Males			Females		
	coeff	p	R square	coeff	p	R square
Resp	-0.40	0.00	18.46	-0.47	0.00	16.98
IHD	-0.51	0.00	18.07	-0.48	0.00	17.92
CVD	-0.18	0.00	7.70	-0.29	0.00	11.10
All cause	-1.34	0.00	26.15	-1.69	0.00	28.84

The coefficients β are a measure of the degree of association between daily temperature and daily mortality. The regression coefficients are highly significant, indicating a strong relationship but the R-square values indicate that the models are not a good fit of the data.

A simple interpretation of the results would lead us to believe that a drop in temperature of 1°C in a day is related to an increase in male mortality from all causes of 1.4 deaths per day. The simple linear regression model depends on three assumptions; constant variance, independence and normality. The assumption of constant variance means that for any temperature the population of the potential values of y_t, the number of deaths, has a constant variance which does not depend on the value of the temperature. The independence assumption assumes that any one value of the dependent variable y_t is statistically independent of any other value of y_t. The final assumption used in simple linear regression is that for any value of temperature, the corresponding population of values of the dependent variable y_t has a normal distribution. Residual analysis and exploratory methods such as probability plots can test the validity of the assumptions.

Traditional methods of examining the residuals from a fitted model involve plotting the residuals ε_t against the fitted values $\hat{y}_t$, as in Figure 3.9.

Figure 3.11 Scatter plot of residuals from the linear model

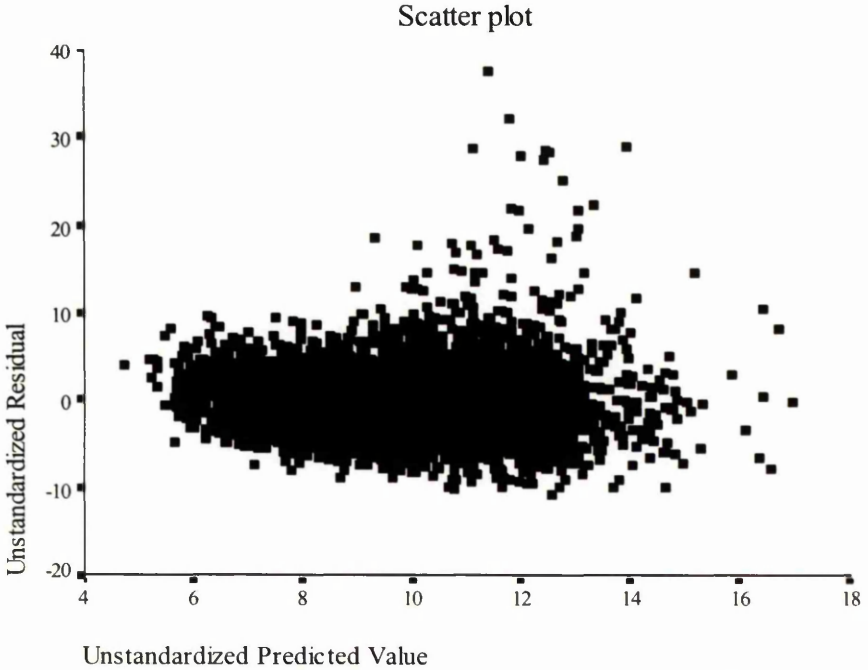


Figure 3.11 shows that apart from a few outliers the residuals seem to form a random scatter about 0. This simple inspection of the residuals may lead us to believe that the linear regression model is an appropriate model to use with the exception of a few outliers. However, prior knowledge of the data would lead to a rejection of the assumptions. Intuitively if the data points are related in time the independence assumption is likely to be violated; residuals close together in time may be similar. Another likely reason that the simple linear model is not appropriate is that the dependent variable is a series of counts; prior to any testing it is likely that these data follow a Poisson distribution. This feature may not be a problem as a Poisson with a large mean is well approximated by a Normal distribution, however when the data are split by cause, age and social class this may become a problem. In order to look at the residuals in more detail, time series methods were used to assess the presence of autocorrelation. These methods are described in section 3.3.3. Details of the development of the methods are provided for male deaths from respiratory disease only, however the results of the application of the final model are supplied for both male and female deaths from respiratory disease, IHD and CVD.

3.3.3 Time Series Methods

Diggle[79] provides an excellent introduction to the use of time series methods for biological data in which he uses several example data sets. Cox[80], Chatfield[81], Box & Jenkins[82] and Kendall[83] all provide a detailed and well structured account of time series methodology.

3.3.3 (a) Autocorrelation, stationarity and differencing

To test for autocorrelation in time series the series must be stationary. The concept of stationarity in a series is that the probabilistic structure of the values in the series is unaffected by a shift in the time origin. However, this assumption is, in practice, an assumption that cannot be checked, and generally a check on second-order stationarity is carried out. Second-order stationarity occurs if there is no trend in the series and if the correlation between any two values in the series depends only on their distance apart and not their actual position in the series. Using algebraic notation a series y_t is second-order stationary if $\mu_t = \mathrm{E}(y_t) = \mu$ for all t and if $\gamma(n,m)$ depends only on $|n-m|$ where $\gamma(n,m)$ denotes the auto-covariance function and equals $Cov(y_{t-n}, y_{t-m})$.

One method of obtaining a stationary series is differencing. Differencing involves analysing the differences between the values of adjacent observations rather than the observations themselves. The first difference of a time series y_t written Dy_t is defined by the transformation $Dy_t = y_t - y_{t-1}$. In this thesis stationarity refers to second order stationarity.

3.3.3 (b) ARIMA models

There are two main classes of time series models. Firstly there are autoregressive (AR) processes where each y_t is defined in terms of its predecessors y_s for $s < t$. We have

$$y_t = \sum_{l=1}^{p} \alpha_l y_{t-l} + Z_t \qquad (3.1)$$

where Z_t is a white noise sequence with variance σ^2.

The other main categories of time series models are moving average (MA) models. A moving average model is simply a linear filter applied to the white noise process Z_t in the form

$$y_t = Z_t + \sum_{j=1}^{q} \beta_j Z_{t-j} \qquad (3.2)$$

An autoregressive moving average process combines both equations and gives

$$y_t = \sum_{l=1}^{p} \alpha_l y_{t-l} + Z_t + \sum_{j=1}^{q} \beta_j Z_{t-j} \qquad (3.3)$$

The clearest way of describing a model of this form is to write $y_t \sim ARMA(p,q)$ where p and q are known as the orders of the process. If either p or q equals zero we can write $y_t \sim MA(q)$ or $y_t \sim AR(p)$. This notation can be expanded to include information on the degree of differencing of the series y_t required to obtain a stationary series. If the series was differenced d times then we can write $y_t \sim ARIMA(p,d,q)$.

Currently we have a time series of residuals from a linear regression fit to the data y_t, and we want to determine which, if any, time series model these residuals follow. This is done using a plot of the auto-correlation function and partial auto-correlation function over lag times. If the residuals do in fact fall into an ARIMA model category then it is likely that time series methodology will have to be applied when analysing the data y_t. If the residuals do not display any autocorrelation than we can assume that there is no need to pursue the time series path in the analysis.

3.3.3 (c) Model Identification

The autocovariance function of a stationary process y_t is $\gamma(k) = Cov(y_t, y_{t-k})$ and $\gamma(k)$ does not depend on t. Since $\gamma(0)$ is the variance of each y_t then the autocorrelation function is defined as $\rho(k) = \gamma(k)/\gamma(0)$. A method of identifying the properties of a time series is to plot the sample autocorrelation coefficients $\rho(k) = \gamma(k)/\gamma(0)$ against the corresponding lags k. For large n the approximate sampling distribution of each $\rho(k)$ is Normal with mean zero and variance $1/n$. We can

create confidence limits to assess the significance of each $\rho(k)$ but in time series analysis the values of each $\rho(k)$ is not as important as the overall pattern of all the $\rho(k)$'s. The simplest pattern to detect from the plot of the autocorrelation function (ACF) is a cut off. If for all k greater than some integer q, the $\rho(k)$ are approximately zero then this indicates an $MA(q)$ process as a possible model for the data. If, the ACF plot shows exponential decay then this may point us in the direction of an $AR(p)$ process.

A variant of the autocorrelation function known as the partial autocorrelation function or PACF provides us with a cut off procedure with which we can identify an $AR(p)$ process. We define the partial autocorrelation coefficient, a_k as the estimate of the value α_k in an $AR(p)$ process. If the underlying process is $AR(p)$ then for all $k > p$ $a_k = 0$. A plot of a_k against k should show a cut off at lag p. Durbin[84] developed a way to calculate a_k or α_k using the recursive formula

$$\hat{\alpha}_{p+1,p+1} = (\rho_{p+1} - \sum_{j=1}^{p} \hat{\alpha}_{jp}\rho_{p+1-j})/(1 - \sum_{j=1}^{p} \hat{\alpha}_{j,p}\rho_j) \qquad (3.4)$$

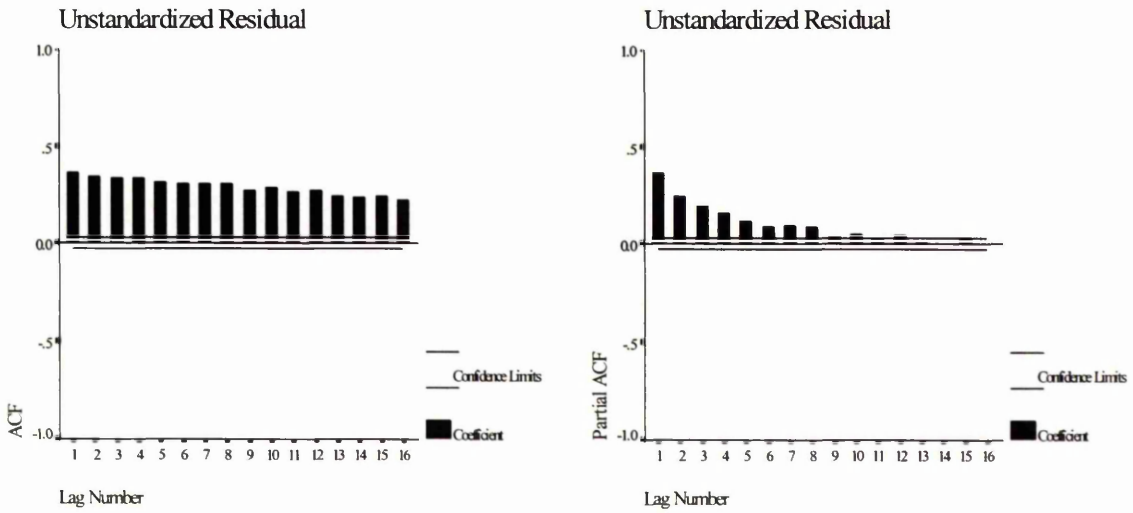
As with the ACF, interpretation of the PACF involves using the limits to assess the cut off point.

So we now have a method of classifying a series into an $ARIMA(p,d,q)$ model by assessing the cut off level in the ACF and the PACF respectively. In practice the cut off is usually no larger than 2 or 3 lags. If neither cuts off at a sufficiently small lag then we should consider a mixed process. That is a model that contains both moving average and autoregressive properties. If both the ACF and the PACF show no cut off then the data may be non-stationary. If this is the case the data should be differenced and a new ACF and PACF plot produced. Again, in practice it is unlikely that a series will have to be differenced more than twice before stationarity is obtained. Finally, parsimony of the model is optimal and if in doubt fewer rather than more parameters should be included in the final model.

3.3.3 (d) Residual analysis

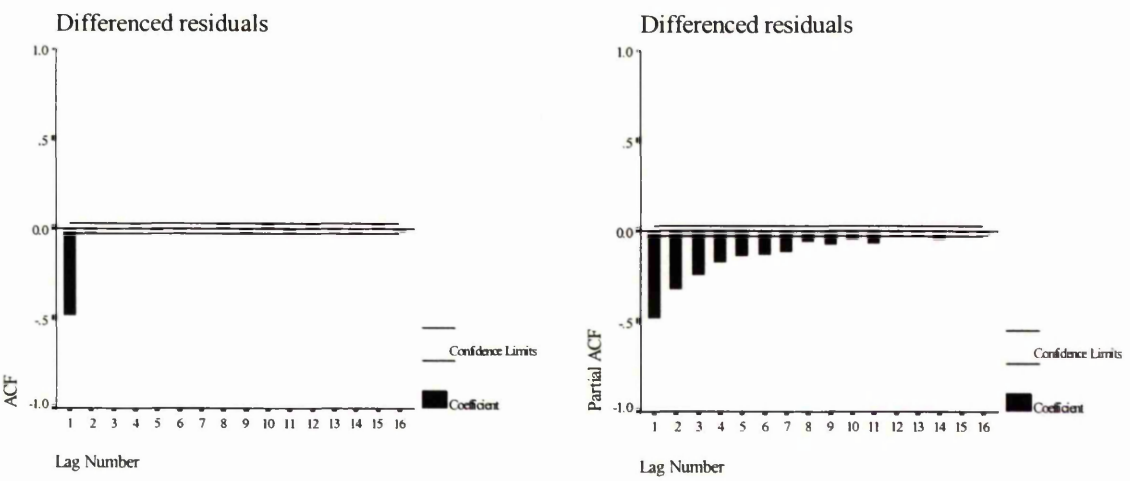
The ACF's and PACF's for the residual series from the linear regression of male respiratory deaths and temperature are shown in figure 3.12.

Figure 3.12 ACF's and PACF's for the residual series from the linear regression



As can be seen, the ACF plot does not die out after 16 lags and the PACF plot appears to show exponential decay. These features point to the fact that the series may require to be differenced. The second two plots in Figure 3.13 are the ACF and PACF for the differenced residuals.

Figure 3.13 ACF's and PACF's for the differenced residual series from the linear regression



The ACF plot cuts off at lag 1 and the PACF plot displays exponential decay. This pattern indicates that the differenced residuals follow an *MA*(1) process and are therefore not independent. The simple linear regression model was not an adequate method of relating the two data sets because the residuals produced from the model were not independent.

Having established that the residuals from the simple linear fit were not independent they were then tested to see if they followed the normal distribution. The results of normal probability tests (figure 3.14) indicate that the residuals do not follow a normal distribution, again violating one of the original assumptions for the linear regression model. The histogram of the residuals (figure 3.15) indicates that they follow a Poisson distribution.

Figure 3.14 Normal and Box plots of the unstandardised residuals from the linear regression

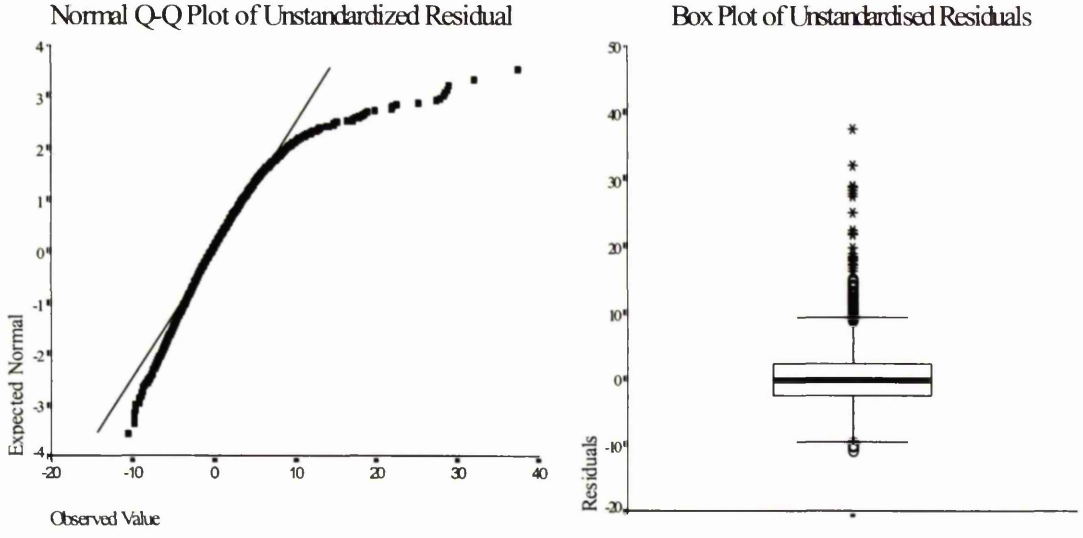
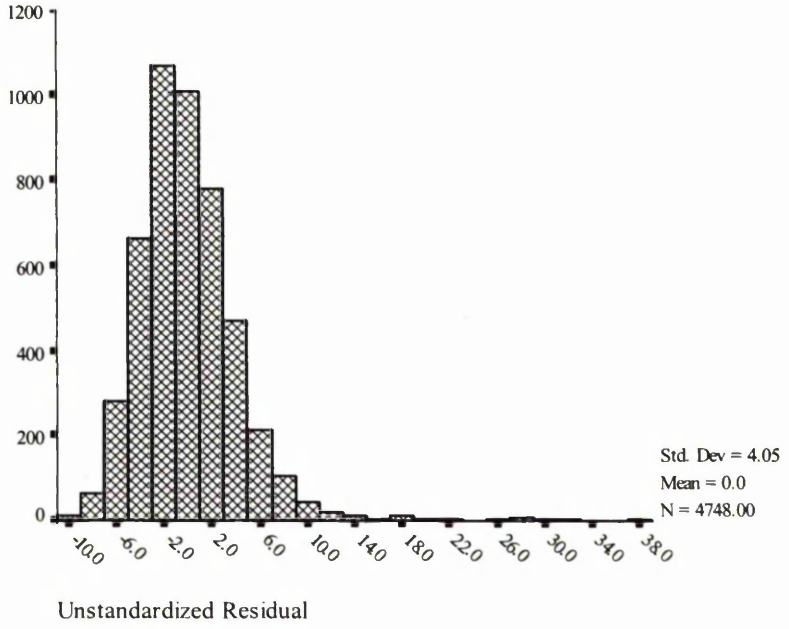


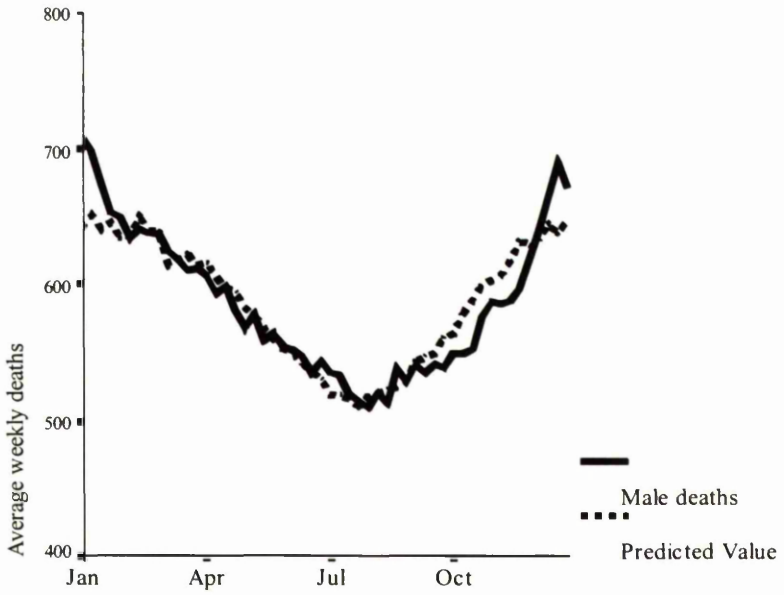
Figure 3.15 Histogram of the unstandardised residuals from the linear regression



3.3.3 (e) Lag effects

A plot of the predicted values and the observed values averaged over a year (figure 3.16) show that there appears to be a lag effect of temperature.

Figure 3.16 Fitted values from the linear regression and observed values averaged over a year



The fitted values anticipate an event to happen earlier than it actually does. If we consider around autumn time when the temperature is getting colder, say in late October, the observed data tells us that approximately 580 deaths will occur each week at this time, whereas the fitted data predicts it will be more in the region of 600 deaths per week. The observed data indicates that around 600 deaths per week occurs later, in Mid November. The fitted model is underestimating the number of deaths when the temperature is getting colder and is overestimating the number of deaths when the temperature is getting warmer. This is because the fitted model uses the absolute temperature values with no measure of the lag effects. As was established previously, the effect of temperature on deaths is not immediate and this was not accounted for in the simple linear regression.

In this exploratory analysis we have determined that a relationship between temperature and deaths exists but is complex. There is a lagged effect of temperature on deaths, which may indicate a causal link. A change in the number of deaths occurs after a change in temperature. There is autocorrelation in the data that requires the use of time series models and further the data is in the form of counts, it does not follow a normal distribution.

Using the ARIMA modelling facilities in SPSS three different methods were tried in the analysis of the data. The first two methods are essentially details of how progress was made to reach the final more satisfactory method.

3.3.4 Residual regression (Method 1)

Initial work described in Chapter 2 had established that the number of deaths from respiratory disease, ischaemic heart disease and cerebrovascular disease followed a seasonal pattern. An examination of the daily average temperature showed that it too followed a seasonal pattern. Analysis of the ACF and PACF plots of the residuals from a linear regression indicated that there was autocorrelation in the residuals and that a straightforward linear regression analysis would be inappropriate. To establish a relationship between two series which both display the same general seasonal pattern short term fluctuations from the identified pattern in one series must be shown to follow

short term deviations from the identified pattern in the other series. In order to do this type of analysis the seasonal variation in both series was removed and the residuals series from both seasonal fits were compared.

The initial methodology involved fitting a detailed Poisson regression model of the seasonal pattern of deaths with the model containing as many harmonics as were significant. The motivation behind this procedure was that if one could remove the entire seasonal pattern within the mortality data and all the seasonal variation in the temperature data then the two residual series that were left would represent the deviations from the normal course of events. If there was some relationship between these two series then we could estimate the way in which temperature influenced the death rate.

The initial model was of the form

$$\ln(y_t) = \alpha + \beta t + a_1(\cos(\varpi t)) + b_1(\sin(\varpi t)) \qquad (3.5)$$

where $\varpi = 2\pi * 1/365.25$

In this model the coefficients a_1 and b_1 can be used to estimate the amplitude and the phase of a seasonal curve that peaks once a year. The next harmonic term to be included in the model represented a peak that occurred twice a year. This term

$$a_2(\cos(2\varpi t)) + b_2(\sin(2\varpi t))$$

was added to the model and if its addition resulted in a reduction of the deviance which was significant at the 5% level then the next term to be added was

$$a_3(\cos(3\varpi t)) + b_3(\sin(3\varpi t))$$

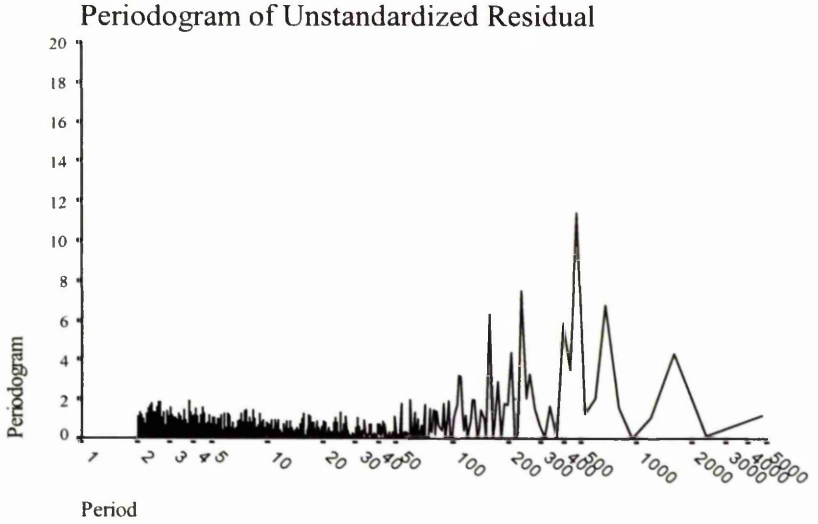
which represented a peak three times a year (every four months) and so on until no more harmonics were required in the model. The harmonics were added in pairs and if either of the terms was significant both were included in the model. This method of modelling can model a complicated pattern of variation, however the coefficients cannot be used directly to compute an amplitude and phase for the seasonal curve. Table 3.7 shows the results for male deaths from respiratory disease, the terms 'cos2day' and 'sin2day' indicate the first harmonic after the sinusoid, 'cos3day' and 'sin3day' the second harmonic etc.

Table 3.7 Results from the Poisson harmonic model for male deaths from respiratory disease

Variable	Coeff	SE	t-value
Const	2.2796	0.0230	99.04
Trend	-0.0002	0.0001	-1.28
Cos1day	0.2768	0.0067	41.46
Sin1day	0.1269	0.0159	8.00
Cos2day	0.0523	0.0067	7.81
Sin2day	-0.0197	0.0099	-1.99
Cos3day	0.0135	0.0067	2.01
Sin3day	-0.0175	0.0083	-2.11
Cos4day	0.0328	0.0067	4.87
Sin4day	-0.0161	0.0077	-2.10
Cos5day	0.0218	0.0067	3.24
Sin5day	-0.0068	0.0074	-0.92
Cos6day	0.0170	0.0067	2.56
Sin6day	-0.0029	0.0072	-0.40

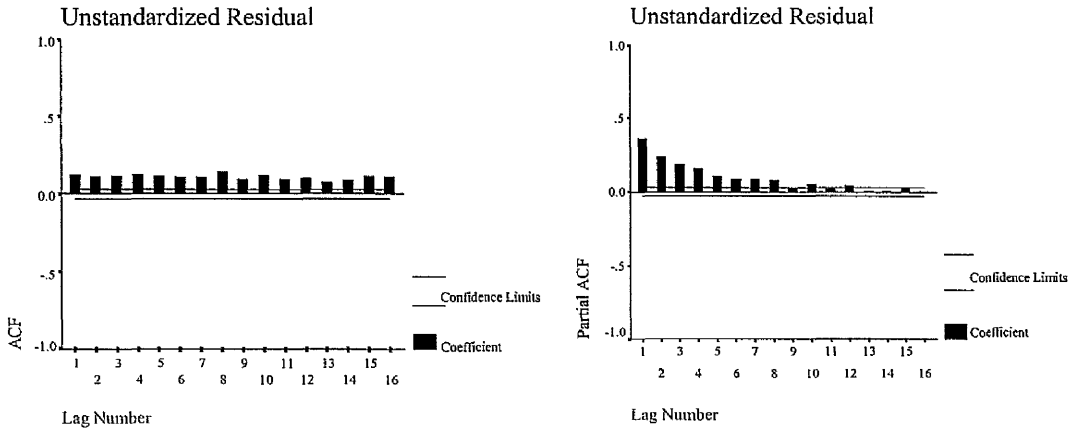
Before the residuals from the above model were related to the residuals from the fit to temperature they were examined to see whether there was any remaining seasonal variation which had not been accounted for in the modelling procedure (figure 3.17). The residual series $\sim N(0,1)$ did not display any seasonal pattern (figure 3.17).

Figure 3.17 Periodogram of unstandardised residuals from Poisson harmonic fit to male deaths from respiratory disease.



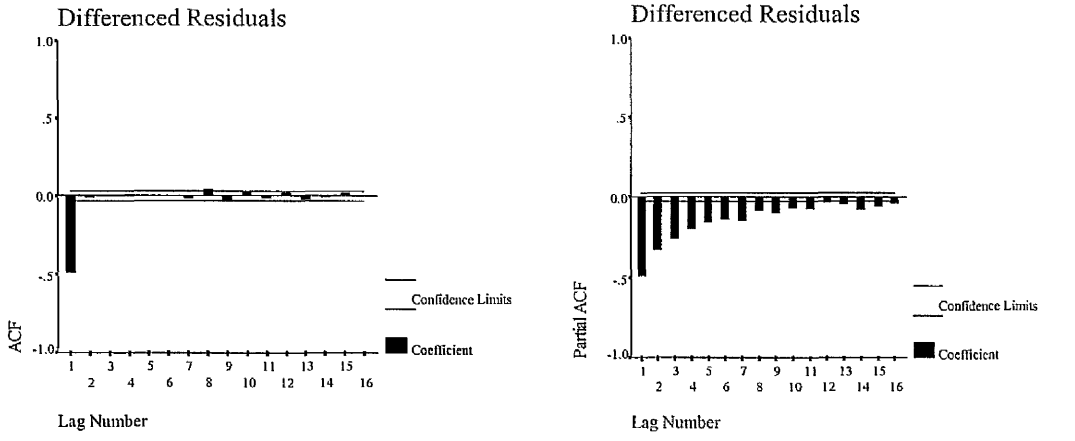
The ACF and PACF plots of the harmonic residuals indicated that the residual series was not stationary (figure 3.18)

Figure 3.18 ACF's and PACF's for residual series from Poisson harmonic model



When the differenced series was examined (figure 3.19) the ACF plot showed a cut-off at lag 1 while the PACF plots showed exponential decay, indicating a moving average process of lag 1. Therefore the model which removed the autocorrelation from the residual series for male deaths from respiratory disease was an *ARIMA*(0,1,1) model.

Figure 3.19 ACF's and PACF's for differenced residual series from a Poisson harmonic model



The time series model that removed all autocorrelation from the residual series for temperature was an *ARIMA*(1,0,1) model.

Once all seasonality and autocorrelation had been removed from the data it was possible to investigate any remaining relationship between the two series. A stepwise regression modelling approach was used with the residuals from the mortality series as the dependent variable and lagged values of the temperature residuals as the explanatory variables. This modelling procedure enabled the lag at which temperature had the greatest effect on mortality for each cause of death to be established. In the first model the data was lagged by up to eight days.

Table 3.8 shows, for both male and female deaths from the three causes of mortality, the lagged days in which there was a significant relationship between the temperature residuals and the mortality residuals.

For most causes of death there was a lag of at least one day before the effect of a change in temperature was reflected by a change in the number of deaths. The relationship with temperature was greatest for male deaths from IHD. For deaths from CVD and female deaths from IHD there was a fairly strong relationship with temperature. Deaths from respiratory disease however did not appear to show much of a relationship with temperature within the 8 day period.

Table 3.8 Significance of relationship between deaths and temperature at lag days

	Males			Females		
Lag day	Resp	CVD	IHD	Resp	CVD	IHD
same day				*		
1 day later			**		*	***
2 days later	**	***	*		**	***
3 days later	*	*	**		***	*
4 days later		**	***			
5 days later			***		***	
6 days later		**	***	*		**
7 days later			**		**	
8 days later		***		*		

'*' represents $p<0.05$, '**' represents $p< 0.01$ and '***' represents $p<0.001$.

In all of the cases above, the coefficient in the stepwise model was negative, indicating an inverse relationship between the two residual series. A negative deviation from the yearly pattern of temperature, a cold spell, was associated with a positive deviation from

the yearly mortality pattern (an increase in mortality). The model also showed at which lag the effect was strongest.

This modelling procedure provided an insight into the data, it showed that there was a negative relationship between deaths and temperature once the seasonal pattern had been removed from the series and that this relationship was strongest a few days after the day in which the temperature was recorded.

This method however does not establish the size of the relationship between mortality and temperature. While we know there is a relationship, we cannot establish how this relationship affects the population in terms of the number of deaths attributable to short term temperature changes. The reason that the effect cannot be measured is that we are dealing with two series that are, in effect, residuals of residuals. It would be almost impossible to arrive at a sensible interpretation of the results of the stepwise regression modelling. Other faults that occur in this modelling procedure are that the significance of the harmonic terms were assessed before any autocorrelation was removed. The effect of assessing significance before removing autocorrelation may result in terms appearing significant when in fact they are not strictly significant. Thus a modified version of the modelling procedure was used to re-model the data.

3.3.5 Residual regression (Method 2)

Although the analysis of mortality using all harmonics appeared to give an insight into the seasonal patterns present in the data, a more appropriate method would have been to remove the autocorrelation first, fit the seasonal curves, and then compare the two residual series. Thus method 2 uses the same procedures as method 1 but they are performed in a different order.

An *ARIMA*(0,1,1) was again found to be the best time series model for the mortality data. Obtaining a time series model for the temperature data however proved more difficult, all combinations of models up to and including an *ARIMA*(3,2,3) were fitted to the data and the residuals were examined, but none of the models gave residual series free of autocorrelation. A plot (Figure 3.1) of the original data showed a clear seasonal pattern. It is likely that even if we accounted for the autocorrelation in the data a seasonal pattern would still be statistically significant. Thus it was felt that fitting the seasonal term

before accounting for autocorrelation in this case was not a problem as it was clearly highly significant. Thus for the modelling of temperature a simple sinusoid was fitted to the data before the autocorrelation was examined. It was found that after accounting for the sinusoidal fit an *ARIMA*(1,0,1) model removed the autocorrelation from the residuals.

Now that the autocorrelation had been removed from the mortality data the harmonic terms which may be significant were entered into the models. For male deaths from respiratory disease none of the previously significant harmonic terms were significant. The data could be modelled straightforwardly by an ARIMA (0,1,1) model and a simple sinusoidal curve. For the temperature data the first harmonic was significant and was therefore included in the model. Thus in terms of the temperature variable method 2 was only very slightly different from method 1. It was decided to use the residual series from method 2 as it used the ARIMA(1,0,1) modelling before the harmonic, thus following the principle that autocorrelation should be removed first before considering any other terms. For the mortality data the use of the ARIMA modelling first meant that there were no significant harmonics. The patterns observed in daily deaths from respiratory disease, CVD and IHD for both males and females could be described adequately by a moving average and sinusoidal term. This feature of the data demonstrates the effect of not accounting for autocorrelation in a modelling procedure.

By removing the autocorrelation first, the data to be modelled by the seasonal curve were no longer in the form of 'counts', therefore a model of the form $r_t = a + \lambda\cos(\omega t + p)$ was fitted to the data where r_t are the residuals from the ARIMA model. This model assumed that the r_t followed a normal distribution. A histogram of the residuals (figure 3.20) showed this assumption to be valid. The residuals r_t^* from the sinusoidal fit on the residuals r_t from the ARIMA fit were used for regression analysis with the temperature residuals. The results are shown in Table 3.9.

Figure 3.20 Histogram of residuals from ARIMA(0,1,1) fit to male deaths from respiratory disease

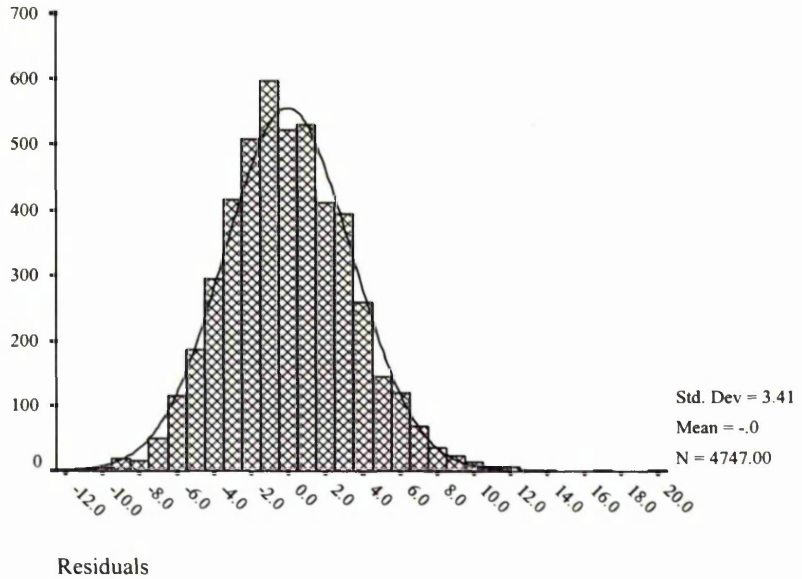


Table 3.9 Significance of relationship between deaths and temperature at lag days

	Males			Females		
Lag day	Resp	CVD	IHD	Resp	CVD	IHD
same day	*					
1 day later			*		*	**
2 days later		***		**	**	***
3 days later		*	*		**	
4 days later		*	***			
5 days later			***		**	
6 days later	*	*	*			
7 days later						
8 days later		***				

As can be seen from the table one of the effects of removing the autocorrelation first was that fewer days were significant and that the strength of the significance of the effects was diminished. The fact that fewer results were significant using this method would indicate that we were perhaps incurring Type I errors using method 1. However as with method 1 the interpretation of the results from method 2 is complex. We are relating the

'residuals of the residuals of deaths' to the 'residuals of the residuals of temperature'. It is virtually impossible to make any sensible interpretation of the results. A method that accounted for all the features of the data but still kept the data in its raw form was required.

3.3.6 Arima Modelling

A method which accounted for the autocorrelation, the non-normality and the seasonal aspects of the mortality data and which related this data directly to the temperature data was developed. The data, being in the form of counts, followed a Poisson distribution. As is the case in most circumstances concerning count data, when the data were transformed using the log transformation the transformed data did not deviate substantially from a normal distribution. Thus the modelling procedure was now concerned only with the logged data. The ACF plot of the logged data showed slow exponential decline, indicating that the logged data required to be differenced before any ARIMA models could be fitted. The resulting data set after differencing displayed the properties of normality (figure 3.21) and stationarity (figure 3.22) with the ACF and PACF plots, indicating that the data followed an ARIMA(0,1,1) model.

Figure 3.21 Histogram of logged differenced male respiratory deaths

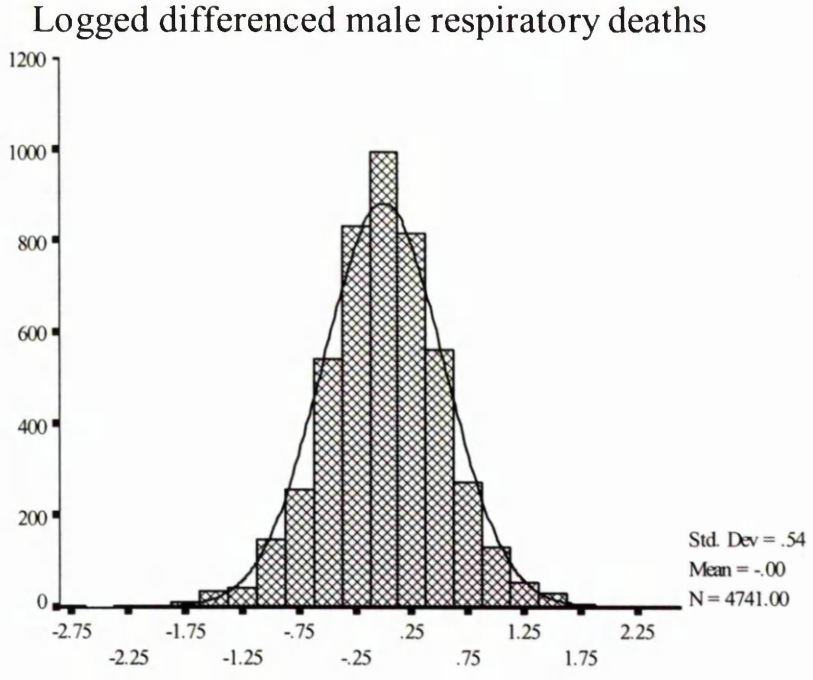
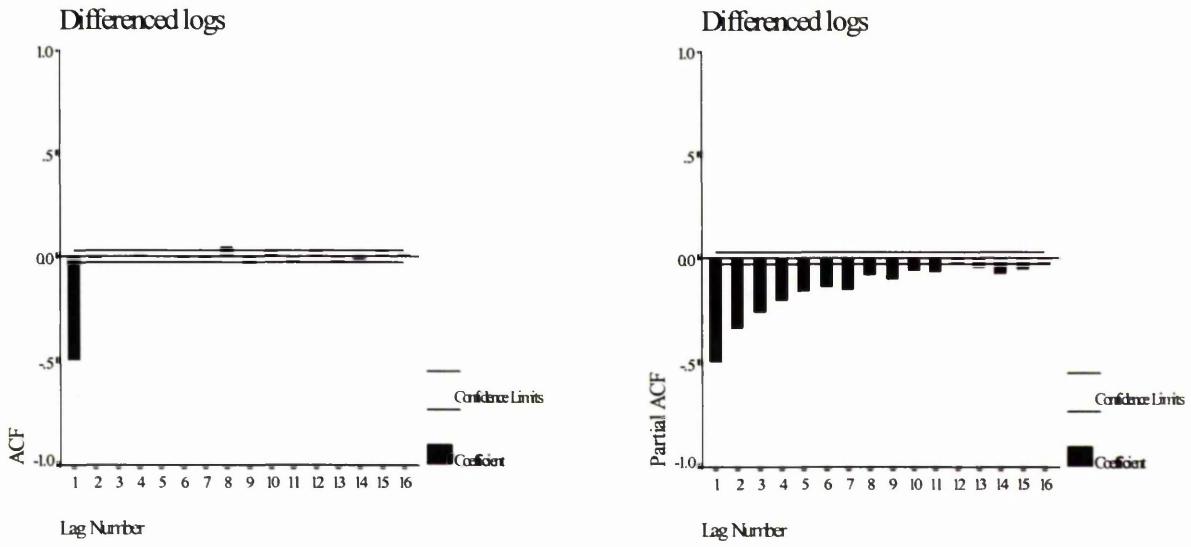


Figure 3.22 ACF's and PACF's of differenced logged data



This replicated the model chosen in 3.3.4. Thus an MA(1) model was fitted to the logged, differenced data.

The inclusion of a sine and a cosine term in the model to represent a simple sinusoidal fit was significant and so these terms remained in the final model. The lagged temperature terms were also added to the model. A model that includes differencing of the dependent variables means that each explanatory term used in the model also has to be differenced. For example, in the simple case, instead of a model of the form $y_t = \alpha + \beta_t x_t$ we have $y_t - y_{t-1} = \beta_t(x_t - x_{t-1})$, the constant term reduces to zero due to differencing. In a stationary series we would expect the mean value of $y_t - y_{t-1}$ to equal zero. The fitted model was therefore of the form

$$\ln(y_t) - \ln(y_{t-1}) = \alpha_1(\sin(t) - \sin(t-1)) + \alpha_2(\cos(t) - \cos(t-1)) + \beta_0(x_t - x_{t-1}) \\ + \beta_1(x_{t-1} - x_{t-2}) + \beta_2(x_{t-2} - x_{t-3}) + \ldots\ldots\ldots + \beta_n(x_{t-n} - x_{t-n-1})) + z_t + \theta z_{t-1} \qquad (3.6)$$

where the y_t are the number of deaths at time t, x_t represents the temperature at time t, and the z_t are the error terms which represent a moving average or $MA(1)$ model.

The model can be thought of as comprising three components, a seasonal part (the sine and cosine terms), a temperature part and a moving average part. The moving average part serves to remove all autocorrelation from the model; it has no explanatory value but is a necessary feature in the model.

The coefficients of the sine and cosine terms describe the amplitude and phase of the seasonal variation in deaths in the same way as in Chapter 2. In this case the seasonal part is the seasonal pattern in mortality that is apparent after temperature and autocorrelation have been accounted for. This could be seen as unexplained seasonality in mortality. Interpretation of the temperature coefficients is best tackled by considering four scenarios.

Scenario 1 - If we consider the first scenario of a constant temperature over all lags except for lag two '2 days ago' where the temperature dropped by 1 degree from its previous constant temperature. Thus all the changes in temperature from day to day were zero except at lag 2 where it was -1. If we ignore the moving average term and the seasonality term then the effect of this temperature drop two days ago is

$$\ln(y_t) - \ln(y_{t-1}) = -\beta_2$$
$$\ln(y_t / y_{t-1}) = -\beta_2$$
$$y_t / y_{t-1} = \exp(-\beta_2)$$

To express this as a percentage increase from y_{t-1} to y_t we have

$$\begin{aligned} \% \text{ increase} &= ((y_t - y_{t-1}) / y_{t-1}) * 100 \\ &= (y_t / y_{t-1} - y_{t-1} / y_{t-1}) * 100 \\ &= ((y_t / y_{t-1}) - 1) * 100 \\ &= (\exp(-\beta_2) - 1) * 100 \end{aligned} \qquad (3.7)$$

Using this method we can attach a percentage increase in deaths at day t (today) due to a drop in temperature 2 days ago, day $t-2$.

Scenario 2 - In the case of constantly decreasing temperature at a rate of 1°C per day for the past n days the overall increase in deaths from n days ago to today would be

$$\% \text{ increase} = (\exp(-\beta_0 - \beta_1 - \beta_2 - \ldots\ldots - \beta_n) - 1) * 100$$

Scenario 3 - If we had constant increasing temperature we would experience a percentage decrease in deaths from day to day.

Scenario 4 - If temperature was constant then all the β terms would disappear and the percentage change in deaths would be determined purely by the underlying seasonal pattern and the moving average term.

The effect of the seasonal component of the model is that the percentage increase in deaths from day to day does not depend purely on the temperature changes in previous days but also on the time of year that these changes are taking place. If the underlying seasonal pattern is increasing towards its peak then the actual increase in deaths due to a drop in temperature will be greater than at a time when the seasonal curve is declining.

The seasonal curve represents patterns in deaths which are not temperature dependent, as does the moving average term. These could be the presence of flu epidemics, national holidays e.g. Christmas and other seasonally fluctuating variables. Generally, however, the coefficients of these terms were very small and the actual effect that these terms had on the percentage change in deaths from day to day was negligible. In the rest of the analyses scenario 1 is adopted and the effect, on mortality, of a drop in temperature of 1°C at n lag days is measured.

3.3.6 (a) Flu deaths

The size of the winter peak in deaths varied considerably from year to year and large increases in deaths from all causes appeared to correspond with influenza epidemics.

Several authors have described the extent to which an influenza epidemic increases the death rate from causes not recorded as influenza [5,9,8,33]. In this analysis the number of recorded influenza deaths were included in the model in order to reduce the possibility of confounding. Influenza epidemics tend to occur in winter, and account for an increase in

deaths from all causes. If influenza deaths were not included in the model we might attribute more winter mortality to temperature than actually existed.

However, there is also the argument that influenza epidemics are related to climate and if the epidemic in turn increases the death rate then this increase is also related to climate. There is no way to separate the effects of influenza epidemics and temperature on the overall death rate. However, the term for influenza deaths was highly significant in each model and was included in the modelling procedure. There are, on average over 100 recorded deaths per year from flu while there are on average over 7000 deaths per year from respiratory disease. It is difficult to estimate the average number of deaths from respiratory disease that are associated with a flu death. The final model now included a moving average component, a seasonal component a temperature component and a flu component and was of the form

$$\ln(y_t) - \ln(y_{t-1}) = \alpha_1(\sin(t) - \sin(t-1)) + \alpha_2(\cos(t) - \cos(t-1)) + \beta_0(x_t - x_{t-1}) \quad (3.8)$$
$$+ \beta_1(x_{t-1} - x_{t-2}) + \beta_2(x_{t-2} - x_{t-3}) + ... + \beta_n(x_{t-n} - x_{t-n-1})) + \gamma flu_t + z_t + \theta z_{t-1}$$

The residuals from this model were checked for normality, seasonality and autocorrelation (figure 3.23).

Figure 3.23 Residual checks on final ARIMA model

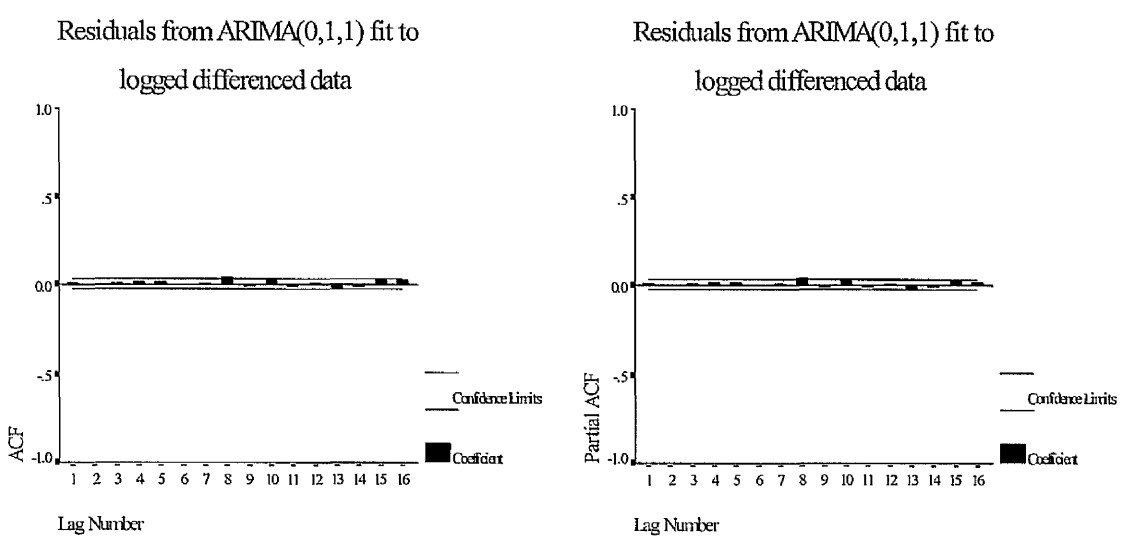
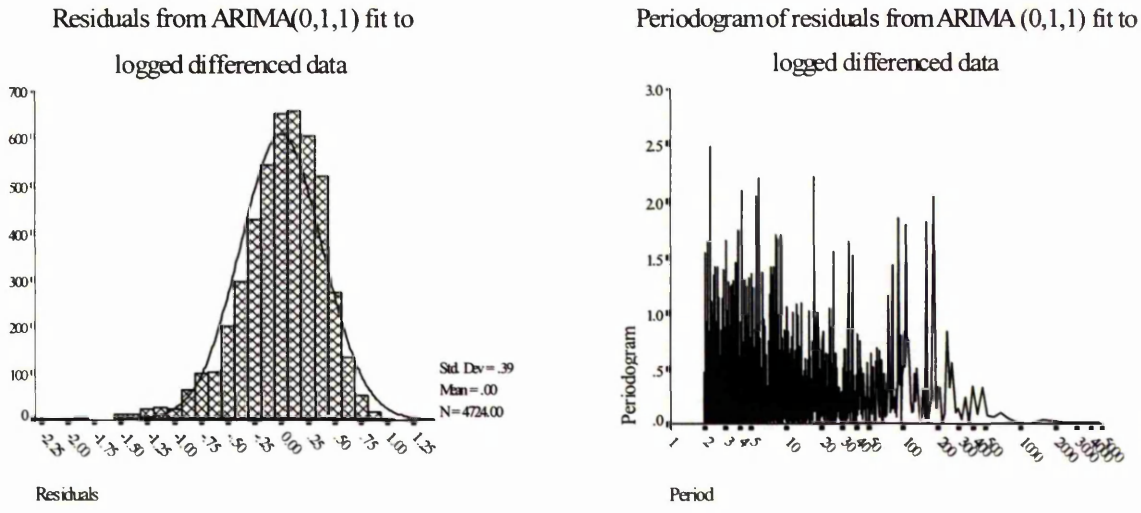


Figure 3.23 (cont) Residual checks on final ARIMA model



The residuals were found to display no seasonal pattern, no autocorrelation, were stationary and normally distributed. This method of modelling the data therefore accounted for the autocorrelation and provided a model in which the coefficients could be easily interpreted and converted into a meaningful figure.

Using this model the data were analysed by cause of death and sex. This analysis was performed for each of the causes of interest, respiratory disease, ischaemic heart disease and cerebrovascular disease. The lagged temperature up to 20 days was included in the model. The number of days included was greater than previous analysis because the previous simple correlation work described in 3.3.2 indicated that the lag effect may be around 15-20 days for deaths from respiratory disease. All lag days were included in the model even if they were not significant. This was because of the possible cumulative nature of the results and because the variables were closely related to each other. It would be wrong to have a model where the effect of temperature at lags of say 2 and 17 days were the only effects included in the model simply because they were the only ones which were significant. The model provided a detailed but slightly complicated picture of the relationship between temperature and deaths. The results from the model applied to daily male deaths from respiratory disease are shown in table 3.10.

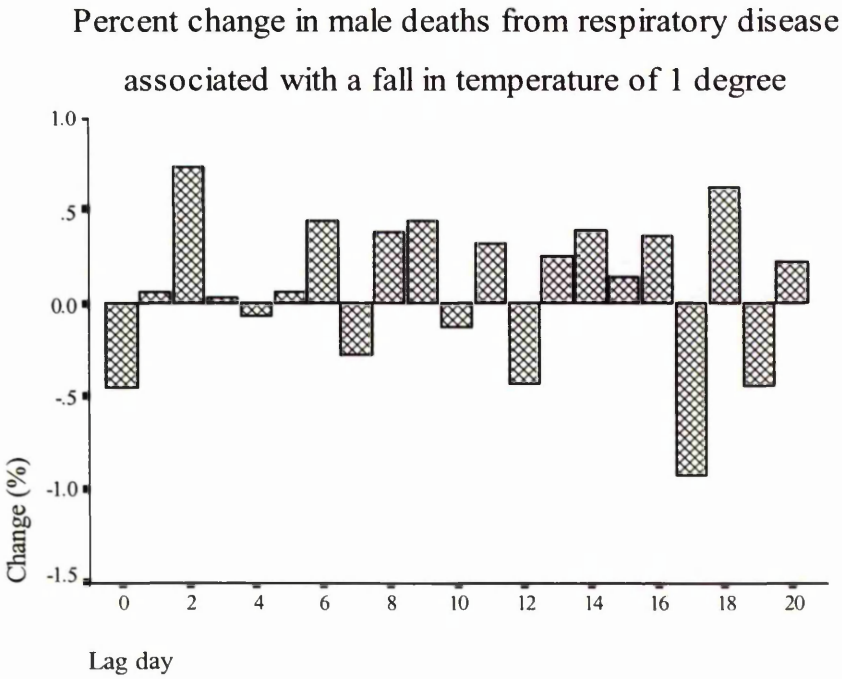
Simply looking at these results in the form of a table it is difficult to appreciate their implication. The moving average term, the cosine and the sine terms are significant as is the daily number of flu deaths, however the relationship between daily temperature and male deaths from respiratory disease is not clear.

Table 3.10 Results from the ARIMA model for male deaths from respiratory disease

Model term	coeff	se	p-value
MA	0.9568	0.0043	0.0000
cos	0.1682	0.0342	0.0000
sin	0.1025	0.0260	0.0001
day 0	0.0045	0.0041	0.1627
day 1	-0.0006	0.0041	0.8865
day 2	-0.0072	0.0041	0.0754
day 3	-0.0003	0.0041	0.9411
day 4	0.0007	0.0041	0.8705
day 5	-0.0006	0.0041	0.8718
day 6	-0.0044	0.0041	0.2829
day 7	0.0028	0.0041	0.4867
day 8	-0.0038	0.0041	0.3431
day 9	-0.0044	0.0041	0.2801
day 10	0.0013	0.0041	0.7551
day 11	-0.0032	0.0041	0.4273
day 12	0.0043	0.0041	0.2990
day 13	-0.0025	0.0041	0.5266
day 14	-0.0039	0.0041	0.3375
day 15	-0.0014	0.0034	0.6675
day 16	-0.0036	0.0032	0.2484
day 17	0.0093	0.0038	0.0145
day 18	-0.0062	0.0038	0.1033
day 19	0.0044	0.0038	0.2442
day 20	-0.0022	0.0031	0.4660
Flu	0.1359	0.1327	0.0000

Figure 3.24 displays the results from the ARIMA modelling procedure in terms of the percentage increase in deaths from respiratory disease due to a drop in temperature of 1°C at each lagged day.

Figure 3.24 Results from the ARIMA model daily male deaths from respiratory disease



As can be seen from the graph there is considerable day to day variation. The effect of a drop in temperature of 1°C appears to result in an increase in male deaths, the most significant increase in deaths occurs at a lag of 17 days, p=0.015. This is close to the value of 15 days lag discussed in 3.2.2

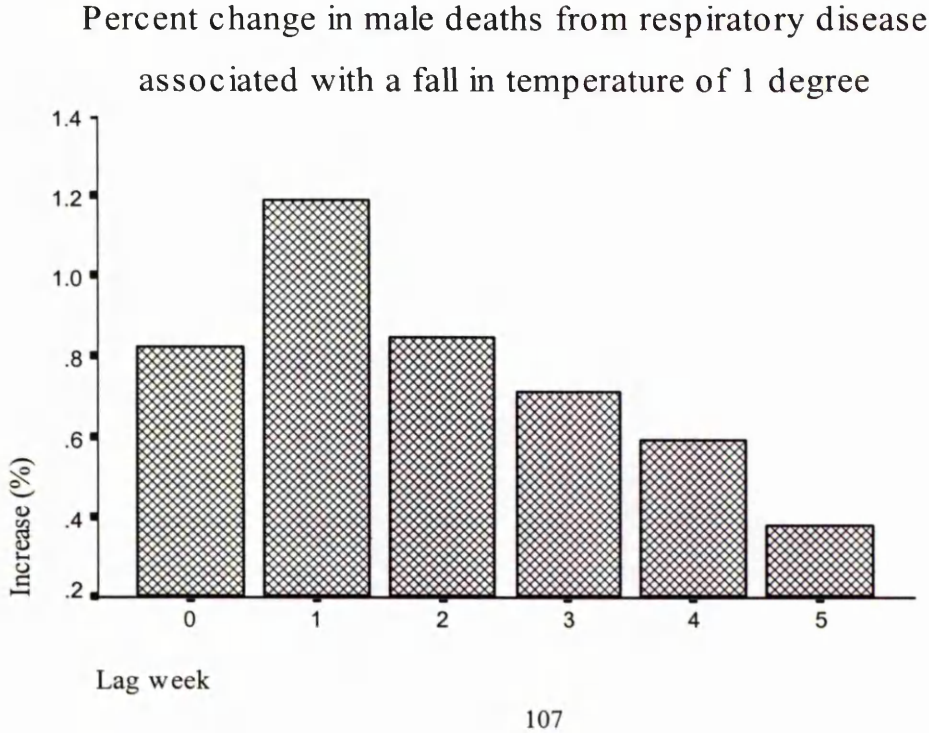
There is considerable noise present in the daily data and to try to reduce the amount of noise in the results various combinations of days were used. These included 3 or 5 day moving average mortality and temperature data, or simply grouping the days into groups, for example 0-2, 3-5, 6-8 days etc. However, it was felt that for the purpose of explanations of patterns and further development of methods it may be more sensible to use weekly numbers of deaths. Using the weekly number of deaths the pattern arising from the modelling procedure was clearer and stronger. The cosine and sine terms were not significant in the model and this made the process more easy to interpret. The analysis was carried out using the logged differenced ARIMA(0,1,1) method for male respiratory deaths. The results from this modelling procedure are given in table 3.11 and are represented graphically in figure 3.25. Tables 3.12 to 3.15 show the results of applying the model to male and female deaths from respiratory disease, IHD, CVD and

deaths from all causes. In these models week 0 indicates the week in which the fall in temperature occurred, this is equivalent to days 0 to 6, week 1 is equivalent to days 7-13 and week 2 is equivalent to days 14-20.

Table 3.11 Results from the ARIMA model weekly male deaths from respiratory disease

Lag week	Coeff	se	prob
Week 0	-0.0082	0.0027	0.0031
Week 1	-0.0118	0.0031	0.0002
Week 2	-0.0084	0.0031	0.0069
Week 3	-0.0071	0.0031	0.0228
Week 4	-0.0059	0.0031	0.0560
Week 5	-0.0038	0.0027	0.1649
Flu	0.0307	0.0023	0.0000
MA/AR	0.8390	0.0221	0.0000

Figure 3.25 Results from the ARIMA model weekly male deaths from respiratory disease



The interpretation of the weekly results is clearer. Using equation 3.7 a fall in temperature of 1°C is associated with an increase in male deaths from respiratory disease of (exp(0.0113)-1)100 = 1.14% one week later. The effect persists for several weeks however at a reduced level at later lags. One recorded death from flu is associated with an increase in male deaths from respiratory disease of over 3%. Using the weekly model, the cumulative effect for each group of seven days is modelled and therefore single days such as at lag 17 are not identified, but a more general picture of the effect of temperature on mortality is achieved. The greatest effect for females occurs at a lag of two weeks, when the increase in mortality associated with a fall in temperature of 1°C is over 1.5%.

Table 3.12 Results from ARIMA model for weekly deaths from respiratory disease

		Males			Females		
		coeff	se	p	coeff	se	p
Resp	Week0	-0.0082	0.0027	<0.01	-0.0044	0.0030	ns
	Week1	-0.0113	0.0031	<0.01	-0.0105	0.0031	<0.01
	Week2	-0.0086	0.0031	<0.01	-0.0156	0.0031	<0.01
	Week3	-0.0067	0.0031	<0.01	-0.0115	0.0031	<0.01
	Week4	-0.0063	0.0031	<0.05	-0.0014	0.0031	ns
	Week5	-0.0041	0.0027	ns	-0.0062	0.0031	<0.05
	Flu	0.0307	0.0023	<0.01	0.0141	0.0013	<0.01
	MA	0.8390	0.0221	<0.01	0.5696	0.0328	<0.01

For deaths from ischaemic heart disease (table 3.13) the increases in mortality were slightly lower than for respiratory disease in the week of the fall in temperature (week 0) and one week later (week 1). A significant increase in mortality from this cause only lasted for these two weeks for male deaths and for a further week for female deaths.

Deaths from cerebrovascular (table 3.14) disease showed a different pattern. The increase in mortality was much less than that for respiratory disease and ischaemic heart disease. A fall in temperature of 1°C is associated with an increase in male and female deaths from cerebrovascular disease of around 0.8% one week later. This effect persisted for longer than that for male deaths from ischaemic heart disease.

Table 3.13 Results from ARIMA model for weekly deaths from ischaemic heart disease

		Males			Females		
		coeff	se	p	coeff	se	p
IHD	Week0	-0.0094	0.0016	<0.01	-0.0087	0.0017	<0.01
	Week1	-0.0107	0.0020	<0.01	-0.0089	0.0020	<0.01
	Week2	-0.0008	0.0019	ns	-0.0047	0.0020	<0.05
	Week3	-0.0005	0.0019	ns	-0.0034	0.0020	ns
	Week4	-0.0020	0.0019	ns	0.0030	0.0020	ns
	Week5	0.0016	0.0016	ns	-0.0009	0.0017	ns
	Flu	0.0060	0.0012	<0.01	0.0043	0.0006	<0.01
	MA	0.9605	0.0112	<0.01	0.9606	0.0111	<0.01

Table 3.14 Results from ARIMA model for weekly deaths from cerebrovascular disease

		Males			Females		
		coeff	se	p	coeff	se	p
CVD	Week0	-0.0057	0.0024	<0.05	-0.0063	0.0020	<0.01
	Week1	-0.0066	0.0029	<0.05	-0.0090	0.0024	<0.01
	Week2	-0.0058	0.0029	<0.05	-0.0024	0.0024	ns
	Week3	-0.0052	0.0029	ns	-0.0024	0.0024	ns
	Week4	-0.0067	0.0029	<0.05	0.0004	0.0023	ns
	Week5	0.0060	0.0024	<0.05	-0.0044	0.0020	<0.05
	Flu	0.0093	0.0018	<0.01	0.0054	0.0007	<0.01
	MA	0.9668	0.0104	<0.01	0.9696	0.0101	<0.01

An overall summary of the effect of falls in temperature on mortality can be obtained by considering deaths from all causes (table 3.15). The most significant increase in mortality occurs one week after the fall in temperature although there is also an immediate associated increase in deaths. The average increase in mortality associated with a fall in temperature of 1°C is around 1% however this varies by cause of death and lag week.

Table 3.15 Results from ARIMA model for weekly deaths from all causes

		Males			Females		
		coeff	se	p	coeff	se	p
All	Week0	-0.0056	0.0010	<0.01	-0.0046	0.0010	<0.01
	Week1	-0.0080	0.0011	<0.01	-0.0086	0.0011	<0.01
	Week2	-0.0013	0.0011	ns	-0.0032	0.0011	<0.01
	Week3	-0.0014	0.0011	ns	-0.0033	0.0010	<0.01
	Week4	-0.0027	0.0011	<0.05	0.0001	0.0010	ns
	Week5	0.0011	0.0010	ns	-0.0005	0.0010	ns
	Flu	0.0104	0.0008	<0.01	-0.0005	0.0004	<0.01
	MA	0.7899	0.0249	<0.01	0.6589	0.0304	<0.01

Residual checks on the models involving ACF and PACF plots, histograms and periodograms implied that the models were good fits of the data. This method overcame the problem of Poisson data by logging the data and provided an adequate means of analysis. However a more appropriate way to deal with Poisson data is to use Poisson regression methods.

3.3.7 Poisson regression methods

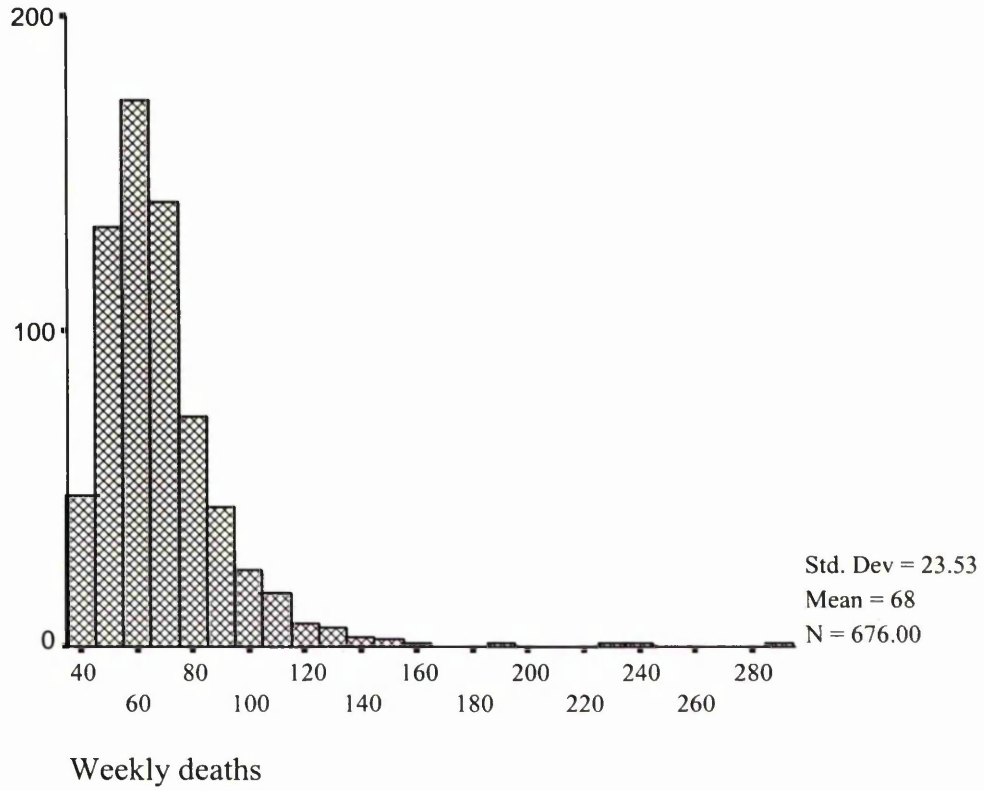
Traditional regression methods for continuous data involve fitting a model of the form $Y = X\beta + e$ where the elements of e are assumed to be independent and identically distributed, iid $N(0,\sigma^2)$, the components of Y are independent normal variables with constant variance σ^2 and

$$E(y_i) = \mu_i \text{ where } \mu_i = x_i^T \beta \qquad (3.9)$$

When dealing with counted data Poisson methods are more appropriate. Poisson regression models are a specific example of generalised linear models. In the case of generalised linear models we estimate a set of parameters β such that a linear combination of the β_i is equal to a function of the expected value of Y_i. For example $g(\mu_i) = x_i^T \beta$ where g is a monotone, differentiable function called the link function and $\mu_i = E(Y_i)$. In the case of Poisson models, the link function $g(\mu_i)$ is the logarithmic function.

Mortality data can be modelled by considering independent random variables $Y_1,.....,Y_N$ to be the number of deaths occurring at successive time intervals $t = 1,.....,N$ where $E(Y_t) = \mu_t$. If we assume that the deaths among different individuals are independent events then the number of deaths, Y, in a fixed time period can be modelled by a Poisson distribution. A histogram of weekly mortality figures for male deaths from respiratory disease in Scotland showed that the data we are dealing with followed the Poisson distribution (figure 3.26). However, the variance is considerably larger than the mean, there were a mean of 68 male deaths from respiratory disease in a week but variance of the distribution was 554. With Poisson data this feature is known as overdispersion. In a Poisson process a general assumption is that of constant variance, which is equal to the mean i.e. $\mathrm{E}(y_i) = \mu_i = Var(y_i)$. Overdispersion occurs when the variance is in fact greater than the mean i.e. $Var(y_i) > \mathrm{E}(y_i)$.

Figure 3.26 Histogram of weekly male deaths from respiratory disease



In a standard Poisson model using the log link we have $\log(\mu_i) = x_i^T \beta$, giving $y_i = \exp(x_i^T \beta) + e_i$ where $E(e_i) = 0, \text{var}(e_i) = \mu_i$ and $\mu_i = E(Y; \beta)$. When overdispersion occurs $\text{var}(e_i) = \tau \mu_i$ and $\tau > 1$. For a more detailed discussion of over-dispersion in Poisson models see Cox[85], Breslow[86] and McCullagh & Nelder[87].

Thus the weekly mortality data has been identified as a Poisson time series. Previously, after transformation, ARIMA time series methods had provided a suitable method of analysis, but this method did not recognise the Poisson nature of the data. However, traditional Poisson regression analysis assumes that the events are independent, and, having identified that the data are a time series, it is clear that the events are not independent. One possible method would be to use Poisson modelling after removal of outliers however Zeger[59] proposed a method of dealing with time series data when the data to be modelled were Poisson.

3.3.8 Zeger's method

We have a time series of counts i.e. the data follow a Poisson distribution and display serial autocorrelation. Overdispersion, was found to be a feature of all of the data sets used in this analysis and tends to be the norm rather than the exception in epidemiological studies.

Cox[85] described two types of models for time dependent data, the first 'observation driven' where the conditional distribution of y_t is driven by previous values $y_{t-1},....,y_1$ and secondly 'parameter driven' where the autocorrelation in the data is introduced due to some latent process. In the case of this project the parameter driven model is appropriate. The number of deaths one week does not directly affect the number of deaths the following week. The data are correlated because they are ordered in time, and because other factors which may be driving the process, are also ordered in time. This process introduces both autocorrelation and overdispersion into the mortality data y_t. In the case of climate related deaths the latent process is the unobservable noise process ε_t present in the temperature data.

Instead of the usual Poisson regression model (equation 3.9) where

$$E(y_t) = \mu_t = \exp(x_t'\beta) \text{ and } Var(y_t) = \mu_t$$

the y_t are no longer independent and the distribution of y_t is conditional on ε_t. So we have

$$u_t = \mathrm{E}(y_t \mid \varepsilon_t) = \exp(x_t'\beta)\varepsilon_t \text{ and } \varpi_t = Var(y_t \mid \varepsilon_t) = u_t \qquad (3.10)$$

Suppose ε_t is an unobserved stationary process with $\mathrm{E}(\varepsilon_t) = 1$ and $Var(\varepsilon_t) = \sigma^2$ then

$$\mu_t = \mathrm{E}(y_t) = \exp(x_t'\beta) \qquad (3.11)$$

and

$$v_t = Var(y_t) = \mu_t + \sigma^2\mu_t^2 \qquad (3.12)$$

The proof of (3.12) can be found in Williams[68]. The basis is the identity

$$\begin{aligned} Var(y_t) &= \mathrm{E}(\mathrm{var}(y_t \mid \varepsilon_t)) + Var(\mathrm{E}(y_t \mid \varepsilon_t)) \\ &= \mathrm{E}(u_t) + \mathrm{var}(\exp(x_t'\beta)\varepsilon_t) \\ &= \mu_t + \sigma^2\mu_t^2 \end{aligned}$$

The latent process introduces both overdispersion and autocorrelation into y_t

To estimate the coefficients β we use a time series analogue to the method discussed by McCullagh & Nelder[87] who use the estimating equation

$$U(\hat{\beta}) = D^T V^{-1}(y - \hat{\mu}) = 0 \qquad (3.13)$$

to give an approximately unbiased and asymptotically normally distributed estimate of β. In this case V is $Cov(y)$ and $D = \dfrac{\partial \mu_t}{\partial \beta_t}$

With independent data, V is diagonal however with time series data V will contain off diagonal elements. Liang and Zeger[88] proposed a method for dealing with repeated measures data, the time series case follows a similar development.
The covariance matrix consists of two elements, the variance associated with β and the variance associated with θ, the autocorrelation.

So we have

$$U(\hat{\beta}) = D^T V^{-1}(\beta, \theta)(y - \hat{\mu}) = 0 \qquad (3.14)$$

To generalise the estimating equation to be applicable to time series data we specify

R_ε is $n \times n$ with j,k element $\rho_\varepsilon(|\, j-k \,|)$

i.e. the autocorrelation matrix.

Then in the parameter driven model

$$V = Var(Y) = A + \sigma^2 AR_\varepsilon A \text{ where } A \text{ is diag}(\mu_1,....\mu_n).$$

So if there was no autocorrelation present in the data we have

$$V = \mu + \sigma^2 \mu^2 = Cov(y)$$

The estimating equation is $D^T V^{-1}(y-\mu)$ if we can estimate ε or θ when β is known.

Liang and Zeger[88] show that

$$V_\beta = \lim_{n\to\infty} n(D^T V^{-1} D)^{-1}(D^T V^{-1} Cov(Y_t) V^{-1} D)(D^T V^{-1} D)^{-1} \qquad (3.15)$$

if $Cov(Y_t) = V$ then

$$V_\beta = \lim_{n\to\infty} n(D^T V^{-1} D)^{-1}(D^T V^{-1} V V^{-1} D)(D^T V^{-1} D)^{-1}$$

$$= \lim_{n\to\infty} n(D^T V^{-1} D)^{-1}$$

Now to develop an iterative procedure to estimate $\hat{\beta}$ for a given value of $\hat{\theta}(\beta)$ we can refer back to the Newton-Raphson method of quasi-likelihood given in McCullagh & Nelder[87]. They show

$$\hat{\beta}_1 = \beta + (D^T V^{-1} D)^{-1} D^T V^{-1}(y-\mu) \qquad (3.16)$$

from this we can get

$$\hat{\beta}_1 = (D^T V^{-1} D)^{-1}(D^T V^{-1} D\beta + D^T V^{-1}(y-\mu))$$

$$= (D^T V^{-1} D)^{-1}(D^T V^{-1} D\beta + (y-\mu))$$

$$= (D^T V^{-1} D)^{-1}(D^T V^{-1} Z) \qquad (3.17)$$

where $Z = D^T \beta + (y-\mu)$

From this method we have a simple iterative procedure for estimating $\hat{\beta}$. The whole procedure can be described as 'given an estimation procedure for $\hat{\theta}(\beta)$, $\hat{\beta}$ is found by alternately solving above for $\hat{\beta}_{j+1}$ given $\hat{\theta}_j$, then using the updated $\hat{\beta}_{j+1}$ to find $\hat{\theta}_{j+1}$ until convergence'.

However, an underlying problem is the variance matrix V which is difficult to invert as it contains off-diagonal elements relating to serial autocorrelation.

To simplify the inversion of V put

$$C = diag(\mu_t + \sigma^2 \mu_t^2)$$

and approximate V with $V_R = C^{1/2} R(\alpha) C^{1/2}$ where $R(\alpha)$ is the autocorrelation matrix for a stationary autoregressive process.

Now $\hat{\beta}_R$ is the solution of the estimating equation

$$D^T V_R(\hat{\theta}_R)^{-1}(y - \mu) = 0$$

To enable the inversion of V_R we create a matrix L that applies the autoregressive filter, i.e. the elements of Ly are

$$y_t - \alpha_1 y_{t-1} - \alpha_2 y_{t-2} ..... - \alpha_p y_{t-p} \qquad (t > p)$$

We have

$$Ly = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ -\alpha_1 & 1 & 0 & \cdots & 0 \\ -\alpha_1 & -\alpha_2 & 1 & 0 & \cdots & 0 \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ -\alpha_1 & -\alpha_2 & \cdots & -\alpha_{n-1} & -\alpha_n & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix}$$

Thus we can substitute $L^T L$ for $R(\alpha)$ which is our symmetric autocorrelation matrix.

For example, if we only had a series of 4 times with each displaying autocorrelation then we have

$$L^T L = \begin{bmatrix} 1 & -\alpha_1 & -\alpha_1 & -\alpha_1 \\ 0 & 1 & -\alpha_2 & -\alpha_2 \\ 0 & 0 & 1 & -\alpha_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ -\alpha_1 & 1 & 0 & 0 \\ -\alpha_1 & -\alpha_2 & 1 & 0 \\ -\alpha_1 & -\alpha_2 & -\alpha_3 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1+3\alpha_1^2 & -\alpha_1+2\alpha_1\alpha_2 & -\alpha_1+\alpha_1\alpha_3 & -\alpha_1 \\ -\alpha_1+2\alpha_1\alpha_2 & 1+2\alpha_2^2 & -\alpha_2+\alpha_2\alpha_3 & -\alpha_2 \\ -\alpha_1+\alpha_1\alpha_3 & -\alpha_2+\alpha_2\alpha_3 & 1+\alpha_3^2 & -\alpha_3 \\ -\alpha_1 & -\alpha_2 & -\alpha_3 & 1 \end{bmatrix}$$

Thus we have $V_R = C^{1/2} R(\alpha) C^{1/2}$, $V_R^{-1} \cong C^{-1/2} L^T L C^{-1/2}$ (3.18)

The iterative weighted least squares procedure now has the form

$$\begin{aligned} \hat{\beta}_{j+1} &= (D^T V^{-1} D)^{-1} (D^T V^{-1} Z) \\ &= (D^T C^{-1/2} L^T L C^{-1/2} D)^{-1} (D^T C^{-1/2} L^T L C^{-1/2} Z) \\ &= [(LC^{-1/2} D)^T (LC^{-1/2} D)]^{-1} (LC^{-1/2} D)^T LC^{-1/2} Z \end{aligned} \quad (3.19)$$

This can be applied to the data using a fairly simple iterative weighted and filtered least squares algorithm.

The algorithm involves the following steps -

(i) weight the current values of $D = \frac{\partial \mu_t}{\partial \beta_t}$ and Z by the inverses of the standard deviations, $C^{-1/2}$;

(ii) filter the normalised values, $C^{-1/2}\frac{\partial \mu_t}{\partial \beta_t}$ and $C^{-1/2}Z$, with a filter for an autoregressive process of order p

(iii) solve the least squares equations

(iv) iterate (i) to (iii) to convergence.

The parameters σ^2 the measure of overdispersion and θ_p the vector of autocorrelation coefficients can be estimated from

$$Var(y_t) = \mu_t + \sigma^2 \mu_t^2$$

We have

$$\sigma^2 = (Var(y_t) - \mu_t)/\mu_t^2$$

$$= \sum_{t=1}^{n} \{(y_t - \hat{\mu}_t) - \hat{\mu}_t\} / \sum_{t=1}^{n} \mu_t^2$$

and $\rho_\varepsilon(\tau)$ can be estimated as

$$\rho_\varepsilon(\tau) = \hat{\sigma}^{-2} \sum_{t=\tau+1}^{n} \{(y_t - \hat{\mu}_t)(y_{t-\tau} - \hat{\mu}_{t-\tau})\} / \sum_{t=\tau+1}^{n} \hat{\mu}_t \hat{\mu}_{t-\tau} \qquad (3.20)$$

The extra Poisson variation can be removed by fitting a Gamma model. If we have $g(y_t) = x_i^T \beta + \varepsilon_t$ where $\varepsilon_t \sim N(0, \tau^2\sigma^2)$ then we fit to the vector of residuals ε_t, a Gamma model and get fitted values v_t, then the ratio $\varepsilon_t / v_t = \psi_t \sim N(0, \sigma^2)$. The values obtained from this ratio are free from overdispersion.

The autocorrelation is removed by filtering. Filtering in time series involves establishing the form of the autocorrelation and transforming the data to remove the autocorrelation. The data y_t are transformed into z_t using a transformation of the form $z_t = \sum_{t,j=1}^{n} \phi_j y_t$ where the ϕ_j are obtained from the autoregressive fit to ψ_t the adjusted residuals.

An S-PLUS function was written by Scott Zeger to apply this methodology to real data. The function involved fitting a generalised linear model using the Poisson link function, estimating the overdispersion by fitting a gamma model to the residuals and removing the overdispersion as described above. An auto-regressive process of order AR(1) was then fitted to the adjusted residuals to obtain the filtering parameters. The linear model was re-fitted to the new data z_t and updated estimates of β and V the covariance matrix were obtained. The percentage change in the value between the original and the updated values of β determined whether the iterative process was repeated again or not.

The results of the above modelling procedure tell the same story as the results from the ARIMA modelling described in section 3.3.5. The greatest effects of temperature on deaths from respiratory disease were experienced after 1 week for males and after two weeks for females and these effects remained significant for up to 4 or 5 weeks. For deaths from IHD the effect is more immediate for both males (1% increase after 1 week) and females (0.9%) and for deaths from CVD the relationship with temperature is not as strong as for the other two disease categories. When considering deaths from all causes the effect is more immediate in males than in females but there is still strong evidence of a causal lagged relationship between outdoor temperature and mortality. A more detailed comparison of the results from the different methods is described in section 3.3.9(b).

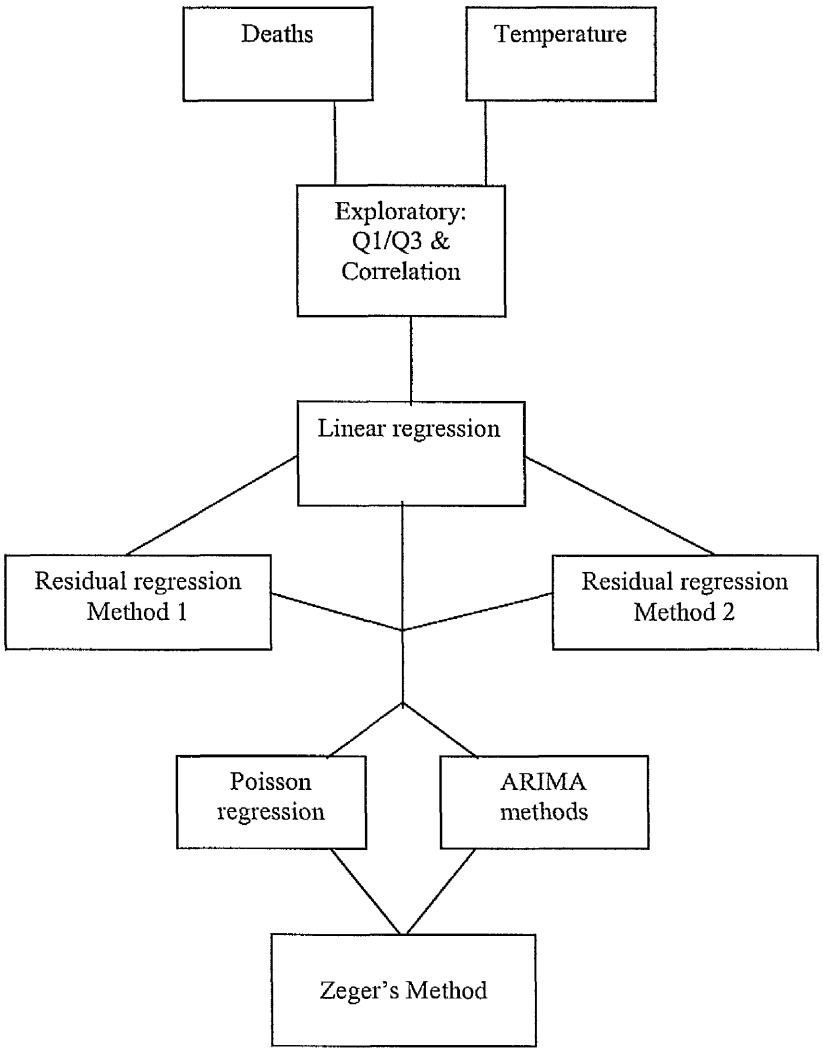
3.3.9 Comparison of methods

The modelling procedures can be compared by assessing the validity of the assumptions made by each modelling procedure and by comparing the results from the various methods in more detail. In assessing the validity of the assumptions made one can study the features of the series to be modelled. When comparing results one can look at the value of the coefficients, the size of the standard errors, the overall fit of the models or can use various tests on the residuals.

The detailed comparison of methods concerns only the ARIMA method, Poisson regression and Zeger's method. The two residual regression methods are not used in the comparison, as the results from these methods were a stepping stone towards the

ARIMA method. Figure 3.27 shows the development of the models in the form of a flow chart.

Figure 3.27 Flow chart of method development



3.3.9 (a) Validity of assumptions

When assessing the appropriate method to use in the analysis of any data set the first considerations should be the assumptions about the data that will be made during analysis. The most common assumption concerns the statistical distribution from which the data arise and most methodology has been developed for the exponential family of

distributions. Other frequently made assumptions concern the distribution of the residuals. The residuals should be normally distributed and display no temporal pattern such as increasing in value over time or autocorrelation.

ARIMA modelling assumes that the data come from a Normal distribution, which is not the case for these data. Generally it can be assumed that count data will arise from the Poisson distribution. McCullagh & Nelder[87] state that 'Even with counted data it is often wise to assume that overdispersion is present unless the data or prior information indicate otherwise.' The data we are dealing with arise from a Poisson distribution with overdispersion. Thus the data does not meet the distributional assumptions necessary for the ARIMA method to be applied appropriately. However, the logged data could be assumed to be normal and this logged data does meet the distributional assumptions. This method, however, could not adjust the regression estimates to control for the overdispersion present in the data.

The Poisson regression method recognised that the data came from a Poisson distribution and was able to adjust the estimates to account for overdispersion, however this method could not deal with the autocorrelation in the data.

Zeger's method was designed specifically to deal with a time series of counts. The modelling procedure was created for data from a Poisson distribution with overdispersion and autocorrelation. The assumptions of the method were entirely appropriate for the data in this case, as the method was developed for data of the type used in this thesis.

One major difference between the ARIMA method and Zeger's method is that Zeger's method models the data as a AR process, while using the ARIMA modelling techniques the MA process was found to more adequately describe the data. In the context of this work the time-series component of the modelling procedure is included to prevent terms in the model appearing significant when in fact they are just significant due to the autocorrelation present in both series. While standard statistical tests indicated that the autocorrelation followed an MA process, either an MA or an AR process will remove the majority of the autocorrelation and will prevent the likelihood of incurring Type II

errors. An examination of the results of the modelling procedures and the residuals demonstrate this.

3.3.9 (b) Comparison of methods & results

The development of the methods concentrated mainly on ARIMA and Zeger's method. However, as the Zeger method can be seen as a Poisson time series, results from a Poisson regression without a time series element were included for comparative purposes. Sinusoidal terms were not included in the models as these were found to be non-significant earlier on in the modelling procedure. Table 3.16 provides a comparison of the results from the three methods.

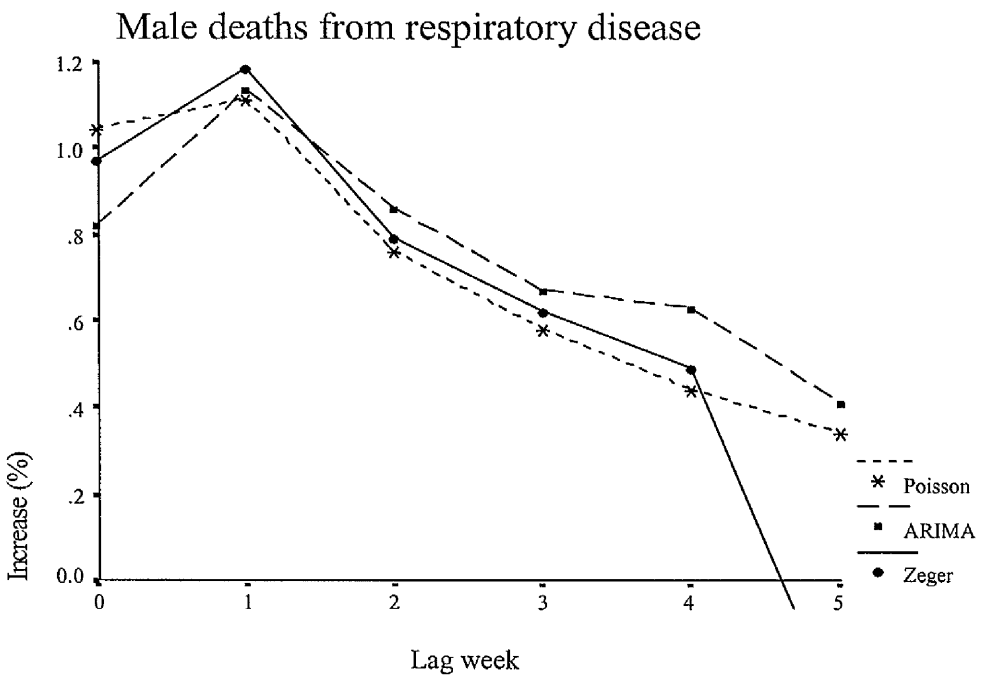
Table 3.16 Comparison of methods for weekly male deaths from respiratory disease

Model	ARIMA			ZEGER			Poisson Regression		
Term	coeff	se	p	coeff	se	p	coeff	se	p
Const	-	-	-	4.1892	.0067	<0.01	4.1898	.0056	<0.01
Trend	-	-	-	2.6*e-7	3.4*e-5	ns	6.2*e-6	.2*e-4	Ns
Week 0	-.0082	.0027	<0.01	-.0097	.0026	<0.01	-.0104	.0027	<0.01
Week 1	-.0113	.0031	<0.01	-.0118	.0030	<0.01	-.0111	.0032	<0.01
Week 2	-.0086	.0031	<0.01	-.0079	.0030	<0.05	-.0076	.0032	<0.05
Week 3	-.0067	.0031	<0.05	-.0062	.0030	<0.05	-.0058	.0032	<0.10
Week 4	-.0063	.0031	<0.10	-.0049	.0030	ns	-.0044	.0032	Ns
Week 5	-.0041	.0027	ns	.0032	.0026	ns	-.0034	.0027	Ns
Flu	.0307	.0023	<0.01	.0490	.0040	<0.01	.0555	.0037	<0.01
MA/AR	0.8390	0.0221	<0.01	0.1648	-	-	-	-	-

It is clear that there is little difference between the three methods in terms of the size of the coefficients, the pattern of the coefficients through the weeks, or the size of the standard errors. However, one difference in terms of the coefficients is that with the Zeger and Poisson methods a constant and a trend term are included in the model. These terms are not included in the ARIMA method because the differencing has made the series stationary with a mean of zero. Thus using Zeger's or the Poisson method one can determine the size of the trend and the mean number of deaths per week.

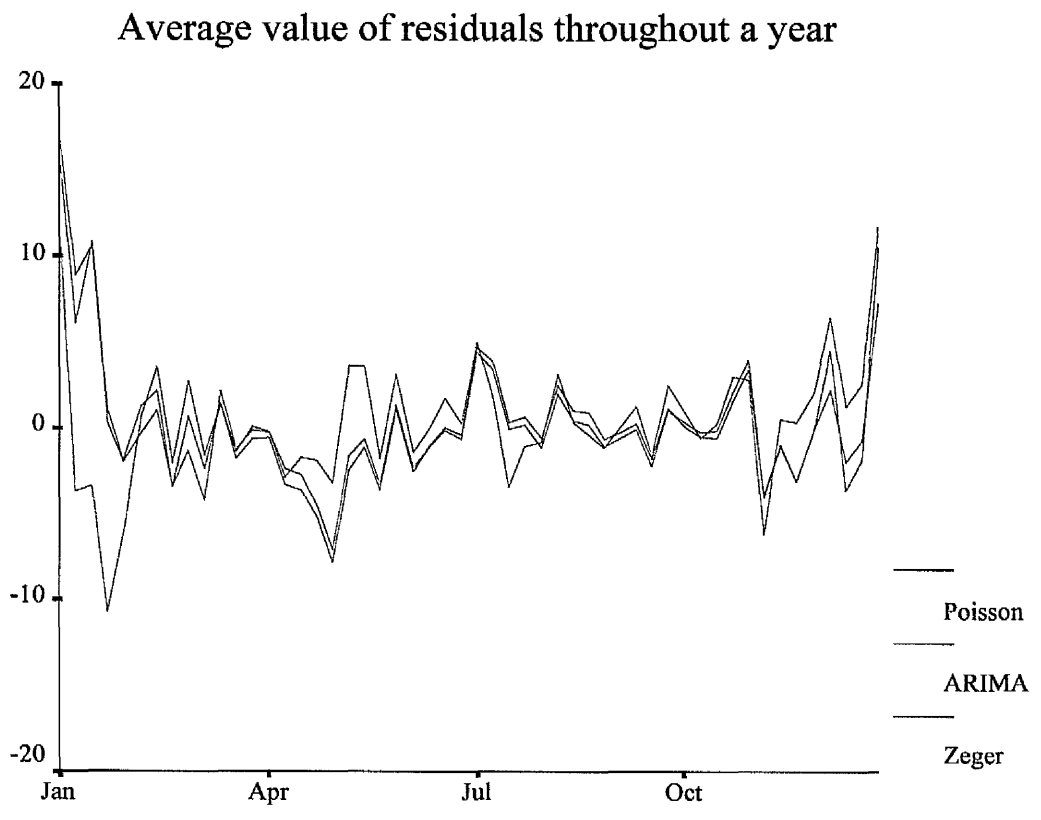
Firstly, if we consider the p-values we can see that Zeger's method and the Poisson regression method are less sensitive; terms which were found to be significant in the ARIMA analysis were not significant when overdispersion in the Poisson distribution was accounted for. One term, temperature at week 3 was found to be more significant in Zeger's method than the Poisson method. In general, the standard errors were greater in the Poisson method, slightly lower in the ARIMA method and lowest in Zeger's method. Thus Zeger's method, having accounted for both forms of extra variation, namely auto-correlation and overdispersion, gave more precise estimates of the coefficients. However, while the p-values and the standard errors are of interest, the coefficients are the more relevant measure. It is the coefficients from which estimates of the size of the relationship between temperature and mortality will be made. Generally, the coefficients from each model show the same pattern, a fairly large value in week 0 rising to a peak in week 1 then tailing off through weeks 2-5. Figure 3.28 demonstrates the interpretation of the coefficients from each model.

Figure 3.28 Percentage increase in male mortality from respiratory disease associated with a fall in temperature of 1°C



The Poisson method gives a coefficient that is larger than the other two methods in week 0 but after that the coefficients are always lower. The coefficients from the ARIMA model are lower than for the Zeger model for weeks 0 & 1, then are consistently higher for the rest of the time. Figure 3.29 shows a plot of the residuals from each method averaged over 1 year. The residuals from the ARIMA method are considerably different to those from the Poisson method and Zeger's method. The ARIMA residuals appear to follow the same pattern as the others, but are generally larger. All the methods under-estimated the size of the unusual peak in male respiratory deaths in the winter of 1989/90. The residuals from these weeks were removed from the data for Figure 3.29.

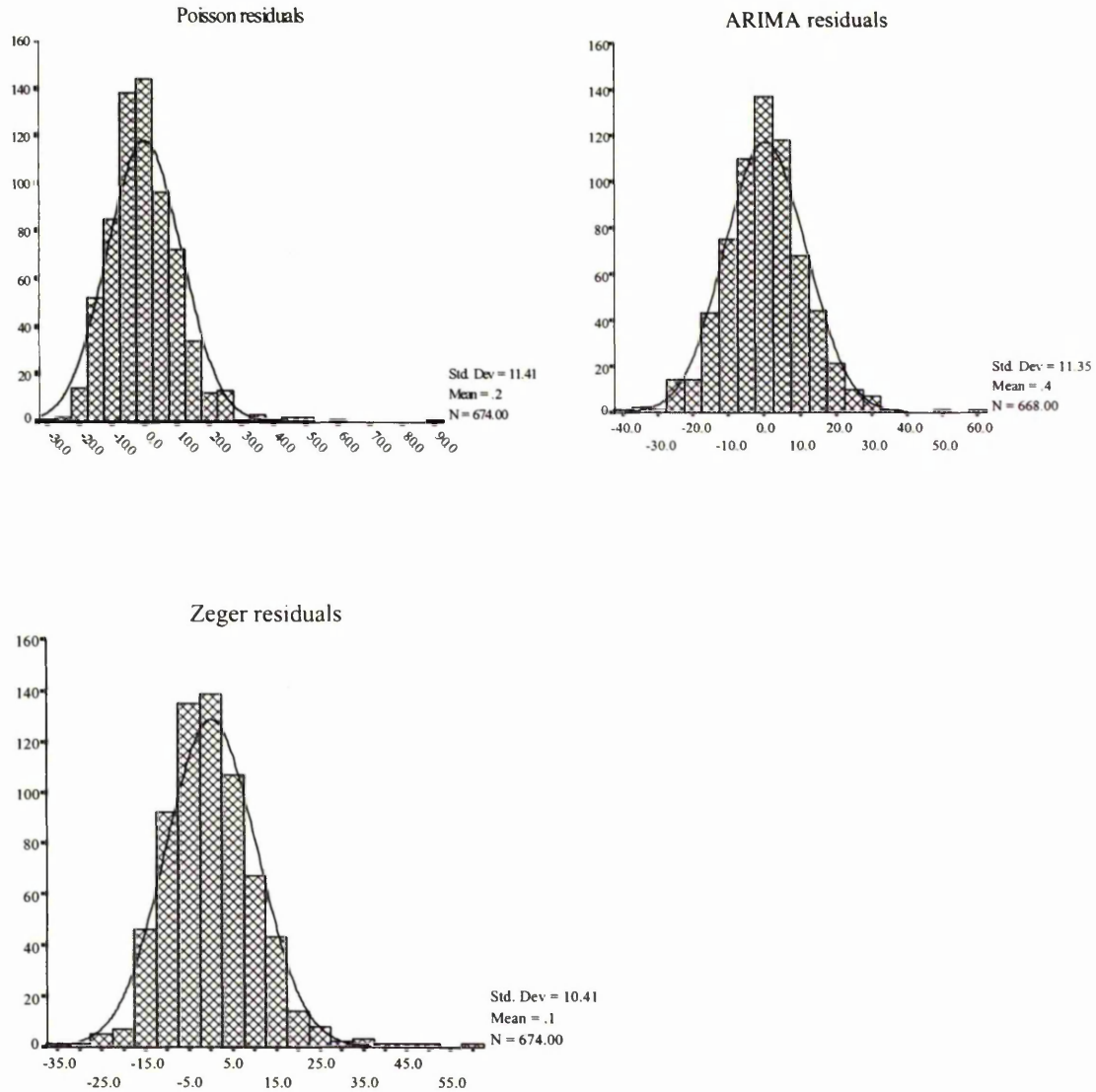
Figure 3.29 Average value of residuals throughout a year



Zeger's method and the Poisson method did not cope well with extreme outliers but on a comparison of p-values, standard errors and coefficients, Zeger's method appears to fare better than either of the other two methods. The residuals from each method appeared to follow a Normal distribution as shown in figure 3.30.

From the histograms the residuals from Zeger's method have the lowest standard deviation and have a mean nearer to 0 than the other methods.

Figure 3.30 Distribution of residuals from each method



Zeger's method was developed to deal with a time series of count data, which is exactly the data we are dealing with. Overall, there was little to choose between the models, but due to the lower standard errors and the fact that the method suits the data, Zeger's method was the one which was used in the further analysis of climate related ill health.

In a further comparison of the methods, the effect of including harmonic terms in the models was assessed. These terms had been found to be non-significant earlier in the modelling process. They were significant when daily deaths were analysed but when the analysis changed to focus on weekly deaths they became non-significant. Table 3.17 compares the results from Zeger's method with and without the harmonic terms.

Table 3.17 Further comparison of Zeger's model male deaths from respiratory disease

Model	Zeger (AR)			Zeger (harmonic)		
Term	coeff	se	p	coeff	se	p
Const	4.1892	.0067	<0.01	4.1887	0.0066	<0.01
Trend	2.6*e-7	3.4*e-5	ns	-8.9*e-6	3.4*e-5	ns
Sin(ωt)	-	-	-	0.0147	0.0239	ns
Cos(ωt)	-	-	-	0.0547	0.0254	<0.05
Week 0	-.0097	.0026	<0.01	-0.0053	0.0031	ns
Week 1	-.0118	.0030	<0.01	-0.0100	0.0032	<0.01
Week 2	-.0079	.0030	<0.05	-0.0061	0.0031	<0.05
Week 3	-.0062	.0030	<0.05	-0.0057	0.0031	ns
Week 4	-.0049	.0030	ns	-0.0038	0.0031	ns
Week 5	.0032	.0026	ns	-0.0043	0.0030	ns
Flu	.0490	.0040	<0.01	0.0500	0.0041	<0.01
AR	0.1648	0.0385	<0.01	0.1521	0.0385	<0.01

For this particular cause of death one of the harmonic terms was found to be significant. The harmonic model was also tried for male and female deaths from ischaemic heart disease and cerebrovascular disease however for both of these causes the harmonic terms were not significant. The detailed comparison was carried out for male deaths from respiratory disease as in the previous comparisons. There was no difference between the constant and the trend term for either of the models in Table 3.17. One of the harmonic terms is significant but it is difficult to tell if the overall inclusion of both terms simultaneously is significant. A comparison of deviance is not possible as both models were modelled separately rather than one being a sub-model of the other.

The effect of including the harmonic terms is to reduce the size of the coefficients and to increase the size of the standard errors of the weekly temperature terms. This results in the effect of temperature in week 0 and week 3 becoming non-significant however the pattern in temperature effects over the 6 week period remains. The assertion that a 1°C fall in temperature is associated with a 1% increase in mortality from respiratory disease one week later and that this effect persists for several weeks is true for both models. The introduction of the harmonic terms into the model for male deaths from respiratory disease would indicate that not all the seasonal variation in mortality from this cause is mediated by temperature. Introduction of these terms reduces the size of the effect of temperature on mortality.

The serial autocorrelation in the model is still highly significant after including the harmonic terms although the actual value of the coefficient is slightly reduced. Introducing the harmonic terms does not negate the use of a term for serial autocorrelation. However it is possible that the effect of temperature on mortality could be adequately measured using Poisson regression methods with numerous harmonics to control for autocorrelation and seasonal patterns not associated with temperature. Overall, Zeger's method adjusting for both serial autocorrelation and overdispersion provides the best method to model these data. The models used in the rest of the section do not include harmonic terms. This may slightly inflate the temperature coefficients if significant harmonic terms were excluded but in many of the models they may not have been significant.

3.4 The association between mortality and temperature

Having developed a method which was suited to time series data that followed a Poisson distribution, the method was applied to male and female weekly deaths for respiratory disease, ischaemic heart disease, cerebrovascular disease and for deaths from all causes as shown in tables 3.18-3.21.

Table 3.18 Results from Zeger's method for deaths from respiratory disease

		Males			Females		
		coeff	se	p	coeff	se	p
Resp	Constant	4.1892	0.0067	<0.01	4.2167	0.0097	<0.01
	Trend	2.55*e-7	3.42*e-5	ns	4.1*e-4	4.9*e-5	<0.01
	Week0	-0.0097	0.0026	<0.01	-0.0021	0.0029	ns
	Week1	-0.0118	0.0030	<0.01	-0.0096	0.0030	<0.01
	Week2	-.0079	0.0030	<0.01	-0.0162	0.0031	<0.01
	Week3	-0.0062	0.0030	<0.05	-0.0104	0.0031	<0.01
	Week4	-0.0049	0.0030	<0.10	-0.0031	0.0030	ns
	Week5	0.0032	0.0026	ns	-0.0054	0.0029	<0.10
	Flu	0.0490	0.0040	<0.01	0.0432	0.0036	<0.01
	AR	0.1648	0.0385	<0.05	0.3139	0.0385	<0.01

The results from Zeger's method can easily be interpreted as the percentage increase in mortality associated with a fall in temperature of 1°C using equation 3.7: % increase = $(\exp(-\beta)-1)*100$. The results are shown graphically in figures 3.31 and 3.32. The charts show the percentage increase in mortality associated with a fall in temperature of 1°C occurring at week 0, assuming that the temperature then returns to its previous value.

Table 3.19 Results from Zeger's method for deaths from ischaemic heart disease

		Males			Females		
		coeff	se	p	coeff	se	p
IHD	Constant	5.2164	0.0039	<0.01	5.0451	0.0043	<0.01
	Trend	-2.4*e-4	1.9*e-5	<0.01	4.9*e-5	2.2*e-5	<0.05
	Week0	-0.0098	0.0015	<0.01	-0.0082	0.0017	<0.01
	Week1	-0.0103	0.0017	<0.01	-0.0091	0.0018	<0.01
	Week2	-0.0010	0.0018	ns	-0.0043	0.0019	<0.05
	Week3	-0.0005	0.0018	ns	-0.0041	0.0019	<0.05
	Week4	-0.0027	0.0017	ns	0.0031	0.0018	<0.10
	Week5	0.0023	0.0015	ns	0.0007	0.0017	ns
	Flu	0.0053	0.0016	<0.01	0.0041	0.0006	<0.01
	AR	0.1693	0.0385	<0.05	0.1928	0.0385	<0.05

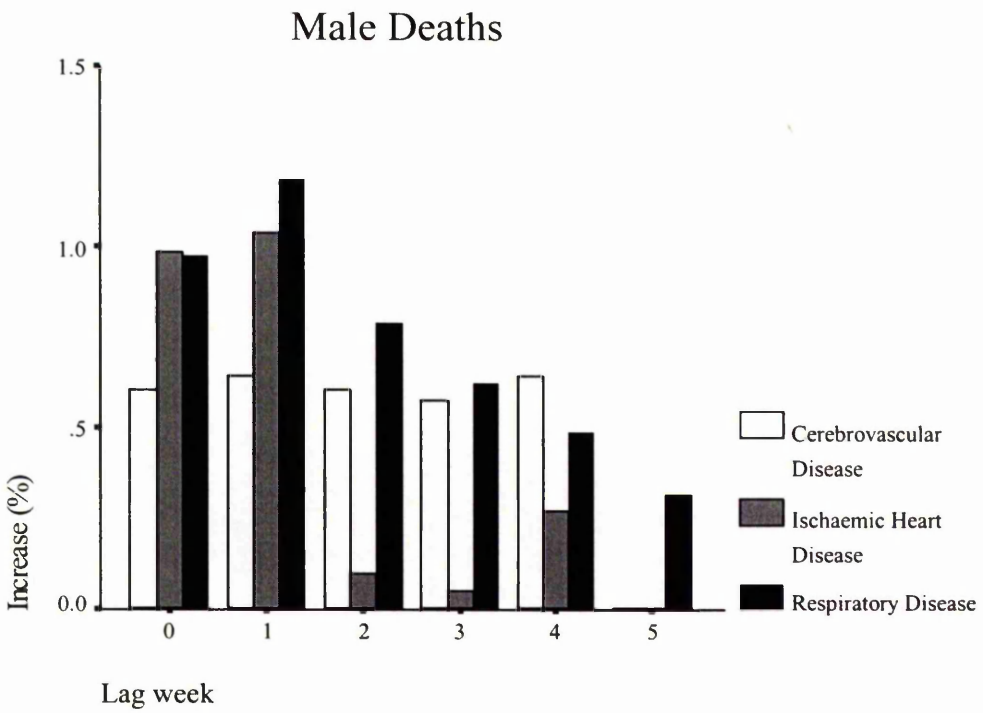
Table 3.20 Results from Zeger's method for deaths from cerebrovascular disease

		Males			Females		
		coeff	se	p	coeff	se	p
CVD	Constant	4.0843	0.0054	<0.01	4.6075	0.0047	<0.01
	Trend	2.3*e-4	2.8*e-5	<0.01	1.1*e-4	2.4*e-5	<0.01
	Week0	-0.0060	0.0023	<0.01	-0.0063	0.0019	<0.01
	Week1	-0.0064	0.0027	<0.05	-0.0085	0.0022	<0.01
	Week2	-0.0060	0.0027	<0.05	-0.0026	0.0022	ns
	Week3	-0.0057	0.0027	<0.05	-0.0024	0.0022	ns
	Week4	-0.0064	0.0027	<0.05	0.0006	0.0022	ns
	Week5	0.0061	0.0023	<0.05	-0.0044	0.0019	<0.05
	Flu	0.0093	0.0017	<0.01	0.0064	0.0009	<0.01
	AR	0.0835	0.0385	ns	0.1351	0.0385	ns

Table 3.21 Results from Zeger's method for deaths from all causes

		Males			Females		
		coeff	se	p	coeff	se	p
All	Constant	6.3689	0.0026	<0.01	6.4243	0.0026	<0.01
	Trend	-0.0001	0.00001	<0.01	0.00002	0.00001	<0.10
	Week0	-0.0052	0.0010	<0.01	-0.0039	0.0009	<0.01
	Week1	-0.0079	0.0011	<0.01	-0.0091	0.0010	<0.01
	Week2	-0.0014	0.0011	ns	-0.0029	0.0011	<0.01
	Week3	-0.0014	0.0011	ns	-0.0029	0.0011	<0.01
	Week4	-0.0022	0.0011	<0.05	0.0002	0.0010	ns
	Week5	0.0009	0.0010	ns	-0.0006	0.0095	ns
	Flu	0.0140	0.0017	<0.01	0.0137	0.0011	<0.01
	AR	0.2062	0.0385	<0.01	0.2255	0.0385	<0.01

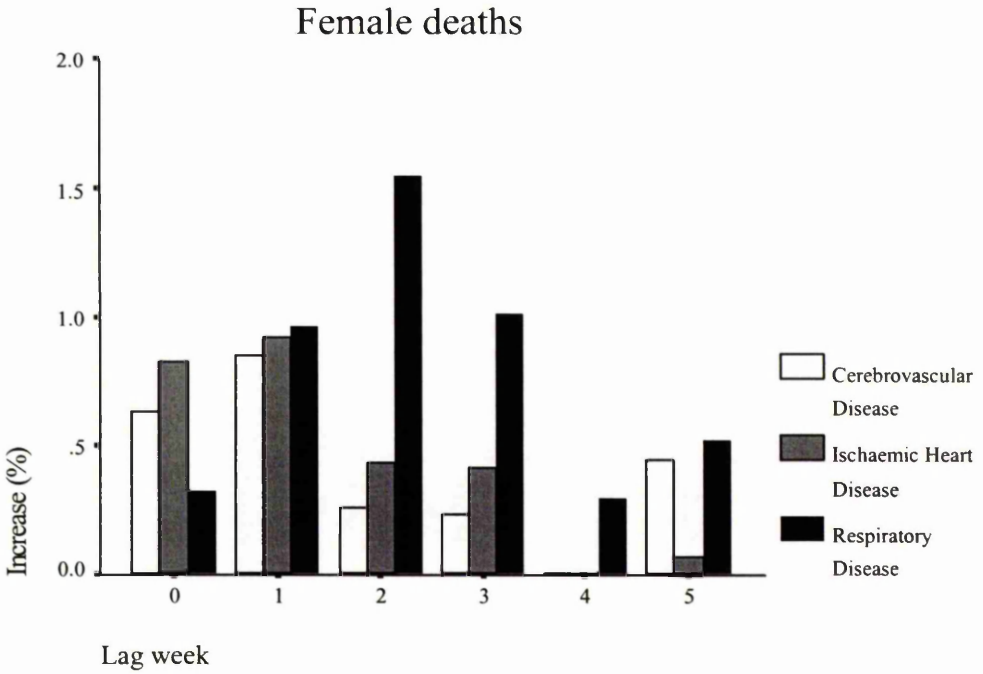
Figure 3.31 Percentage increase in male mortality from cerebrovascular disease, ischaemic heart disease and respiratory disease associated with a fall in temperature of 1°C



For males, mortality from IHD increases by about 1% during the week in which the fall in temperature occurred and remains at this increased level for a further week. At two or more weeks after the fall in temperature there is no further increase in mortality from IHD. Similar initial increases are observed for deaths from respiratory disease with a further increase of 0.2% experienced the following week, this increase in mortality falls away slowly and is still apparent after 5 weeks.

The increase in deaths from CVD is less dramatic, an increase of 0.5% is experienced and this level remains constant for up to 4 weeks. Over the 6 week period the increase in male deaths from respiratory disease associated with a fall in temperature of 1°C is over 4%, for IHD the overall increase is 2.5% and this figure is 3% for male deaths from CVD. This translates into an increase of 3 deaths from respiratory disease, 2 deaths from CVD and 4.5 deaths from IHD.

Figure 3.32 Percentage increase in female mortality from cerebrovascular disease, ischaemic heart disease and respiratory disease associated with a fall in temperature of 1°C

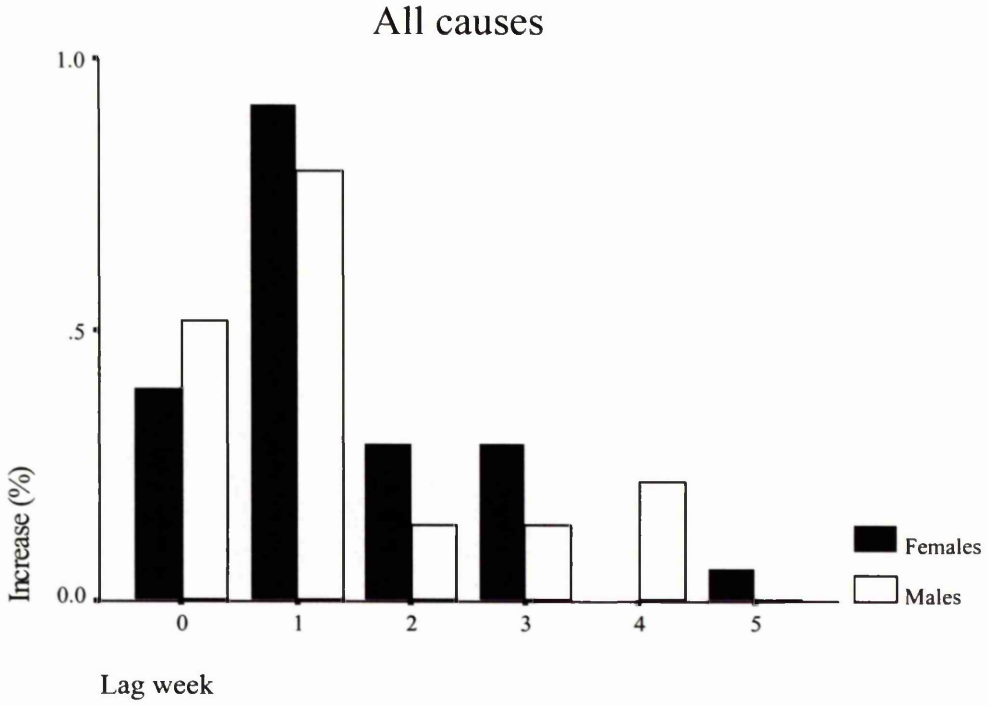


For females, IHD mortality increases by 0.5% on the week that the fall in temperature occurred, then by 1% one week later and falls away after two weeks. This pattern is similar to that for male mortality but the increase is not as large. The greatest increase in

female deaths from respiratory disease occurs two weeks after the fall in temperature, with an increase in mortality of 1.5%, as with male mortality the increase in deaths from respiratory disease persists for several weeks. Female deaths from CVD show a similar pattern to deaths from IHD and unlike male deaths display an increase of around 0.75-1% in the first weeks which then falls away after two weeks.

Figure 3.33 shows the percentage increase in deaths from all causes for males and females associated with a 1°C fall in temperature. An increase in mortality of around 0.5% occurs in the week that a fall in temperature occurs and this figure increases to 0.8% one week after the fall in temperature. The effect of temperature on mortality remains for several weeks but at a much reduced level. The effects are slightly greater for females than for males.

Figure 3.33 Percentage increase in male and female mortality from all causes associated with a fall in temperature of 1°C



3.5 The association between emergency admissions and temperature

Results from Chapter 2 showed that seasonal variation in hospital admissions was not as great as the seasonal variation in mortality. These results suggest that the relationship between temperature and emergency hospital admissions will not be as strong as it was for mortality. Zeger's method was used in the analysis of the relationship between temperature and emergency hospital admissions, residual checks such as those done for the mortality data were carried out and similar results were found. The results of the weekly fits from Zeger's method for both male and female emergency hospital admissions from respiratory disease, ischaemic heart disease and cerebrovascular disease are given in tables 3.22 to 3.25 and are displayed graphically in figures 3.34 and 3.35.

Table 3.22 Results from Zeger's method for emergency admissions for respiratory disease

		Males			Females		
		coeff	se	p	coeff	se	P
Resp	Constant	5.8484	0.0120	<0.01	5.6253	0.0092	<0.01
	Trend	0.0007	0.0001	<0.01	0.0010	0.0001	<0.01
	Week0	0.0006	0.0016	ns	-0.0043	0.0018	<0.05
	Week1	-0.0118	0.0016	<0.01	-0.0117	0.0018	<0.01
	Week2	-0.0092	0.0016	<0.01	-0.0106	0.0018	<0.01
	Week3	-0.0044	0.0016	<0.01	-0.0045	0.0018	<0.05
	Week4	-0.0040	0.0016	<0.05	-0.0029	0.0018	ns
	Week5	-0.0005	0.0016	ns	-0.0007	0.0018	ns
	Flu	0.0045	0.0017	<0.01	0.0134	0.0018	<0.01
	AR	0.7202	0.0385	<0.01	0.5966	0.0385	<0.01

Table 3.23 Results from Zeger's method for emergency admissions for ischaemic heart disease

		Males			Females		
		coeff	se	p	coeff	se	p
IHD	Constant	5.5227	0.0033	<0.01	5.1518	0.0042	<0.01
	Trend	0.0005	0.0001	<0.01	0.0007	0.0001	<0.01
	Week0	0.0030	0.0014	<0.05	0.0022	0.0016	<0.01
	Week1	-0.0080	0.0016	<0.01	-0.0115	0.0017	<0.01
	Week2	-0.0018	0.0016	ns	0.0009	0.0018	ns
	Week3	-0.0020	0.0016	ns	-0.0000	0.0018	ns
	Week4	0.0010	0.0016	ns	0.0010	0.0018	ns
	Week5	-0.0010	0.0014	ns	-0.0013	0.0016	ns
	Flu	-0.0012	0.0009	ns	-0.0003	0.0008	<0.05
	AR	0.1300	0.0385	ns	0.1980	0.0385	<0.01

Table 3.24 Results from Zeger's method for emergency admissions for cerebrovascular disease

		Males			Females		
		coeff	se	p	coeff	se	p
CVD	Constant	4.5066	0.0046	<0.01	4.6870	0.0041	<0.01
	Trend	0.0005	0.0001	<0.01	0.0004	0.0001	<0.01
	Week0	-0.0045	0.0020	<0.05	-0.0054	0.0018	<0.01
	Week1	-0.0041	0.0024	<0.10	-0.0064	0.0021	<0.01
	Week2	0.0022	0.0024	ns	0.0012	0.0022	ns
	Week3	-0.0013	0.0024	ns	0.0007	0.0022	ns
	Week4	0.0022	0.0024	ns	-0.0007	0.0022	ns
	Week5	0.0020	0.0020	ns	0.0008	0.0018	ns
	Flu	0.0014	0.0010	ns	0.0021	0.0009	<0.05
	AR	0.0603	0.0385	ns	0.0505	0.0385	ns

Figure 3.34 Percentage increase in male emergency admissions from cerebrovascular disease, ischaemic heart disease and respiratory disease associated with a fall in temperature of 1°C

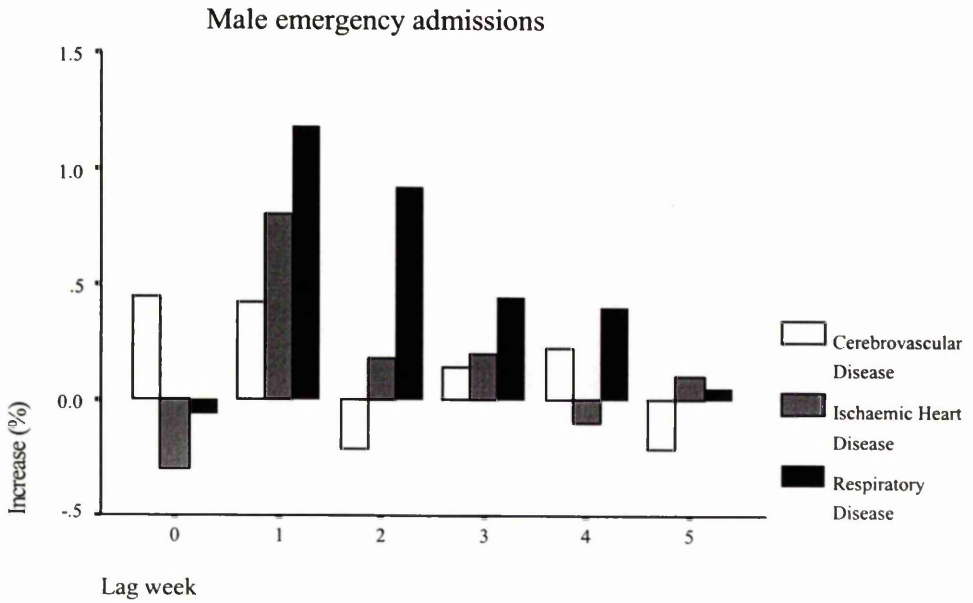
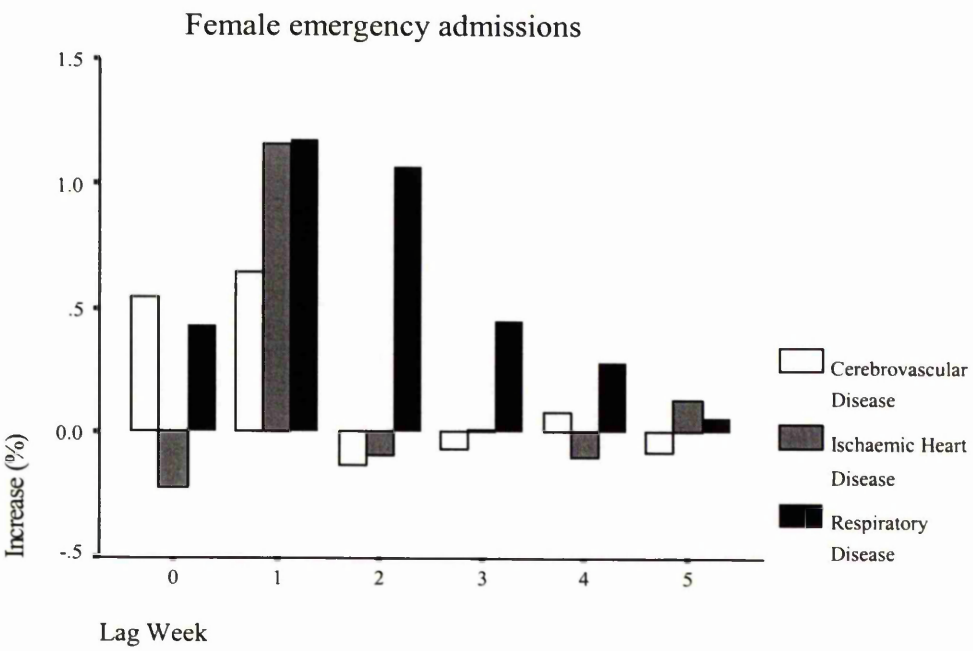


Figure 3.35 Percentage increase in female emergency admissions from cerebrovascular disease, ischaemic heart disease and respiratory disease associated with a fall in temperature of 1°C



The graphs demonstrate the increase in emergency admissions, for the three diagnostic groups, associated with a fall in temperature of 1°C occurring at week 0 assuming that the temperature then returns to its previous value. For reasons given in section 2.5.1 the analysis was not performed on emergency admissions for all causes.

For male emergency hospital admissions increases in emergency admissions for CVD of around 0.5% are seen in the week that the fall in temperature occurred and one week later after which there is little effect of temperature on emergency admissions for CVD. Emergency admissions for IHD increase by 0.7% one week after the fall in temperature but show little evidence of any other effects of a fall in temperature. For respiratory disease an increase of around 1.2% is experienced one week after the fall in temperature and this increase persists (at a reduced level) for up to four weeks. Female increases in emergency admissions behave in a similar manner to that for males, however in general the size of the increase is greater in the first 2 weeks and the remaining effects after 2 weeks are reduced. Increases in female emergency admissions occur more rapidly than male increases in emergency admissions associated with a fall in temperature.

3.6 Summary

Assessing the relationship between ill health and climate and establishing some degree of causality was problematic due to the fact that many other factors change seasonally. In order to be able to infer causality we needed to establish that a short-term fluctuation in temperature was associated with a short-term fluctuation in mortality or morbidity. To do this an appropriate method had to be developed, accounting for the fact that the mortality data was not normally distributed and was ordered in time. There has been much development of time series methods for normally distributed data and many methods for the analysis of Poisson data, however at the time of this analysis there was only one method to deal with a time series of count data, Zeger's method. This chapter was primarily concerned with the application of time series methods, Poisson regression methods and Zeger's method to mortality and morbidity data and temperature data for Scotland from 1981 to 1993. The methods were applied to male deaths from respiratory disease and the results compared.

Overall methodological conclusions were that Zeger's method should generally be used for this type of analysis, however conclusions as to the nature of the relationship between ill health and temperature were invariant under any of the three methods. When several models perform the same purpose all to an adequate degree, as was the case in this analysis, the choice of model can become a subjective one. In this analysis all three models gave similar results but, primarily due to the smaller standard errors and the fact that Zeger's method was developed specifically for this type of data, it was felt that this method was the most appropriate to use. Recently, a method based on Zeger's method that does not require prior specification of the autocorrelation structure of the time series has been developed[89]. This method would be useful in Poisson regression models where the presence of overdispersion or autocorrelation is less easy to quantify.

The results from the analysis in epidemiological terms can be summarised by considering the increase in mortality associated with a fall in temperature. Generally, a 1°C fall in temperature was associated with an increase in mortality and emergency hospital admissions of around 1% one week later. The time lag varied for different causes of death and for different diagnostic groups. This is equivalent to an increase in mortality of 3.5 deaths from ischaemic heart disease, 1.6 deaths from cerebrovascular disease and 1.4 deaths from respiratory disease per week for each 1°C fall in temperature. Deaths from all causes increased by around 12 deaths per week for each 1°C fall in temperature, indicating that these three causes contributed to over half of the temperature associated deaths. The models also accounted for the effect of flu epidemics. Over the full 13 year period one death from flu was associated with an increase in deaths from all causes by around 1.4%, this is equivalent to an extra 16 deaths not recorded as flu. The effect was greater during flu epidemics and much reduced at other times of the year.

These results demonstrate a similar if slightly greater effect of falls in mortality associated with temperature in the rest of Europe. McKee[29] demonstrated an increase in mortality of around 0.7% for deaths from ischaemic heart disease and cerebrovascular disease in Moscow during 1993-1995. Kunst[49], after controlling for season and influenza epidemics, showed an increase in mortality in the Netherlands of 0.43% per 1°C fall in

temperature during the following 3-6 days. While the Eurowinter group[30] showed that percentage increases in all cause mortality per 1°C fall in temperature were greater in warmer regions than colder regions, with Athens showing an increase of 2.15% while south Finland showed an increase of 0.27%.

Various physiological hypotheses as to why there may be a degree of association between cold weather and mortality were discussed in section 1.2.4. This analysis has shown that the lag in the effect is between 0-1 week for deaths from ischaemic heart disease and cerebrovascular disease and around 1-2 weeks for deaths from respiratory disease. Other researchers have assessed these relationships using different lag time periods. The Eurowinter[30] group used daily data and looked at lags in terms of days, Kunst[49] used groups of days from 1-2 days, 3-6, 7-14 and 15-30 days while McKee[29] looked at 4 weekly periods but observed no lag effect and Bowie[54] looked at monthly data and observed no lag effect. The analysis performed by Kunst[49] looking at deaths from all causes found the greatest effect of temperature to occur at lags of 3-6 and 7-14 days.

The mechanism by which deaths from respiratory disease occur 1 to 2 weeks after a fall in temperature is fairly straightforward. Respiratory diseases can be both viral and bacterial. Some viruses survive for longer at colder temperatures and the spread of respiratory viruses is more easy at colder temperatures because people tend to stay indoors and are in closer contact with others. The delayed effects can be due to the incubation period of the virus which usually takes a few days to a week, it is also possible that after contracting the viral infection a secondary bacterial respiratory infection may occur.

For deaths from ischaemic heart disease and cerebrovascular disease, causes that show a more immediate effect of temperature on mortality, the two possible explanations are an increase in blood pressure and viscosity leading to increased stress on the heart or possible clotting. This could happen in patients with long standing narrowing of the arteries and other cardiovascular risk factors. However clots tend to form in a matter of hours so if the clots were caused by the fall in temperature then deaths that occurred due to this mechanism would occur within days. A possible hypothesis for the delay of death for 1 week follows the 'harvesting' hypothesis whereby it is those that are already sick that are dying during cold weather and the fall in temperature simply speeds up the process by putting more stress on the heart.

Chapter 4 Other influences on the effect of climate on mortality and morbidity

4.1 Introduction & Aims

The results from Chapter 3 in which three methods of analysis were compared suggested that Zeger's method was the most suitable method for the analysis of a time series of counts. Three weather variables were considered at the beginning of Chapter 3 and only one, temperature, was associated with the seasonal pattern in deaths and emergency hospital admissions in Scotland. Deaths from influenza were found to explain a proportion of the weekly variation in deaths not recorded as influenza deaths.

The aim of this chapter is to build on these results, using Zeger's methodology on subsets of the data. This will involve producing separate models for different age groups, socio-economic groups and for residents of different geographical areas. Specific moments of interest in the series of daily temperatures will be examined in more detail and other possible explanatory variables such as air pollution will be considered in the analysis.

One problem with an analysis such as this is that the modelling procedure attempts to establish an association between two variables, temperature and mortality, using data that have been collected independently and without thought to this particular application. It is this reason that sets environmental epidemiology apart from classical epidemiological studies such as case-control and longitudinal studies, in which the data tend to have been collected specifically for that particular type of study.

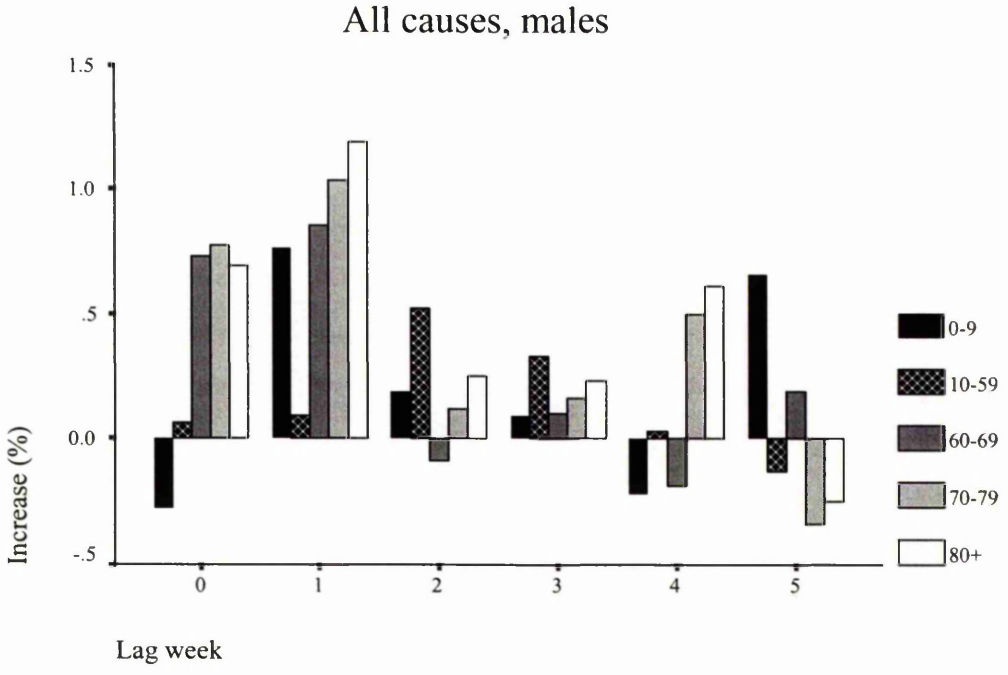
In this study the temperature data have been collected from three sites in Scotland, whereas the mortality and morbidity data were collected from all over Scotland. In the previous chapter the temperature data were averaged and applied to the all Scotland mortality data, producing a relevant analysis of the relationship between the two variables. In this chapter, however, both the temperature and the mortality data are subdivided into demographic or climatological categories. As a result, the power of the statistical methods to detect any association is reduced.

4.2 Effect of socio-demographic variables on the relationship between climate and ill health

4.2.1 Age group

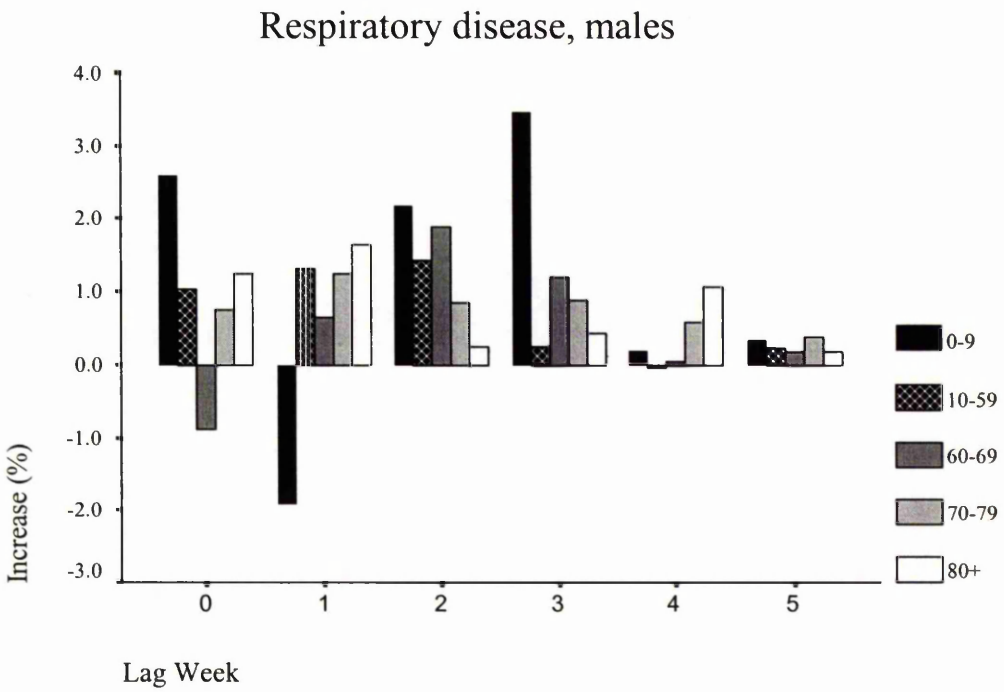
In section 2.5 it was shown that the size of the seasonal variation in mortality and morbidity varied by age group, with older age groups experiencing a greater percentage increase in mortality and morbidity in winter. Having now established a method of assessing the direct effect of temperature, it was of interest to see whether the short term effect of temperature change also varied by age group. The data were split into the same age groups as before and Zeger's method was used to assess the effect of temperature on deaths in each age group. Flu deaths per week in each age group were used as an estimate of the presence of flu epidemics. The analysis was performed for age groups 0-9, 10-59, 60-69, 70-79 and 80+ for male and female deaths from all causes, respiratory disease, IHD and CVD. Figure 4.1 shows the successive percentage change in deaths following a fall in temperature of 1°C for male deaths from all causes in each of the age groups.

Figure 4.1 Successive percentage change in male deaths from all causes associated with a fall in temperature of 1°C



On the week that the fall in temperature occurred there was virtually no change in the number of deaths in age groups 0-9 and 10-59 years, however there was an increase in mortality of around 0.75% in each of the other age groups. One week after the fall in temperature there was again little effect in the 10-59 age group, an increase of 0.72% in the 0-9 age group and a pattern of increasing percentage increases in mortality with increasing age group. The increase in mortality associated with a fall in temperature of 1°C was significantly different from zero ($p<0.05$) in week 2 for ages 10-59, weeks 0 and 1 for ages 60-69 and in weeks 0, 1 & 4 for age groups 70-79 and 80+. However, in terms of consistent patterns there was little effect of temperature on patterns of mortality at two or more weeks after the fall in temperature.

Figure 4.2 Successive percentage change in male deaths from respiratory disease associated with a fall in temperature of 1°C

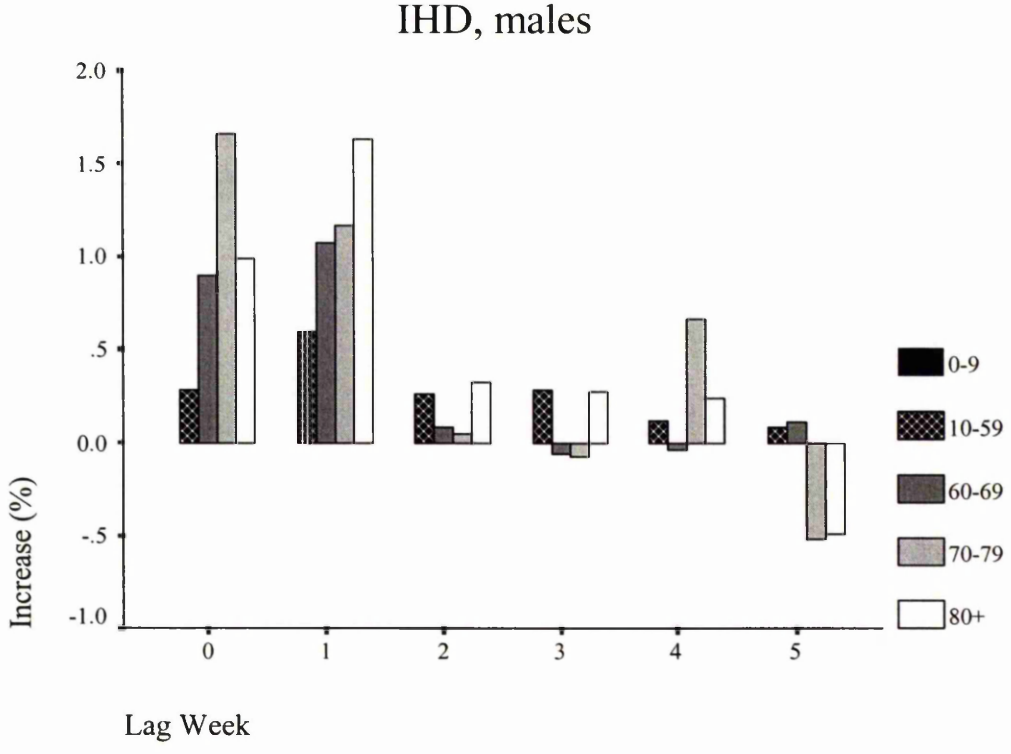


For male deaths from respiratory disease (figure 4.2) the large increase in mortality in the 0-9 age group at week 0 and the large fall in mortality in week 1 were not significant. The only significant increase in mortality in week 0 occurred in the 80+ age group and in week 1 the significant increases in mortality occurred in the 70-79 and 80+ age groups. Two weeks after

the fall in temperature an increase of 2% in deaths in the 0-9 age group was experienced, the size of the increase was lower in the 10-59 age group and then rose to just under 2% in the 60-69 age group and fell away with increasing age. This pattern was similar at a lag of 3 weeks, however the increase in mortality in the 0-9 age group was over 3% after 3 weeks. The effect of a fall in temperature diminished after 4-5 weeks. The only significant increase in the later weeks occurred in week 2 for ages 60-69 and in week 4 for ages 80+.

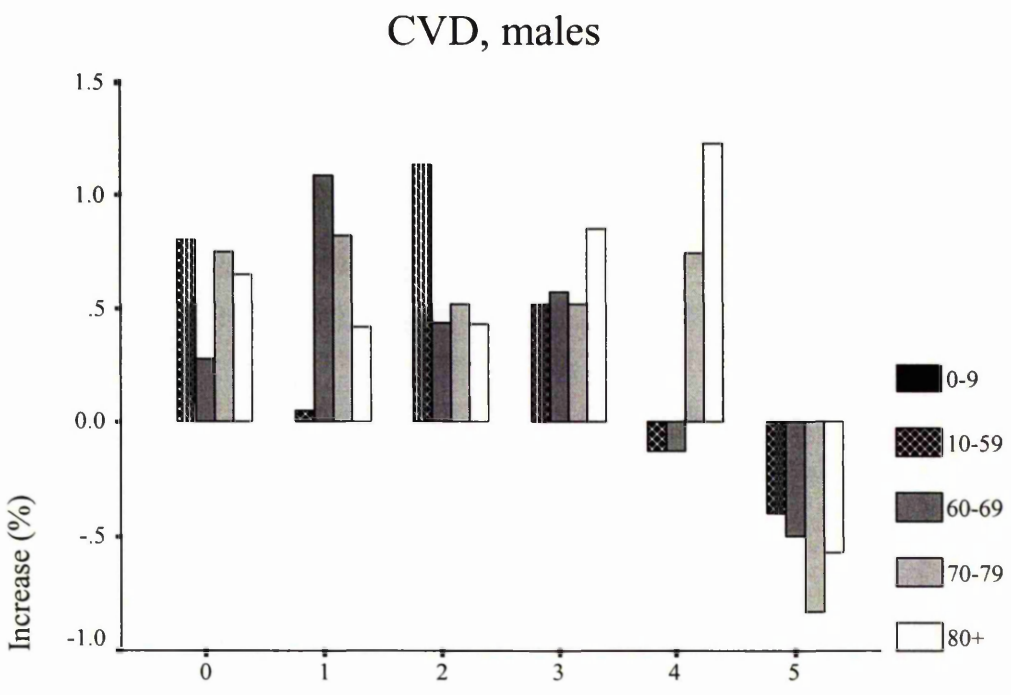
There were very few deaths per week from IHD and CVD in the 0-9 age groups and no results were estimated for these age groups. Male deaths from IHD (figure 4.3) display a more immediate increase associated with a fall in temperature. On the week of the fall in temperature the increase is greatest in the 70-79 age group (1.6%) and 1 week after the fall in temperature the size of the increase increases with increasing age to around 1.6% in the 80+ age group. The increase is significant in weeks zero and 1 for ages 60-69, 70-79 and 80+ and at weeks 4 and 5 for age 70-79.

Figure 4.3 Successive percentage change in male deaths from ischaemic heart disease associated with a fall in temperature of 1°C



With male deaths from CVD (figure 4.4) it was difficult to establish a pattern within the age groups, there was a greater association between mortality and temperature at older age groups but only the increase in ages 70-79 was significantly different from zero.

Figure 4.4 Successive percentage change in male deaths from cerebrovascular disease associated with a fall in temperature of 1°C



The analysis by age group has shown that for deaths from all causes the effect of temperature on mortality was most pronounced in the older age groups and occurred immediately after the fall in temperature. For deaths from each of the three causes; respiratory disease, IHD and CVD, the effect was similarly greater at older age groups. However, for deaths from respiratory disease significant results were found several weeks after the fall in temperature, whereas for IHD the effect was greatest soon after the fall in temperature. The values of the parameter estimates, with associated standard errors, obtained from the modelling procedure for male and female deaths and emergency admissions for each cause are given in appendix IV (I).

4.2.2 Deprivation category

The analysis in chapter 2 demonstrated that there was virtually no deprivation effect when considering the degree of seasonality in deaths experienced by different sections of the population. Zeger's method was used to determine if the effect of a temporary fall in temperature on death rates varied by deprivation category. In this case the data used in the modelling was age standardised to remove any potential bias caused by the effect of age group. A discussion of age standardisation and the need for it in this case can be found in section 2.1.3.

Figure 4.5 Successive percentage change in deaths following a fall in temperature of 1°C for male deaths from all causes by deprivation category

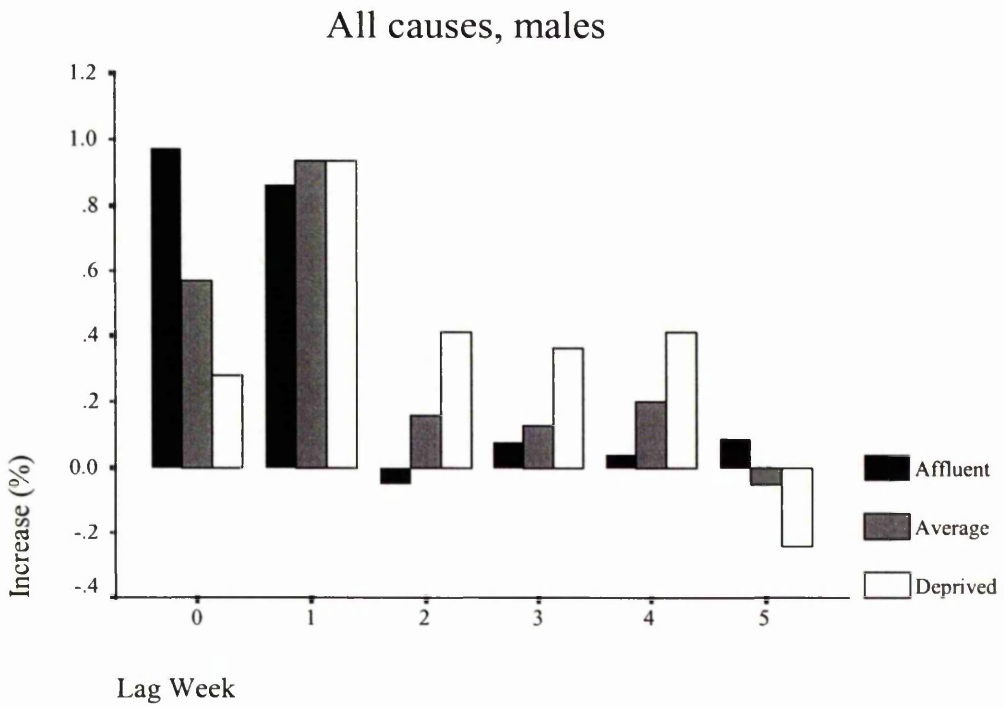


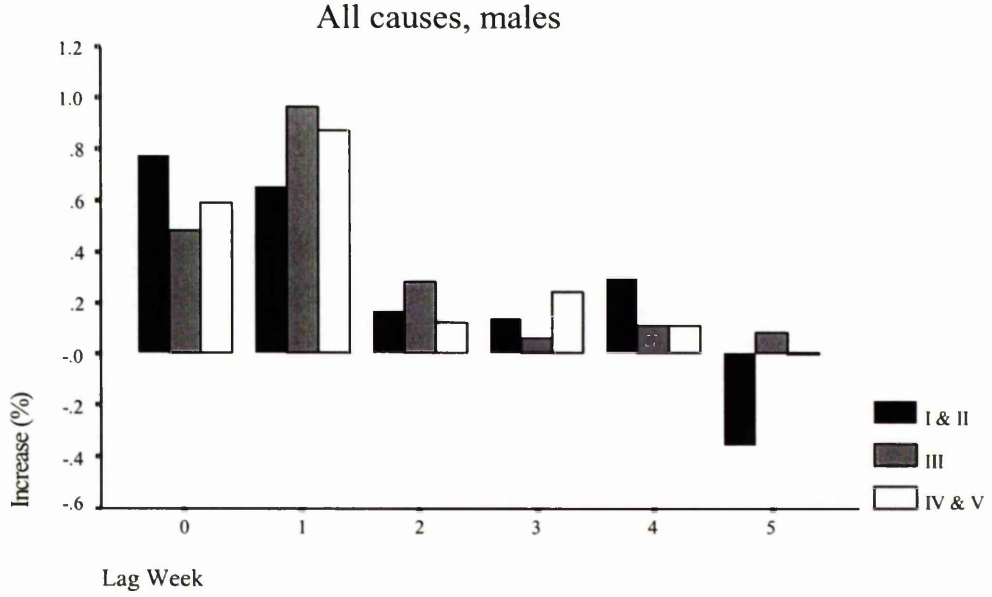
Figure 4.5 shows the successive percentage change in deaths by deprivation category following a fall in temperature of 1°C for deaths from all causes. In the week of the fall in temperature the greatest increase in mortality occurred in the affluent category, with the smallest increase in the deprived category. The percentage increase in mortality after one week was similar for all deprivation categories and then in the following weeks the increases were smallest in the affluent category and greatest in the deprived category. In week 0 the

increase was significantly different from zero for the affluent and average categories and in week 1 the percentage increase in mortality was significantly different from zero in each of the deprivation categories. It appears from the chart that a fall in temperature effects those who live in affluent areas more immediately than those who live in deprived areas. During the week in which temperature falls an increase in mortality of 1% is experienced in the 'affluent' category whereas this is around 0.6% in the 'average' category and 0.3% in the 'deprived' category. Details of the parameter estimates and standard errors from the deprivation analysis are given in appendix IV (II).

4.2.3 Social class

Social class information was available for the mortality data only and when used in Chapter 2 it was found that while there was no difference in the degree of seasonal variation amongst deprivation categories, increased excess winter mortality was associated with belonging to a lower social class. Figure 4.6 shows how the effect of a fall in temperature varies by social class.

Figure 4.6 Successive percentage change in deaths following a fall in temperature of 1°C for male deaths from all causes by social class.



On the week that the fall in temperature occurred the increase in mortality was slightly greater in social classes I & II compared to social classes III, IV & V. One week after the fall in temperature the increase in mortality was greater in social classes III, IV and V compared to social classes I & II. The overall effects were greater in social classes III, IV and V and the delay in effect was longer in this group, however all the percentage increases in mortality in weeks 0 and 1 were significantly different from zero.

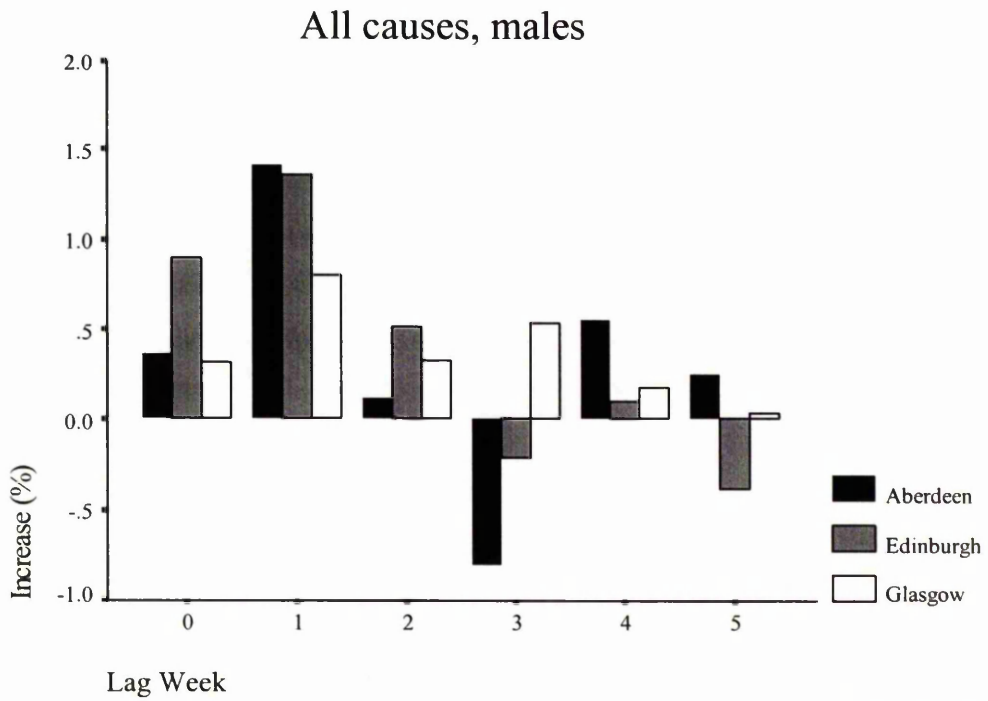
Carstairs[71] deprivation scores and the Registrar General's social classes are closely related to each other, both being measures of socio-economic status. In this analysis and the previous seasonal analysis in Chapter 2 it has been of interest to look at the effects of each measure of socio-economic status separately. From the outset it was felt that low socio-economic status would be related to increased seasonal patterns and a more significant effect of temperature on mortality. Neither the deprivation scores or the social class analysis have demonstrated this, the implications of these results will be discussed later. The results of the social class analysis for male deaths from respiratory disease, IHD and CVD are given in appendix IV (III).

4.2.4 City of residence

It was also of interest to determine whether the effect of temperature varied by area of residence. In Chapter 2 the data were analysed by both health board and by city of residence. In this Chapter an analysis of temperature effects by city seemed more appropriate than by health board as the temperature data were recorded in three cities Aberdeen, Edinburgh and Glasgow. In an analysis by health board a decision would have had to be made as to which temperature series to use for each health board or whether to use the Scottish average. The results from the seasonal analysis in Chapter 2 indicated that an analysis by health board might not give a clear picture of any area effects in the relationship between temperature and mortality. Age standardised mortality rates were computed for each of the three local government districts of Aberdeen City, Edinburgh City and Glasgow City and age standardised flu deaths were also computed for use in the models.

This analysis may also be able to determine the effect of temperature in more detail as we are no longer averaging temperature over the whole country but are looking at temperature in a specific area and mortality in the same specific area. The results from this analysis are shown in figure 4.7.

Figure 4.7 Successive percentage change in deaths following a fall in temperature of 1°C for male deaths from all causes by city of residence.



In the week of the fall in temperature the increase in mortality in Edinburgh was twice as great as the increase in mortality in either Aberdeen or Glasgow, and Edinburgh was the only city in which the increase was significant. One week after the fall in temperature the increase in mortality was twice as great in Aberdeen and Edinburgh as it was in Glasgow, however in all three cities the increase was significant ($p<0.05$). At two or more weeks after the fall in temperature the effects on mortality were negligible. It could be suggested from figure 4.7 that the effect of outdoor temperature on mortality in Glasgow is lower than in Edinburgh or Aberdeen. The results from the analysis by city of residence for deaths and emergency admissions from respiratory disease, IHD and CVD can be found in appendix IV (IV).

4.2.5 Change in the seasonal variation in mortality between 1981 and 1993

A paper published in the BMJ in November 1997[90] described reductions in excess winter deaths in South East England from 1979 to 1994. This led to an investigation of the stability of the observed seasonal increase in mortality in Scotland during our similar time period. Having developed a satisfactory method of assessing the seasonal percentage increase in mortality the data was split into two time periods, 1981-1986 and 1987–1993, to examine if there was any difference in the seasonal pattern between these two time periods. The results from this analysis are shown in table 4.1

Table 4.1 Seasonal percentage increase in mortality by time period

Cause of Death	1981-1986 % Inc	95% CI	1987-1993 % Inc	95% CI	Percentage Change
All causes	34.6	(33.9,35.4)	26.5	(25.9,27.2)	23.4
Resp	122.5	(119.6,125.5)	90.9	(88.4,93.3)	25.8
IHD	37.6	(36.2,39.0)	27.9	(26.6,29.2)	25.8
CVD	40.2	(38.1,42.2)	30.7	(28.8,32.6)	23.6

As can be seen from the table there was a considerable fall in the size of the percentage increase in mortality from summer to winter over the two time periods. The seasonal percentage increase had decreased by around 25%. In order to look more clearly at the change in seasonal variation over time a seasonal percentage increase in mortality from summer to winter was calculated for each year of the study. Figure 4.8 shows the seasonal increase by cause of death and year from 1981 to 1993. The seasonal increase in respiratory disease in 1989 was around 400%. This figure is not shown on figure 4.8 but exerts an influence on the fitted regression line. Figure 4.9 shows the same data with the seasonal increase for 1989 having been recorded as missing.

Figure 4.8 Seasonal increase in mortality each year between 1981-1993

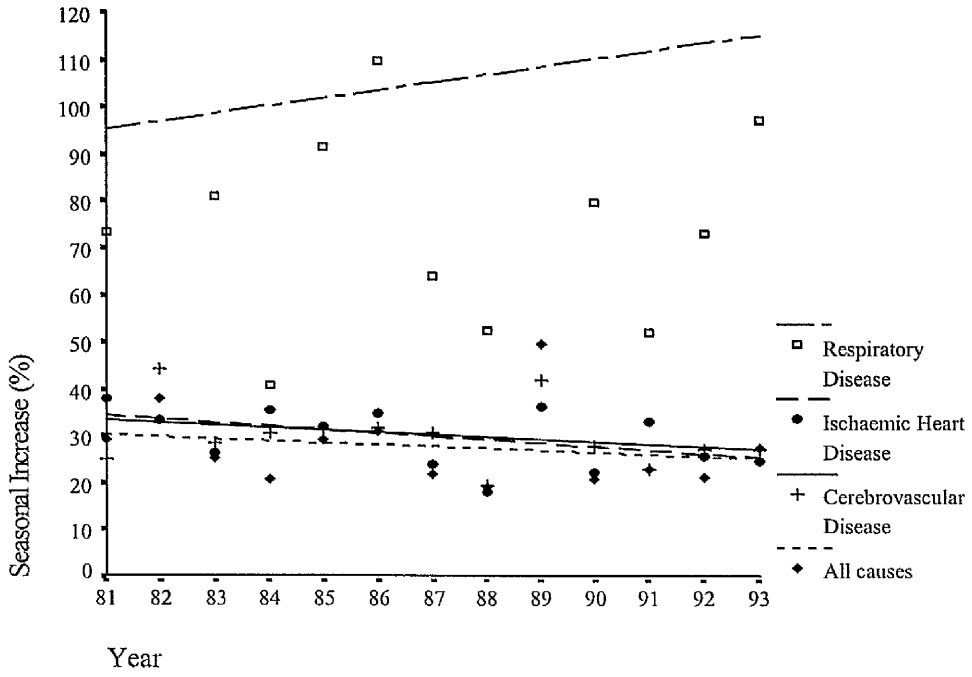


Figure 4.9 Seasonal increase in mortality each year between 1981-1993 (excl 1989)

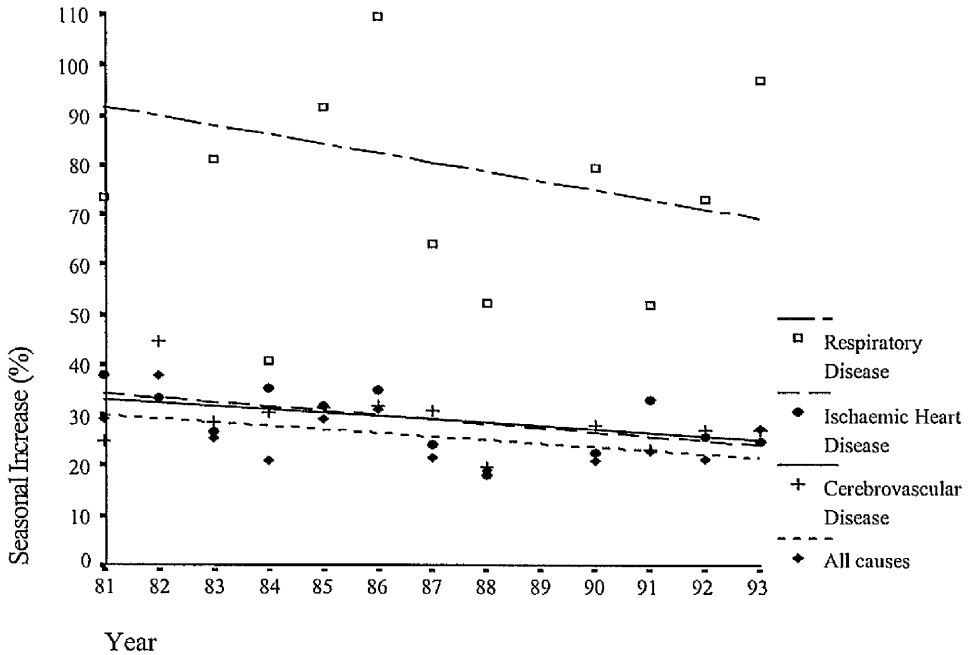


Figure 4.9 shows that there is a clear pattern of a reduction in the size of the seasonal increase in mortality during the time period 1981 to 1993 in Scotland. Details of the fit of the regression lines are given in table 4.2.

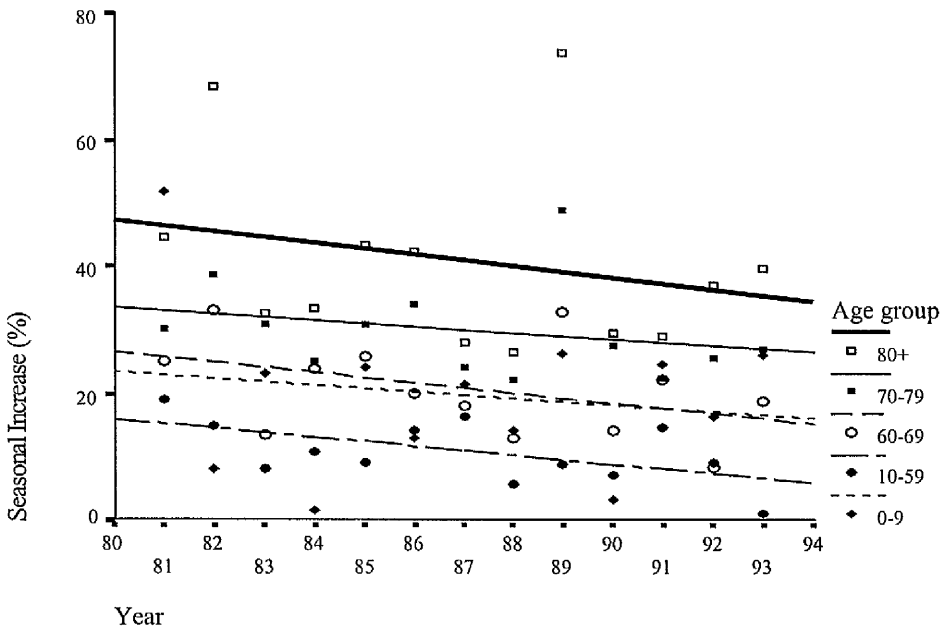
Table 4.2 Regression coefficients and p-values for fits in Figure 4.9

Cause of death	Coefficient	p-value
All causes	-0.365	0.093
RD	-0.034	0.432
IHD	-0.357	0.063
CVD	-0.296	0.143

As can be seen from the table, although there is a consistent downward trend in the regression slopes, none are statistically significant at the 5% level. However, the slopes for all causes and for deaths from IHD are significant at the 10% level.

The data were then examined to see if the slope was the same in each age group. Figure 4.10 shows the seasonal increase in mortality for each year for males by age group.

Figure 4.10 Seasonal increase in mortality each year for males by age group



An analysis of covariance indicated that there was no difference between the slopes for each age group. A similar analysis of covariance to test if the decrease in seasonal variation from 1981 to 1993 in females varied between age groups also showed no difference between age groups.

Given that this trend over time was consistent, but not significant in itself, and that the size of the decrease did not vary between age groups, it was felt that it would be unlikely to vary according to any other socio-demographic variables. Previous work has shown that variation in seasonal mortality varies very little by deprivation category, social class and city of residence. Falls in seasonal mortality have been demonstrated in other countries[26;27] however an inconstistancy in methods and the fact that different methods can produce different results (as demonstrated above) mean the evidence in the UK is at present inconclusive.

4.3 Detailed analysis of temperature effects

4.3.1 Introduction

In this section the relationship between temperature and deaths will be explored in more detail using both Zeger's method and simple summary statistics. In chapter 3 the effect of a change in temperature was the main focus of the analysis, here other features of temperature will be examined. These include changes in temperature at different average temperatures, the effect of changes in temperature during trends in temperature change, the effect of different sizes of temperature change at different average temperatures and the effect of extreme values or outliers in the temperature series.

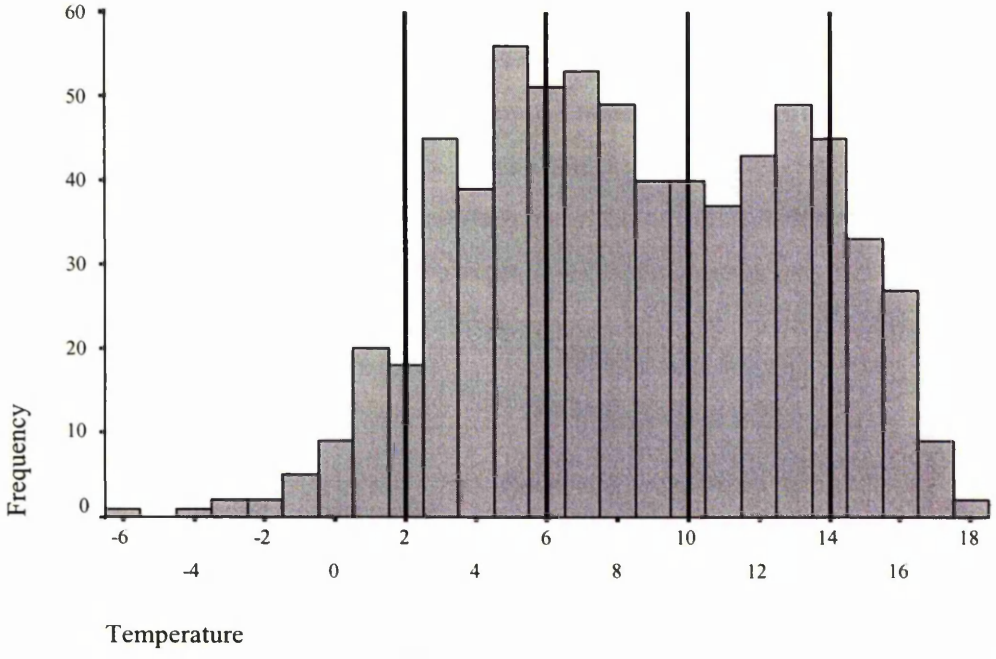
4.3.2 The temperature data

In section 2.2, descriptive statistics of the temperature data for each of the three airports were provided. In this section the average weekly temperature for the whole of Scotland has been used. The aim of this part of the analysis is to examine the effect of changes in temperature at different underlying levels of temperature and to evaluate the effect of temperature change when temperature is generally increasing or decreasing. Both the actual value of the average weekly temperature and the change in value from one week to the next

are involved in the analysis at this stage. In order to carry out this analysis a method of splitting the temperature into categories was required.

There were 3 factors to be considered when determining a suitable range of temperature values for each category. The most important criterion was to ensure that there was a reasonable amount of data in each category. The percentiles of the distribution of temperatures were used to provide indicators of appropriate groupings. Another consideration was to examine the tails of the distribution and not simply divide the distribution into quartiles as the most extreme temperature values would then be overlooked. Two other criteria involved were to have similar temperature ranges in each category and to try to use whole numbers as category boundaries. Analysis was done using groupings of three and five categories for both the change in temperature and the absolute temperature for weekly and daily temperatures. For the weekly data these temperature categories were <2°C, 2-6°C, 6-10°C, 10-14°C and >14°C, and each contained 6%, 27%, 28% 25% and 14% of the temperature values in the full series. This categorisation is shown in figure 4.11.

Figure 4.11 Histogram of weekly temperature data



Most of the analyses described use the categorisation above, however other categorisations were tried and are described further in section 4.3.5.

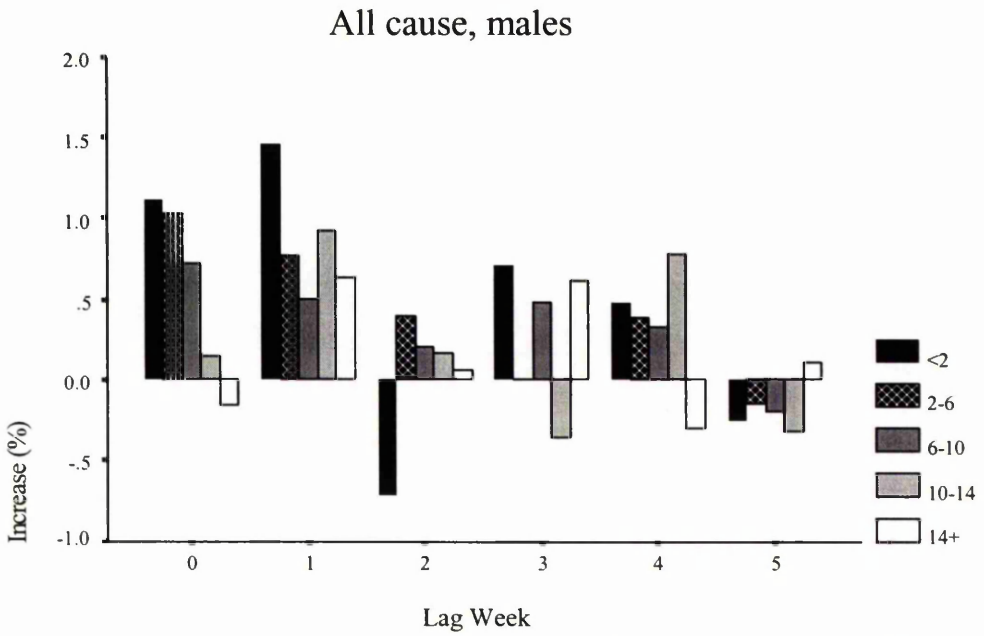
4.3.3 Change in deaths at different temperatures (Zeger's method)

Analysis of the effect of a change in temperature at different underlying temperatures was performed using Zeger's method on weekly data. In order to use Zeger's method effectively the lag variables up to 5 weeks were computed before the data were split into temperature groups. One problem involved in the analysis of the effect of temperature at different temperature categories was that when the data were split into temperature categories the data was broken up into five separate series, the time series was no longer continuous. The autocorrelation structure of the individual series no longer existed in the <2°C and in the >14°C groups due to the fact that they contained less data and therefore contained less data points which had a neighbour to which they were correlated. This may have made Zeger's method unnecessary and a simple Poisson regression method might have been sufficient. However, the autocorrelation structure remained in the other 3 groups and mainly for consistency Zeger's method was used to assess the effect of a fall in temperature of 1°C in each of the underlying temperature categories. There were no methodological problems encountered while using Zeger's method for the <2°C and >14°C temperature categories.

Figure 4.12 shows that a 1°C fall in temperature when the average weekly temperature was less than 2°C was associated with an initial increase in deaths of over 1% and a further increase in deaths of 0.5% in the following week. When the underlying temperature was between 2°C and 6°C a fall in temperature of 1°C was associated with an increase in mortality of 1% initially which fell away 1 week later. A smaller increase in deaths was associated with a fall in temperature of 1°C when the temperature was between 6°C and 10°C. The increase in mortality associated with a fall in temperature of 1°C was significant at the 5% level for all temperature categories one week after the fall and for the <2°C, 2-6°C and 6-10°C categories on the week that the fall in temperature occurred. Two weeks after the fall in temperature when the underlying temperature was <2°C and 2-6°C significant changes in mortality were also found. This was also the case at three weeks after the fall when the underlying temperature was 6-10°C and more than 14°C and at four weeks after when the temperature was 2-6°C and 10-14°C. However, as with the previous charts of this type clear patterns were only really apparent in the first two weeks following the fall in

temperature. When the temperature was greater than 14°C a fall in temperature was associated with a fall in deaths in week 0.

Figure 4.12 Successive percentage change in male deaths from all causes associated with a fall in temperature of 1°C by underlying temperature.



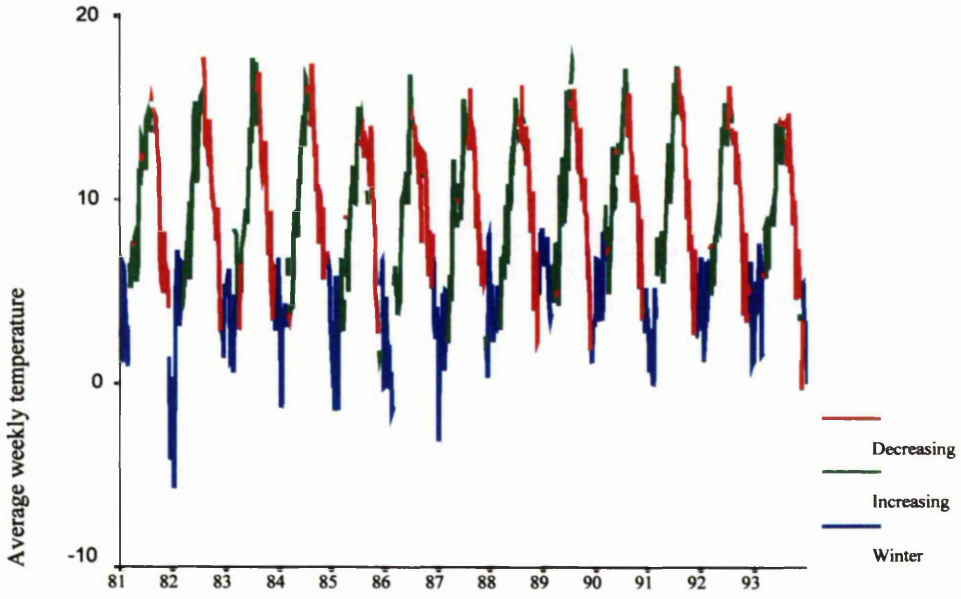
The results demonstrate a U shaped relationship between temperature and mortality which has been discussed by other authors [31; 41; 46; 91]. When temperatures are colder deaths are increased and while the relationship is linear in nature for the most part with increasing temperature leading to fewer deaths this relationship changes when temperatures are very warm where mortality increases.

After the first two weeks in the chart the pattern becomes confusing; this may be due to the fact that the series are broken and contain many missing values or perhaps may be due to a 'harvesting' effect. When the underlying temperature was <2°C, large initial increases in deaths were experienced following a fall in temperature, however after two weeks there was an apparent fall in the number of deaths of around 0.7%; the vulnerable people may have died in the first two weeks and therefore in the following weeks there were relatively fewer deaths than expected. Details of the parameter estimates and standard errors from the above model are in appendix IV (V).

4.3.4 Effect of temperature when types of changing temperature are considered

When one considers the different combinations of temperature that may be detrimental to health there are many scenarios, however from the point of view of statistical modelling of the effect of temperature, these climatic events have to be strictly defined and have to occur frequently in the series. The aim of this section is to determine whether the effect of a fall in temperature of 1°C degree is different in a period of decreasing temperature as compared to a period of increasing temperature. Or, put another way, does a fall in temperature have a greater effect in the Autumn-Winter period as compared to the Winter-Spring period. Methods that divided up the time series into categories based solely on time were undesirable, due to the fact that they were arbitrarily decided upon and were generally determined by month of year. This method of categorisation of a series of temperature values has no specific relation to temperature. In order to measure the temperature effects when there is a particular pattern in temperature the series should, if possible, be categorised into categories determined by temperature. Thus, a method of categorising the series into periods of increasing and decreasing temperature was required. Figure 2.1 showed the weekly average temperature over the time period 1981-1993. While there appeared to be a clear summer peak in weekly average temperature the pattern seemed more erratic in winter. The temperature did not appear to continually decrease to a winter low, then continually increase to a summer peak. To study the effect of a change in temperature due to the general temperature pattern it was decided that three areas needed to be considered. These were periods of increasing temperature, decreasing temperature and erratic winter temperatures. Figure 4.13 shows average weekly temperature in Scotland 1981 to 1993 when categorised as increasing, decreasing or winter. Winter was simply determined as 1st Dec to 28th Feb. In order to get a stable estimate of the general pattern of temperature seven-week moving averages were used to determine whether the temperature was increasing or decreasing. If the change in the seven-week moving average from one week to the next was positive then temperature was said to be increasing, if it was negative then temperature was said to be decreasing.

Figure 4.13 Average weekly temperature (°C)



Initial summary analysis involved looking at the average percentage change in deaths from one week to the next according to the direction of the change of temperature. These results for weekly deaths are shown on table 4.3. The table shows the average change in deaths from one week to the next in each temperature category expressed as a percentage of the average number of deaths per week in the temperature category.

Table 4.3 Average weekly percentage change in deaths in each temperature category

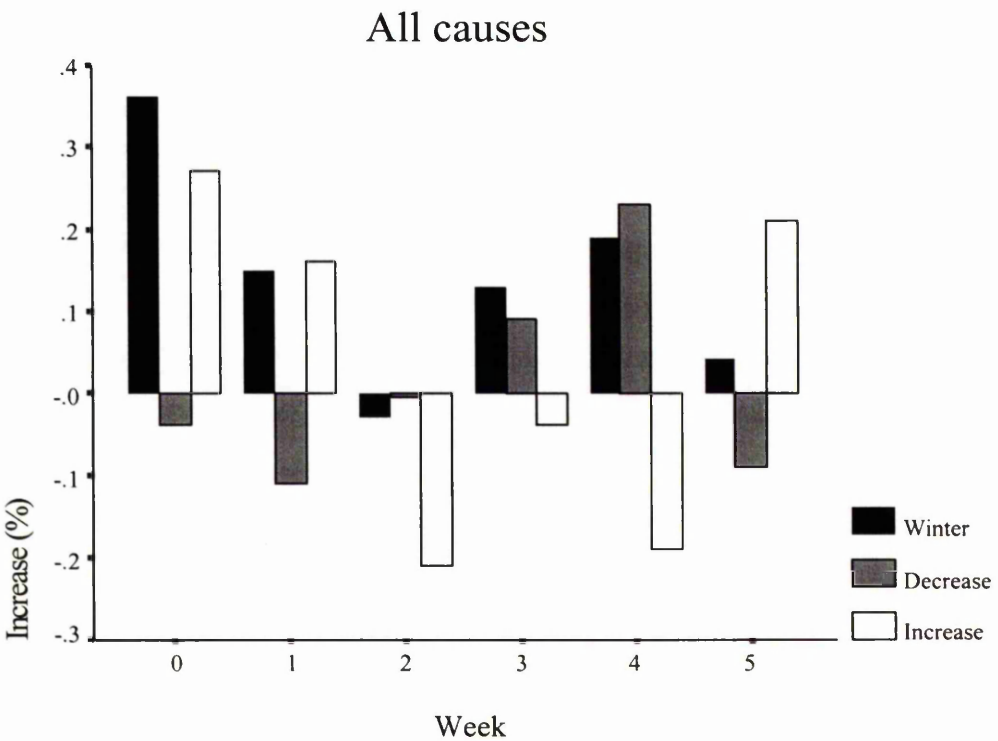
	Males			Females		
Cause of Death	Temperature category (°C)			Temperature category (°C)		
	Increasing	Decreasing	Winter	Increasing	Decreasing	Winter
All causes	-1.08	1.06	0.11	-1.08	1.10	0.12
IHD	-0.91	1.12	-0.29	-1.45	1.53	0.04
CVD	-0.93	0.45	0.56	-0.85	0.79	0.28
Resp	-3.71	3.39	0.99	-3.28	3.28	0.86

When the temperature was increasing the percentage change in mortality was negative and when the temperature was generally decreasing the percentage change in mortality was positive. During winter the change was generally positive although much smaller than in the

periods of decreasing temperature. The percentage change in mortality was greatest for deaths from respiratory disease, followed by deaths from IHD. In this analysis deaths from CVD showed a smaller change than deaths from all causes. These results were expected following previous examination of the seasonal patterns for each cause of death.

To provide a comparison between the effect of a fall in temperature of 1°C in the three categories of changing temperature Zeger's method was used. The results from this analysis are shown in figure 4.14.

Figure 4.14 Successive percentage change in male deaths from all causes associated with a fall in temperature of 1°C by changing temperature.



At first glance the chart seemed to show the opposite to what was expected, however on closer inspection the features shown in the chart reveal an interesting picture. Initial increases in mortality associated with a fall in temperature of 1°C were experienced in winter and in periods of increasing temperature but were not seen when the temperature was decreasing.

Prior to the analysis it was felt that the greatest increase in deaths associated with a fall in temperature would be experienced when the temperature was decreasing. However, in this analysis we are considering the direct effects of a single 1°C fall in temperature. On average when the temperature was decreasing the change in temperature was -0.6°C and when the temperature was increasing the average change in temperature was +0.5°C. In winter the average change in temperature was –0.05°C. Thus in periods of decreasing temperature a fall in temperature of 1°C has little overall initial impact. The impact of this fall is felt a few weeks later, possibly due to the overall lag pattern which exists between temperature and mortality. In periods of increasing temperature a fall in temperature of 1°C when generally temperatures are increasing by 0.5°C has a greater effect on mortality. In winter when the temperatures are cold and variable a fall in temperature of 1°C has a fairly immediate effect and is associated with an increase in mortality of 0.4% in the week that the fall in temperature occurred and a sustained increase in mortality of around 0.1% for several weeks. It would appear from this analysis that falls in temperature have most impact on mortality during cold weather and at periods where the temperature is generally increasing. While this was an unexpected result further thoughts indicate that the observed result is logical. A fall in temperature when temperature is generally increasing has a greater individual impact on mortality than a similar fall in temperature when temperature is generally falling. None of the changes in mortality associated with a fall in temperature of 1°C were significant when the temperature was split into these categories.

4.3.5 Effect of changing temperature at different temperatures

Intuitive feelings about the effect of temperature on health throw up many hypotheses. It is likely that if the temperature is cold and then falls by a large amount this will have a greater effect on health than if the temperature is mild and falls by a large amount or if the temperature is cold and falls by a small amount. However, investigating these theories using real data is not straightforward as we have to consider lag effects and the particular temperature feature must occur often in the series. Table 4.4 shows the number of occasions

in the weekly series that fall into each cell of a 5 by 5 categorisation of temperature and table 4.5 shows the number of weeks that fall into each cell of a 3 by 3 categorisation.

Table 4.4 Number of weeks in each (5 by 5) temperature category

Temperature change (°C)	Temperature category (°C) <2	2-6	6-10	10-14	14+
<-3.5	2	8	17	9	2
-3.5- -1.5	4	30	45	25	26
-1.5-1.5	16	77	99	99	56
1.5-3.5	8	55	21	30	11
>3.5	16	11	5	3	0

Table 4.5 Number of weeks in each (3 by 3) temperature category

Temperature change (°C)	Temperature category (°C) <5	5-12.5	>12.5
<-1.5	27	99	42
-1.5-1.5	70	176	101
>1.5	71	67	22

In table 4.4 there were very few weeks which fell into the more extreme categories such as a large increase from a hot temperature and a large decrease from a cold temperature. It would be wrong to make generalisations of the effect of a fall in temperature at certain average temperatures based on such a small number of cases in some cells. Using the 3-category method shown in table 4.5 the number of cases in each cell was larger, however with this categorisation the extremes of the distribution are hidden.

Using these various different classifications of temperature effects, the average change in deaths or emergency admissions, over one week or for up to five weeks, in each category can be examined. Similar categorisations were produced for the daily data, however in this case the lag effects could be from one day up to more than fifteen days and producing a separate table for each lag time would produce many different results and may not provide a clear answer. The daily data were examined using lag times of three days for CVD, five

days for IHD and fourteen days for respiratory disease (based on the correlation plots in section 3.4). The results from this analysis (not shown) were inconclusive and were likely to be strongly influenced by the choice of lag time. It was decided to concentrate the analysis on the weekly data with a lag time of one week used for each cause of death. Tables 4.6 to 4.9 show the average weekly change in deaths in each of the temperature categorisations while tables 4.10 to 4.14 show the effects of temperature changes at different underlying temperatures on emergency hospital admissions.

Table 4.6 Average weekly change in deaths from all causes by underlying temperature and temperature change

	Males			Females		
Temperature	Underlying temperature			Underlying temperature		
Change	<5°C	5-12.5°C	>12.5°C	<5°C	5-12.5°C	>12.5°C
fell more than 1.5°C	8.61	2.60	-0.2	13.09	4.90	-12.26
other changes	17.22	-2.63	-3.97	4.17	1.22	-0.69
Increased by more than 1.5°C	-14.08	1.69	-2.32	-8.47	-8.03	14.05

When the temperature was cold (<5°C) and fell there was an increase of around 9 male and 13 female deaths per week. When temperature was cold and increased there was a fall in deaths consisting of 14 males and 8 females less than normal. The U shaped relationship is again demonstrated in female deaths, an increase in temperature when the underlying temperature was hot resulted in an increase of 14 deaths per week. The patterns for respiratory disease (table 4.7) and ischaemic heart disease (table 4.8) are similar to that for all causes. One unusual result however is an apparent decrease in female deaths from both respiratory disease and ischaemic heart disease when the temperature was cold and fell. For male deaths from cerebrovascular disease there was little change in any of the categories. For females a large increase in deaths was experienced when the temperature was cold. A fall in the number of deaths was experienced when the temperature was hot and fell and when it was average and increased.

Table 4.7 Average weekly change in deaths from respiratory disease by underlying temperature and temperature change

Temperature change	Males Underlying temperature			Females Underlying temperature		
	<5°C	5-12.5°C	>12.5°C	<5°C	5-12.5°C	>12.5°C
fell more than 1.5°C	1.84	0.77	-1.60	-1.74	2.41	-4.21
other changes	4.28	-0.40	-1.13	5.82	0.47	-0.08
increased by more than 1.5°C	-3.54	0.43	2.36	-3.49	-1.27	1.14

Table 4.8 Average weekly change in deaths from ischaemic heart disease by underlying temperature and temperature change

Temperature change	Males Underlying temperature			Females Underlying temperature		
	<5°C	5-12.5°C	>12.5°C	<5°C	5-12.5°C	>12.5°C
fell more than 1.5°C	1.42	3.87	2.19	-1.65	2.10	1.45
other changes	8.50	-1.30	-2.36	2.88	-0.66	-1.29
increased by more than 1.5°C	-6.94	-3.01	-0.82	-2.98	-2.18	6.45

Table 4.9 Average weekly change in deaths from cerebrovascular disease by underlying temperature and temperature change

Temperature change	Males Underlying temperature			Females Underlying temperature		
	<5°C	5-12.5°C	>12.5°C	<5°C	5-12.5°C	>12.5°C
fell more than 1.5°C	1.44	1.20	-1.40	6.73	1.17	-3.29
other changes	1.35	-1.10	-0.16	-1.27	1.32	0.27
increased by more than 1.5°C	-0.58	0.24	0.95	-1.35	-2.79	-1.95

The results for emergency admissions for respiratory disease (table 4.10) show the opposite to what was expected, with falls in temperature being associated with falls in admissions and rises in temperature being associated with increases in admissions.

Table 4.10 Average weekly change in emergency admissions for respiratory disease by underlying temperature and temperature change

Temperature change	Males Underlying temperature			Females Underlying temperature		
	<5°C	5-12.5°C	>12.5°C	<5°C	5-12.5°C	>12.5°C
fell more than 1.5°C	-2.97	-6.67	-1.88	-7.31	-4.12	2.19
other changes	5.24	-3.36	-2.07	7.33	-3.29	-1.26
increased by more than 1.5°C	7.12	7.04	13.77	5.34	4.63	3.05

One possible reason for this could be that a lag of one week was too short a time to see any effects. Table 4.11 shows the effect on emergency admissions from respiratory disease two weeks after the temperature changes.

Table 4.11 Average two week lagged change in emergency admissions for respiratory disease by underlying temperature and temperature change

Temperature change	Males Underlying temperature			Females Underlying temperature		
	<5°C	5-12.5°C	>12.5°C	<5°C	5-12.5°C	>12.5°C
fell more than 1.5°C	14.27	6.54	3.07	8.00	5.99	6.26
other changes	8.60	-4.34	3.65	14.09	-5.62	3.22
increased by more than 1.5°C	-12.62	-3.69	2.00	-9.76	-1.64	-10.91

Table 4.11 now shows the change in respiratory disease a week after the underlying temperature has changed between any two given weeks. This table shows that the lag time is

indeed longer for emergency admissions for respiratory disease with the U shaped relationship between lagged changes in emergency admissions and temperature again being demonstrated.

Table 4.12 shows the average weekly change in emergency admissions for ischaemic heart disease by underlying temperature and temperature change. Again, this does not appear to demonstrate the pattern expected but if we consider a week after the temperature change (Table 4.13) we find a similar pattern to that for respiratory admissions. The pattern for cerebrovascular disease is shown in table 4.15

Table 4.12 Average weekly change in emergency admissions for ischaemic heart disease by underlying temperature and temperature change

	Males			Females		
Temperature	Underlying temperature			Underlying temperature		
Change	<5°C	5-12.5°C	>12.5°C	<5°C	5-12.5°C	>12.5°C
fell more than 1.5°C	-3.02	-1.55	-3.40	-3.33	-4.88	-5.52
other changes	0.03	-0.83	-1.44	2.53	1.68	-0.13
Increased by more than 1.5°C	8.24	1.93	-5.32	1.59	3.18	0.14

Table 4.13 Average weekly lagged change in emergency admissions for ischaemic heart disease by underlying temperature and temperature change

	Males			Females		
Temperature	Underlying temperature			Underlying temperature		
Change	<5°C	5-12.5°C	>12.5°C	<5°C	5-12.5°C	>12.5°C
fell more than 1.5°C	1.92	5.17	3.21	5.04	1.43	2.33
other changes	2.84	-0.04	-0.50	-0.68	0.34	1.88
Increased by more than 1.5°C	-0.35	-9.73	-8.18	-0.77	-3.78	-10.45

Overall this analysis of the effects of changing temperature at different underlying temperatures has shown that when the temperature was cold and fell there was an increase in deaths. Increases in deaths also occurred when the temperature was mild and fell or was cold and changeable. Furthermore, the number of deaths also increased when the temperature was hot and increased. This pattern is more pronounced for deaths from all causes, respiratory disease and ischaemic heart disease than deaths from cerebrovascular disease. For emergency admissions the effect was lagged by one week for respiratory disease and ischaemic heart disease and negligible for cerebrovascular disease.

Table 4.14 Average weekly change in emergency admissions for cerebrovascular disease by underlying temperature and temperature change

	Males			Females		
Temperature	Underlying temperature			Underlying temperature		
Change	<5°C	5-12.5°C	>12.5°C	<5°C	5-12.5°C	>12.5°C
fell more than 1.5°C	1.82	0.96	-1.02	-0.21	2.87	-4.10
Other changes	-2.00	1.59	0.30	0.49	0.04	-0.50
Increased by more than 1.5°C	-1.69	-2.15	-2.14	-1.36	-0.46	-1.18

4.4.6 Effects of extreme changes in temperature

The previous section was concerned with general changes in temperature, this section of the thesis examines the effect on mortality of extreme temperatures and extreme temperature changes. An extreme temperature change was defined as a change in the average weekly temperature of more than 4 degrees. There were 19 weeks (2.8%) in which the weekly average temperature dropped by more than 4 degrees from the previous weekly average temperature and there were 21 weeks (3.1%) in which the weekly average temperature increased by more than 4 degrees from the previous weekly temperature. The average temperature reached after a fall in temperature of 4°C was 2.45°C, after an increase of 4°C the temperature was on average 8.07°C. The average percentage change in deaths from one week to the next in each of these categories of temperature change is shown in table 4.15. The greatest percentage change in deaths from one week to the next following a fall in

temperature of 4°C was for male deaths from cerebrovascular disease, when an increase of 18% was experienced. Percentage increases for respiratory disease and IHD were around 3-5% and female deaths from CVD increased by around 9%. It would appear that while in previous analyses cerebrovascular disease has failed to show any strong direct relationship with temperature it is only at the extreme temperature changes that effects are noticed in mortality from CVD.

Table 4.15 Percentage change in deaths after one week

Cause of death	Males			Females		
	Temperature change (°C)			Temperature change (°C)		
	<-4.00	>4.00	other	<-4.00	>4.00	other
All causes	3.80	-0.97	0.15	2.98	-0.48	0.15
Resp	2.83	3.14	1.74	4.95	-1.63	1.97
IHD	5.18	-4.46	0.65	3.81	-2.39	0.70
CVD	7.91	-0.4	1.18	8.86	-1.63	0.90

The percentage change at a lag of two weeks is given in table 4.16. Again, after a lag of two weeks the percentage change in mortality associated with an extreme fall in temperature is around 10% for deaths from CVD, 5% for female deaths from IHD and negligible for deaths from respiratory disease.

Table 4.16 Percentage change in deaths after two weeks

Cause of death	Males			Females		
	Temperature change (°C)			Temperature change (°C)		
	<-4.00	>4.00	other	<-4.00	>4.00	other
All causes	2.45	1.29	0.20	4.60	2.14	0.15
Resp	3.14	9.20	2.02	5.58	3.66	2.67
IHD	2.16	-1.33	0.77	7.71	-2.93	0.77
CVD	13.75	1.29	1.35	13.01	0.08	0.91

The average number of deaths per week for each of these causes varies, so while the percentage change in deaths makes the figures comparable the actual effect on the health of the population is not calculated. The average number of deaths per week in each of the temperature categories is shown in table 4.17.

Table 4.17 Average number of deaths per week

Cause of death	Males			Females		
	Temperature change (°C)			Temperature change (°C)		
	<-4.00	>4.00	other	<-4.00	>4.00	other
All causes	623.48	644.33	583.29	661.22	675.86	616.26
Resp	75.23	84.90	67.30	76.12	85.95	70.43
IHD	200.79	206.71	184.55	171.98	170.14	155.42
CVD	69.33	69.33	59.40	113.29	108.29	100.29

The increase in mortality of 18% following a fall in temperature of over 4°C for deaths from CVD is equivalent to over 12 extra male deaths and an increase of 9% is equivalent to 10 extra female deaths in a week. For IHD the increase in deaths following a fall in temperature of over 4°C is 10 male deaths and over 6 female deaths, with corresponding figures of 2 male deaths and around 4 female deaths from respiratory disease. Table 4.17 also shows that the average weekly number of deaths is higher when the extremes of temperature occur than at other levels of temperature change. This would indicate that the extreme increases and decreases in temperature occur in the winter months.

4.4 The effect of pollution on seasonality of deaths in Glasgow

4.4.1 Introduction and literature review

There have been many studies in the USA concerned with environmental pollution and ill health, however there have been considerably fewer similar studies carried out in Britain and, of those in the UK, the majority are concerned with London. McFarlane[92] used data on daily average concentrations of suspended particles (smoke) and sulphur dioxide(SO_2), minimum and maximum daily temperature and daily deaths in greater London to study the association between pollution and mortality. The method used to determine an association was to visually examine the time series and identify peaks and troughs occurring around the same time. Using this method no association was found between pollution and mortality, however an association between cold weather, influenza epidemics and mortality from all causes was identified. Another study[93] using daily data from 1958 to 1972 in London used regression methods and time series methods and found a highly significant relationship between mortality and either particulate matter or sulphur dioxide after controlling for temperature and humidity. More recently Bremner[94], using a method adopted by the APHEA project, demonstrated no association between the main pollutants and all cause mortality in London between 1992 and 1994. However, the authors did find an association with each pollutant and respiratory mortality and with some pollutants and cardiovascular mortality.

Bates[12] reviewed the evidence for a causal relationship between pollution and ill health in a BMJ editorial. While many studies, particularly in North America, have consistently shown a statistical relationship, he found no evidence of a biological mechanism for the relationship between ill health and PM_{10} particulates. PM_{10} is defined as particles less than 10 microns in size. However, a biological mechanism for ozone was more plausible since ozone at very low concentrations induces lung inflammation, and this may also be true for nitrogen dioxide(NO_2). In the same issue Buchdahl[95] used Poisson regression to examine the relationship between pollution and the incidence of acute childhood wheezy episodes in a

London hospital. He showed that after seasonal adjustments daily average concentrations of ozone(O_3) and sulphur dioxide were significantly associated with the incidence of acute wheezy episodes. Anderson[96] used Zeger's method to identify the association between particles, sulphur dioxide, ozone and nitrogen dioxide and daily mortality in London between 1987 and 1992. The analysis was adjusted for trend, season, day of the week, influenza epidemics, temperature, humidity, autocorrelation and overdispersion. Ozone levels and black smoke were significantly associated with increased mortality from all causes, the effects were smaller and less consistent for nitrogen dioxide and sulphur dioxide. Livingstone[97] studied asthma rates and proximity to main roads (a proxy for traffic related pollution) and found no relationship while in Scotland, Mackay[98] also found no relationship between monthly variation in atmospheric pollution and variation in asthma mortality. In a recent study Atkinson[99] analysed daily counts of visits to A&E departments for respiratory complaints. Using a Poisson regression model which adjusted for seasonal patterns, meteorological conditions and influenza epidemics the authors assessed the effect of various pollutants on the number of A&E visits. They found strong associations between visits for all respiratory complaints and increases in SO_2 and PM_{10}. They also found significant associations between visits for asthma and SO_2, NO_2 and PM_{10}.

Schwartz[100-103]has conducted several studies in the USA concerned with hospital admissions and air pollution. He showed that in both Minneapolis-St.Paul, Minnesota and Birmingham, Alabama, PM_{10} and ozone were significantly associated with hospital admissions for pneumonia and chronic obstructive pulmonary disease in the elderly. PM_{10} and carbon monoxide(CO) were significantly associated with hospital admissions for cardiovascular disease in the elderly in Tucson, Arizona. There was no significant association for sulphur dioxide, ozone or nitrogen dioxide. PM_{10}, ozone and sulphur dioxide were all found to be significantly associated with daily admissions to hospital for respiratory disease in the elderly in New-Haven, Connecticut and Tacoma, Washington. However, the magnitude of the risk was small for all three pollutants and the effect of SO_2 was diminished when the model controlled for the other pollutants. In each of these studies Schwartz used Poisson regression including time trends, seasonal fluctuations, weather variables and lagged

pollution variables. In a similar paper Moolgavkar[104] investigated the association between air pollution and hospital admissions in both Mineapolis-St.Paul and Birmingham from 1986 to 1991, but included more pollution data than Schwartz had analysed. The authors reported little evidence of association between air pollution and respiratory causes in Birmingham, but found a strong relationship between ozone and hospital admissions in Mineapolis-St.Paul as well as some association between hospital admissions and PM_{10}, SO_2 and NO_2. Moolgavkar[105] also considered air pollution and daily mortality in Philadelphia from 1973 to 1988. He concluded that weather, total suspended particles, sulphur dioxide and ozone were all associated with mortality, but because of the high correlation between the pollution variables commented that it was difficult to single out one component as being responsible for the observed association between air pollution and mortality. Morris[106] examined the relationship between air pollution and hospital admission for congestive heart failure among elderly people in seven US cities. Using a negative binomial regression model to account for over-dispersion, congestive heart failure admissions were modelled accounting for temperature, year, month, day of the week and exposure variables lagged for 7 days. The pollution variables used were sulphur dioxide, carbon monoxide, nitrogen dioxide and ozone. There was found to be a significant relationship between CO_2 and hospital admissions in both a single-pollutant and a multi-pollutant model, no other pollutant showed a consistent relationship with hospital admissions in both models and in all cities.

There have been relatively few studies of the effects of air pollution on health in Europe. One study[91] considered the association between mortality from all causes and SO_2 levels between 1979 and 1987 in the Netherlands. The log linear regression model included confounders such as temperature, precipitation, humidity, wind speed, influenza incidence, and calendar year, month and day. The authors found that when all the confounders were taken into account SO_2 did not seem to have any short term effect on mortality in the Netherlands. A study in Amsterdam[107] using data from 1986 to 1992 used Poisson regression to examine the association between mortality and black smoke, PM_{10}, sulphur dioxide, carbon monoxide and ozone. The regression model controlled for trend, seasonal patterns, year, month, day, influenza type illnesses, temperature and humidity. The greatest

associations were found for black smoke and PM_{10}, ozone was less strongly linked to mortality while there was no association with SO_2 and CO. Ponka[108] investigated the relationship between air pollution and hospital admissions for ischaemic heart disease and cerebrovascular disease in Helsinki from 1987 to 1989. Poisson regression analysis controlling for temperature and humidity were used to assess the effect of nitric oxide, nitrogen dioxide, sulphur dioxide, ozone and total suspended particles. The authors found that emergency admissions for IHD were significantly associated with nitric oxide and ozone, while emergency admissions for CVD were significantly associated with nitrogen dioxide. Long-term transient myocardial ischaemic attacks were related to particulates and short-term ischaemic attacks were related to nitrogen dioxide. Using time series methodology, Touloumi[109] investigated the association of air pollution with all cause mortality in Athens from 1984 to 1988. The authors adjusted for year, season, day of week, temperature and relative humidity as well as for serial autocorrelation in mortality. The results showed that while SO_2 and smoke were related to mortality, this relationship was weaker than that for temperature and humidity. After accounting for these variables there was no apparent effect of CO on mortality.

The majority of the papers concerned with environmental pollution and mortality have shown a relationship between PM_{10}, O_3 and mortality, with many also showing a relationship between other pollutants such as SO_2, NO_2 and CO. Several of the studies produce conflicting results, this may be because of the different methods used or the fact that in some cases important confounders were not included in the models. The investigation of the association between mortality and pollution in Glasgow is restricted to two pollutants, NO_2 and CO, as these were the only variables available during the time period.

4.4.2 Description of variables and summary statistics

Daily measurements of nitrogen dioxide in Glasgow were available for the period January 1987 to Dec 1993 (2380 days) and carbon monoxide measurements were available for the period July 1989 to March 1993 (1245 days). Measurements were collected at the City Chambers in Glasgow City centre. Figure 4.15 shows a plot of the CO measurements while figure 4.16 shows the levels of NO_2 in Glasgow.

Figure 4.15 CO in Glasgow City (parts per billion)

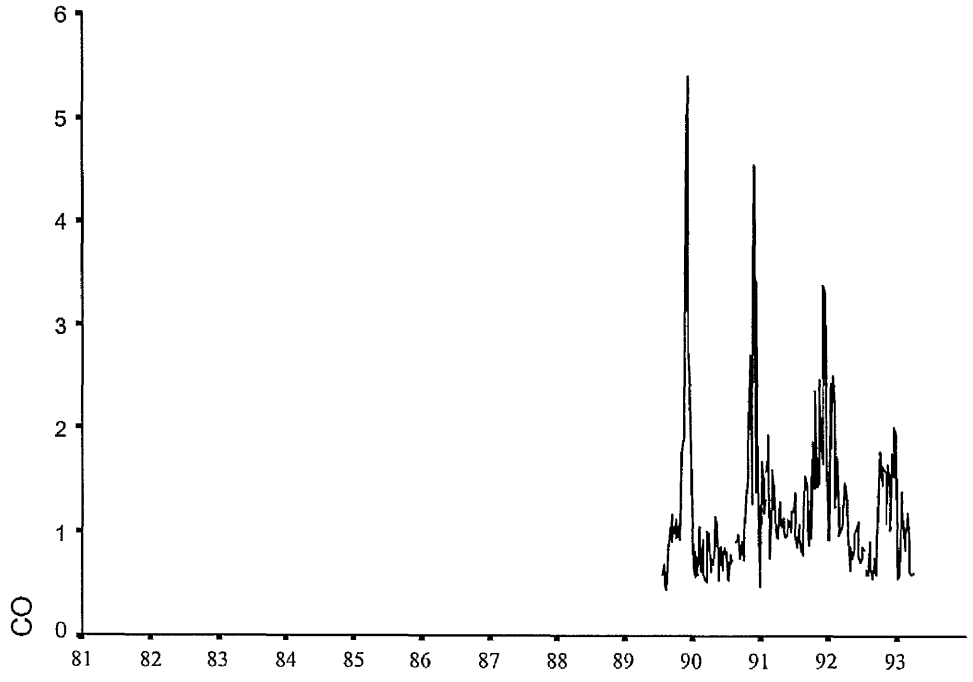
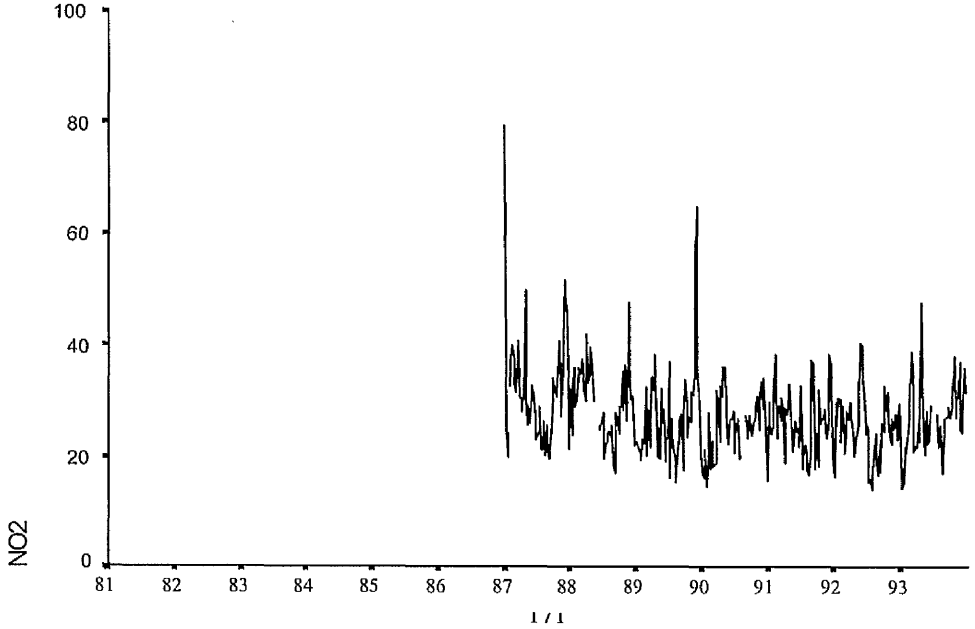


Figure 4.16 NO2 in Glasgow City (parts per billion)



When considering the deaths and emergency admissions data for this work, cases were included in the data set if they were resident in the Glasgow City Local Government District. The weekly average and total number of deaths in this area over the time period January 1981 to December 1993 from the specific causes of interest are detailed in table 4.18

Table 4.18 Weekly average and total deaths between 1981 and 1993

	Males		Females	
Cause of death	Average	Total	Average	Total
Resp	13.73	9279	14.77	9985
IHD	29.91	20220	26.85	18151
CVD	9.07	6133	16.25	10983
Flu	1.44	52	2.13	143

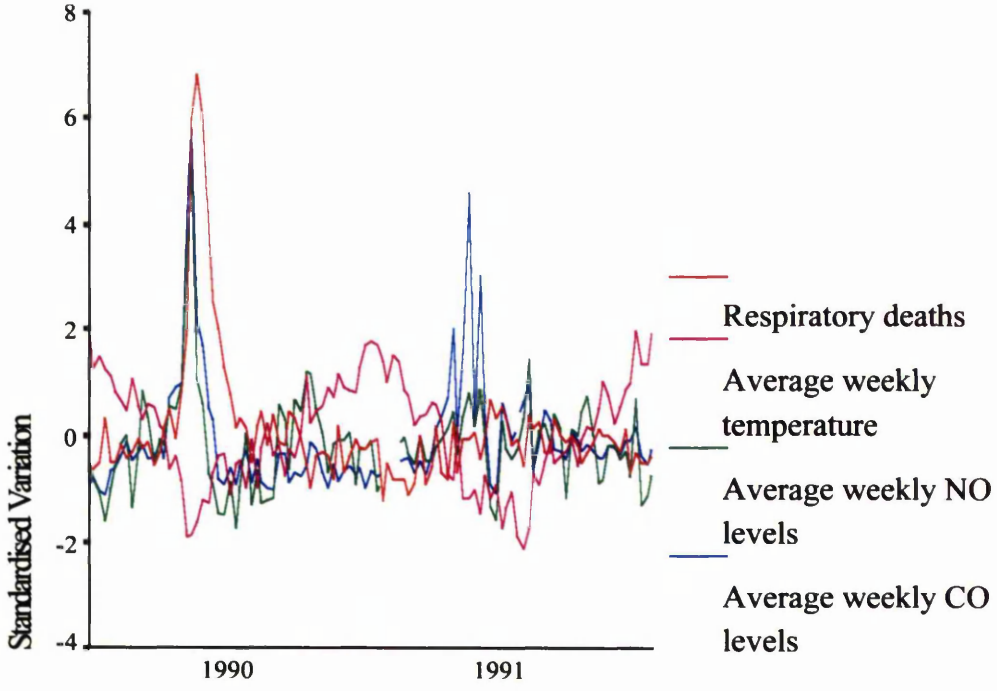
It was felt that this size of area would tend to experience a level of pollution close to that measured in the city centre and would be large enough to provide a sufficient number of cases. Table 4.19 shows the average daily values of CO and NO_2 emissions (parts per billion)

Table 4.19 Average daily values of CO and NO2 emissions (ppb)

	Average	Minimum	Maximum
CO	1.24	0.46	5.40
NO2	27.46	14.29	79.50

Figure 4.17 shows a plot of the standardised variation around the mean for each of the four variables; respiratory mortality, average weekly temperature (measured at Glasgow airport), average weekly nitrogen dioxide levels and average weekly carbon monoxide levels.

Figure 4.17 Standardised variation around the mean



Average weekly temperature rarely exceeds 2 standard errors from the mean throughout the whole time period. Both the average weekly nitrogen dioxide levels and carbon monoxide levels show a large positive deviation from their mean around December 1989 and there is a correspondingly large increase in deaths from respiratory disease in Glasgow City at this time. However, if we look at December 1990 there is a large peak in carbon monoxide levels but little change in the number of deaths from respiratory disease in Glasgow City. Also around January 1991 there is a sharp peak in nitrogen dioxide levels, but again no change in the number of deaths from respiratory disease. An influenza epidemic occurred in the winter of 1989/90 and this may help to explain part of the large peak in respiratory deaths around this time.

From a plot such as this it may be possible to select a small time period and hypothesise about the relationship between pollution and mortality, but a modelling technique which accounted for all confounders was necessary to establish the true nature of the relationship between the variables of interest.

The two pollutants were highly correlated with each other as well as both being highly correlated with average daily temperature. The correlation between CO and NO_2 was 0.6126, for CO and temperature it was -0.5346 and for NO_2 and temperature it was -0.3791, these were all significant at the 1% level. As was done with the initial analysis of the temperature data, the values of the pollutants were lagged and the lagged values were correlated with daily deaths from the three causes of interest for both males and females. The results of this lagged correlation are shown in figures 4.18 and 4.19. For CO there seems to be a stronger relationship with deaths from respiratory disease than from any other cause, with a lag of around 30-40 days. The relationship between CO levels and deaths from IHD and CVD are weak and display no real patterns. It can be seen from the plot for NO_2 that there is very little correlation between deaths from CVD and NO_2 levels. Deaths from respiratory disease and IHD seem to display a lagged effect of NO_2, with the peak in the strength of the relationship occurring at a lag of around 20 days. However, the size of the correlation coefficient in each case is less than 0.1, so any pattern which does exist is weak.

Figure 4.18 Correlation between CO and mortality

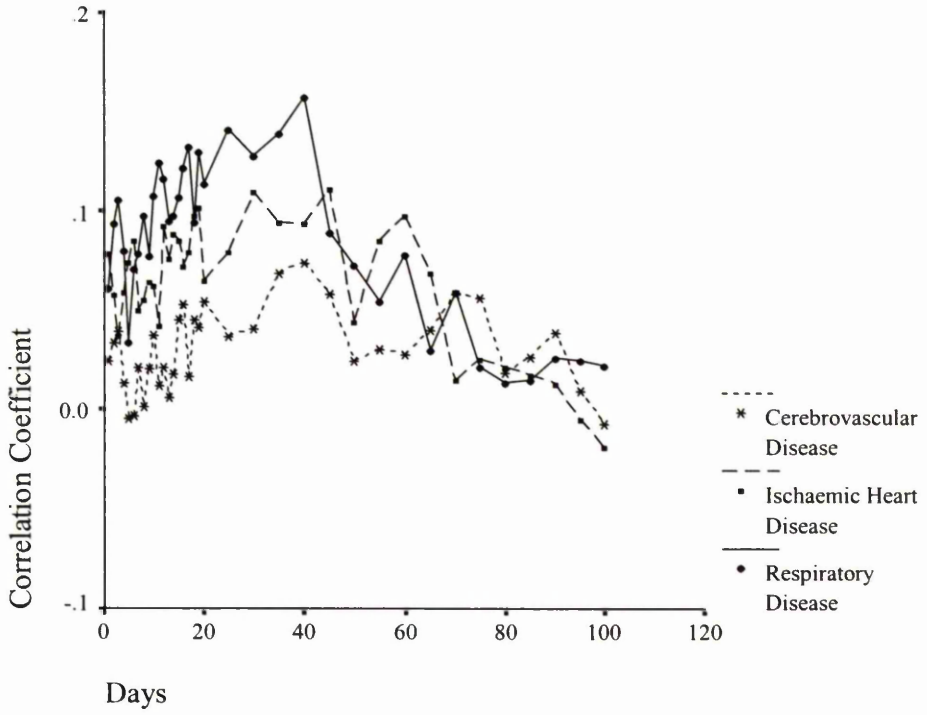
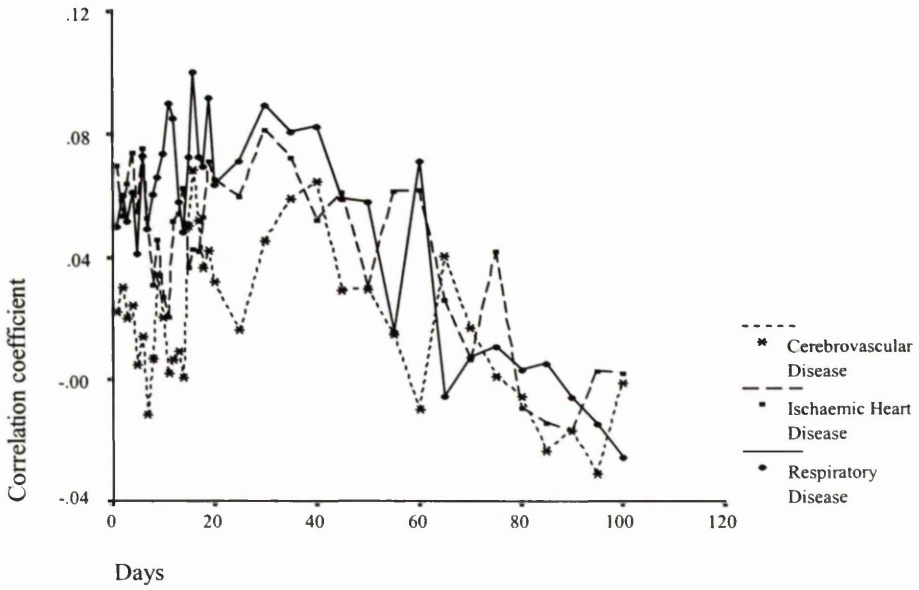


Figure 4.19 Correlation between NO2 and mortality



The patterns for the pollutants are not as clear as those for the temperature data. This may be to do with the number of data points that contribute to each correlation.

Autocorrelation plots and partial ACF plots showed that the pollution data and the temperature data for Glasgow City followed an ARIMA(0,1,1) model. Similar methods to those used in section 3.3.3 were used to assess the structure of the data and as a result Zeger's method was felt to be the most appropriate for use in the analysis.

4.4.3 Results of modelling

The relationship between the three variables, temperature, NO_2 and CO meant that modelling the effect of the variables on the pattern of deaths was complicated. Initial analysis involved using Zeger's method for the weekly data and fitting each variable lagged for 5 weeks in a separate model that also included flu deaths. When assessing the relationship between temperature and mortality, a fall in temperature of 1°C was used as a summary measure. When modelling the relationship between temperature and pollutants it would be inappropriate to compare the effect of a fall of 1 unit in each of the variables as the scales on which the variables are measured are different. A change in levels of CO of 1 part

per billion is a large change, whereas a fall in temperature of 1°C is a fairly small change. Therefore the models for the pollution analysis are concerned with the relative effect of an average change in each of the three variables. These are a fall in temperature of 1.9 degrees, an increase in nitrogen dioxide levels of 5 parts per billion and an increase in carbon monoxide levels of 0.4 parts per billion. The results from the models are shown in Figure 4.20. These results were estimated using a separate model for each variable. There was little effect at week zero from all variables, however a fall in temperature of 1.9 degrees was associated with an increase in mortality from respiratory disease in Glasgow City of 6% after 1 week. Average increases in NO_2 and CO levels were associated with an increase in deaths from respiratory disease in Glasgow City of around 3-4%, occurring 2-3 weeks after the increase in levels.

Having assessed the effect of the variables using separate models, a combined model including all three variables; temperature, NO_2 and CO was fitted. The results from this model are shown in figure 4.21 for deaths from respiratory disease in Glasgow City.

Figure 4.20 Percentage change in respiratory mortality, Glasgow City

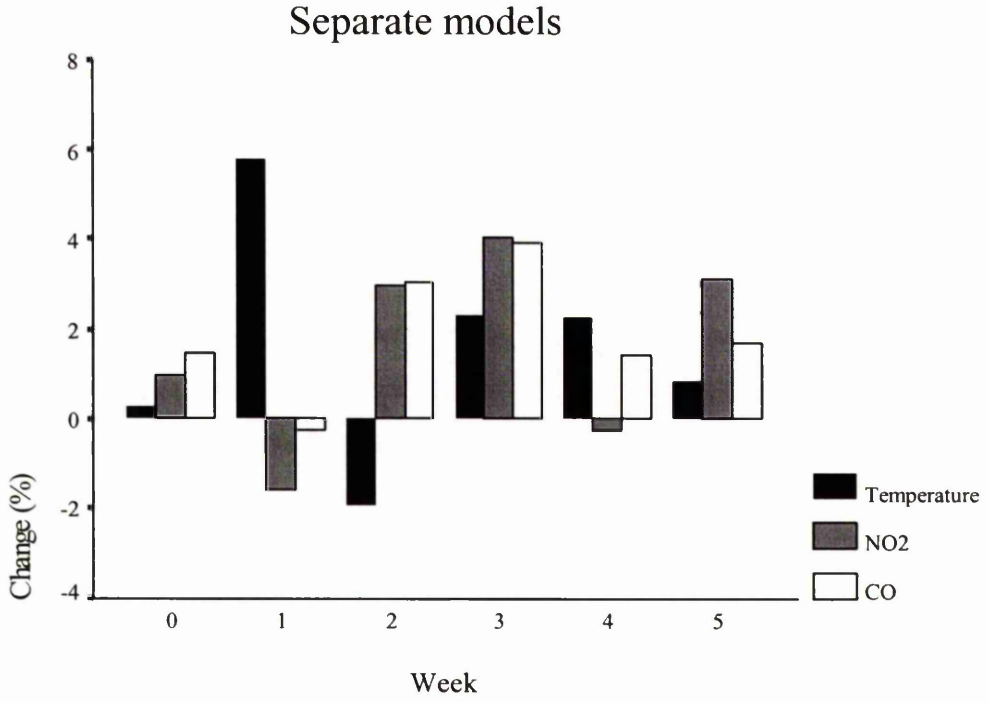
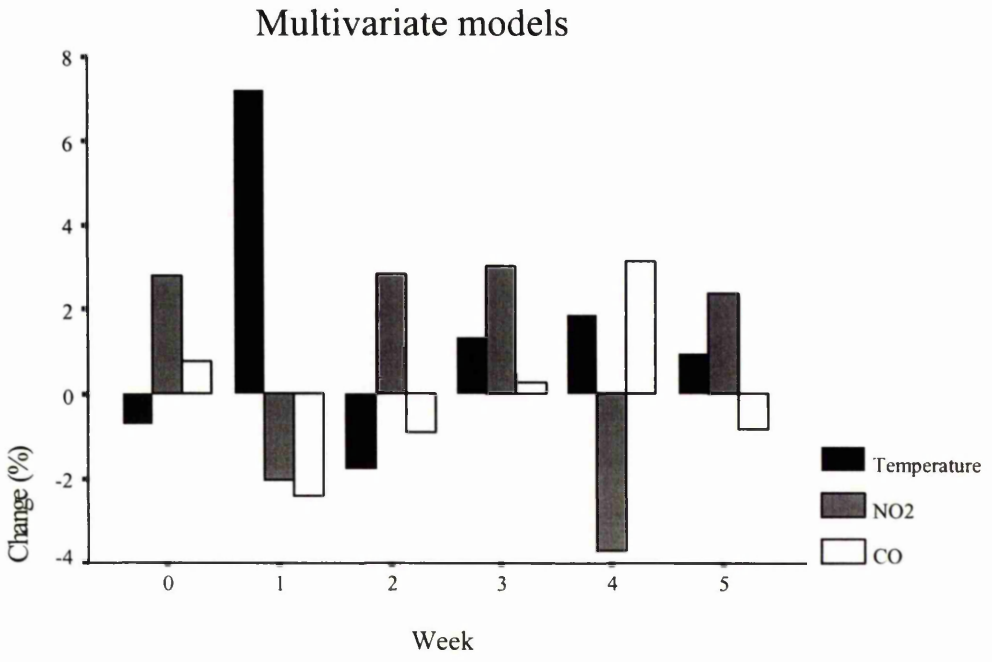


Figure 4.21 Percentage change in respiratory mortality, Glasgow City



The relationship between respiratory mortality and outdoor temperature was considerably stronger than the relationship with pollution levels. A fall in average weekly temperature of around 2°C in Glasgow Citywas associated with an increase in respiratory mortality of over 6%. Changes in pollution levels, having accounted for the effects of temperature, were associated with variations in mortality from respiratory disease of around 2%. These results show that while considered individually, pollutants such as nitrogen dioxide and carbon monoxide may be found to be associated with respiratory mortality, their effect is diminished when confounders such as outdoor temperature are included in the analyses.

4.5 Summary & Discussion

The analysis described in this chapter examined how the relationship between temperature and mortality and morbidity varied according to demographic and social variables. The relationship between temperature and mortality varied according to age group, the effects of a fall in temperature were most pronounced one week after the fall in temperature and in the very young age group and the older age groups. The effect was greatest in the very old age

group. This pattern was evident in deaths from all causes, respiratory disease, ischaemic heart disease and cerebrovascular disease, but for these latter two causes the very young age group, 0-9, was excluded due to the small number of deaths. There was also a general pattern of a greater effect of temperature on emergency admissions in the older age groups.

The effects of deprivation and social class on the relationship between temperature and mortality were similar. The effect of temperature was slightly greater and occurred later in social classes III, IV and V compared to social classes I & II and in the deprived category compared to the affluent category for deaths from all causes. For deaths from respiratory disease the effect of a fall in temperature was greatest in social classes IV & V in the week that the fall in temperature occurred, in the week after the fall in temperature the effect was greatest in social classes I, II & III. There was little difference according to social class for deaths from CVD and IHD.

Overall, however, the effect of socio-economic status was minimal. The initial measure used for this analysis was Carstairs[71] deprivation categories. The results from this measure, both in the effect on seasonal patterns and on the direct temperature effect, were contrary to prior hypotheses which expected that seasonal patterns and the effect of temperature on mortality would be greater amongst people living in deprived areas. When no evidence of this was found, social class data, for males only, was used to determine if there was a socio-economic effect that was not being highlighted by deprivation category analysis. The main difference between the two scores is that one is an area based measure and the other is measured at the individual level. The social class analysis demonstrated more of a gradient than the deprivation analysis, but again this was not significant. These findings support a recent paper by Shah who demonstrated that there was no relationship between deprivation and seasonal patterns in mortality in London[78].

For male deaths from all causes it appeared that the effect of temperature on mortality was greater in Aberdeen and Edinburgh than it was in Glasgow, however there was no consistent pattern existing in each causal group and the patterns in females did not complement the patterns seen in males. It is likely that much of the variation (or city effects) could be simply random variation.

A recent paper by Dominici[110] describes a method of analysis in which evidence of temporal relationships between two variables in different cities can be combined. The paper uses two models, one in which the city specific coefficients are considered independent and the other which introduces a spatial correlation between city coefficients. The paper uses data on air pollution and mortality in 20 US cities and combines the evidence to give an average effect for the whole of the USA. This modelling approach may have been appropriate for the analysis of temperature and mortality in the three Scottish cities. It may have provided different results to the analysis using averaged data however, it was considered inappropriate to introduce a new modelling method at this stage.

There was a U shaped relationship between temperature and mortality, the effect of a fall in temperature was greatest at colder temperatures and fell as the underlying temperature fell. At warmer underlying temperatures a fall in temperature resulted in a fall in mortality. Falls in temperature had the greatest effect on mortality in winter and in periods when the temperature was getting warmer.

The analysis in this chapter considered the effects of temperature on mortality and morbidity in a variety of different scenarios. These were; temperature changes at different underlying temperatures and at changing patterns of temperature; the effect of the size of change of temperature at different underlying temperatures and the effect of extreme changes in temperature. The analysis demonstrated that different patterns of temperature changes have different effects on mortality. However, there are so many possible combinations of temperature variability that it is virtually impossible to quantify the effect of types of temperature changes on mortality.

Pollution data was available on two pollutants nitrogen dioxide(NO_2) and carbon monoxide(CO) for Glasgow City. Initial analysis showed that pollution had a delayed effect on mortality with an increase in deaths of 3-4% occurring 2-3 weeks after the increase in pollution levels. However, the pollution levels were highly correlated with temperature levels and when temperature was included in the model the effect of pollution was greatly reduced. Information on other pollutants such as PM_{10} and SO_2 would be necessary for a

thorough investigation of the relationship between pollution, temperature and mortality in Glasgow City to be carried out.

Overall, this analysis has demonstrated a consistent effect of falls in temperature being associated with increases in mortality and morbidity, however attempting to assess how this relationship varies according to other variables is not clear. Only the pattern with age and the U shaped relationship with underlying temperature were clear.

Chapter 5 Space-Time Analysis

5.1 Introduction

5.1.1 Introduction & aims

Various methods that could be used to describe the temporal relationship between temperature and ill health were discussed in Chapter 3 and Zeger's method was found to be the most suitable for this type of problem. This method was then used in Chapter 4 to examine the relationship in more detail using other demographic features such as age at death/admission, socio-economic status and city of residence. There was found to be no real difference in the relationship between temperature and mortality/morbidity when comparing the three main Scottish cities, Glasgow, Edinburgh and Aberdeen.

Having assessed the temporal relationship between ill health and temperature it was of interest to examine how this temporal relationship varied spatially throughout Scotland. The relationship may be stronger in some areas of the country compared to others, for example northern areas compared to southern areas or rural compared to urban areas. There may also be a space-time relationship whereby the strength of the relationship varies spatially throughout time.

Knox[111] used a χ^2 test to try to detect a space–time relationship in the incidence of cleft lip and palate in Northumberland and County Durham. The method involved producing a contingency table of pairs of incident cases grouped according to their distance apart, measured both in time and space, and comparing the observed and expected values. However, Knox commented that since the pairs were not independent of each other the χ^2 test was not an ideal method of testing for space-time interaction. In a further paper Knox[112] used a similar method to assess the space-time variation in childhood leukaemia in Northumberland and Durham, his results indicated a possible space-time variation. Since his original work, the Knox method has been modified to various degrees and applied in the examination of space-time interactions in incidence of rare diseases[113,114]. However, one weakness of this method is its dependence on the space and time limits set by the investigators. More recently Bhopal[115] used an extension of Knox's method to examine the space-time clustering of non-outbreak cases of legionnaires' disease in

Scotland. This method developed by Diggle[116] involved estimating a function of space and time where $\lambda K(s,t)$ is the expected number of further events within distance s and time t of an arbitrary event and comparing this to the observed number of events. The authors found evidence of clusters in space and time in Glasgow and Edinburgh. This method was also used in an analysis of space-time clustering of new cases of rheumatoid arthritis in East Anglia[117]. The authors found no evidence of space-time clustering but some evidence of spatial clustering in one area.

Generally, the methods described above have been developed to deal with relatively rare diseases where each incident case was recorded at a specific location. A further epidemiological application of space time modelling has been in the analysis of outbreaks of infectious disease such as measles[118; 119] and rubella[120]. The most common model being the SEIR (Susceptible/Exposed/Infectious/Recovered) model which has been studied extensively in mathematical epidemiology[121-123].

However, the data used in this thesis are not records of rare events or general practice notifications of infectious disease epidemics, but records of all cases of mortality and morbidity that can be summarised by area. Historically the space-time analysis of epidemiological data aggregated to an area level was restricted to mapping methods and involved looking at the changing patterns seen in the maps over time. However, maps of disease rates often simply consist of shaded areas representing disease rates in regions. Problems with this technique are that the rates do not take account of the effect of the size of the underlying population in the region. A rate for a region with a large population will be more reliable than a rate for a region with a small population. These maps can also be misleading, as larger geographical areas are more dominant visually. The population in a large area may be smaller than that in a geographically smaller but more densely populated region. In order to provide a more analytical approach to dealing with this type of data, time-series methods such as data transformation, detrending and autocorrelation plotting have been applied to spatial data. These methods have been used in the analysis of Scottish lip cancer data[124], North Carolina sudden infant death syndrome data[125] and lung cancer rates in Sardinia[124; 126].

The aim of this chapter is to extend the work described in chapters 2, 3 & 4 taking into account the spatial aspect of the data. It has been shown that mortality and emergency

hospital admissions for respiratory disease, ischaemic heart disease and cerebrovascular disease vary over time. It is also of interest whether they vary over space and finally whether there is a space-time interaction. It was felt unlikely that death rates from ischaemic heart disease or cerebrovascular disease would display a time varying spatial component. Emergency admissions for respiratory disease were felt to be the most suitable of the disease groups to use in a space-time analysis as these may display more spatial variation, possibly reflecting patterns of infectious disease spreading through the population.

In section 5.2 the spatial aspects of the pattern of emergency admissions for respiratory disease are examined using summary methods such as spatial lag correlation and mapping techniques. Section 5.3 introduces spatial trend and spatial correlation techniques, using weekly death rates in each of the 56 local government districts in Scotland, while section 5.4 uses these methods in Glasgow City and considers the extent to which emergency hospital admissions varied according to postcode sector. Finally, Carstairs deprivation score is introduced as a predictor variable in the postcode level analysis.

5.1.2 Types of Spatial Data

There are three main categories of spatial data; geostatistical data, spatial point patterns and lattice data. With geostatistical data, measurements are usually taken at fixed locations that are spatially continuous. Examples of this type of data are mineral concentrations, rainfall at weather stations or concentrations of pollutants at monitoring stations. Spatial point patterns occur when the locations themselves are of interest and often spatial randomness or clustering is the main outcome of interest. Examples of this type of data are locations of earthquake epicentres or analysis of cancer incidence around nuclear facilities. Lattice data are data that consist of observations associated with spatial regions, where the regions can be regularly or irregularly spaced. An example of this type of data is disease rates in regions of a country. The lattice is defined by a set of vertices and edges, in some cases the lattice can be referenced by the centroids of the regions. Neighbours can be defined by distances or by common boundaries and neighbour relationships are often weighted based on distance between the centroids, the length of

the common boundary or the number of common boundaries a region possesses. The data analysed in this chapter can be described as lattice data, the spatial regions being either health boards, local government districts or postcode sectors.

Having determined the spatial nature of the data, the level of aggregation at which the data is to be analysed had to be considered. In a space-time analysis the data can be aggregated in both the spatial and the temporal dimension. At the highest level a simple figure which described the mean rate over time and space could summarise the data and at the lowest level of aggregation each individual record could be examined separately. With the Scottish morbidity data the highest level of aggregation from which useful information could be gained would be to assess the spatial variation in disease rates by health board in Scotland for each year from 1981 to 1993, or the trend over time (years) in each health board. This type of analysis gives a broad picture of the data, shows long term trends over time and identifies areas which have disease rates which were generally greater or lower than the Scottish average between 1981 and 1993. As the level of aggregation decreases, the generalisability of the results is reduced but the level of detail is increased. Other levels of aggregation that can be considered are weekly disease rates by health board and weekly disease rates by local government district. The lowest level of spatial aggregation that was sensible to consider in this study was the analysis of disease rates for postcode sectors in Glasgow.

5.1.3 Epidemiological mapping of lattice data

The spatial aspects of disease incidence or prevalence can be represented very simply on maps, examples can be found in the Scottish Cancer Atlas[127] and the European Mortality Atlas[128]. Most epidemiological maps use political or governmental boundaries such as international borders, administrative regions, health boards and postcode areas as the boundaries of the areas for which the disease rate is to be mapped. Epidemiological maps fall into two categories, firstly there are maps which show the disease distribution at a single point in time. For these maps, standardised mortality rates (SMR's) or age standardised incidence rates are assessed for each area and plotted on the map. This is often a simple but effective means of arriving at a hypothesis about the aetiology of a

disease. The other form of mapping diseases involves constructing maps that show the spread and retreat of a disease over time. The first type of mapping is used mainly for non-infectious diseases such as cancer whereas the second type of map is used for mapping infectious diseases such as measles, influenza and respiratory disease.

However, when mapping the raw rates one needs to be aware that areas with small populations will have larger standard errors and therefore produce less reliable estimates of disease rates than more densely populated areas. Another complaint about these maps is that each of the small areas stand alone which is not the case in the real world. Methods of overcoming these problems include smoothing the maps so that the influence of surrounding areas is accounted for. These methods involve techniques where the estimate for an area is the combination between the SMR of the small area and the SMR of the larger area in which the small area lies[129-132]. The mapping of the spread and retreat of a disease generally involves mapping the disease incidence over successive time periods or plotting the difference between values in successive periods to produce maps which show rates of change[118; 133].

Mapping methods are used in spatial analysis as tools to indicate where there may be spatial variation. More advanced statistical techniques are necessary to investigate the nature and strength of any observed spatial variation. Cliff[118] provides a good introductory text to spatial comparisons over time in which he uses the example of a measles outbreak in Cornwall. Simple summary statistics such as spatial lag correlation profiles are computed for this data before more complex statistical techniques are used.

5.2 Emergency respiratory admissions by Health Board (HB)

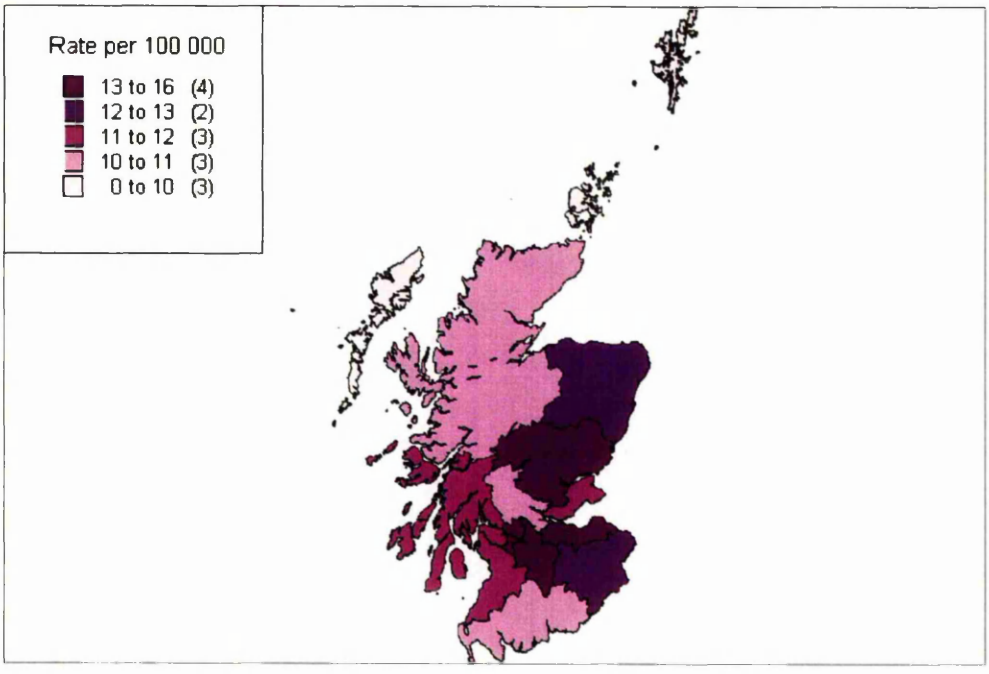
5.2.1 Introduction

It was felt that, of all the disease groups that had been studied in the course of the analysis, emergency admissions for respiratory disease may be the most likely to display any space-time component, since they incorporated the effects of influenza epidemics. Due to small numbers it was not possible to study emergency admissions for influenza alone. There were almost 100 000 emergency admissions for respiratory disease in Scotland between 1981 and 1993 and only around 3 000 emergency admissions recorded

as influenza. The age standardised rates for each of the 15 health board areas had already been calculated for use in the analysis reported in Chapter 2 and it was fairly straightforward to plot these figures on a map using the mapping package MapInfo.

Figure 5.1 shows the age and sex standardised rate for emergency admissions for respiratory disease during the whole time period 1981-1993. The map shows that there are relatively low emergency admission rates for respiratory disease in the Islands, average rates in the Western side of the country with slightly higher rates in Grampian and the Borders. The highest rates were found in Greater Glasgow, Lothian, Lanarkshire and Tayside. Initially it seems that there is a west-east trend with the rate being higher in the east. However it is more likely that what we are observing here is an urban/rural effect with rates being higher in the urban areas. This map provides a good starting point from which to consider the spatial as well as temporal aspects of emergency respiratory admissions.

Figure 5.1 Age & sex standardised emergency admission rate for respiratory disease in Scotland 1981-1993

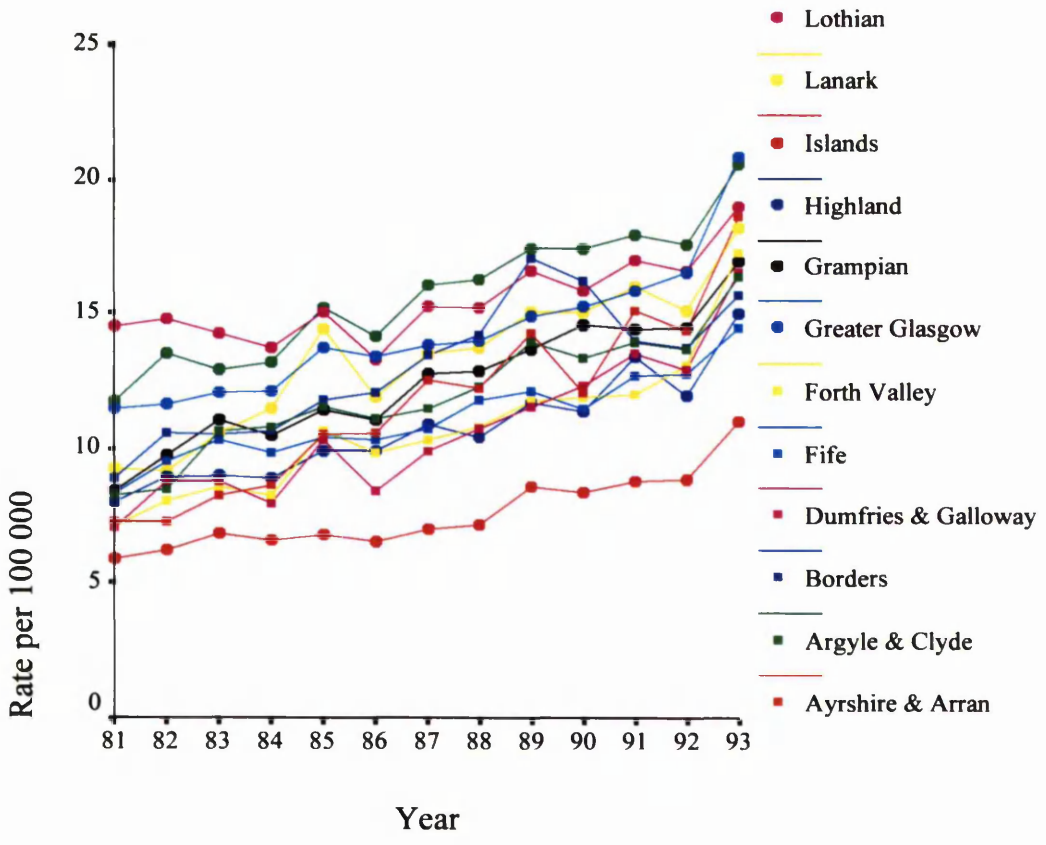


5.2.2 Yearly Trends

The next step of the exploratory analysis was to look at the pattern of emergency hospital admissions for respiratory disease by health board and year. As can be seen from Figure 5.2 the emergency admission rate was increasing in each health board area between 1981 and 1993.

The lowest rates were in the Islands for each year and rates did not seem to be increasing as quickly in these health boards as in the others. All health boards, apart from the Islands, experienced an increase in emergency admissions for respiratory disease in 1985 and 1989 with a reduction in rates experienced in most health boards in 1992. Generally each health board appears to be following a similar temporal pattern but each at its own underlying rate.

Figure 5.2 Age & sex standardised emergency admission rate for respiratory disease by year and health board.



5.2.3 Weekly maps during an Influenza epidemic

Having examined the emergency admission rate in each health board by year, the temporal aggregation of the data was reduced to a weekly level in order to unearth more complex patterns. However, due to potential difficulties in identifying patterns between each of the 13 health boards, when considering all 676 weeks in the 13-year period a specific time period was chosen for analysis. In the winter of 1989/90 there was a substantial influenza epidemic that was associated with the large increase in mortality and morbidity which occurred in that winter. The 8 weeks from mid November 1989 to mid January 1990 were chosen for the weekly analysis of emergency admissions for respiratory disease by health board because it would be during these weeks that a space-time pattern would be most likely to be identified, if indeed one did exist.

The 8 maps showing the age standardised rate per 100 000 population in each of the 15 health boards from week 46 in 1989 to week 1 in 1990 are shown in Figure 5.3. In week 46 the emergency admission rate is less than 20 per 100 000 in the rural areas and between 20-25 /100 000 in the more urban central belt. In week 47 most health boards had experienced an increase in emergency admissions, with the increase greatest in Tayside and Dumfries & Galloway (D&G). There was not much change between weeks 47 and 48, Fife and Forth Valley still had a very low emergency admission rate, most areas had a rate of between 20–30/100 000, Tayside was still experiencing a larger number of emergency admissions, while the initial increase in D&G had moved northwards towards Ayrshire & Arran(A&A) and Greater Glasgow HB. In week 49 most areas had increased emergency admissions for respiratory disease, again Tayside displayed the highest rate of 50-60/100 000, in the Borders and Argyle & Clyde the rates were 40-50/100 000 and the rate in D&G had fallen back to 10-20/100 000. By week 50 the rate in Tayside was beginning to fall, Borders had a rate of 50-60/100 000 and the rate in A&A was 40-50/100 000. In week 51 the epidemic had died down in most health boards apart from a slight resurgence in Tayside and in week 52 the only area with a rate of more than 10-20/100 000 was Borders HB which had a rate of 40-50/100 000.

Figure 5.3a Standardised emergency admission rate for respiratory disease in week 46, 1989

Figure 5.3b Standardised emergency admission rate for respiratory disease in week 47, 1989

Figure 5.3c Standardised emergency admission rate for respiratory disease in week 48, 1989

Figure 5.3d Standardised emergency admission rate for respiratory disease in week 49, 1989

Figure 5.3e Standardised emergency admission rate for respiratory disease in week 50, 1989

Figure 5.3f Standardised emergency admission rate for respiratory disease in week 51, 1989

Figure 5.3g Standardised emergency admission rate for respiratory disease in week 52, 1989

Figure 5.3h Standardised emergency admission rate for respiratory disease in week 1, 1990

By the start of January 1990 the epidemic had almost completely died out and the emergency admission rate for respiratory disease was less than 20/100 000 in all health boards except Tayside and Lothian, where the rate was 20-30/100 000.

From this mapping procedure it is possible to build a picture of how emergency admissions for respiratory disease rose and fell in each health board area over the 8 week period in which there was a substantial influenza epidemic. The main restriction with this analysis however is the size of the areas, some such as Highland cover a large area with a sparse population, whereas Greater Glasgow covers a small area but includes a large population. These variable sizes of areas can result in a map inaccurately reflecting the situation in the Scottish population as the map is visually dominated by large areas with small populations.

5.2.4 Spatial comparison of weekly time series

A spatial lag comparison[118] was carried out to see whether the data displayed any space-time component. A spatial lag comparison allows the comparison of spatial data without knowing the distance between the two areas, all that is required is the number of 'steps' or 'lags' between the two areas. Figure 5.4 shows the health boards in the form of a node where the common boundary, a spatial lag, is shown as a link with weight one.

Figure 5.4 Graphical representation of the 15 health boards of Scotland

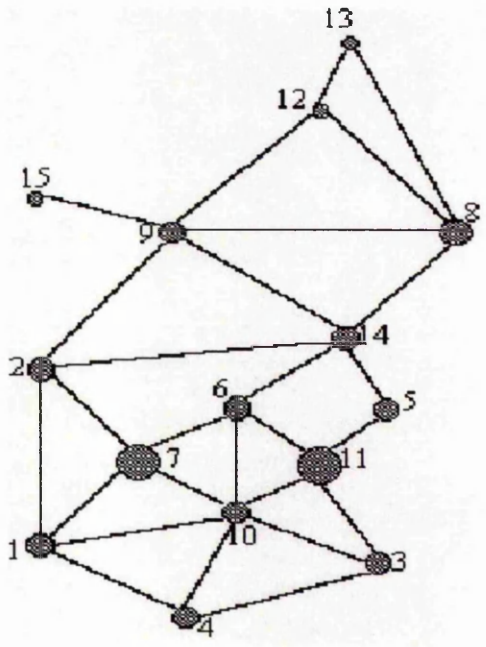


Table 5.1 shows the 'distances' between the 15 Scottish health boards in terms of spatial lags. Two health boards are one spatial lag apart if in figure 5.4 there is a direct link between them, they are two spatial lags apart if, to get from one to the other, two links have to be used. Or in terms of an area map, such as figure 5.1, health boards are one spatial lag apart if they share a common boundary and are two spatial lags apart if, to get from one to the other a different health board area has to be crossed.

Table 5.1 Order of spatial lags separating health board areas in Scotland

Health Board	HB	Health Board Number														
Name	No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A & A	1	0	1	2	1	3	2	1	4	2	1	2	3	4	2	3
A & C	2		0	3	2	2	2	1	2	1	2	3	2	3	1	2
Bord	3			0	1	3	2	2	4	4	1	1	5	5	3	5
D & G	4				0	3	2	2	4	3	1	2	4	5	3	4
Fife	5					0	1	2	2	2	2	2	3	4	1	3
FV	6						0	1	2	2	1	1	3	3	1	3
GG	7							0	3	2	1	2	3	4	2	3
Gram	8								0	1	3	3	1	1	1	2
High	9									0	3	3	1	2	1	1
Lanark	10										0	1	4	4	2	4
Lothian	11											0	4	4	2	4
Orkney	12												0	1	2	2
Shetland	13													0	2	3
Tayside	14														0	1
Western	15															0

The 15 health board areas for Scotland allow 105 pairwise comparisons to be made between the time series of emergency admissions for respiratory disease in each area. For each pair of areas a correlation coefficient which corresponds to the correlation between the two time series is calculated.

Figure 5.5 Mean correlation between series of male emergency admissions by HB

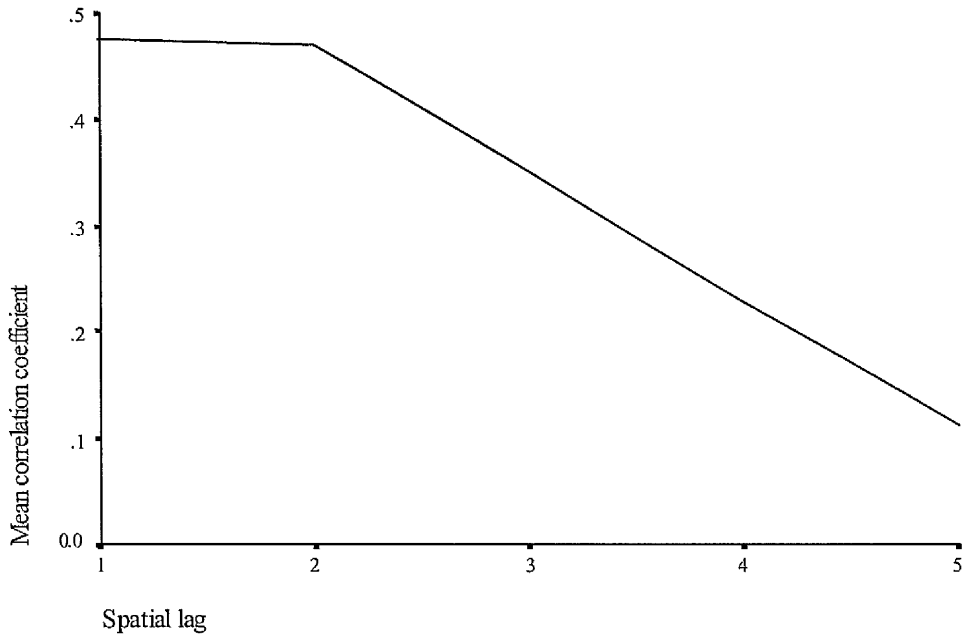


Figure 5.5 shows a plot of the mean correlation between the time series of all pairs of health boards located at each spatial lag. Health boards which share the same boundary or are separated by two spatial lags have a mean correlation of around 0.475, then the size of the average correlation between the time series decreases steadily as the spatial lag increases. At the maximum distance apart, spatial lag 5, the correlation between the time series of emergency hospital admissions is around 0.1.

This fairly straightforward analysis of spatial variation has indicated that there may be a spatial aspect to the change in the numbers of male emergency admissions over time. In order to look more closely at this possible space-time relationship the data were analysed by local government district for the period during the influenza epidemic.

5.3 Emergency respiratory admissions by Local Government District

5.3.1 Weekly Maps during an Influenza Epidemic

In the following analysis the aggregation remained at the weekly level in terms of the temporal component however the spatial component was reduced from health boards to local government districts (LGD's). The weekly age and sex standardised rate of emergency admissions for each of the 56 local government districts was calculated and a similar analysis was performed. The maps shown in figures 5.6a-h show the emergency admission rate for respiratory disease in each of the LGD's over the same 8 week period.

In week 46 of 1989 the rate of emergency admissions for respiratory disease was highest in western parts of Scotland, with rates of 20-30 per 100 000. The Northeast area experienced average rates of emergency admissions for respiratory disease of 11-16 per 100 000, while in the rest of the country the rates were between 0 and 16 per 100 000. By week 47 the emergency admission rate has increased in most areas, the areas showing the highest rates are in the centre of the country, and there are few areas with rates of less that 10 per 100 000. In week 48 the effect of the flu epidemic is becoming apparent, there are now several areas with emergency admission rates for respiratory disease of between 28 and 40 per 100 000. These areas, however, appear to be dispersed and there is little evidence of a spatial pattern in the maps. In week 49 the areas with the highest rates are experiencing rates of 32–50 per 100 000, these areas tend to be in the central highland area of Scotland. Immediately around this area the local government districts are experiencing fairly high emergency admission rates of 23-32 per 100 000, with the more outlying areas experiencing the lowest rates of 0-23/100 000. Only one area, the Western Isles has not been affected by the general increase in emergency admissions over Scotland. By week 50 there are still some areas with emergency admission rates of 30-50/100 000, these areas are those which in the previous week were in the 2nd highest category. The areas that previously were in the highest category now have reduced rates. In week 51 the highland areas with the highest rates in week 49 now have the lowest rates and those in which the epidemic started later now have the highest rates.

Figure 5.6a Standardised emergency admission rate for respiratory disease in week 46, 1989

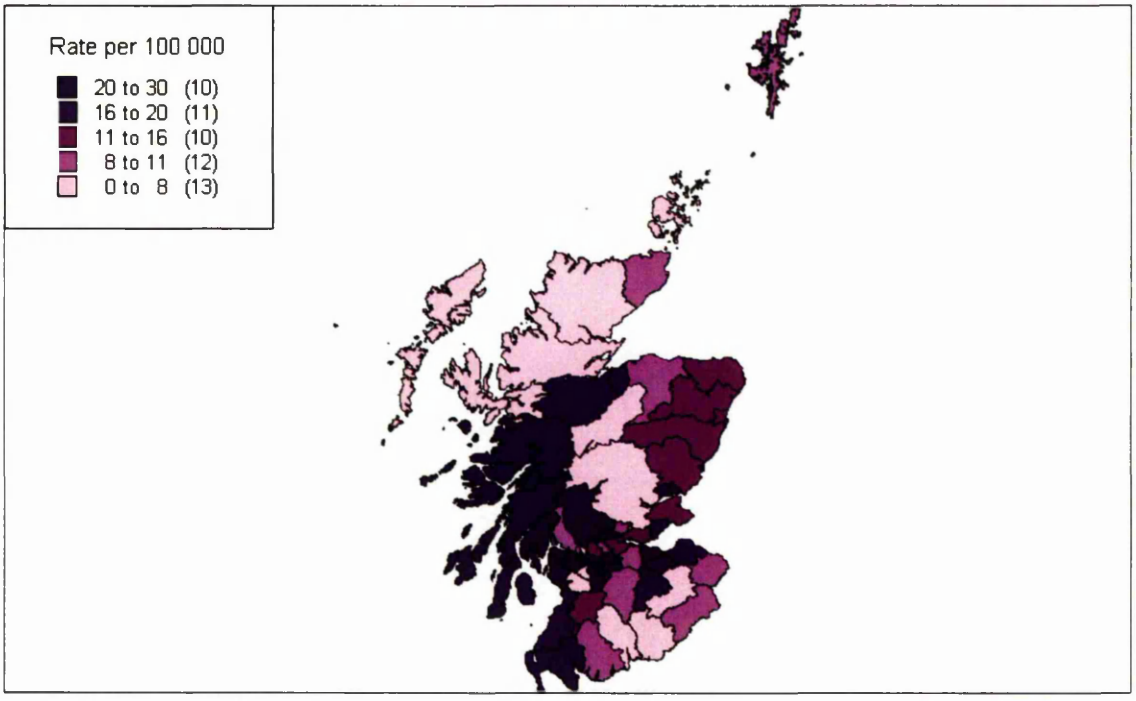


Figure 5.6b Standardised emergency admission rate for respiratory disease in week 47, 1989

Figure 5.6c Standardised emergency admission rate for respiratory disease in week 48, 1989

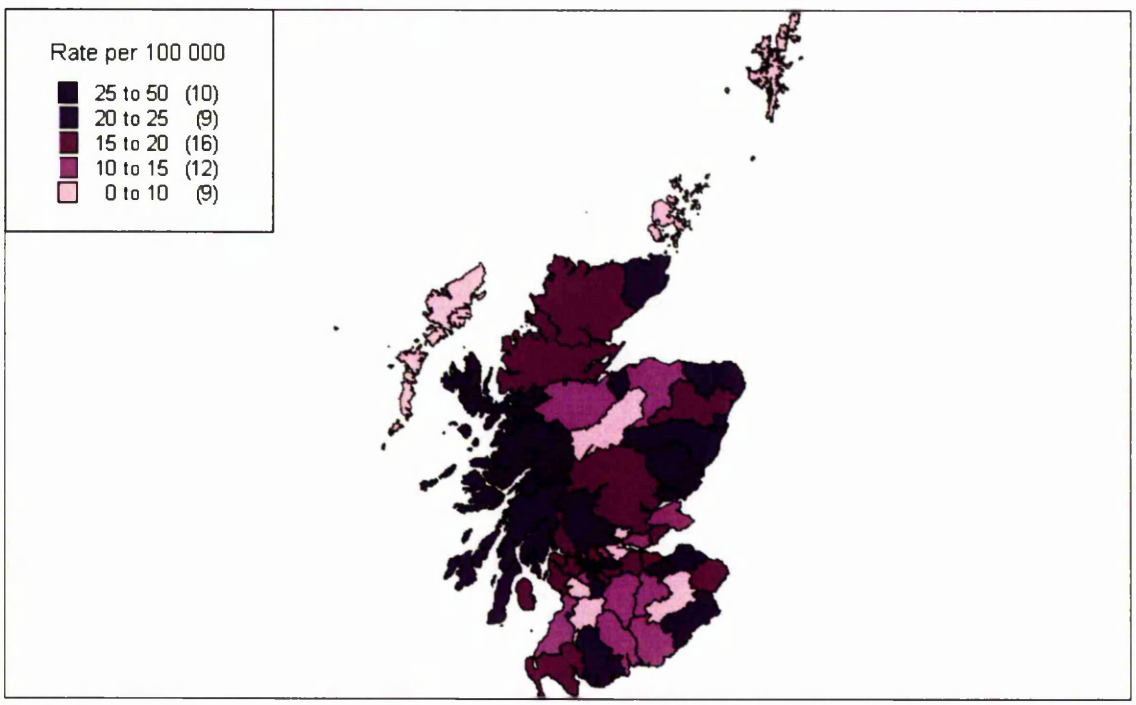
Figure 5.6d Standardised emergency admission rate for respiratory disease in week 49, 1989

Figure 5.6e Standardised emergency admission rate for respiratory disease in week 50, 1989

Figure 5.6f Standardised emergency admission rate for respiratory disease in week 51, 1989

Figure 5.6g Standardised emergency admission rate for respiratory disease in week 52, 1989

Figure 5.6h Standardised emergency admission rate for respiratory disease in week 1, 1990



By week 52 the rates are slightly higher in the NorthEast, the Northwest and the Borders. By the first week of 1990 the rate of emergency admissions for respiratory disease is below 25 per 100 000 population in most areas and the epidemic appears to have subsided.

The information gathered from studying these maps over the 8 week period is more valuable than the health board maps. Making the assumption that the flu epidemic that occurred during the winter of 1989/90 was driving the pattern of emergency admissions around the country, these maps show that the epidemic was greater in different areas throughout the 8 weeks. The epidemic spread throughout the whole country, with all areas showing a higher level than normal at some point, but it was strongest in a few distinct but separate areas which experienced a large increase in emergency admissions. Finally, by week 1 of 1990 the epidemic appeared to be in its later stages and had reduced to almost normal winter levels. In order to study the changing pattern of the epidemic it can also be of interest to study the change in the rate of emergency admissions over the 8 week time period. Maps showing the change in rates in each area from the previous week were produced, but added little to the information gained from Figures 5.6a-h. A better way to examine the spatial properties of the rate of emergency admissions for respiratory disease is to use spatial statistical methods.

5.3.2 Spatial Analysis of lattice data

The first step in beginning the statistical analysis is to introduce a model on which to base the analysis. One such model is

data = spatial trend + error

This model makes as few assumptions about the data as possible, however the errors should be second order stationary. For spatial data the conditions of second order or weak stationarity are equivalent to those for time series data i.e. the error has constant mean zero, constant variance and the covariance between two areas is only dependent on the distance (and perhaps direction) between the areas.

Before any spatial analysis was performed the distributional properties of the data were checked. Normal probability checks were carried out on the age standardised emergency admission rates for respiratory disease for each of the 56 regions over each of the 8 weeks, the data were found to be normally distributed after a log transformation. This transformation is the most commonly used transformation in the process of normalising data that is based on count or Poisson data. While the techniques used in the subsequent spatial analysis do not strictly require the data to be normal, a log transformation of Poisson data also helps to stabilise the variance[134]. The first step in the fitting of this model was to identify the spatial trend.

5.3.2(a) Identifying spatial trend

Cressie[124] described a method of testing for spatial trend using lip cancer incidence rates for the 56 Scottish local government districts from 1974-1980. The method involved overlaying a grid of rows and columns and identifying a district to a particular node of the grid if its centre was closest to that node. The grid locations of the Local Government District's are given in table 5.2, and they are numbered as shown in table 5.3.

Table 5.2 Grid locations of local government districts

	1	2	3	4	5	6	7
1	25,56	27	-	21	54	-	55
2	26	22	-	24	19	17	16
3	23	-	-	20	-	18	15
4	-	-	-	53	-	51,52	-
5	32	38	7	5	12,13	14	-
6	-	34,43	33,35,50	6	29	28	-
7	-	49	41	31,47,48	30	-	-
8	-	37,44	39,40,42	-	-	-	1
9	-	45	36	46	4	2,3	-
10	-	11	10	9	8	-	-

Table 5.3 Labels for table 5.2

No.	Local Government District	No.	Local Government District
1	Berwickshire	29	City of Edinburgh
2	Ettrick & Lauderdale	30	Midlothian
3	Roxburgh	31	West Lothian
4	Clydesdale	32	Argyll & Bute
5	Clackmannan	33	Bearsden & Milngavie
6	Falkirk	34	Clydebank
7	Stirling	35	Cumbernauld & Kilsyth
8	Annandale & Eskdale	36	Cummnock & Doon Valley
9	Nithsdale	37	Cunninghame
10	Stewartry	38	Dumbarton
11	Wigtown	39	East Kilbride
12	Dunfermline	40	Eastwood
13	Kirkcaldy	41	City of Glasgow
14	North East Fife	42	Hamilton
15	City of Aberdeen	43	Inverclyde
16	Banff & Buchan	44	Kilmarnock & Loudon
17	Gordon	45	Kyle & Carrick
18	Kincardine & Deeside	46	Clydesdale
19	Moray	47	Monklands
20	Badenoch & Strathspey	48	Motherwell
21	Caithness	49	Renfrew
22	Inverness	50	Strathkelvin
23	Lochaber	51	Angus
24	Nairn	52	City of Dundee
25	Ross & Cromarty	53	Perth & Kinross
26	Skye & Lochalsh	54	Orkney
27	Sutherland	55	Shetland
28	East Lothian	56	Western Isles

From the information provided from the grid representation of the regions it was possible to model the data by fitting additive row and column effects:

$$Y_i = a + r_{k(i)} + c_{l(i)} + \varepsilon_i \qquad \text{where } k(i) = 1,...,10, l(i) = 1,..,7 \qquad (5.1)$$

The i^{th} district is located at grid node ($k(i)$,$l(i)$), a is the overall mean, r_k is the k^{th} row effect and c_l is the l^{th} column effect, Y_i is the logged emergency admission rate for respiratory disease in the i^{th} district. If there are two or more districts at the same node each is assumed to have the same additive mean, $a+r_k+c_l$. This model was fitted to the logged emergency admission rate for respiratory disease data for each week, the results are shown in table 5.4.

Table 5.4 Results of trend analysis

Week	Rows		Columns	
	Coeff	p-value	Coeff	p-value
46	0.015	0.602	0.034	0.468
47	0.066	0.107	0.081	0.218
48	0.031	0.126	**-0.068**	**0.040**
49	0.015	0.368	-0.011	0.698
50	0.020	0.425	-0.029	0.475
51	**0.053**	**0.007**	-0.022	0.473
52	-0.003	0.898	-0.001	0.984
1	0.002	0.933	-0.039	0.320

There are two significant trends apparent from this analysis. There is a trend in the emergency admission rate for respiratory disease in week 48 where the rate increases from the east to the west of the country. In week 51 a significant trend also exists but in this case the rate increases from the north to the south. When we look back at the maps this is backed up by the visual pictures that the maps provide. This method of analysing the grid data has identified spatial trend in two of the eight weeks in the winter of 1989/90 and the next step in the spatial analysis is to fit this spatial trend.

5.3.2(b) Fitting the Spatial Trend

The median polish[124] was the method used to fit the trend. This method involves sweeping through the spatial matrix removing firstly the row medians and then the column medians and iterating this process until a predetermined stopping criterion is satisfied. The process consists of starting with a p*q matrix of cells, then creating a further p+q+1 cells. The first step of the algorithm is to remove the row medians from the q cells in each row and create a further column of length p to hold the median values that have been removed. Then remove the column medians from the p cells in each of the (q+1) columns and store these column medians in a new row p+1, the cell (p+1)(q+1) will hold the overall median, p+1 the column effects and q+1 the row effects.

The algebraic algorithm is given below.

For $i = 1,3,5,.....,$ define

$$Y_{kl}^{(i)} = Y_{kl}^{(i-1)} - med\{Y_{kl}^{(i-1)} : l = 1,....., q\}, \qquad k = 1,....., p+1, l = 1,....., q,$$

$$Y_{kq+1}^{(i)} = Y_{kq+1}^{(i-1)} + med\{Y_{kl}^{(i-1)} : l = 1,....., q\}, \qquad k = 1,....., p+1,$$

and for $i = 2,4,6,.....,$ define

$$Y_{kl}^{(i)} = Y_{kl}^{(i-1)} - med\{Y_{kl}^{(i-1)} : k = 1,....., p\}, \qquad k = 1,....., p, l = 1,....., q+1,$$

$$Y_{p+1l}^{(i)} = Y_{p+1l}^{(i-1)} + med\{Y_{kl}^{(i-1)} : k = 1,....., p\}, \qquad l = 1,....., q+1,$$

where $med\{y_1,....., y_n\}$ is the median of $y_1,....., y_n$. To start the algorithm assume

$$Y_{kl}^{(0)} = \begin{Bmatrix} Y_{kl}, k = 1,..., p, l = 1,....., p, \\ 0, elsewhere \end{Bmatrix}$$

Table 5.5 shows the matrix containing the original logged rates of emergency admissions for respiratory disease $Y_{kl}^{(0)}$ for week 48, a week in which spatial trend was found.

Table 5.5 Grid matrix for $Y_{kl}^{(0)}$

	1	2	3	4	5	6	7
1	3.42,2.7	2.44	-	-	2.74	-	2.13
2	1.90	3.18	-	3.22	2.25	2.29	2.72
3	3.54	-	-	2.91	-	2.96	2.86
4	-	-	-	3.31	-	3.63,3.5	-
5	3.18	3.19	2.81	2.48	3.03,3.0	1.98	-
6	-	3.16,3.5	3.37,3.2	3.01	3.12	3.11	-
7	-	3.26	3.35	3.24,3.1	2.87	-	-
8	-	3.33,2.9	2.85,2.8	-	-	-	1.72
9	-	3.41	2.91	3.31	3.06	2.95,3.4	-
10	-	2.69	2.89	2.76	3.32	-	-

Table 5.6 shows the overall median, row and column effects and the residuals following the median polish.

Table 5.6 Grid matrix for $Y_{kl}^{(n)}$ after median polish

	1	2	3	4	5	6	7	row
1	-.01	-.29	-	-	.36	-	.01	-.61
2	-1.08	.19	-	.73	-.39	-.19	.34	-.34
3	.08	-	-	-.07	-	.01	-.01	.14
4	-	-	-	-.11	-	.11	-	.59
5	.02	.00	.00	-.21	.24	-.68	-	-.15
6	-	.03	-.12	.00	-.03	.13	-	.17
7	-	-.20	.26	.20	-.23	-	-	.12
8	-	.00	.51	-	-	-	-.59	-.42
9	-	.00	-.13	.40	.00	-.13	-	.08
10	-	-.57	.00	.00	.42	-	-	-.08
Column	.35	.37	.00	-.13	.01	-.15	-.24	2.96

These matrices contain the figures for each of the local government districts as shown in table 5.3. The new model can be created from the overall median 2.96 and the row effects from column q+1, the column effects from row p+1 and the individual cell residuals from the cells values $Y_{kl}^{(n)}$.

Now we have the model

$Y_i = \tilde{a} + \tilde{r}_k + \tilde{c}_l + Y_{kl}^{(n)}$ where $\tilde{a} = 2.96$, $\tilde{r}_k$ is the vector of row effects, $\tilde{c}_l$ is the vector of column effects and $Y_{kl}^{(n)}$ are the matrix of residuals after trend has been removed following n iterations. A stopping criterion has to be used in the algorithm and this is either that the change in the residuals is less than a certain value ε, or until a specified number of iterations have been carried out. The function **twoway** in S-PLUS which was used to carry out the median polish iterates until convergence occurs, which is usually within 3 to 4 iterations.

In order to check that the median polish had removed the spatial trend the model given in equation 5.1 was fitted to the residuals $Y_{kl}^{(n)}$ for each week and the results are shown in table 5.7; as can be seen there are no spatial patterns left in the residuals.

Table 5.7 Results of trend analysis after median polish

Week	Rows		Columns	
	Coeff	p-value	Coeff	p-value
48	0.009	0.545	0.018	0.474
51	0.008	0.530	-0.001	0.964

5.3.2(c) Checks for stationarity

Having now removed the trend or large scale spatial variation in the data we are left with the error or small scale spatial variation. Before any statistical analysis can be carried out the errors have to be checked for stationarity. This can be done by plotting the median and squared interquartile range by row and column[124]. If the data are stationary then the plots will be level. These are shown in figures 5.7 – 5.10.

Figure 5.7 Median by row number for each week

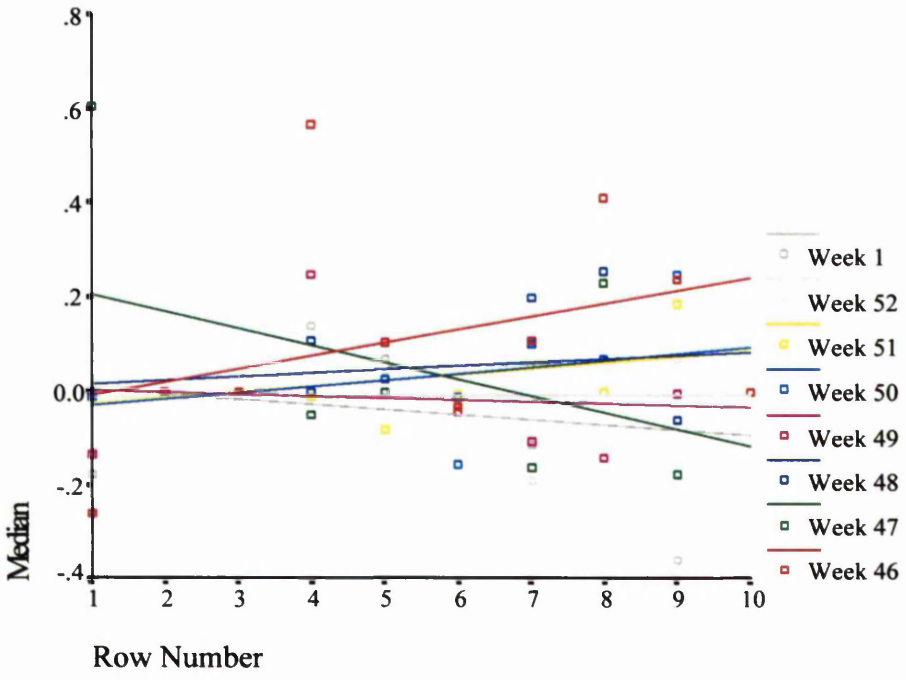


Figure 5.8 Median by column number for each week

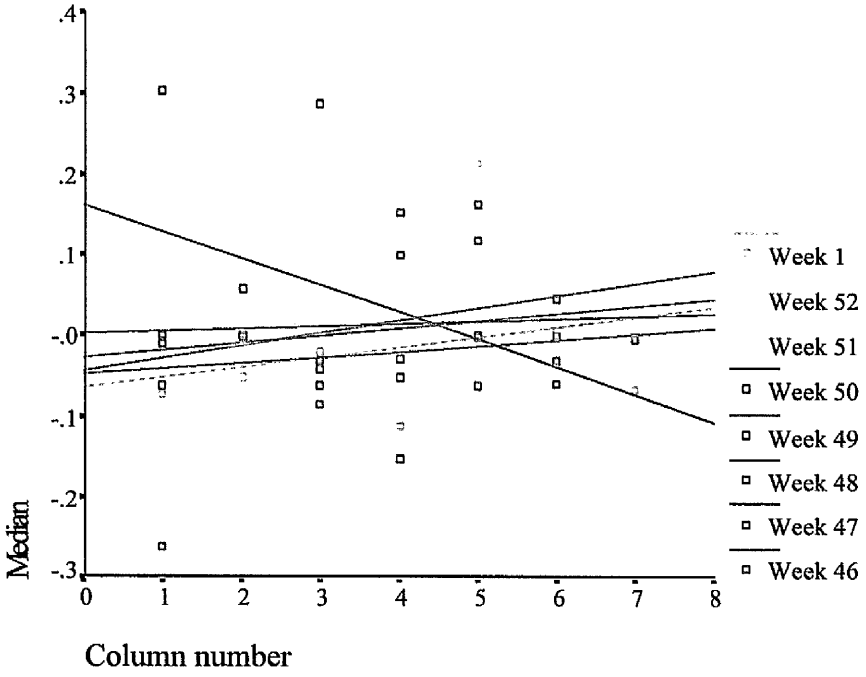


Figure 5.9 Squared interquartile range by row number for each week

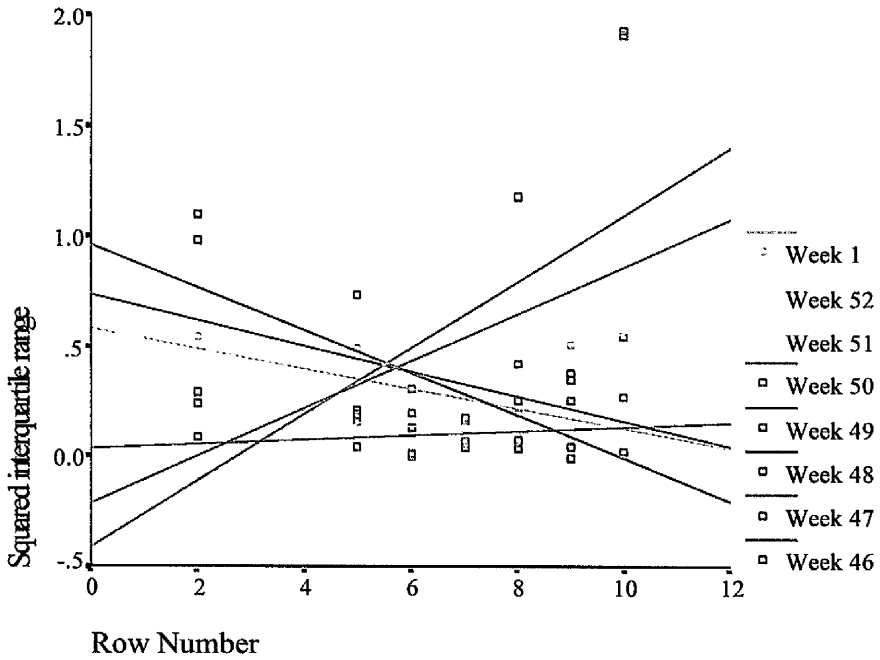
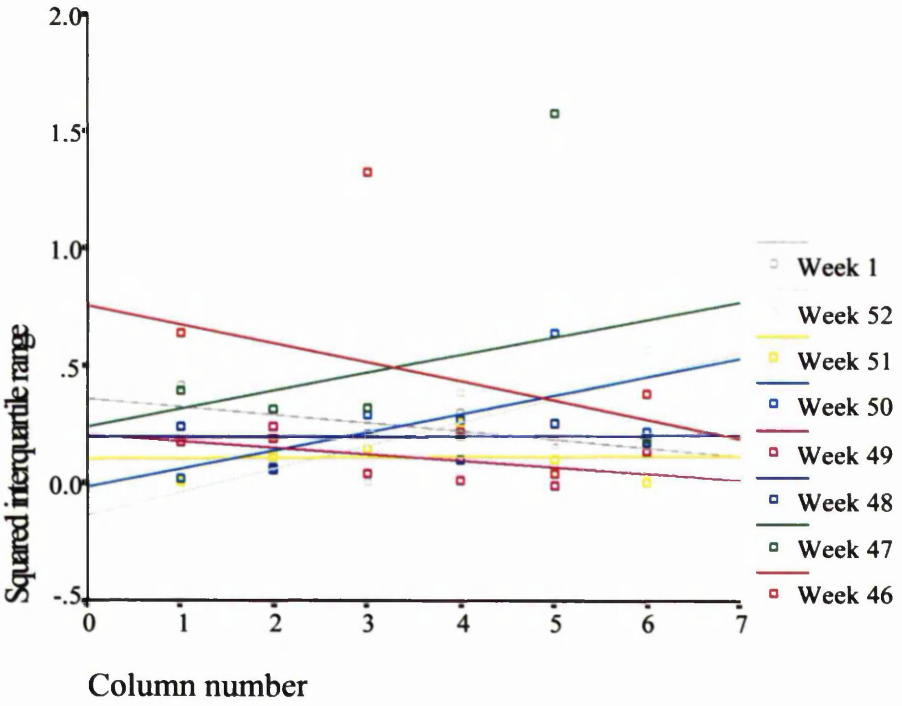


Figure 5.10 Squared interquartile range by column number for each week



As can be seen, there appear to be several trends in the data, since the lines do not seem level. However, regression analysis (table 5.8) showed that in fact there are no significant relationships between the median and interquartile range and row and column for the residuals from the fit for spatial trend. We can assume that these residuals are stationary.

Table 5.8 Results of the regression analysis for median and squared interquartile range

Measure	Week	Rows		Columns	
		Coeff	p-value	Coeff	p-value
Median	48	0.007	0.558	0.003	0.840
	51	0.012	0.240	-0.002	0.846
Squared inter-	48	-0.006	0.874	0.032	0.727
Quartile range	51	-0.027	0.132	0.050	0.186

The median polish residuals have constant mean zero and are stationary. It is now possible to assess these residuals for any small-scale spatial autocorrelation.

If we return to the model

data=spatial trend + error

we have identified spatial trend in weeks 48 and 51 and have removed this trend to leave residuals which are stationary. While spatial trend implies that the pattern of emergency admissions for respiratory disease increases consistently from one area of the country to the other, spatial autocorrelation implies that areas that are close to each other in location experience similar emergency admission rates.

5.3.2(d) Tests for Spatial Autocorrelation

The aim of this section was to test for spatial autocorrelation in the weekly emergency admission rates for respiratory disease in the winter of 1989/90. There are two main test statistics which have been developed to assess the degree of spatial autocorrelation between the $\{x_i\}$ in joined areas, where x_i is the value of X in the ith area. These are Moran's[135] I statistic defined as

$$I = \frac{n}{S_0} \frac{\sum_i \sum_j W_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \tag{5.2}$$

and Geary's[136] statistic c, defined as

$$c = \frac{n-1}{2S_0} \frac{\sum_i \sum_j W_{ij}(x_i - x_j)^2}{\sum_i (x_i - \bar{x})^2} \tag{5.3}$$

where $S_0 = \sum_i \sum_j W_{ij} (i \neq j)$, W_{ij} is a measure of the spatial proximity of two locations i and j and x_i and x_j are the observed values at locations i and j.

Both Moran's I and Geary's c statistic have well established distributions and are commonly accepted as the standard spatial autocorrelation test statistics, however it has been shown that the I test is generally better than the c test although the margin of advantage is slight[137]. In the following analysis Moran's I is used.

The null hypothesis is that there is no spatial autocorrelation present in the data and the alternative is specifically defined by a weighted neighbour matrix W_{ij}. The neighbourhood matrix combined with the choice of weights determines the precise form of the alternative hypothesis. Neighbours can be defined as regions whose centroids are within a certain distance of each other, as regions which border each other or by a variety of different methods based on the available data. Due to the fact that the data on centroids was not available for the local government districts, the definition of neighbours in this analysis was any two regions that shared a common boundary.

The alternative hypothesis in this case is determined by the weighting function, this choice introduces ambiguity into the investigation of spatial autocorrelation, but on the other hand introduces flexibility because it means that alternative weighting functions can be used to investigate alternative hypotheses. Cliff & Ord[137] suggested that when carrying out a spatial autocorrelation test the investigator must specify, in advance, the set of weights in accordance with the kind of spatial pattern that he wishes to detect. However, others[138] have said that more than one set of neighbour weights should be tried in order to determine the sensitivity of the model results.

Odland[139] discusses a variety of weighting functions that are available for area data. The simplest is a binary function that takes the value 1 if a pair of areas are neighbours and 0 if they are not. An extension to this is to weight neighbours according to the length of the boundary that they share under the presumption that areas which share long boundaries are 'closer' than those that share short boundaries. The distance between the centres of areas can also be used in conjunction with the length of their common boundaries to provide a weighting system which gives greater weight to areas whose centres are separated by shorter distances and that also share long boundaries.

Gatrell[140] discussed more complex weighting schemes based on hierarchical levels, or using areas classified as urban or rural, other methods could be based on the number of road or rail links between two areas. Generally speaking however weighting functions are dependent on the data available.

In the case of the local government districts in Scotland, in this analysis, where there is no information available on area centroids or common boundary distances, there are few weighting functions to choose from. A simple binary method or row-standardisation are the only real options. Row-standardisation[138] involves basing the weights on the number of neighbours of each region. In the neighbourhood matrix, a region (row) with 5 neighbours would have a weight of 1/5 for each neighbour pair (non-zero column). The row-standardisation method was preferred over the simple binary function as it made more use of the available information.

The choice of weights has been found to be a fairly crucial one in spatial analysis and an incorrect specification of weights can inflate the standard error[141]. Some geologists have developed 'optimal' schemes in which they determine a weighting system which minimises the error variance however these have been found to perform no differently than a priori weights[142].

For this analysis the neighbours of each of the 56 local government districts were determined and each pairing was given a weighting according to the number of pairs. Morans I statistic was then used to test for spatial autocorrelation in the logged emergency admission rates for the 8 weeks during the winter of 1989/90. The logged rates themselves were used for weeks 46,47,49,50,52 and 1 as they were approximately normal, displayed no spatial trend and had relatively constant variance. For weeks 48 and 51, for which spatial trend was found, the spatial autocorrelation test can be carried out on the median polish residuals, which were also found to be normally distributed and stationary. When concentrating mainly on the error term meeting stationarity assumptions spatial trend removal was carried out, this ensured that the errors represented only small scale spatial variation. Table 5.9 shows the first 10 regions and their pairs, again the same numbering system of areas applies.

Table 5.9 First ten pairs of regions and their neighbours

Area	Neighbour									Weight
1	2	3	28							1/3
2	1	3	4	8	28	30				1/6
3	1	2	8							1/3
4	2	8	46	31	29	30				1/6
5	7	12	53							1/3
6	7	35	47	31						1/4
7	5	6	35	50	33	34	38	32	53	1/9
8	2	3	4	9	46					1/5
9	8	10	36	46						1/4
10	9	11	36							1/3

In the case of the Island local government districts, Shetland was a neighbour of Orkney, Orkney had two neighbours, Shetland and Caithness and the Western Isles neighboured Skye & Lochalsh and Ross & Cromarty.

Having defined the weighted neighbourhood matrix it is fairly straightforward to compute Moran's I statistic in S-PLUS. The variance can be calculated in two ways, one which assumes normality of the data and the other which does not assume a specific distribution. Due to the fact that the data are originally from a Poisson distribution but have been transformed to approximate a normal distribution, it was decided to use the method that assumes no specific distribution.

The Moran's I statistic was computed for the logged emergency admission rate for each week and the results are shown in table 5.10.

Table 5.10 Moran's I for logged emergency admissions by week

Week	Moran's I	Normal statistic	p-value
46	0.154	1.856	0.063
47	0.197	2.808	0.005
48*	0.037	0.603	0.547
49	0.118	1.459	0.145
50	0.056	0.822	0.411
51*	-0.002	0.174	0.862
52	0.201	2.355	0.019
1	0.127	1.563	0.118

** indicates that residuals from the median polish were used instead of the log rates*

There appears to be some evidence of spatial autocorrelation in the data in weeks 46,47 and 52. However, the results of these tests have to be interpreted with caution. A non-significant result means that there is no spatial autocorrelation given the particular neighbourhood structure and choice of weights.

One other reason why we may be finding a significant result in the spatial autocorrelation test is that there could be a trend in the data that was not identified earlier. The analysis at the start of section 5.5.2 used the model $Y_i = a + r_{k(i)} + c_{l(i)} + \varepsilon_i$ to identify spatial trend in the logged emergency admission rate for respiratory disease in each week, and only found evidence of a trend for weeks 48 and 51. For these weeks the trend was removed and the spatial autocorrelation test showed that there was no spatial autocorrelation in the residuals. Although the tests for trend showed evidence of spatial trend for only weeks 48 and 51, it was decided that a simple median polish should be performed for each week. The results of this median polish are shown on table 5.11.

Table 5.11 Results of trend analysis after median polish

Week	Rows		Columns	
	Coeff	p-value	Coeff	p-value
46	0.010	0.623	0.025	0.438
47	0.015	0.678	0.017	0.767
48	0.009	0.545	0.018	0.474
49	0.001	0.901	-0.004	0.810
50	0.018	0.315	-0.009	0.769
51	0.008	0.530	-0.001	0.964
52	-0.015	0.387	-0.010	0.714
1	-0.032	0.091	0.002	0.940

Checks for stationarity were performed on the residuals and the results of these are shown in tables 5.12 and 5.13

Table 5.12 Results of stationarity checks on median values

Week	Rows		Columns	
	Coeff	p-value	Coeff	p-value
46	0.028	0.321	0.015	0.676
47	-0.036	0.163	0.092	0.164
48	0.007	0.558	0.003	0.840
49	-0.004	0.770	0.007	0.649
50	-0.014	0.251	0.009	0.585
51	0.012	0.240	-0.002	0.846
52	0.001	0.956	0.008	0.552
1	-0.011	0.520	0.012	0.596

Table 5.13 Results of stationarity checks on squared interquartile range

Week	Rows		Columns	
	Coeff	p-value	Coeff	p-value
46	0.102	0.197	-0.098	0.296
47	-0.080	0.081	0.038	0.720
48	-0.006	0.874	0.032	0.727
49	-0.014	0.335	0.026	0.492
50	0.058	0.411	0.077	0.157
51	-0.027	0.132	0.050	0.186
52	-0.009	0.925	0.089	0.020
1	-0.139	0.072	-0.026	0.186

The column residuals for week 52 may not be stationary, the trend coefficient of 0.089 is significant at the 2% level indicating that the size of the residuals increases from the west to the east of Scotland. There are a few other weeks where the residuals may not be stationary, these are weeks 47 and 1 which show a decrease in the value of the residuals from the north to the south of Scotland, the p-values for the row coefficients are significant at the 10% level. Due to the fact that the data do not appear to be stationary, Moran's I statistic cannot be used on the residuals for these weeks. Thus the results from the tests for autocorrelation which were performed on the log rates will be used for weeks 47, 52 and 1. Tests for spatial autocorrelation were carried out for the remaining weeks for which the median polish residuals were stationary and no spatial trend had been found i.e. Weeks 46, 49 and 50, the results are shown in table 5.14

Table 5.14 Results of Moran's I statistic for median polish residuals

Week	Moran's I	Normal statistic	p-value
46	0.016	0.368	0.713
49	-0.013	0.060	0.952
50	-0.109	-0.996	0.319

Table 5.14 shows that there is no evidence of spatial autocorrelation in either of these weeks. This result for week 46 is contrary to the result using the log rates that indicated there might be spatial autocorrelation. It may have been the case that in week 46 there was unidentified spatial trend in the data that caused the initial test for spatial autocorrelation to be significant.

This spatial analysis of the emergency admission rate for respiratory disease in the 56 local government districts involved initially identifying spatial trend. Significant spatial trend (or large scale spatial variation) was found for weeks 48 and 51. Having established spatial trend in these weeks, this spatial trend had to be removed using the median polish technique before analysis of spatial autocorrelation could be carried out. After the trend was removed the median polish residuals were tested for stationarity and found to be stationary. Moran's I statistic requires that the data are approximately normal and stationary. This was found to be the case for the log rates for weeks 46,47,49,50,52 and 1, and for the median polish residuals for weeks 48 and 51. Moran's I statistic identified significant spatial autocorrelation in weeks 46,47 and 52. However, Moran's I statistic may produce a significant result if the effects of spatial trend have not been completely removed. In an additional analysis a median polish was performed for all weeks, instead of just those which had previously shown evidence of spatial trend and the new residuals were tested for spatial trend and stationarity. This analysis showed that the residuals for weeks 52, 47 and 1 might not be stationary. Therefore, for these weeks the autocorrelation analysis which was carried out on the log rates was the most appropriate for testing for the presence of spatial autocorrelation, for the remaining weeks 46, 49 and 50 it was appropriate to test for spatial autocorrelation in the median polish residuals. The results of this analysis suggested there was no spatial autocorrelation for each of these weeks, indicating that the initial observation of spatial autocorrelation for week 46 on the log rates may in fact represent some spatial trend that was then removed by the median polish.

The spatial analysis of this data has shown that the patterns observed in each week were not the same, some displayed spatial trend, others spatial autocorrelation and in some weeks there was no spatial pattern present at all. There is little evidence of any consistent space-time variation in the emergency admission rate for respiratory disease during the

influenza epidemic in the winter of 1989/90. Weekly emergency admissions for respiratory disease varied considerably over time and varied spatially at individual weeks of the year, but there was little evidence that there was any pattern over time in the spatial variation. There was no evidence that the increased rate of emergency respiratory admissions during the influenza epidemic of 1989/90 spread around the country increasing initially in one area before spreading to surrounding areas.

5.4 Emergency respiratory admissions for Glasgow by postcode

5.4.1 Spatial analysis of lattice data

In this section the analysis will focus on Glasgow City local government district and the level of aggregation of the spatial data is further reduced to postcode level. The analysis at postcode sector level is only carried out in a spatial dimension. This is mainly due to availability and accuracy of the data. Weekly postcode sector populations were not available and the numbers of deaths per week in each area were too small to warrant a weekly analysis. Age standardised emergency admission rates for respiratory disease were calculated for each postcode sector in Glasgow City during the whole of 1991. This year was chosen because the population data were more accurate as 1991 was a census year. Also there did not appear to be a significant influenza epidemic during 1991 which would influence the stability of the conclusions of any purely spatial analysis. The population weighted (x,y) co-ordinates of the centroid of each of the postcode areas were available for this analysis as well as the postcode population.

The first step of the analysis was to map the data. Figures 5.11a-b show the age standardised emergency admission rate per 1000 population for respiratory disease for males and females in Glasgow City. Both maps seem to show low rates of disease occurring in the south east of the city compared to higher rates in the north west of the city.

Figure 5.11a Age & sex standardised emergency admission rate for males

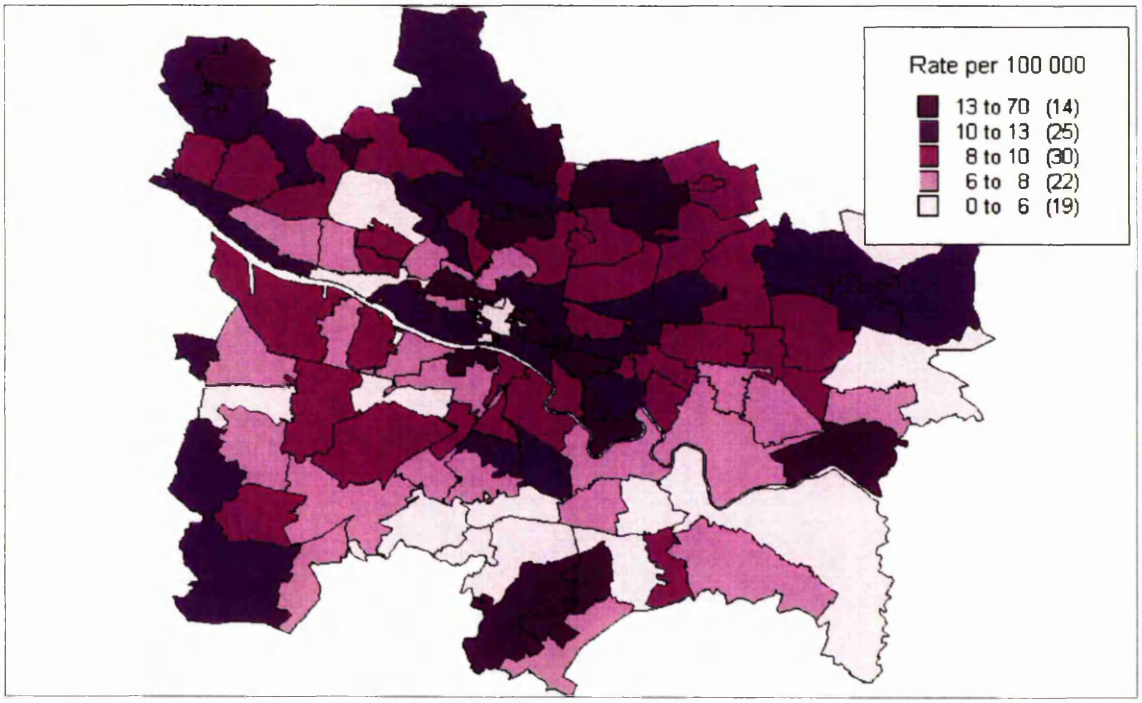
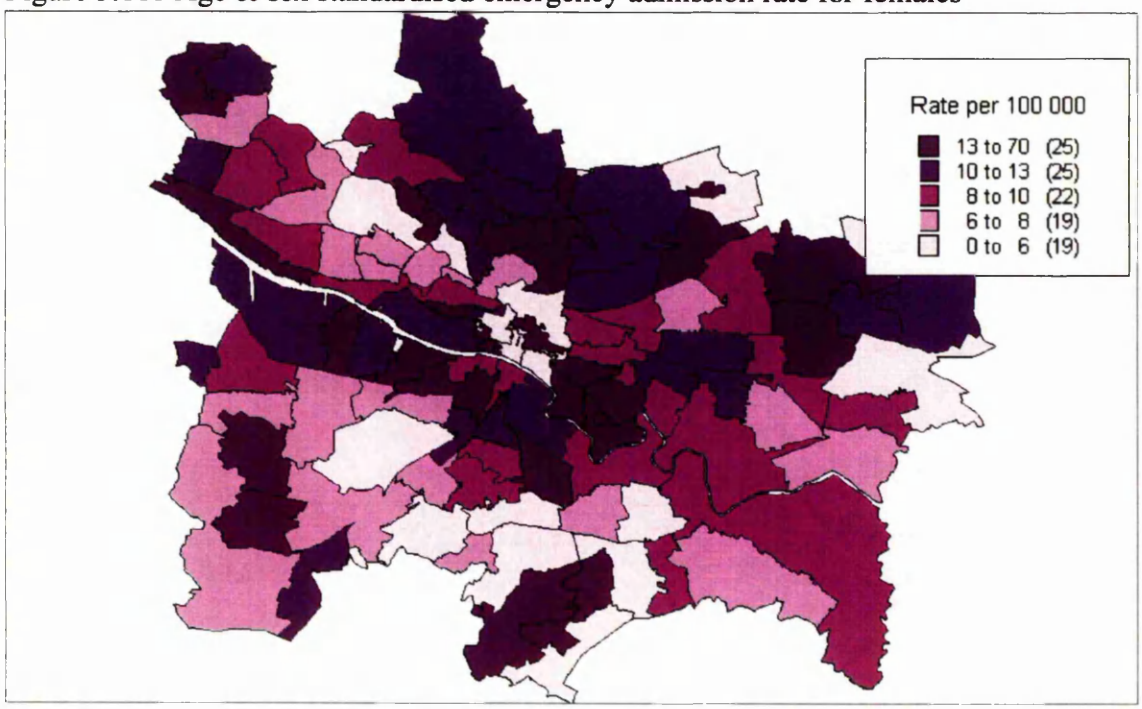


Figure 5.11b Age & sex standardised emergency admission rate for females



Before any analysis was carried out on these data they were tested for normality. The emergency admissions rate for respiratory disease for males and females was found to

deviate slightly from the normal distribution, however a log transformation ensured that the data were approximately normally distributed. Having transformed the data, the first step in the spatial analysis was to test for spatial trend.

A test for spatial trend was described in section 5.3.2 where the geographical areas are classified in a grid matrix and the data were modelled by fitting additive row and column effects where:

$$Y_i = a + r_{k(i)} + c_{l(i)} + \varepsilon_i \qquad \text{where } k(i) = 1,...,10, l(i) = 1,..,7$$

For the Glasgow postcode data there were 110 postcode sectors (the first 5 digits make up the sector) within the Glasgow City local government district, and because the (x,y) co-ordinates for each postcode sector were available from the census it was not necessary to create a grid to perform this test for spatial trend. A simple regression model based on the co-ordinates was used. The regression coefficients from the centroids model are shown in table 5.15.

Table 5.15 Regression coefficients from the centroids model

	Y co-ordinate		X co-ordinate	
	Coeff	p-value	Coeff	p-value
Males	0.0032	0.008	-0.0002	0.798
Females	0.0029	0.080	-0.0001	0.907

Using the postcode centroid co-ordinates, there was found to be a significant trend in emergency admission rates for males and, to some extent, females. The results of the trend analysis indicate that the emergency admission rate for respiratory disease increases significantly from the south to the north of the city. For both models an interaction term was found to be non-significant.

The residuals from this model were then assessed for spatial autocorrelation using Moran's I statistic. With the postcode data it is possible to try different spatial neighbour matrices and different systems of weighting, as there is information available on the co-ordinates of the postcode sector centroids. Two matrices were used in this analysis, the first based on the common boundary principle where areas were said to be neighbours if they shared a common boundary, and the second was based on the distance between the

centroids of any two postcodes. Postcodes were classed as neighbours if their centroids were less than 2 kilometres apart. For each of these two matrices the weights were based on the inverse of the number of neighbours each area had.

However, with the centroid information it was possible to compute a new weighting vector for both the common boundary matrix and the centroid matrix, here the weights were computed as the inverse of the distance between centroids. The distance between the pairs of centroids was calculated using Pythagoras theorem.

Thus, there were four possible spatially weighted neighbourhood matrices which could be used in this analysis. Moran's I statistic was computed for each matrix. Table 5.16 below gives the results from the spatial autocorrelation test.

Table 5.16 Results from spatial autocorelation test

	Neighbour Definition	Weighting Matrix	Males		Females	
			I	p-value	I	p-value
1	Boundary	Inverse distance	0.710	0.478	1.015	0.310
2	Boundary	Inverse no. of boundaries	1.365	0.107	1.682	0.093
3	Distance	Inverse distance	1.132	0.258	1.552	0.121
4	Distance	Inverse no. of boundaries	2.678	0.007	2.987	0.003

Table 5.16 shows that there was no spatial autocorrelation identified for three of the possible hypotheses, but when the neighbours were defined by the distance between their centroids and were weighted according to the inverse of the number of boundaries there appeared to be some form of spatial autocorrelation. Moran's I statistic is a test of the null hypothesis of no spatial correlation against the alternative hypothesis specified by the weighted neighbour matrix. These results indicate that if there is spatial autocorrelation in the data it can be described by scenario 4, however this is probably the least likely way in which neighbours would realistically be defined and weighted. The most logical choice was to define neighbours as areas that shared a common boundary and to use the inverse of the distance between the centroids of the two areas as the weighting matrix. It is likely that the weighting system based on the distance between the centroids of two areas is more effective in establishing a measure of the 'closeness' of two areas than a value

based on the number of borders each area has. Also, in terms of specifying 'neighbours', areas which border each other are intuitively seen as neighbours rather than areas which are less than some distance apart. The distance method of establishing the neighbourhood matrix identified many more pairs of neighbours (696) than the common border method (524). Griffith[141] recommended that it is better to have too few neighbours than too many and accordingly the common border neighbourhood matrix with distance weights would appear to be the most appropriate framework in which to test for spatial autocorrelation.

These results have identified some spatial autocorrelation, but further analysis was required to get a clearer picture of the spatial patterning of emergency admission rates in postcode sectors in Glasgow City.

Each postcode sector in Scotland has been assigned a Carstairs score[71](described in chapter 2) which reflects the relative affluence or socio-economic deprivation of the area. It is possible that the spatial patterns in the emergency admission rate could be explained by a spatial pattern in socio-economic deprivation. Other authors have consistently shown a relationship between socio-economic deprivation and health[71, 72]. A simple linear model of the form $Y_i = \alpha + \beta x_i$ was fitted to the data where Y_i is the age standardised emergency admission rate for respiratory disease in postcode sector i and x_i is the Carstairs deprivation score for postcode sector i. The results of this fit are shown in table 5.17.

Table 5.17 Results from regression of emergency admission rates and deprivation

	Coeff	p-value
Males	0.032	<0.0001
Females	0.063	<0.0001

There is clear linear relationship between deprivation and emergency admission rate in Glasgow. Could this explain any apparent spatial variation? The residuals from the fits on Carstairs score were examined for possible spatial autocorrelation and the results are shown in table 5.18

Table 5.18 Results for tests of spatial autocorrelation in residuals

	Neighbour Definition	Weighting Matrix	Males		Females	
			I	p-value	I	p-value
1	Boundary	Inverse distance	0.581	0.561	1.590	0.112
2	Boundary	Inverse no. of boundaries	0.220	0.826	1.719	0.086
3	Distance	Inverse distance	1.015	0.310	1.891	0.059
4	Distance	Inverse no. of boundaries	1.952	0.051	2.349	0.019

Again, having removed the effect of Carstairs score from the pattern of emergency admission rates for respiratory disease in Glasgow there still appears to be spatial variation in the data which is not accounted for by deprivation score. The spatial pattern appears to be defined again by scenario 4.

The next step in the analysis was to include both a spatial trend (postcode co-ordinates) and the deprivation scores in the linear model. The results of this analysis are shown in table 5.19. The model is of the form $Y_i = a + \beta_1 x_i + \beta_2 y_i + \beta_3 z_i + \varepsilon_i$ where z_i is the Carstairs score for the ith postcode sector

Table 5.19 Results from postcode co-ordinate and deprivation score analysis

	X co-ordinate			Y co-ordinate			Carstairs score		
	coeff	se	p-value	coeff	se	p-value	coeff	se	p-value
Males	-.0006	.001	.449	.0025	.001	.029	.0300	.008	<.0001
Females	-.0013	.001	.226	.0008	.001	.577	.0630	.010	<.0001

In this model there is still north/south trend in the male emergency admission rate as well as a strong relationship with deprivation score for both males and females. Tests for spatial autocorrelation were carried out for the residuals from the models and the results are described in table 5.20

Table 5.20 Results for tests of spatial autocorrelation in residuals

	Neighbour Definition	Weighting Matrix	Males		Females	
			I	p-value	I	p-value
1	Boundary	Inverse distance	0.124	0.901	1.148	0.251
2	Boundary	Inverse no. of boundaries	-0.430	0.667	1.138	0.255
3	Distance	Inverse distance	0.320	0.749	1.240	0.215
4	Distance	Inverse no. of boundaries	0.635	0.525	1.288	0.198

There now does not appear to be any spatial autocorrelation for any of the four possible weighted matrices in the residuals from the full fit for male and female emergency admissions from respiratory disease. However, having now determined the presence of explanatory variables it was of interest to model these linear models in a spatial framework in which the spatial autocorrelation is fitted within the modelling process.

5.4.2 Spatial regression models

The autoregressive models that have been explored in previous sections have assumed a purely spatial process. These models have been able to take account of the spatial covariance but have required that the mean is constant or that any apparent trend is removed before the spatial modelling is undertaken.

In this section both the large scale and the small scale spatial variation is of interest. The large scale variation may be due to the presence of an explanatory variable that could explain much of the observed spatial variation in the dependent variable. Ordinary least squares techniques would be inappropriate as the residuals would not be independent. Thus methods that account for both a regression component for the mean and an autoregressive component are required but firstly we will concentrate on the purely spatial part of the process. There are three types of spatial covariance structures available for spatial regression models in S+SpatialStats. These are simultaneous spatial regression (SAR), conditional spatial regression (CAR) and moving average (MA) models.

Whittle[143], assuming purely spatial process, proposed a simultaneous autoregressive model for spatial variation of the form

$$Y_i = \sum_{j \neq i} g_{ij} Y_j + \varepsilon_i, \quad i = 1,...,n \tag{5.4}$$

where $E(Y_i) = 0$. The ε_i are uncorrelated error terms with $E(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = \sigma_i^2$, $i = 1,...,n$.

In matrix notation this can be written as $Y = GY + \varepsilon$ where is G a n x n matrix.

$$E(\varepsilon) = 0, \text{ and } \operatorname{var}(\varepsilon) = \Sigma = \begin{bmatrix} \sigma_1^2 & & \\ & \sigma_2^2 & \mathbf{0} \\ \mathbf{0} & \ddots & \\ & & \sigma_n^2 \end{bmatrix}$$

From equation 5.4

$Y = (I - G)^{-1}\varepsilon$, and since $E(Y_i) = 0$,

$$\operatorname{var}(Y) = V = E(YY^T) = (I - G)^{-1} E(\varepsilon\varepsilon^T)(I - G^T)^{-1}$$

$$= (I - G)^{-1}\Sigma(I - G^T)^{-1} \tag{5.5}$$

If we assume normality for ε then Y is multivariate normal with covariance matrix V.

If $Y^* = Y + \mu$, then $Y^* \sim MVN(\mu, V)$.

Assuming a simple spatial relationship as in figure 5.12 where areas are only spatially dependent if they share a common boundary then a first-order simultaneous autoregressive model would be

$$Y_{ij} = \rho(Y_{i-1,j} + Y_{i+1,j} + Y_{i,j-1} + Y_{i,j+1}) + \varepsilon_{ij}, \tag{5.6}$$

where $|\rho| < 1/4$ to ensure stationarity.

Figure 5.12 Example of spatial neighbours in a simple grid layout

		$y_{i-1,j}$		
	$y_{i,j-1}$	$y_{i,j}$	$y_{i,j+1}$	
		$y_{i+1,j}$		

The matrix G could be written in the form ρW, where the elements of W would be one or zero depending on whether the two areas were regarded as neighbours or not.

In a conditional spatial model(CAR) $y_i^* = \{y_j, j \neq i\}$, that is y_i^* denotes y after the deletion of y_i, the model is $E(Y_i \mid y_i^*) = \sum_{j \neq i} g_{ij} y_i$, $Var(Y_i \mid y_i^*) = \sigma_i^2$ for $i = 1,...,n$, from this we get $Y \sim MVN(0,V)$ where $V = \Sigma(I-G)^{-1}$

So, in the simultaneous model the covariance matrix V is defined as

$$V = (I-G)^{-1}\Sigma(I-G^T)^{-1}$$

and in the conditional model it is defined as

$$V = \Sigma(I-G)^{-1}$$

where G must be symmetric to ensure that V is symmetric.

In the moving average example the model is $Y_i = \varepsilon_i + \sum_{j \neq i} g_{ij}\varepsilon_j$, where $E(\varepsilon_i) = 0$, $E(\varepsilon_i^2) = \sigma_i^2$ and $E(\varepsilon_i \varepsilon_j) = 0$ if $i \neq j$. The model can be written in matrix form as

$$Y = (I+G)\varepsilon \qquad \text{from this we get}$$

$$V = (I+G)\Sigma(I+G^T) \tag{5.7}$$

A more in depth discussion of these models can be found in Cliff[144], Haining[145] and Cressie[124].

Having described the spatial processes involved in these models the next step is to include the regression component for the mean. The conditional model can be written as

$$Y = N[X\beta, \sigma^2(I-G)^{-1}]$$

or to include the neighbourhood weights

$$Y = N[X\beta, \sigma^2(I-\rho W)^{-1}] \tag{5.8}$$

The CAR and SAR models correspond to the autoregressive models in time series analysis, while the MA models correspond to moving average models. One problem with the simultaneous model is that the residuals are correlated with the neighbour data values which can result in inconsistent least-squares parameter estimates[124], however for the conditional model the matrix W must be symmetric.

In the case of the Glasgow postcode data this matrix is not symmetric so the SAR model is the most appropriate. The two parts of the model; the linear regression and the spatial autoregressive part are fitted interactively. Table 5.21 shows the coefficients for each of the regression terms in the model containing a trend surface and the deprivation scores for both males and females using the four different weighted neighbour matrices.

Table 5.21 Results from postcode co-ordinate and deprivation score analysis using SAR model

		X co-ordinate			Y co-ordinate			Carstairs score		
		Coeff	se	p-value	coeff	Se	p-value	coeff	se	p-value
1	M	.0000	.002	.992	<.001	.002	.786	.0714	.015	<.001
	F	-.0004	.002	.794	-.0015	.002	.419	.0353	.014	.0118
2	M	-.0004	.002	.768	<.001	.002	.986	.0786	.017	<.001
	F	-.0001	.002	.927	-.0015	.002	.456	.0302	.014	.0320
3	M	.0000	.002	.984	<.001	.002	.859	.0721	.015	<.001
	F	-.0004	.002	.804	-.0015	.002	.421	.0346	.014	.0136
4	M	-.0003	.001	.846	.0008	.002	.612	.0750	.014	<.001
	F	.0000	.002	.976	-.0010	.002	.622	.0298	.014	.0328

The table shows that for this model the effect of the choice of weighted neighbourhood matrix makes little difference to the coefficients of the linear part of the model. However, if we compare the coefficients from this model with the earlier model (Table 5.20) which did not account for spatial autocorrelation we see that the inclusion of the spatial autocorrelation does affect the coefficients. The coefficients for the trend variables (x and y co-ordinates) are smaller and have a larger standard error, and are therefore less likely to reach statistical significance in the autocorrelation model as compared to the ordinary

linear model. For the deprivation score variable the coefficient is greater for males and less for females, both have a higher standard error and are also less likely to reach statistical significance, however in this model Carstairs score is consistently statistically significant.

This modelling method adjusts for any spatial autocorrelation that may be present in the data and reduces the chance of obtaining a spurious significant result when identifying explanatory variables in a spatial context.

The spatial regression model has shown that, accounting for spatial autocorelation, the emergency admission rate for respiratory disease among males is significantly related to the socio-economic deprivation of their postcode of residence. For females this association is also significant but not as strong. When deprivation scores are included as predictor variables in the model there is little evidence of any remaining large scale spatial trend.

5.5 Summary and Discussion

In this chapter the analysis of the temporal patterns in disease rates in Scotland has been extended to cover variation in space as well as in time. Emergency hospital admission rates for respiratory disease were used in this chapter for several reasons (i) if mortality data had been used the number of events in each week would have been very small, (ii) respiratory disease had shown the strongest relationship with temporal variation in previous analyses and (iii) this would be the disease group that may be most likely to display a space-time dimension as it has an infectious element.

Mapping is the first step in any spatial analysis. A map of the weekly emergency admission rate for respiratory disease in each health board during the influenza epidemic of 1989/90 showed that there might be a space-time structure in the data. Further analysis showed that when the weekly time series of emergency admissions for respiratory disease were assessed by health board, the correlation between the time series was greatest for health boards who shared a common boundary or were only 1 spatial lag apart. However the 15 health boards in Scotland are very variable in terms of size and population density. The data were then further analysed by local government district. This analysis showed

that the patterns observed in each week were not the same, some displayed spatial trend others spatial autocorrelation, and in some weeks there was no spatial pattern present at all. Emergency admissions for respiratory disease appeared to show some space-time variation but the form of this variation was hard to quantify. When the data were analysed spatially by postcode sector in Glasgow City it was found that most if not all of the spatial variation could be accounted for by a trend and by Carstairs deprivation scores. After accounting for these two variables there was little evidence of spatial autocorrelation, however including the spatial information in a modelling process ensured that the chance of detecting spurious associations between emergency admissions and the explanatory variables was reduced. Both the male and female emergency admission rates for respiratory disease were significantly related to socio-economic deprivation in Glasgow, even after accounting for spatial trend and spatial auto-correlation. Recent work[146] has also suggested that spatial variation in emergency hospital admissions is related to hospital catchment areas and the admitting policies of individual hospitals, however this information was not included in this analysis.

The spatial regression analysis was only carried out for 1991. A more complete space time examination of the spatial variation in emergency respiratory admissions in Glasgow and its relationship to socio-economic deprivation would involve looking at these patterns over a number of different time periods. However, the Carstairs deprivation scores are computed directly from the 1991 census and will not be updated until the 2001 census. A perceived change in the relationship between emergency admission rates and deprivation may simply reflect a loss of strength of the Carstairs score as the years move away from 1991. Also, it is likely that space-time variation in emergency respiratory admissions would occur over short time periods. An analysis looking at year on year spatial variation at postcode sector level would probably not have detected this variation. Perhaps the most ideal method of space-time analysis would be to look at weekly variation in emergency admission rates by postcode sector in Glasgow by week with adjustment for deprivation scores. However, because weekly postcode sector populations were not available and the number of cases would be very small it was felt that this type of in-depth analysis of space-time variation was beyond the scope of this project.

Finally, a regression model for spatially correlated count data generalising the work by Zeger has been proposed by McShane[147]. In this model the spatial autocorrelation in count data is introduced through a latent process. The authors used the distribution of neuronal cells in a laboratory culture well to test the model and then looked at simulation results to fully assess the properties of their method. Overall, the method was consistent with the performance of the time series model proposed by Zeger. However, while this extension of Zeger's method to spatial data is of interest, the method was not suitable for application in this thesis. Firstly, we would need to have access to temperature data spatially throughout Scotland and secondly, while spatial variation in disease rates may exist it is likely that other factors such as socio-economic and demographic factors may explain a considerable amount of the variation. The space-time analysis was concerned more with the progression of epidemic like patterns rather than identifying spatial variations in disease rates that could be related to temperature.

This chapter has recognised the importance of not only studying temporal variations in disease patterns but also being aware that disease patterns vary spatially. While the work described is by no means a complete spatial-temporal analysis of emergency respiratory admissions in Scotland, the analysis has shown that disease rates vary at both spatial and temporal levels.

Chapter 6 Discussion and Conclusion

6.1 Introduction

This work had two main aims; the first was to examine in detail the seasonal variation in mortality and morbidity in Scotland and to update, using current data and more sophisticated techniques, previously reported measures of seasonal variation in ill health in Scotland. The second main aim was to assess the direct effect of climate and influenza epidemics on the observed seasonal patterns using appropriate statistical methods. The fact that the mortality and morbidity data were a Poisson time series provided an interesting statistical concept, as traditional time series methods assume that the data are normal; furthermore, regression methods which could be used to assess the relationship between climate and ill health assume that the observations are independent. The main focus of this thesis was on the problems encountered when trying to establish the degree of association between two separate time series when one is a time series of counts. A further interesting aspect of the data was the spatial component. The final part of the work examined whether the temporal variation in mortality also existed spatially throughout Scotland. This involved examining the issues of spatial autocorrelation in a similar, but less detailed, manner to the examination of temporal autocorrelation.

6.2 Seasonality methods and results

Other authors[4,3,6,5] have used simple summary measures such as the quarter 1 to quarter 3 ratio and excess winter deaths index to describe seasonal variation. In chapter 2 a method that provided a simple summary statistic, which was not dependent on calendar months and for which a 95% confidence interval could be easily computed, was developed. The model was fitted in a generalised linear model framework with Poisson errors using the trend as an offset term. This method provided an estimate of the seasonal increase expressed as the percentage increase from the trough to the peak of the fitted seasonal curve, taking account of any trend in the data.

From the model, the seasonal percentage increase in all cause mortality in Scotland between 1981 and 1993 was 28% for males and 34% for females. This seasonal increase

was greatest in older age groups, but there was found to be little difference when the data were analysed according to social class or deprivation category. The seasonal increase in mortality was similar in each health board except in the Scottish Islands, where it was lower. Seasonal increases varied according to cause of death. For respiratory disease it was 89% for males and 114% for females, for ischaemic heart disease it was 34% for males and 36% for females, and a seasonal increase of 36% was found for both male and female deaths from cerebrovascular disease. Deaths from cancer showed little seasonal variation, 4% for males and 5% for females.

Emergency hospital admissions showed similar patterns, the seasonal increase was greatest for admissions for respiratory disease, 49% for males and 64% for females. Emergency admissions for ischaemic heart disease increased by 10% for males and 8% for females, while for emergency admissions for cerebrovascular disease these figures were 9% and 12%. Again, there was little seasonal variation in emergency admission rates for cancer, 2% for males and 4% for females. The size of the seasonal increase in emergency admissions was greatest in the older age groups and, in the case of respiratory disease alone, the very young.

There was little variation in the seasonal increase in either mortality or emergency hospital admissions according to socio-economic deprivation or city of residence, and no clear pattern when the data were analysed separately for each health board.

6.3 Poisson regression methods and results

In Chapter 3 the direct effect of climate on mortality and morbidity in Scotland was assessed, the seasonal aspect of the mortality and morbidity data had already been established in Chapter 2 and seasonal variation in temperature is universally acknowledged. A simple regression of mortality and temperature would be virtually guaranteed to show a significant (causal) relationship between temperature and mortality purely due to the fact that both of the variables followed a seasonal pattern. Any method that did not account for the fact that both variables were autocorrelated time series would produce spurious results. The mortality and morbidity data were also Poisson in nature, and the most appropriate method to analyse these data was to use a time series regression model which accounted for both these features. Three methods of analysing these data

were compared. These were ARIMA methods which accounted for the serial auto-correlation but did not allow for the data to be non-normal, Poisson regression methods which allowed for the data to be non-normal but did not account for serial autocorrelation, and the method developed by Zeger which took account of both the serial autocorrelation and the Poisson nature of the data. The comparison showed that Zeger's method gave more precise estimates of the coefficients with smaller standard errors. The residuals from Zeger's method had a lower standard deviation and a mean closer to zero than the residuals from the other two methods.

Since this work was carried out, Kelsall[89] has modified Zeger's method to allow an adjustment for autocorrelation that does not require the prior specification of the autocorrelation structure. This method was compared to other methods that adjusted for overdispersion only and was found to give more precise estimates of the coefficients of interest. It is likely that this method will be useful in Poisson regression models where the presence of overdispersion or autocorrelation is less easy to quantify.

Using Zeger's method the analysis showed that on average a fall in temperature of 1°C resulted in an increase in mortality of around 1% one week later. The size of the increase in mortality and the lag time between the fall in temperature and the increase in mortality or emergency hospital admissions varied by diagnosis. For deaths from respiratory disease the increase was greatest and it persisted, though at a reduced level, for up to 5 weeks. A 1% increase in deaths from ischaemic heart disease occurred on the week of the fall in temperature and one week afterwards, whereas the increase in deaths from cerebrovascular disease was around 0.5% but persisted for several weeks. The pattern was broadly similar for emergency hospital admissions, however the size of the increase was slightly lower and the duration of the effect was slightly shorter.

A more in depth analysis of the association between climate and mortality and morbidity was carried out in Chapter 4. Various demographic variables were considered, such as age, socio-economic deprivation, social class and city of residence. The only demographic variable that appeared to have a consistent effect on the relationship between climate and mortality was age group. The elderly and the very young were most affected by changes in temperature. There was little clear evidence of a consistent effect

on the relationship according to the other demographic characteristics. The size of the seasonal increase in mortality fell between 1981 and 1993, however this fall was only significant for deaths from all causes and for deaths from ischaemic heart disease.

An analysis of the effect of a fall in temperature at different underlying temperatures showed that a fall in temperature when the temperature was colder was associated with a larger increase in deaths, and a fall in temperature when temperature was hot was associated with a fall in deaths. Falls in temperature had the greatest impact on mortality when temperature was cold or when the temperature was generally increasing. Extreme temperature changes appeared to have the greatest effect on mortality from cerebrovascular disease, a cause of death that in previous analyses had shown the weakest relationship with temperature. There was little to be observed when the detailed relationship between emergency hospital admissions and temperature was examined.

With limited data on pollution, an analysis of the relationship between pollution, temperature and mortality in Glasgow showed that when considered separately, the pollutants were found to be related to mortality. However, after accounting for temperature there was little relationship between nitrogen dioxide, carbon monoxide and mortality in Glasgow between 1987 and 1993.

6.4 Space-time analysis

The final chapter of this thesis considered the spatial variation in mortality and morbidity in Scotland. An analysis of the emergency admission rate for respiratory disease by health board showed that for health boards which shared the same common boundary or were only one spatial lag apart, the correlation between the weekly time series of emergency admission rates was around 0.5 and the strength of this correlation diminished as the spatial lag increased.

The spatial variation in weekly emergency admission rates for local government districts was assessed using Moran's I statistic. This assessment of spatial autocorrelation is dependent on the choice of the spatial neighbourhood matrix, but for the data available there was only one possible choice of spatial neighbourhood matrix and weighting matrix. The analysis showed that during the influenza epidemic of 1989/90 there was

little evidence of a consistent space-time element in the emergency admission rate for respiratory disease, however in some weeks there was evidence of significant spatial variation.

An analysis of emergency admissions for respiratory disease by postcode sector for Glasgow City provided more scope for spatial analysis, as there were more data available for each geographical area. There was found to be a spatial trend in the Glasgow emergency admissions data as well as some evidence of significant spatial autocorrelation after trend removal. A spatial regression analysis demonstrated that most of the spatial trend and spatial variation could be explained by the relative socio-economic deprivation of the postcode sectors. A spatial regression was more appropriate for these data as, like the time series regression described in Chapter 3, it took account of the autocorrelation in the data and adjusted the model coefficients and standard errors accordingly.

6.5 Limitations of the study

6.5.1 Routinely collected health data

The data used in the epidemiological research described in this thesis came from large scale routinely collected data bases; the General Register Office (GRO) death records, the SMR1 hospital in-patients records supplied by the Information and Statistics Division of the NHS, the Meteorological Office's daily temperature measures and the 1981 and 1991 censuses, as well as the additional smaller data on pollution in Glasgow City.

Routinely collected health statistics are commonly used in the monitoring of diseases and generation of hypotheses. The method of analysis of routine data is frequently determined by the level of aggregation, in space and time, at which the data are collected[148]. Patterns observed in these data can often lead to more specific studies involving additional data from other routine sources or to studies in which specific information is collected on a sample of the population.

The main advantage of using routinely collected health data such as the GRO death records and the SMR1 hospital in-patient data is that the data are complete. The records include every death and hospital in-patient stay which occurred in Scotland during any

particular time period, however they inevitably contain some errors mainly through inaccurate recording of details or inaccurate coding of the information recorded. In the SMR1 records there has been found to be greater levels of error and omission in subsequent diagnostic fields when compared to the level of error in the principal diagnostic field[149; 150]. In this thesis only the information on principal diagnosis was required and since these diagnoses were aggregated into larger homogeneous categories, such as 'respiratory disease', it is likely that the level of error will be minimal. Another feature of these data is that diagnostic practices and coding of death certificates may vary in space and over time, however there is little that can be done to avoid this type of imperfection in such a large scale routine data set. These features of the data should be borne in mind when the results of any analyses are interpreted.

The limitations of using hospital in-patient statistics as measures of population morbidity also has to be recognised. The data should not be seen as proxies for population morbidity, but should seen purely as hospital in-patient statistics. To be included in these data a patient's condition has to warrant being admitted to hospital, and there are many other factors which influence the admission of a patient to hospital, even as an emergency, other than purely the condition of the patient. With the mortality data the outcome, death, is dependent solely on the condition of the individual.

6.5.2 Combining separately collected data sets

Having identified the seasonal variation in both mortality and emergency hospital admissions in Scotland, other data sets were required to investigate this further. Information from the 1981 and 1991 censuses was used to provide age standardised rates and postcode sector deprivation indices so that patterns of seasonal variation could be compared across different socio-economic groups. The Meteorological Office provided information on temperature, rainfall and wind speed for three sites in Scotland. This data enabled the direct effects of climate on health to be ascertained, however there are problems involved in using two separately collected data sets to try to establish a cause and effect relationship. The climate data was collected at three sites but averaged to give an all Scotland figure, this value for the temperature was then related to all deaths and hospital admissions from all over Scotland. Both data sets contained a spatial element but

this had to be ignored due to the fact that there was no one to one matching between the data sets. If it had been possible to have recorded, on the hospital admission record, the outdoor temperature recorded at the nearest weather station, for each individual patient more conclusive results may have been reached. As it was the data were aggregated to an all Scotland level and, because of noise, were then aggregated to a weekly level when they had been collected on a daily level. It is likely that a smaller, but more in depth study, with a data collection system designed specifically to evaluate the relationship between climate and ill health would produce more reliable results.

One of the most important features of this thesis was the method in which the information from the two independent time series was combined. A simple analysis could have simply plotted the two series and visually compared them, found opposing peaks and troughs in the seasonal curve and assumed causality. Given that the health data was Poisson with overdispersion and the series were autocorrelated, a method that took both of these features into consideration was preferred to those which could only account for either the non-normality of the data or the serial autocorelation.

When the analysis was extended to examine any space time variations in the pattern of emergency admissions for respiratory disease in Scotland further limitations of the data were discovered. Firstly, the data available for analysis were unlikely to show strong space-time variation. Emergency hospital admissions for respiratory disease were the closest approximation to a health event that may have an infectious nature. Secondly, the information available on spatial location was limited. Only at postcode sector level was it possible to use centroid co-ordinates and examine possible confounding variables. However, when this part of the analysis was planned these limitations were recognised and, as the spatial analysis was included to complement the previous, more detailed, temporal analysis it was felt that such limitations were acceptable.

6.5.3 Other data limitations

Other limitations of the study were that individual measures of socio-economic status were only available for male mortality data. Social class is not recorded on the SMR1 hospital in-patient record and for females this type of occupation based measure of social class is often unreliable. The socio-economic analysis was therefore based on area based

measures of deprivation calculated from the census. These area based measures have been proven to ascribe socio-economic status to a population satisfactorily, however it is also useful to have individual based measures. Ascribing area based measures such as deprivation to individuals within the area for the purpose of demonstrating an association between two variables can result in the degree of association between the two variables being more tenuous than if the data was collected solely on an individual level. This problem has been termed the ecological fallacy, however despite the problems of inadequate control of possible confounding ecological studies are useful in describing the effects of various factors on mortality rates within a population.

6.6 Conclusion

Despite the possible problems involved in undertaking this work, the statistical methods employed have enabled a robust and thorough analysis of the data. This thesis has demonstrated that significant seasonal variation in mortality and emergency hospital admissions still exists in Scotland. The extent of the seasonal increase in mortality, around 30%, varies by cause of death, age, and sex and has slightly declined over time. Seasonal variations for emergency hospital admissions are smaller than those observed for mortality. In general, a 1°C fall in temperature is associated with a 1% increase in mortality one week later, and a smaller but still significant increase in emergency hospital admissions is also observed.

These results demonstrate that seasonal variations in mortality in Scotland which were reported using data from 1977-1983[3] are still present today. This seasonal variation is related to outdoor temperature and it is the most vulnerable members of the population (the very young and the elderly) who are most affected. The thesis has also shown that appropriate statistical techniques must be applied when combining routine data sources, as failure to do so will lead to an overestimation of the cause and effect relationship.

Appendix I

Ia Copy of SMR form

Scotland - Inpatient and Day Case Records Summary Sheet — SMR1 Medical In Confidence

Hospital

Time of admission/transfer — Ward (WARD) 157-159

Previously attended this hospital? YES/NO

If YES state year

Address (ADDR) 160-219

Tel.no. — Postcode

Religion

Next of kin (Relationship)

Name

Address

Tel. no. — Postcode

Family Doctor

Name

Address

Tel. no. — Postcode

Provisional Diagnosis on Admission

Hospital Code (HOSP) 1-5

Hospital Case Reference No. (CRN) 6-15

Surname (SN) 16-35

First Forename (FN1) 36-50

Second Forename (FN2) 51-65

Maiden Name (MN) 66-85

Alternative Case Reference No. (ACRN) 86-95

Age ..... Date of Birth (DOB) 96-103

Sex (SEX) 104 — Marital State (MRST) 105

Postcode (PC) 106-112

GP Practice Code (GPPC) 113-117

GP GMC Number (GPC) 118-124

Admitted/transferred from (ADTF) 125 — Type of Admission (TADM) 126

Date placed on Waiting List (DWL) 127-132

Date of Admission (DOA) 133-138

Date of Discharge (DOD) 139-144

Time of Discharge — Discharge Code (DISC) 145

Category of Patient (CAT) 146 — Type of Facility (TOF) 147

Specialty (SPEC) 148-151

Consultant Surname (CONS) 152-171

Consultant Initials 172-173 — Consultant Code (CONC) 174-180

To be completed by doctor on discharge of patient

(1)	Main Condition (DG1)	(DG1C) 1-66
(2)	Other Conditions (DG2)	(DG2C) 67-132
(3)	(DG3)	(DG3C) 133-198
(4)	(DG4)	(DG4C) 1-66
(5)	(DG5)	(DG5C) 67-132
(6)	(DG6)	(DG6C) 133-198

(1)	Main Operation (OP1)	(OP1C) 1-68
(2)	Other Operations (OP2)	(OP2C) 69-136
(3)	(OP3)	(OP3C) 137-204
(4)	(OP4)	(OP4C) 1-68

Date of Main Operation (DOP1) → Date of Main Operation (DOP1) 69-74

ERROR REPORT COMMENT: Data items specified in boxes below are correct (COM)

Enter field abbreviations

75-134

Additional notes:

Contract No. (CTN) 135-144

Local use (LUSE) 145-150

National use (NUSE) 151-156

Ib Copy of death certificate

Counterfoil – Medical certificate of cause of death

Form 11
F(11)138648

Name of deceased:

Date of death:

Place of death:

Please circle the appropriate letters and figures using the information below

Post mortem	PM1	or	PM2	or	PM3
Procurator fiscal	PF				
Extra information	X				
Attendance on deceased	A1		A2		A3
Maternal deaths	M1		M2		

Cause of death

I (a)

(b)

(c)

(d)

II

Date of certificate:

Medical certificate of cause of death

Form 11
F(11)138648

(Section 24(1) of the Registration of Births, Deaths, and Marriages (Scotland) Act 1965)

The completed certificate is to be taken to the Registrar of Births, Deaths and Marriages

For registration office use
Year:
RD number:
Entry number:

Name of deceased

Date of death — Day | Month | Year

Time of death — Fill in an approximate time if you do not know the exact time (Please use the 24-hour clock) — Hour | Minute

Place of death

Cause of death

I hereby certify that to the best of my knowledge and belief, the cause of death was as stated below:

Approximate interval between onset and death — Years | Months | Days

I Disease or condition directly leading to death*
(a)
due to (or as a consequence of)

Antecedent causes Morbid conditions, if any, giving rise to the above cause, stating the underlying condition last
(b)
due to (or as a consequence of)
(c)
due to (or as a consequence of)
(d)

II Other significant conditions contributing to the death, but not related to the disease or condition causing it

* This does not mean mode of dying, such as heart or respiratory failure; it means the disease, injury or complication that caused death.

Please tick the relevant box

Post mortem
PM1 or ☐ Post mortem has been done and information is included above
PM2 or ☐ Post mortem information may be available later
PM3 ☐ No post mortem is being done

Attendance on deceased
A1 ☐ I was in attendance upon the deceased during last illness
A2 ☐ I was not in attendance upon the deceased during last illness: the doctor who was is unable to provide the certificate
A3 ☐ No doctor was in attendance on the deceased

Procurator fiscal
PF ☐ This death has been reported to the procurator fiscal

Maternal deaths
M1 ☐ Death during pregnancy or within 42 days of the pregnancy ending
M2 ☐ Death between 43 days and 12 months after the end of pregnancy

Extra information for statistical purposes
X ☐ I may later be able to supply the Registrar General with additional information

Signature:

Date:

Name in BLOCK CAPITALS:

Official address:

Registered medical qualifications

For a death in hospital
Name of the consultant responsible for deceased as a patient

Appendix II S-PLUS function for seasonal variation

Function created to compute the amplitude of the sinusoidal curve expressed as the percentage increase from the trough to the peak with 95% confidence intervals

ampfit.fun

```
function(x, y)
{
        fit <- lm(y ~ week)
        fit
        fitted <- fit$fitted.values
        fitted <- log(fitted)
        model <- glim(x, y, error = "poisson", link = "log", intercept = F, offset =
                fitted, eps = 0.1)
        sinval <- model$coef[1]
        sinval2 <- sinval * sinval
        cosval <- model$coef[2]
        cosval2 <- cosval * cosval
        sinvar <- (model$var[1, 1])
        cosvar <- (model$var[2, 2])
        cossinvar <- model$var[1, 2]
        amp <- (sqrt(cosval2 + sinval2))
        varamp <- ((sinval/amp) * (sinval/amp) * sinvar) + ((cosval/amp) * (cosval/amp
                ) * cosvar) + (((sinval * cosval)/(sinval2 + cosval2)) * cossinvar)
        if(amp > 1)
                amp <- amp - 1
        ampa <- (2 * amp)/(1 - amp)
        varampa <- (2/((1 - amp) * (1 - amp))) * varamp
        cil <- ampa - 1.96 * sqrt(varampa)
        ciu <- ampa + 1.96 * sqrt(varampa)
        return(cil, ampa, ciu)
}
```

Appendix III Further results from the seasonal analysis

Respiratory disease

Table IIIa. Seasonal percentage increase in emergency hospital admissions by deprivation category (age standardised)

	Males		Females	
Dep Cat	Percent Increase	CI	Percent Increase	CI
Affluent	53.8	(48.7,58.9)	67.3	(61.1,73.6)
Average	52.7	(47.5,58.0)	70.0	(63.5,76.5)
Deprived	51.7	(47.2,56.1)	63.9	(58.5,69.4)

Table IIIb. Seasonal percentage increase in emergency hospital admissions by health board of residence (age standardised)

Health Board	Percent Increase	Males 95%CI	Percent Increase	Females 95%CI
Argyll & Clyde	53.3	(48.1,58.4)	71.3	(65.0,77.6)
Ayrshire & Arran	53.4	(48.2,58.6)	67.0	(60.8,73.3)
Borders	74.6	(69.2,80.0)	81.2	(75.0,87.4)
Dumfries & Galloway	68.8	(63.0,74.6)	91.1	(84.0,98.1)
Fife	46.4	(41.3,51.6)	59.3	(53.0,65.7)
Forth Valley	56.5	(51.0,62.0)	66.8	(60.3,73.3)
Grampian	55.4	(50.4,60.4)	70.0	(63.5,75.9)
Greater Glasgow	49.1	(44.4,53.7)	65.4	(59.8,71.1)
Highlands	60.9	(55.4,66.4)	57.1	(50.6,63.5)
Lanarkshire	45.4	(40.7,50.1)	60.9	(55.2,66.7)
Lothian	51.7	(47.2,56.2)	69.9	(64.4,75.5)
Tayside	52.1	(47.7,56.6)	82.6	(76.9,88.4)
Islands	36.9	(30.9,42.8)	30.5	(23.5,37.4)

Ischaemic Heart Disease

Table IIIc. Seasonal percentage increase in emergency hospital admissions for IHD by deprivation category (age standardised)

	Males		Females	
Dep Cat	Percent Increase	CI	Percent Increase	CI
Aff	12.6	(7.3,18.0)	9.4	(2.7,16.1)
Ave	15.1	(9.9,20.3)	12.4	(5.9,18.9)
Dep	10.0	(5.5,14.5)	10.5	(4.9,16.2)

Table IIId. Seasonal percentage increase in emergency hospital admissions for IHD by health board of residence (age standardised)

Health Board	Percent Increase	Males 95%CI	Percent Increase	Females 95%CI
Argyll & Clyde	15.8	(10.8,20.8)	12.9	(6.7,19.0)
Ayrshire & Arran	19.1	(13.9,24.3)	4.2	(-1.9,10.4)
Borders	19.3	(14.2,24.4)	14.8	(8.5,21.2)
Dumfries & Galloway	14.5	(9.5,19.6)	13.8	(7.4,20.2)
Fife	8.3	(3.3,13.3)	1.0	(-5.4,7.4)
Forth Valley	18.8	(13.7,23.9)	16.5	(10.1,22.8)
Grampian	10.4	(5.2,15.6)	11.7	(5.3,18.1)
Greater Glasgow	11.9	(7.1,16.8)	10.3	(4.3,16.2)
Highlands	8.1	(2.7,13.4)	4.9	(-1.9,11.9)
Lanarkshire	14.3	(9.5,19.2)	12.1	(6.1,18.1)
Lothian	12.7	(8.0,17.4)	12.5	(6.6,18.4)
Tayside	14.4	(9.7,19.0)	16.7	(10.9,22.5)
Islands	6.8	(0.7,12.9)	10.4	(2.0,18.9)

Cerebrovascular Disease

Table IIIe. Seasonal percentage increase in emergency hospital admissions for CVD by deprivation category (age standardised)

	Males		Females	
Dep Cat	Percent Increase	CI	Percent Increase	CI
Afflent	14.3	(5.3,23.3)	16.3	(7.5,25.2)
Average	10.7	(2.3,19.0)	12.8	(4.7,21.0)
Deprived	14.9	(7.0,22.8)	16.0	(8.4,23.7)

Table IIIf. Seasonal percentage increase in emergency hospital admissions for CVD by health board of residence (age standardised)

Health Board	Percent Increase	Males 95%CI	Percent Increase	Females 95%CI
Argyll & Clyde	20.0	(11.7,28.2)	19.2	(11.4,26.9)
Ayrshire & Arran	5.4	(-3.2,14.1)	21.5	(12.6,30.5)
Borders	4.3	(-3.3,11.9)	15.8	(8.3,23.3)
Dumfries & Galloway	8.3	(0.8,15.8)	10.3	(3.3,17.2)
Fife	11.6	(3.1,20.0)	12.5	(4.5,20.5)
Forth Valley	11.8	(3.4,20.2)	22.4	(14.3,30.7)
Grampian	17.9	(9.6,26.1)	13.8	(6.1,21.7)
Greater Glasgow	14.5	(6.4,22.6)	21.4	(13.5,29.2)
Highlands	8.2	(1.2,16.3)	17.6	(9.5,25.7)
Lanarkshire	11.4	(3.4,19.4)	10.6	(3.1,18.2)
Lothian	12.8	(4.9,20.6)	13.4	(5.8,21.0)
Tayside	17.7	(10.0,25.4)	9.6	(2.3,16.9)
Islands	12.9	(3.7,22.1)	8.8	(0.4,17.2)

Appendix IV Further results from Zeger's model

Section I

Cause of death		Age group									
Respiratory Disease		0-9		10-59		60-69		70-79		80+	
		coeff	se	coeff	se	coeff	se	coeff	se	coeff	se
Males	Const	-1.2147	.0749	1.6339	.0178	2.4026	.0150	3.1869	.0094	3.2163	.0097
	Trend	-.0015	.0004	-.0006	.0001	-.0002	.0001	-.0003	.0001	.0005	.0001
	Week 0	-.0256	.0284	-.0102	.0080	.0089	.0060	-.0074	.0039	-.0123	.0039
	Week 1	.0194	.0332	-.0129	.0096	-.0064	.0067	-.0123	.0045	-.0162	.0044
	Week 2	-.0215	.0329	-.0141	.0096	-.0186	.0068	-.0083	.0046	-.0025	.0045
	Week 3	-.0341	.0329	-.0025	.0096	-.0118	.0068	-.0087	.0046	-.0042	.0045
	Week 4	-.0020	.0331	.0003	.0098	-.0004	.0067	-.0057	.0046	-.0105	.0044
	Week 5	-.0035	.0284	-.0022	.0079	-.0018	.0060	-.0038	.0040	-.0017	.0039
	Flu	1.3720	.2714	.2487	.0402	.2003	.0369	.1044	.0129	.0723	.0075
	AR	.0810	.0385	-.0050	.0385	.1600	.0385	.1047	.0385	.1461	.0385
Females											
	Const	-1.6120	.0899	1.3010	.0220	2.0871	.0174	2.8409	.0124	3.6491	.0103
	Trend	-.0015	.0004	-.0005	.0001	.0003	.0001	.0002	.0001	.0006	.0001
	Week 0	-.0611	.0338	-.0058	.0095	-.0157	.0067	-.0030	.0048	-.0036	.0036
	Week 1	-.0372	.0433	-.0315	.0112	-.0069	.0076	-.0152	.0054	-.0064	.0038
	Week 2	.0623	.0445	-.0186	.0113	-.0198	.0077	-.0127	.0055	-.0171	.0039
	Week 3	-.0511	.0433	.0010	.0113	.0045	.0076	-.0163	.0055	-.0122	.0039
	Week 4	-.0205	.0433	.0216	.0113	-.0113	.0076	-.0029	.0054	-.0027	.0038
	Week 5	.0442	.0361	-.0146	.0095	-.0095	.0067	-.0013	.0048	-.0077	.0036
	Flu	-	-	.2048	.1015	.1855	.0183	.1531	.0153	.0609	.0043
	AR	-.0065	.0385	.0642	.0385	.1595	.0385	.1556	.0385	.2540	.0385

Cause of death		Age group									
All causes		0-9		10-59		60-69		70-79		80+	
		coeff	se	coeff	se	coeff	se	coeff	se	coeff	se
Males	Const	2.1149	.0134	4.6388	.0047	4.9077	.0047	5.2670	.0034	4.9006	.0042
	Trend	-.0008	.0001	-.0004	.0000	-.0002	.0000	-.0003	.0000	.0005	.0000
	Week 0	.0028	.0060	-.0006	.0020	-.0073	.0018	-.0077	.0014	-.0069	.0018
	Week 1	-.0076	.0071	-.0009	.0023	-.0085	.0020	-.0103	.0017	-.0118	.0021
	Week 2	-.0019	.0071	-.0052	.0023	.0009	.0020	-.0012	.0017	-.0025	.0021
	Week 3	-.0009	.0072	-.0033	.0023	-.0010	.0020	-.0016	.0017	-.0023	.0021
	Week 4	.0022	.0072	-.0003	.0023	.0019	.0020	-.0050	.0017	-.0061	.0020
	Week 5	-.0066	.0060	.0013	.0020	-.0019	.0018	.0034	.0014	.0025	.0018
	Flu	.1069	.1424	.0156	.0146	.0628	.0149	.0326	.0046	.0290	.0024
	AR	.0206	.0385	.1091	.0385	.2058	.0385	.1006	.0385	.1037	.0385
Females											
	Const	1.7708	.0158	4.1267	.0052	4.5452	.0047	5.1920	.0037	5.5910	.0037
	Trend	-.0009	.0001	-.0004	.0000	-.0002	.0000	-.0002	.0000	.0004	.0000
	Week 0	.0012	.0072	-.0037	.0023	-.0068	.0020	-.0052	.0015	-.0035	.0014
	Week 1	-.0111	.0086	-.0086	.0027	-.0060	.0023	-.0117	.0017	-.0085	.0015
	Week 2	-.0055	.0087	-.0038	.0027	-.0008	.0024	-.0010	.0017	-.0053	.0015
	Week 3	-.0022	.0087	.0020	.0027	-.0053	.0024	-.0037	.0017	-.0041	.0015
	Week 4	-.0019	.0087	.0039	.0027	.0017	.0023	-.0011	.0017	.0000	.0015
	Week 5	.0025	.0072	-.0021	.0023	-.0012	.0020	.0016	.0015	-.0025	.0014
	Flu	.0025	.0072	.0628	.0227	.0597	.0116	.0323	.0049	.0219	.0018
	AR	.0140	.0385	.0507	.0385	.0954	.0385	.1455	.0385	.2116	.0385

Cause of death		Age group									
Ischaemic Heart Disease		0-9		10-59		60-69		70-79		80+	
		coeff	se	coeff	se	coeff	se	coeff	se	coeff	se
Males	Const	-	-	3.4437	.0074	3.9280	.0067	4.1558	.0054	3.5573	.0074
	Trend	-	-	-.0008	.0000	-.0003	.0000	-.0003	.0000	.0005	.0000
	Week 0	-	-	-.0028	.0034	-.0089	.0025	-.0165	.0023	-.0098	.0032
	Week 1	-	-	-.0059	.0041	-.0107	.0028	-.0116	.0027	-.0162	.0037
	Week 2	-	-	-.0026	.0041	-.0008	.0029	-.0005	.0027	-.0032	.0037
	Week 3	-	-	-.0028	.0042	.0006	.0029	.0007	.0027	-.0027	.0037
	Week 4	-	-	-.0012	.0042	.0004	.0028	-.0066	.0027	-.0024	.0037
	Week 5	-	-	-.0008	.0034	-.0011	.0025	.0052	.0023	.0049	.0032
	Flu	-	-	-.0193	.0254	.0392	.0148	.0085	.0059	.0197	.0037
	AR	-	-	-.0151	.0385	.1663	.0385	.0856	.0385	.0770	.0385
Females											
	Const	-	-	2.2141	.0135	3.2590	.0086	3.9944	.0058	4.1589	.0057
	Trend	-	-	-.0007	.0001	-.0004	.0000	-.0003	.0000	.0004	.0000
	Week 0	-	-	-.0043	.0060	-.0109	.0036	-.0109	.0024	-.0073	.0024
	Week 1	-	-	-.0051	.0071	-.0062	.0042	-.0103	.0029	-.0102	.0028
	Week 2	-	-	-.0039	.0071	-.0017	.0042	-.0048	.0029	-.0047	.0028
	Week 3	-	-	.0023	.0072	-.0084	.0042	-.0044	.0029	-.0032	.0028
	Week 4	-	-	.0059	.0071	.0047	.0042	.0024	.0028	.0021	.0028
	Week 5	-	-	-.0118	.0059	.0001	.0036	.0029	.0024	-.0027	.0024
	Flu	-	-	.1697	.0478	.0401	.0140	.0141	.0046	.0069	.0008
	AR	-	-	.0277	.0385	.0921	.0385	.0906	.0385	.0978	.0385

Cause of death		Age group									
Cerebrovascular Disease		0-9		10-59		60-69		70-79		80+	
		coeff	se	coeff	se	coeff	se	coeff	se	coeff	se
Males	Const	-	-	1.5691	.0167	2.3716	.0121	3.1481	.0086	2.9961	.0082
	Trend	-	-	-.0008	.0001	-.0006	.0001	-.0005	.0000	.0004	.0000
	Week 0	-	-	-.0080	.0077	-.0028	.0052	-.0075	.0036	-.0065	.0039
	Week 1	-	-	-.0005	.0094	-.0108	.0061	-.0082	.0042	-.0042	.0047
	Week 2	-	-	-.0113	.0093	-.0044	.0062	-.0052	.0043	-.0043	.0047
	Week 3	-	-	-.0052	.0094	-.0057	.0062	-.0052	.0043	-.0085	.0047
	Week 4	-	-	.0013	.0094	.0013	.0062	-.0074	.0042	-.0122	.0047
	Week 5	-	-	.0040	.0077	.0050	.0052	.0083	.0038	.0057	.0039
	Flu	-	-	.0699	.0506	.0680	.0223	.0291	.0065	.0188	.0039
	AR	-	-	-.0096	.0385	.0452	.0385	.0813	.0385	.0200	.0385
Females											
	Const	-	-	1.5181	.0197	2.2549	.0123	3.3748	.0069	4.0215	.0062
	Trend	-	-	-.0009	.0001	-.0007	.0001	-.0006	.0000	.0003	.0000
	Week 0	-	-	-.0141	.0080	-.0153	.0056	-.0055	.0033	-.0059	.0025
	Week 1	-	-	-.0117	.0094	-.0056	.0068	-.0181	.0040	-.0030	.0029
	Week 2	-	-	.0109	.0094	.0034	.0068	.0071	.0040	-.0091	.0029
	Week 3	-	-	.0140	.0095	-.0049	.0068	-.0037	.0040	-.0036	.0029
	Week 4	-	-	-.0090	.0093	-.0021	.0068	-.0043	.0040	.0034	.0029
	Week 5	-	-	-.0110	.0079	-.0001	.0056	-.0012	.0032	-.0064	.0025
	Flu	-	-	.0139	.0777	.0740	.0123	.0240	.0040	.0077	.0012
	AR	-	-	.0719	.0385	.0020	.0385	-.0274	.0385	.1331	.0385

Emergency admissions

Diagnosis		Age group									
Respiratory Disease		0-9		10-59		60-69		70-79		80+	
		coeff	se	coeff	se	coeff	se	coeff	se	coeff	se
Males	Const	4.7654	.0263	4.4270	.0065	3.8930	.0083	4.1076	.0090	3.5604	.0105
	Trend	.0009	.0001	.0006	.0000	.0006	.0000	.0004	.0001	.0012	.0001
	Week 0	.0040	.0028	-.0011	.0023	.0066	.0031	-.0012	.0029	-.0059	.0036
	Week 1	-.0116	.0028	-.0058	.0025	-.0154	.0034	-.0126	.0031	-.0161	.0038
	Week 2	-.0123	.0027	-.0068	.0026	-.0082	.0035	-.0094	.0031	-.0117	.0039
	Week 3	-.0092	.0027	.0004	.0026	-.0054	.0034	-.0028	.0031	.0038	.0039
	Week 4	-.0054	.0028	.0042	.0025	-.0077	.0033	-.0058	.0031	-.0077	.0038
	Week 5	-.0029	.0028	.0056	.0023	.0023	.0030	-.0035	.0029	-.0058	.0036
	Flu	.0062	.0070	.0250	.0037	.0700	.0153	.0553	.0110	.0657	.0124
	AR	.7541	.0385	.2456	.0385	.2238	.0385	.3289	.0385	.2772	.0385
Females											
	Const	4.2314	.0254	4.3897	.0066	3.6781	.0108	3.8801	.0092	3.6710	.0112
	Trend	.0008	.0001	.0010	.0000	.0010	.0001	.0011	.0000	.0013	.0001
	Week 0	.0011	.0033	-.0104	.0025	.0010	.0035	-.0032	.0033	-.0062	.0036
	Week 1	-.0152	.0033	-.0069	.0027	-.0120	.0037	-.0139	.0037	-.0128	.0039
	Week 2	-.0146	.0032	-.0087	.0027	-.0067	.0038	-.0112	.0038	-.0144	.0039
	Week 3	-.0129	.0032	.0042	.0027	-.0100	.0037	-.0009	.0037	-.0103	.0039
	Week 4	-.0089	.0032	.0031	.0027	-.0072	.0036	-.0064	.0037	-.0041	.0038
	Week 5	-.0023	.0033	.0075	.0024	-.0002	.0034	-.0067	.0033	-.0020	.0036
	Flu	-.0456	.0418	.0328	.0041	.0656	.0140	.0832	.0110	.0848	.0092
	AR	.7207	.0385	.2185	.0385	.3277	.0385	.2315	.0385	.3146	.0385

Diagnosis		Age group									
Ischaemic Heart Disease		0-9		10-59		60-69		70-79		80+	
		coeff	se	coeff	se	coeff	se	coeff	se	coeff	se
Males	Const	-	-	4.5781	.0047	4.3436	.0052	4.0467	.0054	2.8930	.0099
	Trend	-	-	.0003	.0000	.0006	.0000	.0006	.0000	.0015	.0000
	Week 0	-	-	.0076	.0020	.0041	.0022	-.0038	.0024	-.0018	.0043
	Week 1	-	-	-.0072	.0023	-.0100	.0025	-.0034	.0028	-.0132	.0051
	Week 2	-	-	-.0011	.0024	-.0006	.0025	-.0031	.0028	-.0086	.0051
	Week 3	-	-	-.0019	.0024	-.0023	.0025	-.0033	.0028	.0015	.0051
	Week 4	-	-	-.0018	.0023	-.0003	.0025	.0056	.0028	.0056	.0051
	Week 5	-	-	-.0013	.0020	.0006	.0022	-.0037	.0024	.0031	.0044
	Flu	-	-	-.0065	.0031	-.0104	.0079	.0044	.0053	.0178	.0091
	AR	-	-	.0972	.0385	.1112	.0385	.0368	.0385	.0576	.0385
Females											
	Const	-	-	3.6062	.0072	3.8357	.0064	4.0261	.0058	3.4912	.0076
	Trend	-	-	.0004	.0000	.0006	.0000	.0006	.0003	.0014	.0000
	Week 0	-	-	.0091	.0031	.0031	.0028	.0003	.0026	-.0036	.0032
	Week 1	-	-	-.0088	.0037	-.0085	.0032	-.0121	.0031	-.0130	.0037
	Week 2	-	-	.0034	.0037	.0016	.0033	-.0017	.0031	-.0021	.0037
	Week 3	-	-	-.0060	.0037	-.0014	.0032	.0008	.0031	.0035	.0038
	Week 4	-	-	.0046	.0037	-.0024	.0032	.0029	.0031	-.0020	.0037
	Week 5	-	-	-.0039	.0032	-.0034	.0028	.0012	.0026	.0045	.0032
	Flu	-	-	-.0031	.0043	-.0246	.0080	-.0002	.0051	.0016	.0035
	AR	-	-	.0752	.0385	.0794	.0385	.0398	.0385	.0744	.0385

Diagnosis		Age group									
Cerebrovascular Disease		0-9		10-59		60-69		70-79		80+	
		coeff	se	coeff	se	coeff	se	coeff	se	coeff	se
Males	Const	-	-	2.7974	.0099	3.1579	.0080	3.4928	.0072	2.8511	.0089
	Trend	-	-	.0001	.0001	.0003	.0000	.0003	.0000	.0013	.0001
	Week 0	-	-	-.0004	.0045	-.0008	.0038	-.0044	.0033	-.0103	.0043
	Week 1	-	-	-.0012	.0054	-.0042	.0046	-.0062	.0039	-.0056	.0052
	Week 2	-	-	.0054	.0054	.0042	.0046	.0038	.0039	-.0044	.0052
	Week 3	-	-	-.0088	.0054	.0038	.0046	-.0061	.0039	.0027	.0052
	Week 4	-	-	-.0044	.0054	-.0040	.0046	.0021	.0039	-.0033	.0052
	Week 5	-	-	.0029	.0045	-.0015	.0038	.0018	.0033	.0039	.0043
	Flu	-	-	-.0066	.0069	-.0083	.0126	.0153	.0076	.0018	.0088
	AR	-		.0288	.0385	-.0229	.0385	.0437	.0385	-.0364	.0385
Females											
	Const	-	-	2.5207	.0118	2.9162	.0093	3.6512	.0069	3.6548	.0065
	Trend	-	-	.0000	.0001	.0001	.0000	.0002	.0000	.0009	.0000
	Week 0	-	-	-.0086	.0052	-.0034	.0044	-.0058	.0030	-.0048	.0030
	Week 1	-	-	.0024	.0063	-.0007	.0053	-.0072	.0035	-.0108	.0036
	Week 2	-	-	-.0067	.0063	-.0052	.0053	.0004	.0036	.0052	.0036
	Week 3	-	-	.0094	.0063	.0029	.0053	.0012	.0036	-.0022	.0036
	Week 4	-	-	.0021	.0063	.0041	.0053	-.0022	.0036	-.0009	.0036
	Week 5	-	-	-.0015	.0052	-.0063	.0044	.0025	.0030	.0021	.0030
	Flu	-	-	-.0005	.0073	.0107	.0122	.0124	.0067	.0071	.0036
	AR	-	-	.0332	.0385	-.0048	.0385	.0635	.0385	.0151	.0385

Section II

Cause of death		Deprivation Category					
All causes		Affluent		Average		Deprived	
		coeff	se	coeff	se	coeff	se
Males	Const	3.0730	.0055	3.2318	.0033	3.3589	.0063
	Trend	-.0004	.0000	-.0003	.0000	-.0003	.0000
	Week 0	-.0097	.0021	-.0057	.0013	-.0028	.0021
	Week 1	-.0086	.0023	-.0093	.0014	-.0093	.0022
	Week 2	.0005	.0023	-.0016	.0014	-.0041	.0023
	Week 3	-.0008	.0023	-.0013	.0014	-.0036	.0023
	Week 4	-.0004	.0023	-.0020	.0014	-.0041	.0022
	Week 5	-.0009	.0021	.0005	.0013	.0024	.0021
	Flu	.2490	.0405	.2473	.0351	.1116	.0539
	AR	.2030	.0385	.1731	.0385	.3149	.0385
Females							
	Const	3.1184	.0048	3.2097	.0029	3.2729	.0064
	Trend	-.0002	.0000	-.0003	.0000	-.0003	.0000
	Week 0	-.0048	.0018	-.0055	.0011	-.0016	.0021
	Week 1	-.0061	.0021	-.0104	.0013	-.0106	.0023
	Week 2	-.0050	.0021	-.0015	.0013	-.0059	.0023
	Week 3	-.0064	.0021	-.0020	.0013	-.0036	.0023
	Week 4	.0019	.0021	-.0006	.0013	.0003	.0022
	Week 5	-.0023	.0018	.0000	.0011	-.0020	.0021
	Flu	.1778	.0157	.2787	.0205	.0143	.0258
	AR	.1854	.0385	.1745	.0385	.3120	.0385

Cause of death		Deprivation category					
Respiratory Disease		Affluent		Average		Deprived	
		coeff	se	coeff	se	coeff	se
Males	Const	.8548	.0134	1.0391	.0084	1.3546	.0126
	Trend	-.0003	.0001	-.0003	.0000	-.0004	.0001
	Week 0	-.0125	.0057	-.0121	.0034	-.0012	.0055
	Week 1	-.0051	.0066	-.0130	.0038	-.0234	.0059
	Week 2	-.0134	.0066	-.0082	.0039	-.0045	.0060
	Week 3	-.0144	.0066	-.0008	.0039	-.0160	.0059
	Week 4	.0093	.0066	-.0120	.0038	-.0020	.0059
	Week 5	-.0142	.0057	.0008	.0034	-.0028	.0051
	Flu	.8417	.0762	.7789	.0780	.5885	.1000
	AR	.0768	.0385	.1494	.0385	.1284	.0385
Females							
	Const	.9799	.0141	1.0751	.0097	1.3012	.0163
	Trend	.0002	.0001	.0000	.0001	.0001	.0001
	Week 0	-.0068	.0053	-.0058	.0036	-.0024	.0057
	Week 1	-.0109	.0059	-.0094	.0039	-.0095	.0062
	Week 2	-.0026	.0059	-.0159	.0040	-.0251	.0063
	Week 3	-.0174	.0059	-.0100	.0040	-.0087	.0063
	Week 4	-.0040	.0058	-.0038	.0040	.0023	.0062
	Week 5	-.0149	.0053	-.0016	.0036	-.0131	.0057
	Flu	.8056	.0809	.8571	.0666	.5224	.0722
	AR	.1866	.0385	.2199	.0385	.2434	.0385

Cause of death		Deprivation Category					
Ischaemic Heart Disease		Affluent		Average		Deprived	
		coeff	se	coeff	se	coeff	se
Males	Const	1.9122	.0084	2.0890	.0050	2.1425	.0078
	Trend	-.0006	.0000	-.0004	.0000	-.0003	.0000
	Week 0	-.0115	.0034	-.0106	.0020	-.0102	.0033
	Week 1	-.0142	.0040	-.0119	.0022	-.0098	.0038
	Week 2	.0057	.0040	-.0025	.0023	-.0049	.0038
	Week 3	-.0005	.0040	.0003	.0023	-.0035	.0038
	Week 4	-.0043	.0040	-.0018	.0022	-.0032	.0038
	Week 5	-.0006	.0035	.0013	.0020	.0061	.0033
	Flu	.1984	.0509	.0711	.0375	.1043	.0609
	AR	.1147	.0385	.1660	.0385	.1085	.0385
Females							
	Const	1.7854	.0083	1.9238	.0056	1.9881	.0088
	Trend	-.0004	.0000	-.0003	.0000	-.0003	.0000
	Week 0	-.0100	.0033	-.0080	.0022	-.0109	.0036
	Week 1	-.0062	.0039	-.0127	.0024	-.0038	.0040
	Week 2	-.0041	.0039	-.0047	.0025	-.0046	.0041
	Week 3	-.0060	.0039	-.0017	.0024	-.0081	.0041
	Week 4	.0042	.0039	.0034	.0024	.0013	.0040
	Week 5	-.0007	.0034	-.0014	.0022	-.0042	.0036
	Flu	.1063	.0221	.1136	.0162	.0748	.0394
	AR	.1033	.0385	.1847	.0385	.1392	.0385

Cause of death		Deprivation category					
Cerebrovascular Disease		Affluent		Average		Deprived	
		coeff	se	coeff	se	coeff	se
Males	Const	.8649	.0112	.9854	.0065	.9843	.0128
	Trend	-.0004	.0001	-.0005	.0000	-.0005	.0001
	Week 0	-.0040	.0051	-.0082	.0030	-.0034	.0053
	Week 1	-.0136	.0062	-.0047	.0037	-.0083	.0061
	Week 2	-.0021	.0062	-.0055	.0037	-.0107	.0062
	Week 3	-.0064	.0062	-.0065	.0037	-.0029	.0062
	Week 4	-.0024	.0062	-.0065	.0037	-.0117	.0061
	Week 5	.0044	.0052	.0056	.0030	.0114	.0054
	Flu	.1740	.0535	.2473	.0457	.0884	.0829
	AR	-.0084	.0385	-.0039	.0385	.0835	.0385
Females							
	Const	1.4369	.0092	1.5197	.0054	1.4580	.0105
	Trend	-.0003	.0001	-.0005	.0000	-.0005	.0001
	Week 0	-.0080	.0040	-.0091	.0024	.0035	.0048
	Week 1	-.0020	.0048	-.0084	.0028	-.0162	.0056
	Week 2	-.0062	.0048	-.0016	.0029	-.0082	.0056
	Week 3	-.0046	.0048	-.0044	.0029	.0076	.0056
	Week 4	.0006	.0048	.0029	.0028	-.0058	.0056
	Week 5	-.0071	.0040	-.0038	.0024	-.0060	.0047
	Flu	.1198	.0340	.1852	.0268	.1044	.0294
	AR	.0583	.0385	.0569	.0385	.0317	.0385

Emergency admissions

Diagnosis		Deprivation category					
Respiratory Disease		Affluent		Average		Deprived	
		coeff	se	coeff	se	coeff	se
Males	Const	2.6076	.0140	2.5335	.0107	2.8432	.0102
	Trend	-.0001	.0001	.0009	.0005	.0009	.0001
	Week 0	-.0025	.0030	.0001	.0021	-.0056	.0027
	Week 1	-.0122	.0030	-.0150	.0021	-.0110	.0028
	Week 2	-.0078	.0030	-.0086	.0021	-.0138	.0028
	Week 3	-.0071	.0030	-.0041	.0021	-.0033	.0028
	Week 4	-.0051	.0030	-.0005	.0021	-.0057	.0027
	Week 5	.0020	.0030	-.0032	.0021	.0084	.0027
	Flu	.1076	.0398	.1827	.0463	.1331	.0370
	AR	.5382	.0385	.5894	.0385	.4465	.0385
Females							
	Const	2.3026	.0155	2.2042	.0097	2.5127	.0111
	Trend	.0002	.0001	.0012	.0001	.0014	.0001
	Week 0	-.0092	.0033	-.0047	.0023	-.0065	.0031
	Week 1	-.0111	.0034	-.0141	.0024	-.0144	.0032
	Week 2	-.0137	.0034	-.0112	.0024	-.0111	.0032
	Week 3	-.0048	.0034	-.0044	.0024	.0013	.0032
	Week 4	-.0014	.0033	-.0020	.0023	-.0067	.0031
	Week 5	.0001	.0033	-.0009	.0023	.0009	.0031
	Flu	.1203	.0489	.3275	.0452	.1920	.0434
	AR	.5211	.0385	.5016	.0385	.4172	.0385

Diagnosis		Deprivation Category					
Ischaemic Heart Disease		Affluent		Average		Deprived	
		coeff	se	coeff	se	coeff	se
Males	Const	2.1743	.0091	2.2639	.0041	2.5075	.0066
	Trend	.0002	.0000	.0006	.0000	.0005	.0000
	Week 0	.0065	.0032	.0015	.0017	.0022	.0027
	Week 1	-.0142	.0035	-.0090	.0020	-.0038	.0030
	Week 2	.0029	.0036	-.0032	.0020	-.0061	.0031
	Week 3	-.0060	.0036	-.0004	.0020	-.0015	.0031
	Week 4	-.0016	.0035	.0017	.0020	.0004	.0030
	Week 5	.0036	.0032	-.0031	.0017	.0011	.0027
	Flu	.0246	.0345	-.0259	.0264	.0075	.0310
	AR	.2400	.0385	.0929	.0385	.1460	.0385
Females							
	Const	1.7155	.0117	1.7924	.0051	2.0504	.0072
	Trend	.0003	.0001	.0008	.0000	.0008	.0000
	Week 0	.0044	.0038	.0010	.0021	.0028	.0031
	Week 1	-.0131	.0041	-.0089	.0024	-.0172	.0035
	Week 2	.0000	.0042	-.0045	.0025	.0104	.0036
	Week 3	.0046	.0042	.0009	.0025	-.0085	.0036
	Week 4	.0024	.0041	-.0013	.0024	.0076	.0036
	Week 5	-.0073	.0038	.0001	.0021	-.0042	.0031
	Flu	-.0073	.0474	-.0528	.0246	-.0090	.0319
	AR	.3107	.0385	.1213	.0385	.0738	.0385

Diagnosis		Deprivation category					
Cerebrovascular Disease		Affluent		Average		Deprived	
		coeff	se	coeff	se	coeff	se
Males	Const	1.1684	.0100	1.2772	.0059	1.4377	.0094
	Trend	.0002	.0001	.0004	.0000	.0005	.0000
	Week 0	-.0026	.0046	-.0034	.0027	-.0090	.0042
	Week 1	-.0089	.0055	-.0057	.0031	-.0017	.0049
	Week 2	.0015	.0055	.0022	.0031	.0027	.0050
	Week 3	-.0017	.0055	.0014	.0031	-.0087	.0049
	Week 4	-.0077	.0054	-.0034	.0031	.0065	.0049
	Week 5	.0093	.0046	-.0003	.0027	-.0031	.0042
	Flu	.0021	.0051	.0102	.0413	-.0401	.0451
	AR	.0286	.0385	.0503	.0385	.0412	.0385
Females							
	Const	1.2213	.0106	1.3532	.0054	1.4983	2.6076
	Trend	.0002	.0001	.0003	.0000	.0003	-.0001
	Week 0	-.0025	.0045	-.0049	.0025	-.0007	-.0025
	Week 1	-.0123	.0052	-.0045	.0030	-.0091	-.0122
	Week 2	-.0003	.0053	-.0007	.0030	.0005	-.0078
	Week 3	-.0073	.0052	.0037	.0030	-.0010	-.0071
	Week 4	.0077	.0052	-.0039	.0030	.0017	-.0055
	Week 5	.0005	.0045	-.0001	.0025	-.0029	.0020
	Flu	-.0528	.0516	.0413	.0336	.0781	.1076
	AR	.1121	.0385	.0247	.0385	-.0079	.0385

Section III

Males Only		Social Class					
Cause of death		I&II		III		IV&V	
		coeff	se	coeff	se	coeff	se
All Causes	Const	2.6457	.0051	3.4708	.0041	3.3907	.0042
	Trend	.0000	.0000	-.0002	.0000	.0000	.0000
	Week 0	-.0077	.0021	-.0048	.0014	-.0059	.0017
	Week 1	-.0065	.0025	-.0096	.0016	-.0087	.0019
	Week 2	-.0017	.0025	-.0028	.0016	-.0012	.0012
	Week 3	-.0014	.0025	-.0006	.0016	-.0024	.0019
	Week 4	-.0030	.0025	-.0011	.0016	-.0011	.0019
	Week 5	.0036	.0021	-.0008	.0014	.0000	.0017
	Flu	.2556	.0579	.2106	.0388	.2471	.0384
	AR	.0952	.0385	.2562	.0385	.1670	.0385
Respiratory Disease	Const	.1586	.0134	1.3293	.0098	1.2649	.0099
	Trend	.0002	.0001	-.0002	.0000	.0001	.0001
	Week 0	-.0009	.0058	-.0096	.0040	-.0151	.0041
	Week 1	-.0158	.0069	-.0150	.0046	-.0073	.0047
	Week 2	-.0127	.0069	-.0097	.0047	-.0054	.0048
	Week 3	-.0085	.0069	-.0033	.0047	-.0120	.0048
	Week 4	-.0065	.0069	-.0073	.0046	.0018	.0047
	Week 5	-.0051	.0058	-.0046	.0040	-.0039	.0041
	Flu	1.1750	.1233	.6870	.0807	.6377	.0565
	AR	.0284	.0385	.1083	.0385	.1220	0.385

Males Only		Social class					
Cause of death		I&II		III		IV&V	
		coeff	se	coeff	se	coeff	se
Ischaemic	Const	1.3545	.0075	2.1344	.0066	2.0277	.0066
Heart	Trend	-.0004	.0000	-.0005	.0000	-.0002	.0000
Disease	Week 0	-.0098	.0034	-.0123	.0024	-.0099	.0028
	Week 1	-.0110	.0040	-.0110	.0027	-.0134	.0033
	Week 2	.0011	.0041	-.0018	.0028	.0008	.0033
	Week 3	-.0014	.0041	.0011	.0028	-.0034	.0033
	Week 4	-.0063	.0040	-.0002	.0027	.0015	.0033
	Week 5	.0076	.0034	-.0009	.0024	.0006	.0028
	Flu	-.0537	.0922	.0964	.0362	.1014	.0459
	AR	.0365	.0385	.1814	.0385	1070	.0385
Cerebrovascular							
Disease	Const	.3721	.0108	1.2310	.0085	1.1014	.0097
	Trend	-.0001	.0001	-.0004	.0000	-.0001	.0001
	Week 0	-.0053	.0050	-.0057	.0036	-.0058	.0041
	Week 1	-.0062	.0060	-.0085	.0042	-.0056	.0049
	Week 2	-.0049	.0060	-.0073	.0042	-.0069	.0049
	Week 3	-.0134	.0060	-.0056	.0042	.0003	.0049
	Week 4	-.0078	.0060	-.0017	.0042	-.0113	.0049
	Week 5	.0124	.0050	.0043	.0036	.0031	.0042
	Flu	.3007	.1168	.1195	.0536	.2711	.0451
	AR	.0139	.0385	.0937	.0385	.0615	.0385

Section IV

Cause of death		City of Residence					
All causes		Aberdeen		Edinburgh		Glasgow	
		coeff	se	coeff	se	coeff	se
Males	Const	3.1530	.0092	3.1735	.0056	3.1414	.0050
	Trend	-.0003	.0000	-.0003	.0000	-.0002	.0000
	Week 0	-.0037	.0040	-.0090	.0025	-.0032	.0020
	Week 1	-.0140	.0046	-.0135	.0031	-.0080	.0023
	Week 2	-.0012	.0046	-.0052	.0031	-.0033	.0023
	Week 3	.0081	.0046	.0021	.0031	-.0054	.0023
	Week 4	-.0056	.0046	-.0010	.0031	-.0018	.0023
	Week 5	-.0025	.0040	.0039	.0025	-.0003	.0020
	Flu	.1468	.0588	.2122	.0285	.1462	.0447
	AR	.0696	.0385	-.0232	.0385	.0966	.0385
Females							
	Const	3.1194	.0086	3.1380	.0057	3.3635	.0055
	Trend	-.0004	.0000	-.0002	.0000	.0000	.0000
	Week 0	-.0015	.0038	-.0050	.0025	-.0034	.0019
	Week 1	-.0069	.0044	-.0075	.0029	-.0098	.0022
	Week 2	-.0116	.0044	-.0069	.0029	-.0059	.0022
	Week 3	-.0059	.0044	-.0063	.0029	-.0010	.0022
	Week 4	-.0010	.0044	-.0005	.0029	-.0004	.0021
	Week 5	.0056	.0038	.0013	.0025	-.0045	.0019
	Flu	.1240	.0291	.1300	.0110	.1141	.0260
	AR	.0635	.0385	.0658	.0385	.1918	.0385

Cause of death		City of residence					
Respiratory Disease		Aberdeen		Edinburgh		Glasgow	
		coeff	se	coeff	se	coeff	se
Males	Const	.9727	.0257	.9476	.0174	1.4173	.0138
	Trend	-.0003	.0001	-.0005	.0001	-.0003	.0001
	Week 0	.0001	.0116	-.0114	.0076	-.0019	.0052
	Week 1	-.0341	.0137	.0008	.0089	-.0284	.0061
	Week 2	-.0055	.0138	-.0157	.0089	.0043	.0061
	Week 3	-.0099	.0138	-.0083	.0089	-.0118	.0061
	Week 4	-.0086	.0138	-.0093	.0089	-.0049	.0062
	Week 5	.0026	.0117	-.0076	.0074	-.0089	.0052
	Flu	.1716	.2424	.6320	.0389	.4485	.0737
	AR	.0111	.0385	.0337	.0385	.0979	.0385
Females							
	Const	.9074	.0255	.8684	.0206	1.3493	.0161
	Trend	-.0004	.0001	.0000	.0001	.0003	.0001
	Week 0	-.0069	.0113	-.0032	.0077	-.0048	.0057
	Week 1	-.0072	.0131	-.0060	.0086	-.0113	.0064
	Week 2	-.0448	.0133	-.0288	.0087	-.0167	.0065
	Week 3	.0125	.0132	-.0130	.0087	-.0096	.0064
	Week 4	-.0264	.0131	-.0140	.0087	-.0013	.0063
	Week 5	.0149	.0113	.0135	.0078	-.0150	.0057
	Flu	.3625	.0551	.3267	.0435	.3310	.0386
	AR	.0390	.0385	.1520	.0385	.1977	.0385

Cause of death		City of Residence					
Ischaemic Heart Disease		Aberdeen		Edinburgh		Glasgow	
		coeff	se	coeff	se	coeff	se
Males	Const	1.9756	.0156	1.9396	.0100	2.1634	.0078
	Trend	-.0005	.0000	-.0005	.0001	-.0002	.0000
	Week 0	-.0041	.0067	-.0157	.0047	-.0103	.0030
	Week 1	-.0187	.0077	-.0159	.0058	-.0019	.0036
	Week 2	.0074	.0079	-.0117	.0058	-.0092	.0036
	Week 3	.0060	.0079	.0114	.0059	-.0050	.0036
	Week 4	-.0036	.0077	-.0013	.0059	-.0015	.0036
	Week 5	-.0094	.0067	.0071	.0047	.0007	.0031
	Flu	-.0438	.1435	.0691	.0603	.0781	.0397
	AR	.0559	.0385	-.0463	.0385	.0741	.0385
Females							
	Const	1.7315	.0165	1.6776	.0104	1.9372	.0086
	Trend	-.0004	.0001	-.0003	.0001	-.0001	.0000
	Week 0	-.0037	.0070	-.0095	.0045	-.0151	.0036
	Week 1	-.0033	.0081	-.0088	.0054	-.0008	.0042
	Week 2	-.0092	.0082	-.0038	.0054	-.0036	.0042
	Week 3	-.0032	.0082	-.0209	.0054	-.0033	.0043
	Week 4	-.0029	.0082	.0138	.0055	.0011	.0042
	Week 5	.0029	.0710	.0026	.0045	-.0044	.0036
	Flu	.0424	.0519	.1157	.0170	.0742	.0268
	AR	.0543	.0385	.0062	.0385	.0408	.0385

Cause of death		City of residence					
Cerebrovascular Disease		Aberdeen		Edinburgh		Glasgow	
		coeff	se	coeff	se	coeff	se
Males	Const	.7692	.0309	.9156	.0198	1.0004	.0133
	Trend	-.0005	.0002	-.0003	.0001	-.0004	.0001
	Week 0	-.0055	.0134	-.0105	.0080	-.0060	.0055
	Week 1	-.0134	.0154	-.0112	.0092	-.0075	.0067
	Week 2	-.0049	.0156	-.0121	.0092	-.0082	.0067
	Week 3	.0032	.0156	-.0097	.0093	-.0070	.0067
	Week 4	-.0050	.0154	-.0115	.0093	-.0022	.0067
	Week 5	.0022	.0134	.0171	.0080	.0038	.0056
	Flu	.3902	.1994	.0736	.1032	.0858	.0627
	AR	.0624	.0385	.0997	.0385	-.0028	.0385
Females							
	Const	1.2191	.0200	1.3713	.0126	1.4738	.0105
	Trend	-.0008	.0001	-.0002	.0001	-.0002	.0001
	Week 0	-.0166	.0090	-.0073	.0056	-.0022	.0045
	Week 1	.0079	.0108	-.0022	.0066	-.0092	.0053
	Week 2	-.0052	.0011	-.0109	.0066	-.0107	.0054
	Week 3	-.0137	.0011	.0038	.0066	.0012	.0053
	Week 4	.0084	.0108	-.0047	.0066	-.0025	.0053
	Week 5	-.0014	.0090	-.0047	.0056	-.0029	.0045
	Flu	.1653	.0491	.0724	.0264	.4260	.0353
	AR	.0004	.0385	.0411	.0385	.0190	.0385

Emergency admissions

Diagnosis		City of residence					
Respiratory Disease		Aberdeen		Edinburgh		Glasgow	
		coeff	se	coeff	se	coeff	se
Males	Const	2.8358	.0133	2.9598	.0114	2.8677	.0100
	Trend	.0009	.0001	.0001	.0001	.0006	.0001
	Week 0	-.0067	.0049	-.0012	.0036	-.0041	.0030
	Week 1	-.0165	.0053	-.0169	.0037	-.0106	.0031
	Week 2	-.0061	.0055	-.0102	.0038	-.0154	.0032
	Week 3	-.0009	.0055	-.0050	.0038	-.0023	.0032
	Week 4	.0055	.0054	-.0027	.0037	-.0014	.0031
	Week 5	.0011	.0049	.0057	.0035	.0040	.0030
	Flu	.1062	.0320	.0666	.0280	.2035	.0382
	AR	.1887	.0385	.3441	.0385	.3344	.0385
Females							
	Const	2.4869	.0131	2.6218	.0107	2.5840	.0099
	Trend	.0013	.0001	.0004	.0001	.0011	.0001
	Week 0	-.0120	.0055	-.0076	.0038	-.0099	.0032
	Week 1	-.0103	.0062	-.0136	.0041	-.0120	.0035
	Week 2	-.0022	.0064	-.0125	.0042	-.0115	.0035
	Week 3	.0001	.0063	-.0071	.0042	-.0057	.0035
	Week 4	-.0043	.0062	-.0038	.0041	-.0008	.0034
	Week 5	-.0013	.0055	.0071	.0038	.0002	.0032
	Flu	.1873	.0390	.1805	.0314	.2443	.0397
	AR	.0924	.0385	.2349	.0385	.2663	.0385

Diagnosis		City of residence					
Ischaemic Heart Disease		Aberdeen		Edinburgh		Glasgow	
		coeff	se	coeff	se	coeff	se
Males	Const	2.4074	.0129	2.4610	.0090	2.4332	.0099
	Trend	.0000	.0001	.0001	.0000	.0008	.0000
	Week 0	.0014	.0056	-.0054	.0038	.0065	.0032
	Week 1	-.0137	.0062	-.0072	.0045	-.0096	.0035
	Week 2	.0075	.0064	-.0017	.0045	-.0012	.0035
	Week 3	-.0024	.0065	.0068	.0045	-.0031	.0035
	Week 4	.0068	.0063	-.0005	.0045	.0031	.0035
	Week 5	-.0025	.0056	-.0002	.0038	-.0049	.0032
	Flu	.0056	.0424	-.0799	.0313	-.0201	.0329
	AR	.0744	.0385	.0798	.0385	.2329	.0385
Females							
	Const	2.0050	.0143	1.9823	.0104	2.0248	.0089
	Trend	.0003	.0001	.0002	.0001	.0011	.0000
	Week 0	.0097	.0066	-.0035	.0043	.0092	.0033
	Week 1	-.0067	.0077	-.0115	.0049	-.0226	.0037
	Week 2	-.0042	.0078	-.0014	.0050	-.0004	.0037
	Week 3	-.0043	.0077	-.0001	.0050	-.0001	.0037
	Week 4	-.0020	.0077	.0082	.0050	.0034	.0037
	Week 5	.0001	.0066	-.0044	.0043	.0000	.0033
	Flu	.0004	.0458	-.0323	.0326	-.0679	.0345
	AR	.0013	.0385	.1035	.0385	.1394	.0385

Diagnosis		City of residence					
Cerebrovascular Disease		Aberdeen		Edinburgh		Glasgow	
		coeff	se	coeff	se	coeff	se
Males	Const	1.4204	.2100	1.4512	.0137	1.4457	.0113
	Trend	.0009	.0001	.0001	.0001	.0007	.0001
	Week 0	-.0050	.0091	.0095	.0062	-.0095	.0046
	Week 1	-.0011	.0106	-.0123	.0074	-.0057	.0054
	Week 2	-.0137	.0107	.0003	.0074	.0009	.0054
	Week 3	.0160	.0107	-.0053	.0074	-.0020	.0054
	Week 4	-.0202	.0106	-.0050	.0074	.0094	.0054
	Week 5	.0090	.0093	.0004	.0062	-.0018	.0046
	Flu	-.1600	.0872	.0269	.0484	.0761	.0469
	AR	.0156	.0385	.0200	.0385	.0688	.0385
Females							
	Const	1.4386	.0182	1.4989	.0127	1.5880	.0108
	Trend	.0008	.0001	-.0002	.0001	.0008	.0001
	Week 0	.0002	.0088	-.0123	.0056	-.0053	.0041
	Week 1	-.0128	.0107	-.0030	.0067	-.0040	.0047
	Week 2	-.0105	.0107	-.0008	.0068	-.0054	.0048
	Week 3	.0099	.0106	.0083	.0067	-.0009	.0047
	Week 4	.0013	.0106	.0016	.0067	-.0009	.0047
	Week 5	-.0002	.0088	-.0053	.0056	.0005	.0042
	Flu	-.0533	.0613	.0314	.0395	.1115	.0382
	AR	-.0481	.0385	.0311	.0385	.1163	.0385

Section V

Cause of death		Temperature Category									
All causes		<2°C		2-6°C		6-10°C		10-14°C		14°C+	
		coeff	se	coeff	se	coeff	se	coeff	se	coeff	se
Males	Const	6.5300	.0095	6.4436	.0049	6.3648	.0040	6.2813	.0033	6.2469	.0058
	Trend	-.0002	.0001	-.0001	.0000	-.0001	.0000	-.0001	.0000	.0000	.0000
	Week 0	-.0111	.0042	-.0102	.0036	-.0071	.0034	-.0014	.0038	.0016	.0047
	Week 1	-.0145	.0033	-.0077	.0017	-.0050	.0019	-.0092	.0028	-.0063	.0030
	Week 2	.0072	.0032	-.0039	.0019	-.0020	.0019	-.0016	.0025	-.0006	.0032
	Week 3	-.0070	.0036	.0000	.0018	-.0048	.0020	.0037	.0025	-.0062	.0029
	Week 4	-.0048	.0039	-.0038	.0018	-.0032	.0019	-.0078	.0025	.0031	.0029
	Week 5	.0026	.0046	.0015	.0016	.0020	.0016	.0032	.0021	-.0011	.0025
	Flu	.0103	.0004	.0154	.0020	.0109	.0024	.0079	.0141	.0007	.0025
	AR	.1284	.0385	.2205	.0385	.1291	.0385	-.0708	.0385	.2115	.0385
Females											
	Const	6.6159	.0149	6.5198	.0056	6.4320	.0040	6.3344	.0032	6.2958	.0049
	Trend	-.0002	.0001	.0000	.0000	.0001	.0000	.0001	.0000	.0000	.0000
	Week 0	-.0141	.0060	-.0067	.0038	-.0109	.0035	-.0066	.0032	.0160	.0048
	Week 1	-.0067	.0032	-.0080	.0019	-.0078	.0020	-.0084	.0023	-.0127	.0033
	Week 2	-.0044	.0040	-.0045	.0020	-.0015	.0019	-.0017	.0022	-.0039	.0034
	Week 3	.0045	.0039	-.0039	.0020	-.0065	.0019	-.0016	.0021	.0005	.0031
	Week 4	-.0107	.0047	-.0018	.0018	.0033	.0019	.0007	.0022	.0007	.0030
	Week 5	.0123	.0040	-.0003	.0017	-.0018	.0016	-.0021	.0018	-.0013	.0026
	Flu	.0056	.0005	.0122	.0013	.0152	.0015	-.0260	.0105	.0217	.0133
	AR	.4557	.0385	.2310	.0385	.1467	.0385	.0500	.0385	.0643	.0385

Cause of death		Temperature category					
All causes		Increasing		Decreasing		Winter	
		coeff	se	coeff	se	coeff	se
Males	Const	6.3628	.0071	6.3635	.0070	6.3625	.0070
	Trend	.0002	.0000	.0002	.0000	.0002	.0000
	Week 0	-.0027	.0015	.0004	.0015	-.0036	.0024
	Week 1	-.0016	.0015	.0011	.0015	-.0015	.0024
	Week 2	.0021	.0015	.0001	.0015	.0003	.0023
	Week 3	.0004	.0015	-.0009	.0016	-.0013	.0024
	Week 4	.0019	.0015	-.0023	.0016	-.0019	.0024
	Week 5	-.0021	.0015	.0009	.0015	-.0004	.0023
	Flu	.0115	.0015	.0107	.0025	.0108	.0021
	AR	.6152	.0385	.6129	.0385	.6124	.0385
Females							
	Const	6.4246	.0074	6.4249	.0075	6.4248	.0075
	Trend	.0002	.0000	.0002	.0000	.0002	.0000
	Week 0	-.0040	.0015	.0016	.0015	-.0017	.0023
	Week 1	.0001	.0015	.0002	.0015	.0006	.0024
	Week 2	.0012	.0015	.0002	.0015	-.0031	.0024
	Week 3	.0017	.0015	-.0013	.0015	.0027	.0023
	Week 4	-.0004	.0015	-.0017	.0016	-.0018	.0023
	Week 5	.0023	.0015	-.0020	.0015	.0021	.0023
	Flu	.0149	.0016	.0148	.0016	.0149	.0016
	AR	.6410	.0385	.6415	.0385	.6426	.0385

REFERENCES

1. Quetelet MA. *A treatise on man and the development of his faculties*. Edinburgh: William & Robert Chambers, 1842;34-39

2. Langmuir AD. William Farr: Founder of modern concepts of surveillance. *International Journal of Epidemiology* 1976;**5**:13-18.

3. GRO(S). Quarterly variation in mortality in Scotland and 12 selected countries, since 1977. 1987. GRO(S). Vital Events Report.

4. Curwen M. Excess winter mortality: a British phenomenon? *Health Trends* 1991;**22**:169-175.

5. Curwen M, Devis T. Winter mortality, temperature and influenza: has the relationship changed in recent years? *Population Trends* 1988;**54**:17-20.

6. McDowell M. Long term trends in seasonal mortality. *Population Trends* 1981;**26**:16-19.

7. Douglas AS, Allan TM, Rawles JM. Composition of seasonality of disease. *Scottish Medical Journal* 1991;**36**:76-82.

8. Curwen M, Dunnell K, Ashley J. Hidden influenza deaths:1989-1990. *Population Trends* 1990;**61**:31-33.

9. Fleming DM, Cross KW, Crombie DL, Lancashire RJ. Respiratory illness and mortality in England and Wales. *European Journal of Epidemiology* 1993;**9**:571-575.

10. Nicholson KG. Impact of influenza and respiratory syncytial virus on mortality in England and Wales from January 1975 to December 1990. *Epidemiol.Infect.* 1996;**116**:51-63.

11. Khot A, Evans N, Lenney W. Seasonal trends in childhood asthma in south east England. *British Medical Journal* 1983;**287**:1257-1258.

12. Bates DV. Air pollution: time for more clean air legislation? - Ozone and fine exhaust particles make asthma worse. *British Medical Journal* 1996;**312**:649-650.

13. Dunnigan MG, Harland WA, Fyfe T. Seasonal incidence and mortality of ischaemic heart disease. *Lancet* 1970;**2**:793-797.

14. Anderson TW, Harding Le Riche W. Seasonal variation in ischaemic heart disease mortality. *The Lancet* 1970;**i**:1140-1140.

15. Douglas AS, Dunnigan MG, Allan TM, Rawles JM. Seasonal variation in coronary heart disease in Scotland. *Journal of Epidemiology and Community Health* 1995;**49**:575-582.

16. Wilmshurst P. Temperature and cardiovascular mortality. *British Medical Journal* 1994;**309** :1029-1030.

17. Cooke EA, McNally MA, Mollan RAB. Seasonal variations in fatal pulmonary embolism. *British Medical Journal* 1995;**310**:129-129.

18. Lloyd OL. Temperature and cardiovascular mortality. *British Medical Journal* 1995;**310**:467-467.

19. Lloyd EL. Temperature and cardiovascular mortality. *British Medical Journal* 1995;**310**:467-467.

20. Rothwell PM, Wroe SJ, Slattery J, Warlow CP. Is stroke incidence related to season or temperature. *Lancet* 1996;**347**:934-936.

21. Haberman S, Capildeo R, Clifford R. The seasonal variation in mortality from cerebrovascular disease. *Journal of the Neurological Sciences* 1981;**52**:25-36.

22. Bull GM. Meteorological correlates with myocardial and cerebral infarction and respiratory disease. *British Journal of Preventive and Social Medicine* 1973;**27**:108-113.

23. West RR, Lloyd S, Roberts CJ. Mortality from ischaemic heart disease: association with weather. *British Journal of Preventive and Social Medicine* 1973;**27**:36-40.

24. Bainton D, Moore F, Sweetman P. Temperature and deaths from ischaemic heart disease. *British Journal of Preventive and Social Medicine* 1977;**31**:49-53.

25. Mackenbach JP, Kunst AE, Looman CWN. Seasonal variation in mortality in The Netherlands. *Journal of Epidemiology and Community Health* 1992;**46**:261-265.

26. Kunst AE, Looman CWN, Mackenbach JP. The decline in winter excess mortality in The Netherlands. *International Journal of Epidemiology* 1990;**20**:971-977.

27. Lerchl A. Changes in the seasonality of mortality in Germany from 1946 to 1995: the role of temperature. *International Journal of Biometerology* 1998;**42**:84-88.

28. Laake K, Sverre JM. Winter excess mortality: a comparison between Norway and England plus Wales. *Age and Ageing* 1996;**25**:343-348.

29. McKee M, Sanderson C, Chenet L, Vassin S, Shkolnikov V. Seasonal variation in mortality in Moscow. *Journal of Public Health Medicine* 1998;**20**:268-274.

30. Eurowinter Group. Cold exposure and winter mortality from ischaemic heart disease, cerebrovascular disease, respiratory disease, and all causes in warm and cold regions of Europe. *Lancet* 1997;**349**:1341-1346.

31. Saez M, Sunyer J, Castellsague J, Murillo C, Anto JM. Relationship between weather temperature and mortality: A time series analysis approach in Barcelona. *International Journal of Epidemiology* 1995;**24**:576-582.

32. Dales RE, Schweitzer I, Toogood JH, et al. Respiratory infections and the autumn increase in asthma morbidity. *European Respiratory Journal* 1996;**9**:72-77.

33. Sprenger MJW, Mulder PGH, Beyer WEP, Van Strik R, Masurel N. Impact of influenza on mortality in relation to age and underlying disease, 1967-1989. *International Journal of Epidemiology* 1993;**22**:334-340.

34. Anderson TW, Le Riche WH. Cold weather and myocardial infarction. *Lancet* 1970;**i**:291-296.

35. Douglas AS, Russell D, Allan TM. Seasonal, regional and secular variations of cardiovascular and cerebrovascular mortality in New Zealand. *Aust NZ J Med* 1990;**20**:669-676.

36. Marshall RJ, Scragg R, Bourke P. An analysis of the seasonal variation of coronary heart disease and respiratory disease mortality in New Zealand. *International Journal of Epidemiology* 1988;**17**:325-331.

37. Knoler R, Poole WK, Perritt RL. When throughout the year is coronary death most likely to occur? A 12-year population based analysis of more than 220 000 cases. *Circulation* 1999;**100**:1630-1634.

38. Seretakis D, Lagiou P, Lipworth L, Signorello L, Rothman K, Trichopoulos D. Changing seasonality of mortality from coronary heart disease. *Journal of the American Medical Association* 1997;**278**:1012-1014.

39. Green MS, Harari G, Kristal-Boneh E. Excess winter mortality from ischaemic heart disease and stroke during colder and warmer years in Israel - An evaluation and review of the role of environmental temperature. *European Journal of Public Health* 1994;**4**:3-11.

40. Sheth T, Nair C, Muller J, Yusuf S. Increased winter mortality from acute myocardial infarction and stroke: the effect of age. *Journal of the American College of Cardiology* 1999;**33**:1916-1919.

41. Wen-Harn Pan, Lung-An Li, Ming-Jan Tsai. Temperature extremes and mortality from coronary heart disease and cerebral infarction in elderly Chinese. *Lancet* 1995;**345**:353-355.

42. Mannino JA, Washburn RA. Environmental temperature and mortality from acute myocardial infarction. *International Journal of Biometerology* 1989;**33**:32-35.

43. Baker-Blocker A. Winter weather and cardiovascular mortality in Minneapolis-St. Paul. *American Journal of Public Health* 1982;**72**:

44. Rogot E, Padgett SJ. Associations of coronary and stroke mortality with temperature and snowfall in selected areas of the United States, 1962-66. *American Journal of Epidemiology* 1976;**103**:565-575.

45. Anderson TW, Rochard C. Cold snaps, snowfall and sudden death from ischaemic heart disease. *CMA Journal* 1979;**121**:1580-1583.

46. Heunis JC, Olivier J, Bourne DE. Short-term relationships between winter temperatures and cardiac disease mortality in Cape Town. *SAMJ* 1985;**85**:1016-1019.

47. Ohlson C-G, Bodin L, Bryngelsson I-L, Helsing M, Malmberg L. Winter weather conditions and myocardial infarctions. *Scandinavian Journal of Social Medicine* 1996;**19**:20-25.

48. Enquselassie F, Dobson AJ, Alexander HM, Steele PL. Seasons, temperature and coronary disease. *International Journal of Epidemiology* 1993;**22**:632-636.

49. Kunst AE, Looman CWN, Mackenbach JP. Outdoor air temperature and mortality in The Netherlands: A time series analysis. *American Journal of Epidemiology* 1993;**137**:331-341.

50. Lanska DJ, Hoffmann RG. Seasonal variation in stroke mortality rates. *Neurology* 1999;**52**:984-990.

51. Catalano R, Serxner S. Time series designs of potential interest to epidemiologists. *American Journal of Epidemiology* 1996;

52. Giles G. Biometeorological investigations of asthma morbidity in Tasmania using co-spectral analysis of time series. *Soc.Sci.Med* 1981;**15D**:111-119.

53. Hoppenbrouwers T, Calub M, Arakawa K, Hodgman JE. Seasonal relationship of sudden infant death syndrome and environmental pollutants. *American Journal of Epidemiology* 1981;**113**:623-635.

54. Bowie C, Prothero D. Finding causes of seasonal disease using time series analysis. *International Journal of Epidemiology* 1981;**10**:87-92.

55. Alberdi JC, Diaz J, Montero JC, Miron I. Daily mortality in Madrid community 1986-1992: Relationship with meteorological variables. *European Journal of Epidemiology* 1998;**14**:571-578.

56. Albert PS, McFarland HF, Smith ME, Frank JA. Repeated measures time series for modelling counts from a relapsing-remitting disease: Application to modelling disease activity in multiple sclerosis. *Statistics in medicine* 1994;**13**:453-466.

57. Kuhn L, Davidson LL, Durkin MS. Use of Poisson regression and time series analysis for detecting changes over time in rates of child injury following a prevention program. *American Journal of Epidemiology* 1994;**140**:943-955.

58. Campbell MJ. Time series regression for counts: an investigation into the relationship between sudden infant death syndrome and environmental temperatures. *Journal of the Royal Statistical Society A* 1994;**157**:191-208.

59. Zeger SL. A Regression model for time series of counts. *Biometrika* 1988;**75**:621-629.

60. Brennan PJ, Greenberg G, Miall WE, Thompson SG. Seasonal variation in arterial blood pressure. *British Medical Journal* 1982;**285**:919-923.

61. Dobson HM, Muir MM, Hume R. The effect of ascorbic acid on the seasonal variations in serum cholesterol levels. *Scottish Medical Journal* 1984;**29**:176-182.

62. MacRury SM, Muir M, Hume R. Seasonal and climatic variation in cholesterol and vitamin C:effect of vitamin C supplementation. *Scottish Medical Journal* 1992;**37**:49-52.

63. Stout RW, Crawford V. Seasonal variations in fibrinogen concentrations among elderly people. *Lancet* 1991;**338**:9-13.

64. Woodhouse PR, Tee Khaw K, Plummer M, Foley A, Meade TW. Seasonal variations of plasma firbrinogen and factor VII activity in the elderly: winter infections and death from cardiovascular disease. *Lancet* 1994;**343**:435-439.

65. Meade TW, Brozovic M, Chakrabarti RR, et al. Haemostatic function and ischaemic heart disease: Principal results of the Northwick Park Heart Study. *Lancet* 1986;**September 6**:533-537.

66. Keatinge WR, Coleshaw SRK, Cotter F, Mattock M, Murphy M, Chelliah R. Increases in platelet and red cell counts, blood viscosity, and arterial pressure during mild surface cooling: factors in mortality from coronary and cerebral thrombosis in winter. *British Medical Journal* 1984;**289**:1405-1408.

67. Donaldson GC, Robinson D, Allaway SL. An analysis of arterial disease mortality and BUPA health screening data in men, in relation to outdoor temperature. *Clinical Science* 1997;**92**:268

68. Williams DA. Extra-binomial variation in logistic linear models. *Applied Statistics* 1982;**31**:144-148.

69. Monto AS, Sook KL. The Tecumseh study of respiratory illness. III Incidence and periodicity of respiratory syncytial virus and mycoplasma pneumoniae infections. *American Journal of Epidemiology* 1971;**94**:290-301.

70. World Health Organisation. *International classification of diseases*. Geneva: WHO, 1977;

71. Carstairs V, Morris R. *Deprivation and health in Scotland*. Aberdeen: Aberdeen University Press, 1992;

72. McLoone P, Boddy FA. Deprivation and mortality in Scotland:1981 and 1991. *British Medical Journal* 1994;**309**:1465-1470.

73. Halberg F, Tong YL, Johnson EA. Circadian system phase: An aspect of temporal morphology; procedures and illustrative examples. In: H.Von Mayersbach, ed. *The Cellular aspects of biorhythms*. New York: Springer-Verlag, 1967;20-48.

74. Lentz MJ. Time series analysis-cosinor analysis:A special case. *Western Journal of Nursing research* 1990;**12**:408-412.

75. Teicher MH, Barber NI. COSIFIT: An interactive program of simultaneous multioscillator cosinor analysis of time-series data. *Computers and Biomedical Research* 1990;**23**:283-295.

76. Faure A, Nemoz C, Claustrat B. A graphical and statistical method for investigation of time series in chronobiology according to the cosinor procedure. *Comput.Biol.Med.* 1990;**20**:319-329.

77. Bloomfield P. *Fourier analysis of time series: an introduction*. New York: Wiley, 1946;

78. Shah S, Peacock J. Deprivation and excess winter mortality. *Journal of Epidemiology and Community Health* 1999;**53**:499-502.

79. Diggle PJ. *Time series: A biostatistical introduction*. Oxford: Oxford University press, 1990;

80. Cox DR, Lewis PAW. *The statistical analysis of series of events*. London: Chapman & Hall, 1965;

81. Chatfield C. *The analysis of time series: An introduction*. London: Chapman & Hall, 1975;

82. Box GEP, Jenkins GM. *Times series analysis forecasting and control*. San Francisco: Holden-Day, 1970;

83. Kendall. *Time series*. London: Griffin, 1976;

84. Durbin J. The fitting of time series models. *International Statistical Review* 1960;**28**:233-244.

85. Cox DR. Some remarks on overdispersion. *Biometrika* 1983;**70**:269-274.

86. Breslow NE. Extra-poisson variation in log-linear models. *Applied Statistics* 1984;**33**:38-44.

87. McCullagh P, Nelder JA. *Generalised linear models*. London: Chapman & Hall, 1983;

88. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986;**73**:13-22.

89. Kelsall JE, Zeger SL, Samet JM. Frequency domain log-linear models; air pollution and mortality. *Journal of the Royal Statistical Society Series C-Applied Statistics* 1999;**48**:331-334.

90. Donaldson GC, Keatinge WR. Mortality related to cold weather in elderly people in southeast England, 1979-94. *British Medical Journal* 1997;**315**:1055-1056.

91. Mackenbach JP, Looman CWN, Kunst AE. Air pollution, lagged effects of temperature, and mortality: The Netherlands 1979-87. *Journal of Epidemiology and Community Health* 1993;**47**:121-126.

92. MacFarlane A. Daily mortality and environment in English conurbations 1: air pollution, low temperature, and influenza in Greater London. *British Journal of Preventive and Social Medicine* 1977;**31**:54-61.

93. Schwartz J, Marcus A. Mortality and air pollution in London: A time series analysis. *American Journal of Epidemiology* 1990;**131**:185-194.

94. Bremner SA, Anderson HR, Atkinson RW, et al. Short term associations between outdoor pollution and mortality in London 1992-1994. *Occupational and Environmental Medicine* 1999;**56**:237-244.

95. Buchdahl R, Parker A, Stebbings T, Babiker A. Association between air pollution and acute childhood wheezy episodes: prospective observational study. *British Medical Journal* 1996; **312**:661-665.

96. Ross Anderson H, Ponce de Leon A, Bland JM, Bower JS, Strachan DP. Air pollution and daily mortality in London: 1987-1992. *British Medical Journal* 1996;**312**:665-669.

97. Livingstone AE, Shaddick G, Grundy C, Elliot P. Do people living near inner city main roads have more asthma needing treatment? Case-control study. *British Medical Journal* 1996;**312**:676-677.

98. Mackay TW, Wathen CG, Sudlow MF, Elton RA, Caulton E. Factors affecting asthma mortality in Scotland. *Scottish Medical Journal* 1992;**37**:5-7.

99. Atkinson RW, Anderson HR, Strachan DP, Bland JM, Bremner SA, Ponce de Leon A. Short-term associations between outdoor air pollution and visits to accident and emergency departments in London for respiratory complaints. *European Respiratory Journal* 1999;**13**:257-265.

100. Schwartz J. PM10, ozone, and hospital admissions for the elderly in Minneapolis-St. Paul, Minnesota. *Archives of Environmental Health* 1994;**49**:366-374.

101. Schwartz J. Air pollution and hospital admissions for the elderly in Birmingham, Alabama. *American Journal of Epidemiology* 1994;**139**:589-598.

102. Schwartz J. Air pollution and hospital admissions for cardiovascular disease in Tucson. *Epidemiology* 1997;**8**:371-377.

103. Schwartz J. Short term fluctuations in air pollution and hospital admissions of the elderly for respiratory disease. *Thorax* 1995;**50**:531-538.

104. Moolgavkar SH, Luebeck EG, Anderson EL. Air pollution and hospital admissions for respiratory causes in Minneapolis-St. Paul and Birmingham. *Epidemiology* 1997;**8**:364-370.

105. Moolgavkar SH, Luebeck EG, Hall TA, Anderson EL. Air pollution and daily mortality in Philadelphia. *Epidemiology* 1995;**6**:476-484.

106. Morris RD, Naumova EN, Munasinghe RL. Ambient air pollution and hospitalization for congestive heart failure among elderly people in seven large US cities. *American Journal of Public Health* 1995;**85**:1361-1365.

107. Verhoef AP, Hoek G, Schwartz J, van Wijnen JH. Air pollution and daily mortality in Amsterdam. *Epidemiology* 1996;**7**:225-230.

108. Ponka A, Virtanen M. Low-level air pollution and hospital admissions for cardiac and cerebrovascular diseases in Helsinki. *American Journal of Public Health* 1996;**86**:1273-1280.

109. Touloumi G, Pocock SJ, Katsouyanni K, Trichopoulos D. Short-term effects of air pollution on daily mortalty in Athens: A time-series analysis. *International Journal of Epidemiology* 1994;**23**:957-967.

110. Dominici F, Samet JM, Zeger SL. Combining evidence on air pollution and daily mortality from the 20 largest US cities: a hierarchical modelling strategy. *Journal of the Royal Statistical Society A* 2000;**163**:

111. Knox G. Detection of low intensity epidemicity: Application to cleft lip and palate. *British Journal of Preventive and Social Medicine* 1963;**17**:121-127.

112. Knox G. Epidemiology of childhood leukaemia in Northumberland and Durham. *British Journal of Preventive and Social Medicine* 1964;**18**:17-24.

113. Petridou E, Revinthi K, Alexander FE, Haidas S, Koliouskas D. Space-time clustering of childhood leukaemia in Greece: evidence supporting a viral aetiology. *British Journal of Cancer* 1995;**73**:1278-1283.

114. Knox G, Gilman E. Leukaemia clusters in Great Britain. 1. Space-time interactions. *Journal of Epidemiology and Community Health* 1992;**46**:566-572.

115. Bhopal RS, Diggle P, Rowlingson B. Pinpointing clusters of apparently sporadic cases of legionnaires disease. *British Medical Journal* 1992;**304**:1022-1027.

116. Diggle PJ. Space-time interaction. In: Barnett V, Turkman KF, eds. *Statistics for the environment*. Chichester: Wiley, 1993;89-110.

117. Silman, A., Bankhead, C, Rowlingson, B., Brennan, P., Symmons, D, and Gatrell, A. Do new cases of rheumatoid arthritis cluster in time or in space. International Journal of Epidemiology 26, 628-634. 1997.

118. Cliff AD, Haggett P, Ord JK, Bassett KA, Davis RB. *Elements of spatial structure*. Cambridge: Cambridge University Press, 1975;

119. Grenfell BT, Kleczkowski A, Gilligan CA, Bolker BM. Spatial heterogeneity, nonlinear dynamics and chaos in infectious diseases. *Statistical Methods in Medical Research* 1995;**4**:160-183.

120. Anderson RM, Grenfell BT. Quantitative investigations of different vaccination policies for the control of congenital rubella syndrome (CRS) in the U.K. *Journal of Hygiene (Cambridge)* 1986;**96**:305-333.

121. Dietz K, Schenzle D. Mathematical models for infectious disease statistics. In: Atkinson AC, Feinberg SE, eds. *A celebration of statistics*. New York: Springer, 1985;167-204.

122. Anderson RM, May R.M. *Infectious diseases of humans: dynamics and control.* 1991;

123. Soper MA. The interpretation of periodicity in disease prevalence. *Journal of the Royal Statistical Society A* 1929;**92**:34-61.

124. Cressie NAC. *Statistics for spatial data*. New York: John Wiley & Sons Ltd, 1991;

125. Cressie NAC, Chan NH. Spatial modelling of regional variables. *Journal of the American Statistical Association* 1989;**84**:393-401.

126. Clayton D, Bernardinelli L, Montomoli C. Spatial correlation in ecological analysis. *International Journal of Epidemiology* 1998;**22**:1193-1202.

127. Kemp I, Boyle P. *Atlas of cancer in Scotland 1975-1980: Incidence and epidemiological perspective.* IARC Scientific Publications No 72, 1985;

128. WHO. *Atlas of mortality in europe; subnational patterns, 1980/81 and 1990/91.* WHO Regional Publications European Series No 75, 1997;

129. Clayton D, Kaldor J. Empirical bayes estimates of age-standardised relative risks for use in disease mapping. *Biometrics* 1987;**43**:671-681.

130. Cislaghi C, Biggeri A, Braga M, Lagazio C, Mattock M. Exploratory tools for disease mapping in geographical epidemiology. *Statistics in medicine* 1995;**14**:2363-2381.

131. Bernardinelli L, Clayton D, Montomoli C. Bayesian estimates of disease maps: How important are priors? *Statistics in medicine* 1995;**14**:2411-2431.

132. Muller H, Stadtmuller U, Tabnak F. Spatial smoothing of geographically aggregated data, with application to the construction of incidence maps. *Journal of the American Statistical Association* 1997;**92**:61-71.

133. Cliff AD, Haggett P. Mapping respiratory diseases. In: Scadding J.G., Cumming G, eds. *Scientific Foundations of Respiratory Medicine*. 1981;30-43.

134. Cressie NAC. When are relative variograms useful in geostatistics? *Journal of the International Association for Mathematical Geolog* 1985;**17**:693-702.

135. Moran PAP. Notes on continuous stochastic phenomena. *Biometrika* 1950;**37**:17-23.

136. Geary RC. The contiguity ratio and statistical mapping. *The Incorporated Statistician* 1954;**5**:115-145.

137. Cliff AD, Ord JK. *Spatial autocorrelation*. London: Pion, 1975;

138. Mathsoft. *S-PLUS spatial stats user's manual*. Washington: Mathsoft, 1996;

139. Odland J. *Spatial Autocorrelation*. California: SAGE, 1988;

140. Gatrell AC. Autocorrelation in Spaces. *Environment and Planning* 1979;**11**:507-516.

141. Griffith DA. Some guidelines for specifying the geographic weights matrix contained in spatial statistical models. In: Arlinghaus SL, ed. *Practical Handbook of Spatial Statistics*. CRC Press, 1995;65-82.

142. Tukey J. Some further inputs. In: Merriam D, ed. *Geostatistics*. New York: Plenum Press, 1970;163-174.

143. Whittle P. On stationary processes in the plane. *Biometrika* 1954;**41**:434-439.

144. Cliff AD, Ord JK. *Spatial processes models and applications*. London: Pion, 1981;

145. Haining Robert. *Spatial data analysis in the social and environmental sciences*. Cambridge: Cambridge University Press, 1990;

146. Blatchford O, Capewell S, Blatchford M. Are hospitals rather than GPs the 'gatekeepers' for emergency medical admissions? *Journal of Epidemiology and Community Health* 1999;**53**:4-4.(Abstract)

147. McShane LM, Albert PS, Palmatier MA. A latent process regression model for spatially correlation count data. *Biometrics* 1997;**53**:698-706.

148. Inskip H, Beral V, Fraser P, Haskey J. Epidemiological monitoring: Methods for analysing routinely-collected data. *The Science of the Total Environment* 1984;**32**:219-232.

149. Kohli HS, Knill-Jones RP. How accurate are SMR1 (Scottish Morbidity Record 1) data? *Health Bulletin* 1992;**50**:14-23.

150. Harley K, Jones C. Quality of scottish morbidity record (SMR) data. *Health Bulletin* 1996;**54**:410-417.