

16 Discussion

This section interprets the findings and examines their relevance in relation to the aims of the project (as presented in the Introduction, Section 1). The limitations of the routine data used and the methods are discussed. Epidemiological findings for each disease are examined and compared with other work. The role of confounding by social class and smoking and possible approaches to investigating this are detailed. The concluding section summarises the scope of routine data to examine environmental influences on health.

Scope and quality of data

16.1 Surveys – Health Survey for England 1995

16.2 Primary care data

GPRD as a valid source of information on primary care

Commercially released data on CD-ROM

PACT data

16.3 Routine data

Hospital admission data

Mortality

Discussion of methods

16.4 Limitations of methods used

Geographical identifiers

Boundary changes to administrative regions and districts

Postcoded data

Dealing with boundary changes in the absence of postcoded data

Urban rural classifications

ONS group

Geographical correlations

Patients and events

Epidemiological findings

16.5 General results

16.6 Asthma

16.7 COPD

16.8 Hayfever

16.9 Pneumonia

16.10 Acute bronchitis/bronchiolitis

16.11 Tuberculosis

16.12 Sarcoidosis

16.13 Idiopathic fibrosing alveolitis

16.14 Cystic fibrosis

16.15 Pneumothorax

Confounding

16.16 Confounding by smoking and social class

Conclusions

16.17 Scope of routine data to explore environmental influences on disease

16.1 Surveys

A number of surveys are available with information on symptoms of respiratory disease, but the Health Survey for England was chosen because it is representative, nationwide, could be analysed at regional and district health authority level and anonymised, individual level data was publicly available from the Data Archive situated at Essex university.

Health Survey for England 1995

Health Survey for England data was generally of a very high quality in terms of internal consistency and apparent completeness.

The Health Survey for England was used for information on symptoms and on self-reported illness, but three variables were used to indicate asthma prevalence as there was no 'best match'. Symptoms of wheezing in the past year were sensitive but not specific, including patients with other illnesses such as COPD, particularly in later life. Inhaler use partially reflected severity, but again was not specific to asthma. Self-reported asthma as a long-standing illness appeared to be the most specific, but was not particularly sensitive – only 27% of those with wheezing or whistling in the chest self-reported asthma. This is consistent with a literature review of questionnaires[1] which suggested that self-reported asthma from questionnaires had a mean sensitivity of 68% (range 48% to 100%) and mean specificity of 94% (range 78% to 100%) when validated in relation to a clinical diagnosis of asthma.

16.2 Primary care data

The main sources of data on primary care consultations are from the Morbidity Surveys in General Practice (MSGPs) which take place approximately every 10 years, continuous data supplied by computerised GP practices such as the GPRD, the Weekly Returns Service (WRS) which provides information on selected conditions from volunteer 'spotter' practices and from PACT data which contains national data on all prescriptions redeemed.

One of the main limitations of using the GPRD, MSGPs or WRS data is that they involve volunteer practices who may not be representative of all GP practices, particularly single-handed urban practices, and may include a higher proportion of more innovative practices.

GPRD as a valid source of information on primary care

Very little had been published using the GPRD for epidemiological studies of respiratory disease at the start of this project. We found that the GPRD was a valid source of primary care data for respiratory disease, by comparison with the 4th Morbidity Survey in General Practice (MSGP4)[2] and also, for seasonal patterns, with the Weekly Returns Service (WRS) data for asthma (see section 16.6). The GPRD had advantages over the MSGP4 because it was larger, run over a longer time period and prescribing data was also available by diagnosis and by age.

The GPRD holds data on prescriptions issued rather than prescriptions redeemed as held in routine data sources such as PACT. A study published in 1993 suggested that 5.1% of all drugs and 4.6% of respiratory drug prescriptions were not redeemed.[3]

Information recorded in the GPRD may reflect clinical practice closely through use of the clinically orientated OXMIS coding system and because there is no requirement to record a definitive diagnosis (for example, permitting diagnoses such as wheezing, cough or chest infection). A number of epidemiological analyses using the GPRD have now been published.[4-6] However, the number of practices included in the GPRD has been falling from its maximum in 1995, because practices which have started using new software (ViSion) cannot currently be included in analyses.[5] This may limit the generalisability of findings from more recent years.

GPRD data needed careful interpretation. There was some miscoding in cystic fibrosis, which had been confused with fibrocystic disease of the breast, which has now been corrected. Common but imprecise diagnoses such as chest infection needed investigation. We found that chest infection was best interpreted as acute bronchitis or bronchiolitis, but it could potentially have been attributed to a number of more precise diagnoses. Prescriptions plus relevant diagnosis appeared more sensitive at detecting contacts with primary care and therefore indicating the burden of morbidity than consultations alone. However, this measure was not appropriate for investigation of seasonal variations of morbidity because of lack of seasonal variation (inhalers) or apparent 'stocking up' (inhalers prior to Christmas and New Year, hayfever

medication prior to the pollen season). Seasonal variations were better investigated using patient consultations or non-repeat prescribing.

Commercially released data on CD-ROM – GPRD

CD-ROMs containing selected tables of GPRD data are available from EPIC, a commercial company. We explored the use of the CD-ROM covering 1987 to 1995, which allows searches between 2/1/92 and 31/12/95. A CD-ROM covering more recent years has also been issued, but the coverage of the GPRD was less good and less likely to be representative after 1995. We could not have performed the comparison with MSGP4[2] using the EPIC CD-ROM due to differences in definition. Also not all the respiratory conditions we studied were included (Table 16.1).

Table 16.1 Correspondence between conditions considered in this study and those included on the EPIC CD-ROM

Collation and comparison project	EPIC	
Condition	Possible corresponding condition(s) in EPIC	Defined in EPIC as long-term “L” or short-term “S”. (time period over which any mention counts as one episode if short-term).
Asthma	Asthma	L
COPD	(i) Chronic obstructive airways disease (ii) Infection – lower respiratory tract chronic	L L
Hayfever	None	N/A
Pneumonia	Infection – pneumonia	S (31 days)
Acute bronchitis or bronchiolitis	Infection – lower respiratory tract acute	S (14 days)
Tuberculosis	Infection – tuberculosis	L
Sarcoidosis	None	N/A
Idiopathic fibrosing alveolitis	None	N/A
Cystic fibrosis	Cystic fibrosis	L
Pneumothorax	None	N/A

While the graphics were impressive, outputting data as both tables and maps by region and by year or specified period, we found a number of limitations:

- Poor written documentation accompanied the CD-ROM, particularly regarding definition of search terms. Online help was available but was mainly concerned with technical issues. It did not help with general interpretation of the GPRD (as discussed in the comparison of MSGP4 and GPRD in Section 3).
- The data was not clean. For example, small numbers of women aged 40-50 had fibrocystic disease of the breast, which was miscoded as cystic fibrosis.
- Diagnosis only was available – there was no information on prescribing or possibility of combining prescribing and diagnostic information.
- The OXMIS codes used to form each disease group e.g. asthma were not specified on the first disk (this is corrected on the EPIC 2 disk).

- Maps gave an indication of high and low prevalence areas but there was no key for the actual numerical rate ranges corresponding to each colour.

We had to liaise with the company to obtain a precise epidemiological definition of the search terms used. It became apparent that the EPIC measures focused on disease incidence (which was dependent on the duration of recording) while our project focused on period prevalence.

Diseases were specified as either long-term (e.g. asthma) or short-term (e.g. acute lower respiratory tract infection) in the EPIC CD-ROM. For long term conditions, a patient was only counted on the first ever occasion they were recorded as having a problem. For short-term conditions, there was a specified period of time over which any mention counted as one episode (Table 16.1). Search terms used in the EPIC CD-ROM were:

(a) “had an episode in the year under study” This only applied to long-term conditions. The numerator related to a diagnosis entered on the database at a year at any time between 1992 and 1995 *which had not been previously recorded*. The denominator related to the number of patients registered and alive at the beginning of the study time-period. The measure given was therefore an *incidence risk* (not a rate nor a period prevalence as implied by the term). Under-recording of conditions prior to the study period would increase the number of apparent new cases and therefore over-estimate the incidence risk. Under-recording of conditions during the study period would under-estimate the incidence risk. Since measures calculated were risks and not rates, the risk of having had a diagnosis of a long-term condition would tend to increase with the length of the study period, as the longer the GP practice has been on the database, the greater the number of patients who have been diagnosed prior to the study period.

(b) “ever had an episode” This search could be used for short-term and long-term conditions. The numerator was the number of patient registered and alive at the specified date with the condition recorded at any time prior to the specified date and the denominator was the number of patient registered and alive at the specified date. The measure given was therefore the *cumulative incidence at the specified date*. For long term conditions, under-recording (for example, prior to registration) would act to under-estimate the cumulative incidence. For short term conditions, the longer the practice has been submitting data, the more likely it is that its patients were included in the numerator, so the cumulative incidence would tend to increase with duration of follow-up.

(c) “Number of episodes” This applied to short-term conditions over a defined period and, unlike (a) and (b), was not dependent on when the practice joined the database.

- The episode counter could be set to one in which case the numerator was the number of patients who had the condition recorded at least once during the study period and the denominator was the number of patients registered and alive at the beginning of the study period. The result was a measure of *period prevalence*.

This was closest to the measure we used for this report, but was only available from the EPIC CD-ROM for short-term conditions.

- If the episode counter was set to two, the numerator was the number of patients who had two or more episodes and the denominator was as before. The result selected patients with recurrent disease.

Commercially released data on CD-ROM – MSGP4

A CD-ROM of individual level data from the MSGP4 was released shortly before the termination of the project which contained period prevalence data. Unfortunately, we did not have time to evaluate it.

PACT data

The commonly used and main source of prescribing data is PACT analysis, based on redeemed prescriptions. While PACT gives data on prescribing patterns of all UK doctors, its epidemiological value is limited because prescriptions are not linked to age or to diagnosis. On the other hand, although the GPRD provides more detailed information on patient characteristics and indications for therapy, it is based on a self-selected panel of doctors whose prescribing behaviour may be atypical. This study did not compare GPRD and PACT data, but a previous analysis for 1992[99] suggested generally good agreement, apart from some excess GPRD prescribing in older adults probably due to “uncashed” prescriptions and some minor under-recording in chapters 5, 9 and 19, but not in chapter 3 (used in our analysis). However, an analysis of inhaler prescribing in GPRD practices in Northern Region in 1992-3[29] suggested that GPRD practices in this region had lower prescribing rates than the average for all GPs as suggested by PACT data.

16.3 Routine data

Both hospital admissions and mortality data are collected routinely for administrative purposes, but also widely used for epidemiological investigations.

Hospital Episode Statistics (HES)

Quality and completeness of HES

Although HES contained the largest number of events for seven of the 10 conditions and, therefore, hospital admission rates were generally least affected by random fluctuations, the quality of the data was not as good as in other data sources. HES quality appeared variable by year, by region and by clinical area (speciality of consultant) in terms of both completeness (as assessed against KP70 contracting returns) and missing primary diagnoses. Missing diagnoses may also vary on a temporal basis – as shown in handouts of unpublished monthly HES data for 1998/9 at a Department of Health conference HES '98 (held in November 1998).

As discussed earlier, KP70 use has changed over time with the development of contracting and using it as an indicator of completeness to calculate grossing factors, as used in published volumes, may introduce distortions into the data.

Adjusting for missing primary diagnoses in records may also introduce distortions into the data as the assumption made is that missing diagnoses are distributed in the same proportion as coded diagnoses, which may not be justified (i.e. incompleteness within a speciality may not apply to the disease under investigation). Additionally, adjustments are made according to the speciality of the admitting consultant. However, this speciality is designated by the employing Trust and may not be consistent from hospital to hospital or area to area.

Variations in coverage and missing diagnostic codes are less likely to be important when large aggregations of data area used, for example at national and probably at regional level, as the variations in completeness and in missing diagnostic codes at a trust or district level could reasonably be assumed to cancel each other out. Variations in quality of HES data used for daily or weekly time series are unlikely to vary systematically with the exposure variable (for example, daily air pollution levels) and will not need to be considered[9] unless there are large gaps in the data.[10] Investigation of apparent disease clustering can be particularly sensitive to systematic variations in quality and completeness of the data.[11] However, using data from a single trust should be less prone to *systematic* differences across small geographical areas or in coding across time (unless systems or 'rules' are revised, when data would show abrupt changes).

Variations in coverage and missing diagnostic codes become important in situations involving comparison of trusts or aggregated data from a small number of trusts, for example investigation of a disease cluster within a Health Authority or of an environmental hazard. Epidemiological studies may choose to exclude those with

high levels of missing data. For example, a study examining national differences in hospital death rates in 1991/2 to 1994/5[12] excluded hospital trusts where more than 30% of primary diagnoses were missing. Exclusion may not be possible in certain circumstances, for example when comparing trusts for performance management purposes, when local knowledge, trust level data quality reports and liaison with the trust concerned may be needed. Alternatively statistical methods can attempt to allow for variations in HES quality. For example, a study investigating the effect of living near cokeworks on hospital admissions for respiratory disease, used a Poisson regression analysis to model the effect of distance from the Cokeworks on admissions and included hospital as a covariate.[13] Another study investigating the effect of living near a busy road on hospital admissions for asthma estimated the size of effect for each hospital separately (by choosing hospital controls admitted to the same hospital as the case) and then combined them.[14]

Comparability with other data sources

Comparability of HES with other data sources was also complex. We decided not to include 1995 data in this analysis, because of the coding change from ICD9 to ICD10. Bridge coding is not routinely performed for HES as it is for mortality, so it is less easy to assess artefactual changes due to the differences in coding. Mortality will not change to ICD10 until 2001, so analyses comparing mortality and HES between 1995 and 2000 will need careful interpretation. Another difference with HES is that most other data sources are presented by calendar year, but HES is made available by financial year.

Using emergency hospital admissions as an indicator of the burden of disease

Emergency hospital admissions are commonly used as an indicator of the burden of disease in a district, but this assumption is rarely validated.

We found that the predictive value of hospital admissions varied by disease as detailed in the following sections. For asthma, in particular, the local hospital admission rate appears to be a poor predictor of the burden of disease.

Arguably, for some of the rarer conditions such as sarcoidosis, cystic fibrosis and idiopathic fibrosing alveolitis all admissions (rather than just emergency admissions) might be a better indicator of the burden of disease because a large proportion of admissions are elective. On the other hand, elective admission rates may be influenced to a greater extent than emergency admissions by local traditions of clinical management and the availability of hospital beds.

Mortality

In general, mortality data appeared to be of good quality. This project did not assess the accuracy of death certification directly, but published studies provide some insight into the validity of certified causes of death, particularly from chronic obstructive airways disease and to a lesser extent, from interstitial lung disease.

Obstructive airways disease

Asthma as a cause of death has been found to be both over and under attributed on death certificates.[15] However, more information is available on deaths wrongly attributed to asthma as it is logistically much simpler to locate deaths where asthma is mentioned on the death certificate and decide how many have been wrongly attributed than to look at all deaths in order to identify deaths attributed to other causes which should have been attributed to asthma.[16]

Tables 16.2 and 16.3 detail some of the studies which have examined the accuracy of death certificates attributing cause of death to asthma or COPD. The more recent studies suggest that the diagnosis of asthma on the death certificate is highly specific but that sensitivity (and therefore false negativity) varies.

Table 16.2 UK Studies examining the accuracy of the registered cause of death on the death certificate: asthma

Author	Details of study	Sens-itivity	Spec-ificity
UK			
Guite HF [16]	85 deaths in 1988-1992, South East Thames. Deaths from 2382 patients within 3 years of admission for asthma, aged 16-64 on admission	82%	97%
Subcommittee of the BTA research committee [17]	147 deaths in 1979, Mersey and West Midlands. Asthma mentioned anywhere in death certificate, ages 15-64	87%	59%
Smyth ET [18]	350 deaths in 1987, Northern Ireland. Asthma or COPD mentioned anywhere in death certificate, all aged <56 & 50% COPD ages 56-75	40%	97%
Sears MR [19]	101 deaths in 1979, England. Deaths where asthma certified as cause of death.	89%	-
North America			
Hunt [20]	339 deaths 1964-1983, Olmstead County, Minnesota. Deaths from 5241 patients with medical treatment for asthma and related respiratory conditions between 1964-83. Definite asthma, age<66 Definite asthma, age>65	67% 42%	96% 97%

Sensitivity here defined as proportion of true asthma deaths where asthma is identified as the underlying cause of death by the death certificate. **Specificity** here defined as the proportion of true non-asthma deaths where asthma is not identified as the cause of death on the death certificate

Table 16.3 UK Studies examining the accuracy of COPD as the registered cause of death

Author	Details of study	Sens-itivity	Spec-ificity
UK			
Smyth ET [18]	350 deaths in 1987, Northern Ireland. Asthma or COPD mentioned anywhere in death certificate, all aged <56 & 50% COPD ages 56-75	65%	86%

Sensitivity here defined as proportion of true COPD deaths where COPD is identified as the underlying cause of death by the death certificate. **Specificity** here defined as the proportion of true non-COPD deaths where COPD is not identified as the cause of death on the death certificate

There was some crossover between asthma and COPD:

In a 1998-92 study in the UK by Guite HF et al [16] (see Table 16.2)

- 2 (9%) of the 22 deaths certified as asthma were thought by the expert panel to be due to COPD
- 2 (10%) of the 21 deaths certified as COPD were thought by the expert panel to be due to asthma

In a 1987 study in Northern Ireland [18] (see Table 16.3):

- 29 (25%) of the 114 deaths certified as COPD were thought by the expert panel to be due to asthma (2 (22%) of 9 in ages <56, 27 (26%) of 105 in ages 56-75)
- 1 (2%) of 50 deaths in under 75 year-olds certified as due to asthma were thought by the expert panel to be due to COPD.

Idiopathic fibrosing alveolitis (IFA) and interstitial lung disease

A paper published in 1990 [21] examined the validity of deaths from IFA in Nottingham in 1979-88 and found that 19 (83%) of 23 deaths certified as due to IFA had clinical evidence of the disease on record review, but estimated that up to half of patients dying from the disease might have been missed.

Multiple causes of death

This study, in common with most published work, used the underlying cause of death rather than all causes mentioned on the death certificate. However, multiple cause data (all conditions mentioned by the certifier are coded) are available for 1985, 1986 and all years from 1993 onwards.[22] Multiple cause data may be more appropriate for investigating environmental influences on the total burden of mortality. They could also be used to explore regional differences in certification behaviour which may affect the position in which the diagnosis is placed on the death certificate.

16.4 Limitations of methods used

Geographical identifiers

Boundary changes of administrative regions and districts

This analysis concentrated on health authority level analyses as the health authority is currently a key focus for commissioning of local services and surveillance of morbidity data. Health authority is a common geographical identifier across datasets (for example, mortality, the General Practice Research Database and the Health Survey for England) and in published data from the Office for National Statistics and the Public Health Common Dataset. Some data also or alternatively use local government districts as the geographical identifier. However, local government and district health authority boundaries did not become co-terminous until 1996 and standard regions not until 1999 – i.e. after the period analysed.

Health Authority boundaries underwent a large number of changes in 1991-1995, the period covered by the analysis. Further large scale changes occurred in 1996, when Health Authority boundaries were aligned with local government boundaries resulting in boundary changes to 80 of 105 DHAs including one merger[23] and again in January 1999, when the London (NHS) region was created with changes to North Thames, South Thames, Anglia & Oxford and South & West regions. When simple mergers occurred, it was relatively simple to deal with this in the analysis, but a number of the boundary changes were more complex. Different datasets may not be consistent in the boundaries that they use even if they relate to the same year(s), for example, the HSE95 was released using 8 rather than 14 regions, while mortality for 1995 was available to 14 regions.

Postcoded data

Ideally, data for all sources would have been available at a small area level (e.g. postcode, ward or enumeration district) allowing aggregation to construct a data-series allowing comparison of the same geographical area. While specific requests can be made for some data sources such as HES, for data to be allocated to specified boundaries, this is not possible with the HSE95 or the GPRD. Postcoded data has its own problems – for example, in certain districts patients with unknown postcodes have routinely been allocated to a ‘dump’ postcode within the district.[24] Individual health authorities wishing to perform analyses usually have postcoded hospital episode data on file for their own residents related to contracting arrangements, but each authority is unlikely to store more than ten years of back data and data may be incomplete if health authorities have merged or split in that time period. Specific epidemiological analyses may need to combine admissions from several adjacent authorities, even in the absence of boundary changes.

Dealing with boundary changes in the absence of postcoded data

Where postcoded data are not available for all data sources, a number of different options can be considered to deal with boundary changes. Generally, analysis will need to use the district boundaries of the most recent year analysed.

- (i) If mergers only have occurred in the area and time period studied, the data from districts in earlier years can be combined. This involves simple recoding to the district code present in later years.
- (ii) Since most problems occur when districts have been split, artificial 'conglomerate' districts incorporating districts which have been split and recombined can be created. This approach was when we needed to combine data from years 1991-5 in this analysis. Six artificial 'conglomerate' districts incorporating 10 split districts were created, including the 'West London conglomerate' involving five districts and 2.7 million people. However, geographical correlations which involved single years of data made minimal use of conglomerates. None were necessary for the 1991 analyses (which used 1992 boundaries). Two small conglomerates were needed for the 1994 analysis – each involved two Health Authorities (Suffolk + East Norfolk, Kingston & Richmond + Eastern Surrey).
- (iii) Population shifts using data on population by district health authority can be estimated and small movements (e.g. less than 50,000) ignored. This may suffice for certain localities.
- (iv) Discontinuous analysis can be performed where districts have been split. This limits the ability to monitor time trends.
- (v) Estimates of percentage population shifts can be calculated and applied to admission or other counts. This would not be accurate if the population that was recoded is different in some way from the rest of the district (e.g. different socio-economic class, different age composition)

However, any other approach than the 'conglomerate' approach which we adopted would have affected counts in split districts for:

- (i) Hospital admissions starting in the financial year prior to the split and finishing in the next (most likely to affect admissions in March and long hospital admissions). HES is released as *finished* consultant episodes and the admission is coded to the discharge district. This might have had disproportionate effect on rarer diseases such as tuberculosis or sarcoidosis in our analysis as these might be expected to be more prevalent in areas affected by splits such as West London due to ethnic mix.
- (ii) Deaths in January to March prior to the split, because deaths are coded to the district existing at the end of the calendar year, whereas health authority boundaries usually change at the end of a financial year.

Urban rural classification

The urban rural codes used showed increasing population weighted population density (pwpd) from rural to conurbation codes, but there was some overlap between the pwpds of neighbouring codes (Section 4.5).

Additional analyses for asthma by region for hospital admissions and symptoms in the HSE95 were performed using population weighted population density (pwpd) instead of urban rural classification. This preserved the general pattern of regional SERs in both analyses, with the exception of a minor change in SERs for emergency hospital admissions in West Midlands (region 12 in Figures 16.1 and 16.2). Similar urban rural patterns were seen with a gradient in emergency hospital admissions (lower in the lowest pwpd quartile, highest in the highest pwpd quartile) and not urban rural pattern in symptoms (not shown).

Figure 16.1 SERs for emergency hospital admissions for asthma for 1993 adjusted for age, sex and urban rural code

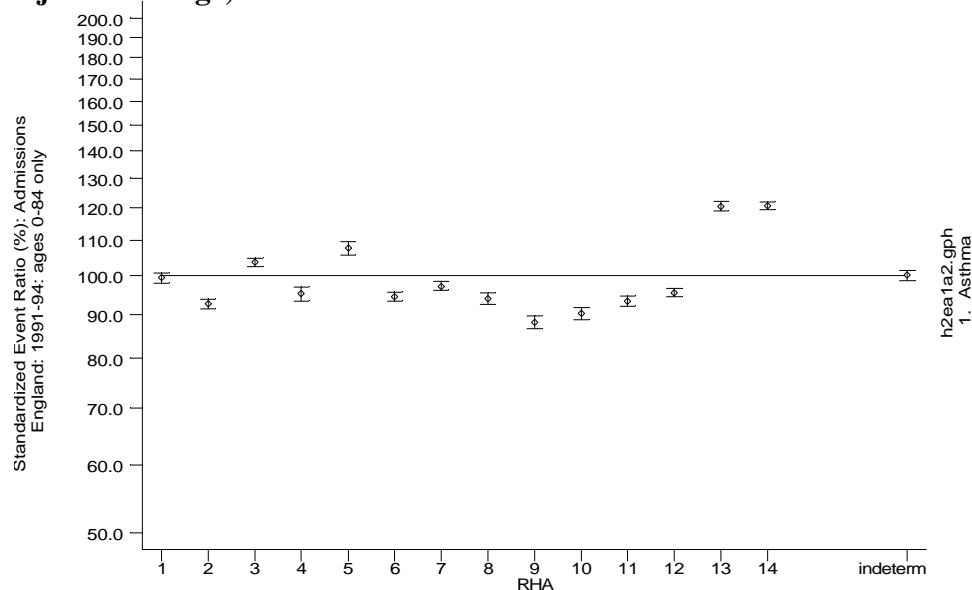
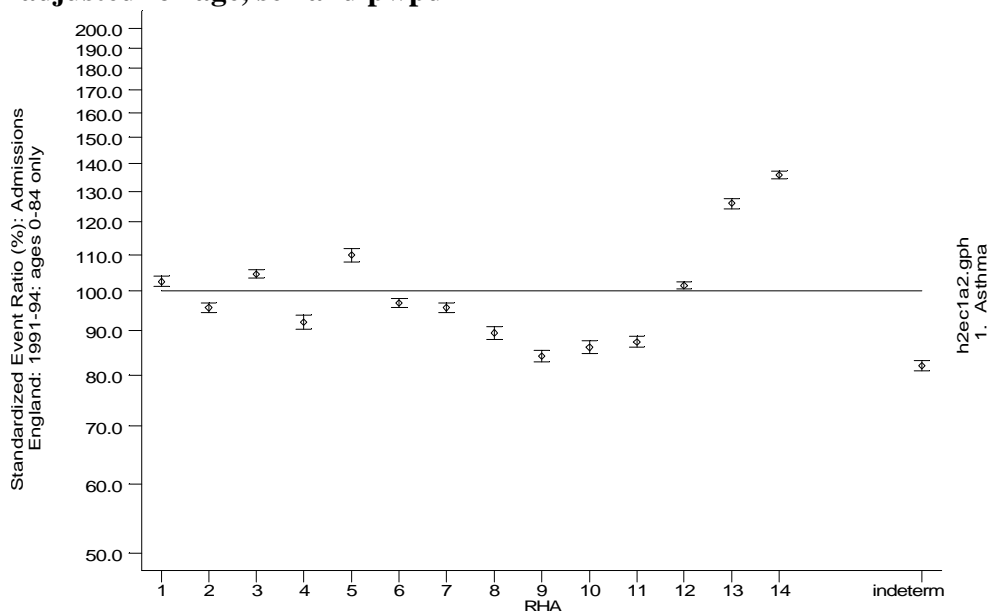


Figure 16.2 SERs for emergency hospital admissions for asthma for 1993 adjusted for age, sex and pwpd



* Key to RHA: 1 = Northern, 2 = Yorkshire, 3 = Trent, 4 = East Anglia, 5 = North West Thames, 6 = North East Thames, 7 = South East Thames, 8 = South West Thames, 9 = Wessex, 10 = Oxford, 11 = South Western, 12 = West Midlands, 13 = Mersey, 14 = North Western

Our results suggested that population weighted population density, either as a continuous or grouped variable could be used as a measure of urbanisation in future analyses. An alternative variable measure for urbanisation that could be considered is the percentage of population living in an urban area as defined in the 1991 census.[25]

An additional urban rural classification to that of districts was available in the Health Survey for England 1995 relating to the urban or rural setting of the households. This suggested slightly higher levels of asthma symptoms, inhaler use and self-reported asthma in households situated in urban areas, which was not seen using the urban rural classification of district or using quartiles of pwpd suggesting that these are crude measures of urbanisation.

Geographical analysis by ONS group

An alternative to analysis by health authority or urban rural classification is to use ONS group.[26] This coding was devised using selected 1991 Census variables to group local authorities with similar census (demographic, employment and socio-economic, household composition and housing variation) characteristics. Each group was given a name which was assigned post hoc and therefore does not truly reflect the physical character of an area such as land use. ONS groups can be further amalgamated into one of six families (Table 16.4). Group and family codes were assigned to health districts by allocating them to the local authority grouping to which they were most similar. A detailed list of ONS group by 1994 health authority can be found in Appendix A5.

ONS group can be used to adjust for socio-economic characteristics, but is relatively crude as it contains a grouping of a number of census variables and is applied to areas rather than to individuals. It can be difficult to interpret the findings as they do not relate directly to socio-economic conditions or to geographical areas. However, it has been used to analyse the GPRD to explore differences in the prevalence of treated selected common (non-respiratory) conditions.[5]

Table 16.4 ONS group and family codes

Group Code	Group name	Family Code	Family name
1	Coast & Country	1	Rural Areas
2	Mixed Urban And Rural	1	
3	Mixed Economies	2	Urban Centres
4	Manufacturing	2	
5	Growth Areas	3	Prospering Areas
6	Most Prosperous	3	
7	Ports & Industry	4	Mining & Industrial
8	Coalfields	4	
9	Inner London	5	Inner London
10	Services & Education	6	Maturer Areas
11	Resort & Retirement	6	

Geographical correlations

Rank correlations were here used as a descriptive statistic to examine the degree of linear association between rankings by geographical area for two variables. A more familiar use of correlations is in investigative analyses where they are used to test the null hypothesis of no association between exposure and outcome. When using rank correlations as a descriptive statistic, the statistical significance becomes difficult to interpret. In any case, statistical significance depends on both the strength of the association and the number of points included in the correlation. We calculated the critical value at which Spearman rank correlation coefficients would become statistically significant for the analyses performed (Table 16.5). For example, in correlations for 1994 involving points from the 14 regions, any coefficient larger than or equal to +0.54 would be statistically significant.

Table 16.5 Critical Values of Spearman's Rank Correlation Coefficient to achieve statistical significance ($p < 0.05$)

Number of Points	Critical Values
n=14 (correlations of 1994 data)	$r_s \leq -0.54$ or $r_s \geq 0.54$
n=33 (correlations of 1991 data, with/within the GPRD)	$r_s \leq -0.35$ or $r_s \geq 0.35$
n=50 (correlations of 1991 data, between Mortality and HES)	$r_s \leq -0.28$ or $r_s \geq 0.28$

Spearman's rank correlation was chosen because it does not make any assumption about the underlying distribution of data and should not therefore give undue weight to outliers which might result from variations in quality of the data, or systematic differences in coding or collecting data in different areas. However, a particular problem with all correlations (rank or other) is that no account is taken of the fact that some points are better estimated than others. Our correlations were calculated from rankings based on rates and therefore on numbers of events. Rankings were arbitrarily deemed unreliable if any rank was based on less than 10 events. Using this criterion, we found that nine of the 10 respiratory diseases had at least one point where the number of events were judged small enough to affect the interpretation of the Spearman rank correlations (all comparisons between data sources for four diseases: tuberculosis, cystic fibrosis, sarcoidosis and pneumothorax, some between data source comparisons for hayfever, acute bronchitis and bronchiolitis and within data source comparisons by age for asthma and pneumonia). Choosing 10 events as the cut-off point still allowed scope for random fluctuations (assuming a Poisson distribution of events, random variation about 10 events would range from 4 to 16 events). Since the number of events differed by region or region plus urban rural combination, the variability would not be equal across the points. These random fluctuations would not be apparent from the value of r_s or the associated p-value.

Further analyses either using more years of data or larger area groupings or both to increase the number of events would be needed to explore correlations for the rarer conditions. However, correlations between very broad regional groupings may

minimise underlying differences between areas and increase the likelihood that even a moderately strong correlation has arisen by chance (as suggested by the increase in critical value with decreased numbers of regional divisions in Table 16.5).

Patients and events

Care is required distinguishing between patients and events in routinely available data. Only in mortality and cross-sectional data is the relationship straightforward. Although Scottish hospital admission data have been routinely 'linked' by a unique identifier for some years, this is not the case for English HES data. Since we did not have access to local patient identifiers, we were not able to link individuals admitted to hospital on more than one occasion. HES data measure the incidence of 'spells' of illness, but overestimate the prevalence because of readmissions for the same condition for both within year analyses and for combined year analyses. The analyses of GPRD data used individual level data to derive annual period prevalence rates. Prevalence may have been overestimated for combined year analyses in the GPRD, because while each patient was only counted once within a single year, the analysis precluded linking patients between years.

16.5 General results

Small numbers

Small numbers limited the planned analyses for all diseases except for asthma. This is likely to be a problem for analyses of single years of data or age-specific analyses for many conditions and even for common diseases such as asthma at a district level. For example, Health Survey for England data at district level for common diseases has recently been published, including one respiratory outcome: proportion of wheeze or diagnosed asthma.[27] Despite the use of three years of pooled data (1994-6) for all ages and both sexes, confidence intervals were wide. In this analysis, the numbers of events were small in rarer conditions such as sarcoidosis, fibrosing alveolitis, cystic fibrosis, tuberculosis and pneumothorax even in combined five years of data and this limited the ability to interpret regional differences.

Emergency or all admissions?

Analyses involving HES were conducted using emergency admissions. However, the breakdown of all admissions by type of admission (Table 5.2 in Section 5) showed that over 40% of admissions in 1994 for fibrosing alveolitis, cystic fibrosis and sarcoidosis and 30% of admissions for tuberculosis were elective. Therefore, for some studies of these conditions, all admission types (elective, emergency and other) might be a better measure of the burden of disease than emergency admissions.

Geographical correlations

Where geographical correlations across datasets were strong, this was interpreted as indicating that, for example, areas with high hospital admission rates would correspond to areas with a high proportion of patients consulting in general practice for that disease.

We considered that inconsistent correlations between datasets for a disease were most likely to reflect differences in non-epidemiological factors such as provision of and ease of access to medical care, its quality and differences in diagnostic labelling and coding. An alternative interpretation would be that inconsistent correlations reflected differences in the severity of disease in different regions (if, for example, asthmatics in one area suffered from more severe disease they would be more likely to be require hospital admission than in another region). Differences in severity might be related to lifestyle factors such as smoking or social factors as well as environmental factors and these might all interact. However, it seems likely that environmental and social factors at a regional level would affect disease across all levels of severity from symptoms to deaths. Asthma and COPD both have wheezing as a prominent symptom and there some diagnostic overlap between them, but asthma showed marked inconsistency across all data sources while COPD showed a marked consistency. It seems unlikely that this could be solely explained as regional differences in severity.

16.6 Asthma

Summary of discussion section for asthma

We found a ‘double peak’ in the prevalence of asthma presenting to general practice, which would correspond to two phenotypes of asthma – ‘wheezy bronchitis’ peaking in infants and atopic asthma peaking at puberty.

Asthma measures were higher in boys than girls until puberty, which may reflect gender-specific differences in severity or in health care utilisation.

The timing of the asthma peak in emergency hospital admissions in children relates closely to the commencement of the school term. This may have implications for health service planning, but argues against important environmental influences.

Differences in the use of medical care for wheezing illness by the young and the elderly would support two opposing interpretations: that high rates of hospital admissions in young children prevent asthma deaths or that high rates of admissions in young children are inappropriate because they are very unlikely to die from wheezing illness.

Urban areas had higher rates of hospital admissions than rural areas, which may be due to service factors such as ease of access and quality of primary care rather than to severity.

The poor consistency between different data sources for asthma indicate that hospital admissions are not a good indicator of the prevalence of asthma locally.

Investigation of environmental influences should use the data source most clearly related to the problem. For example, asthma severity might be better assessed using hospital admissions, while prevalence might be better assessed by survey data on symptoms.

Two phenotypes of asthma?

A 'double peak' in the prevalence of asthma presenting to general practice was revealed by analysis of single years of age. This was an age effect and not a cohort effect because it was consistent from year to year. This would support the hypothesis that there are two different phenotypes of asthma[28]: 'wheezy bronchitis' peaking in infants[29] (wheezing related to viral infections and small airways and to parental smoking habits) and 'atopic asthma' which steadily increases throughout childhood with a peak at puberty.[30] However, the first peak at ages 5-6 is slightly later than observed in a study in the USA[29] (which suggested that 20% of children experienced wheeze before the age of 3, which remitted by the age of 6) and may reflect the lag in recording a definitive diagnosis in general practice.

Male:female differences

Asthma measures were higher in boys than girls in early childhood until the mid teens in all data sources except mortality (where there were very small numbers of events) and the sex differential became smaller with age (Table 16.6). This is consistent with published literature.[31] However, our study was able to demonstrate that the magnitude of the gender difference varied by data source.

Table 16.6 Male:female ratios for childhood asthma in HSE95 (wheeze in the past 12 months), GPRD (inhaler prescriptions plus asthma diagnosis) and HES (emergency admissions) by age group

Age group	HSE95 ratio (M:F prevalence)*	GPRD ratio (M:F rate) †	HES ratio (M:F rate)‡
0-4	1.14 (23.6%: 20.7%)	1.53 (91.4 : 59.8)	1.97 (11,455 : 5,826)
5-9	1.09 (17.1%: 15.7%)	1.35 (120.2 : 88.9)	1.76 (3,587 : 2,037)
10-14	1.05 (19.2%: 18.3%)	1.31 (121.8 : 93.2)	1.20 (2,237 : 1,871)

*HSE95 expressed as % prevalence

† GPRD crude rate per 1000 pyar for 1991-1995

‡ HES crude rate per million population for 1991-1994

Recent questionnaire surveys have also demonstrated changes in the male:female ratios in symptoms in the mid teens, although the age at which this occurs varies (Tables 16.5 and 16.6). An analysis of the 1958 birth cohort followed up at age 16 and age 23[32] (i.e. relating to a population at least 20 years older than the recent questionnaire surveys) found that the changes in male:female ratio of 'asthma or wheezing' in the previous 12 months occurred later between the ages of 16 (when the male:female ratio was 1.4) and 23 (when the ratio was 0.68).

Table 16.7 Male:female (risk) ratio of wheeze in the past year* – comparison of HSE95 with a national survey in 1992

Study	Age in years			
	5-7	8-10	11-13	14-17
Strachan[15], UK 1992 (n=5,472)	1.4	1.2	1.7	1.1
HSE 1995 (n=3,282 in age range)	1.27	0.88	1.24	0.83

Table 16.8 Male:female (risk) ratio of wheeze in the past year* – comparison of HSE95 with a survey in Nottingham in 1996 and a national survey in 1995

Study	Age in years					
	11	12	13	14	15	16
HSE 1995 (n=1,461 in age range)	1.62	0.96	1.26	0.62	0.90	1.11
Venn[33] Nottingham, 1996 (n=27,826)	1.10	0.91	0.82	0.76	0.63	0.66
		12-14				
Kaur[34] UK, 1995 (n=27,507)		0.88†				

† Odds ratio

The differences in male:female ratio in children between survey data, GP consultation/prescriptions and hospital admissions could suggest one or more of the following explanations:

- (i) a difference in health care utilisation i.e. if boys have an attack they are more likely to see the GP and/or be admitted to hospital or
- (ii) Gender specific differences in severity at different ages – a tendency to more severe disease in young boys than young girls which reverses in mid teens

The latter may have several components:

- Airway size. This is consistent with observations that boys are more likely to have smaller airways than girls early life and to develop early asthma which they ‘grow out of’.[31] Part of the change over of male:female ratio could therefore be related to an increase in the size of airways in males, so that respiratory illnesses would be less likely to cause wheezing.
- An increase in prevalence of asthma in girls in adolescence related to hormones and/or environmental exposures such as smoking.[33]
- Gender differences in atopy. Prevalence of allergic phenomena is higher in boys than girls.[31] For example, a higher proportion of males than females aged 9-11 have been shown to have atopy on skin prick testing in a previous study[35] and in unpublished ISAAC UK data. Atopic individuals are more likely to develop allergic asthma and also to be labelled as asthmatic if they develop wheezing.[36]

* Questions used in surveys

- **Strachan 1992**[15]: Has this child *ever* had wheezing or whistling in the chest at any time in the past? If yes, asked Has this child had wheezing or whistling in the chest *in the last 12 months*?
- **HSE95**: Have you ever had wheezing or whistling in the chest at any time, either now or in the past? If yes, asked
 1. Have you ever had this wheezing or whistling when you did not have a cold? and
 2. Have you had wheezing or whistling in the chest in the last 12 months?
- **Venn 1996**[33]: Have you ever had attacks of wheezing in the chest? (a noisy whistling sound from the chest, not the throat causing tightness and breathlessness)
Have you had any wheezing attacks in the last year?

Urban rural differences in asthma

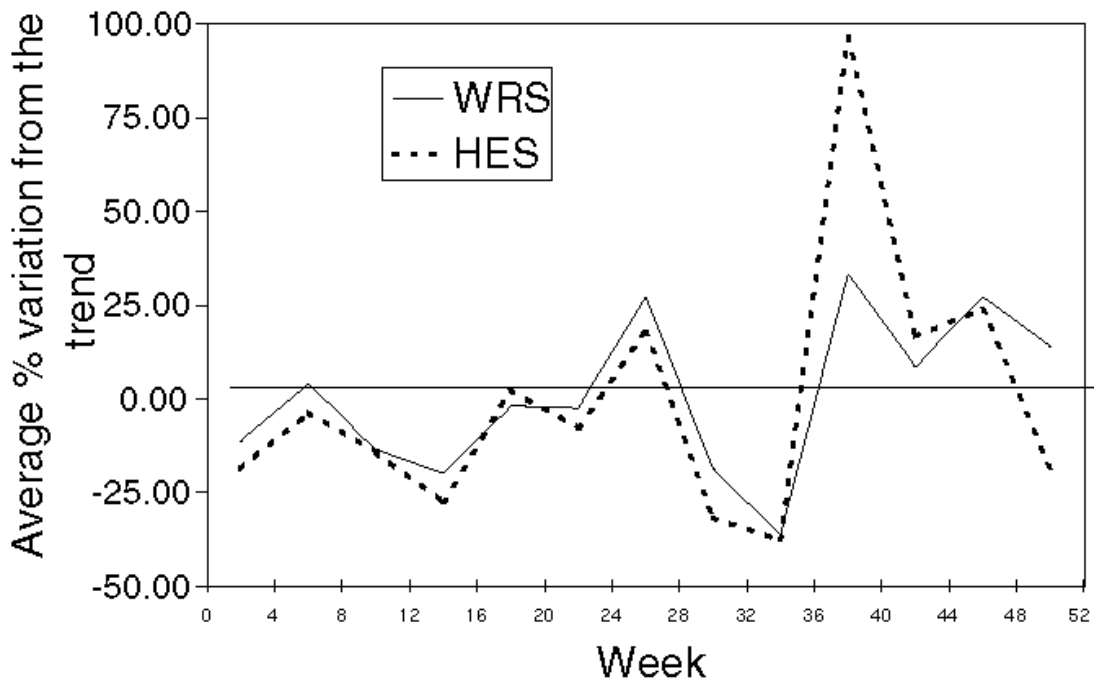
Scatterplots suggested that urban areas and conurbations had higher age standardised rates of hospital admissions than the corresponding age-standardised rates of asthma treated in primary care, while the reverse was true in rural and mixed areas. Possible explanations include:

- (i) Ease of access – it may be easier to access hospitals in urban areas than rural areas and comparatively easier to access the GP in rural areas. This is supported by the HES vs. mortality scatterplots which showed higher age standardised rates of hospital admissions than for asthma mortality in conurbations and urban areas.
- (ii) Differences in quality of care in urban and rural areas. Work in East London[37], an urban area with high levels of single-handed GP practices, suggested that higher admission rates were most strongly associated with small size of practice partnership. Practices with higher rates of night visits also had significantly higher admission rates.
- (iii) Asthma is exacerbated in urban environments.

Seasonal patterns

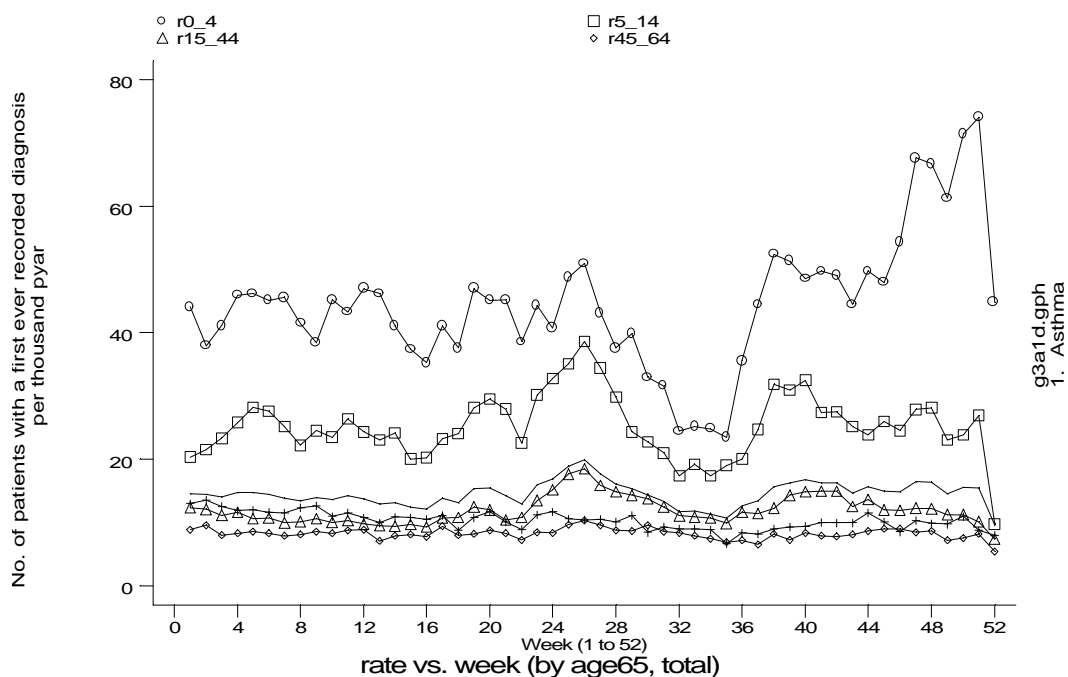
Seasonal patterns in asthma were best interpreted by non-repeat prescribing or first ever diagnosis of asthma. In contrast, the Weekly Returns Service relates to new episodes, defined as consultations for ‘a new diagnosis or for advice and treatment for an exacerbation of pre-existing illness’.[38] However, similar patterns were seen when comparing 1987 to 1992 WRS data (Figure 16.3) and 1991-1995 GPRD data on first ever consultations (Figure 16.4), which is consistent with a similar environmental trigger for first ever episodes and for a mixture of first ever episodes and exacerbations of existing disease.

Figure 16.3 Average 4-weekly percentage variation from the trend in hospital admissions for asthma and new GP episodes of asthma in ages 5-14 for August 1987 to July 1992



Source: LAIA factsheet 93/4[39]

Figure 16.4 Seasonal pattern in first ever consultations for asthma, 1991-5 in the GPRD



Further analyses of the autumn peak in asthma admissions

In order to investigate further the early autumn peak in emergency admissions for asthma, HES data were analysed for the months of August and September 1987-1994 inclusive. Daily counts were derived for males and females in four age groups: 0-4 (pre-school ages), 5-18 (school ages), 19-44 (childbearing ages) and 45+ years. Series were constructed for all England, and for four geographical areas (North, Midlands, Southwest, Southeast). Additionally, information on emergency admissions for asthma among Scottish children aged 0-14 years was available for 1988-1993.

Cumulative sums (cusums)[40] were used to identify the date of onset of the autumn epidemic of asthma admissions. These "turning point" dates were analysed by age, sex, area and year by analysis of variance (ANOVA). In the English series, turning points could be defined for all series for children except one (1992, 5-18-year-olds), but there were fewer years in which a turning point could be discerned for adults (only 3 of 8 years among 19-44-year-olds, and only 1 of 8 years among over 45-year-olds). Further analyses therefore concentrated on the 0-4 and 5-18 year-olds (Figures 16.5 and 16.6, respectively).

In the all-England series for children, all the turning points occurred during the first ten days of September, typically on a Friday, Saturday or Sunday, 11-13 days after the August bank holiday Monday. The date of maximum admissions typically occurred about ten days later (figures 16.5 and 16.6).

Counts for 0-18-year-olds were aggregated to investigate regional differences (Table 16.9). There were no consistent variations in the date of onset of the epidemic among the geographical areas of England, although for some years (1987-1989) the turning point occurred 5-6 days earlier in northern England than in the other regions. More consistent was the observation that the autumn rise in asthma admissions in Scotland started 2-3 weeks earlier than in any of the English series (21-25 August, rather than 3-14 September, table 16.9).

Table 16.9 Regional differences in turning points for asthma admissions among children

Year	North	Midlands	Southwest	Southeast	Scotland*
1987	4 Sep	11 Sep	12 Sep	10 Sep	no data
1988	4 Sep	10 Sep	9 Sep	10 Sep	23 Aug
1989	3 Sep	8 Sep	9 Sep	7 Sep	25 Aug
1990	9 Sep	none	9 Sep	8 Sep	23 Aug
1991	7 Sep	6 Sep	12 Sep	none	22 Aug
1992	10 Sep	11 Sep	1 Sep	11 Sep	22 Aug
1993	12 Sep	14 Sep	none	10 Sep	21 Aug
1994	9 Sep	11 Sep	9 Sep	none	no data

* 0-14-year-olds in Scotland, 0-18-year-olds in areas of England

The most plausible explanation for the difference between Scottish and English time-series is that the school autumn term starts two weeks earlier in Scotland than in most areas of England. Fluctuations in asthma admission rates throughout the year have been linked to school holidays and half-term breaks.[41] Community-acquired viral infections are closely associated with exacerbations of asthma among school children,[42] and it is likely that the renewed circulation of viruses as children mix at the start of the autumn term is responsible for the abrupt rises in daily asthma admission counts at this time.

An alternative explanation relates to differences in climate, but no consistent relationship could be found across years between the timing of the rise in asthma admissions and the patterns of temperature or relative humidity (to be reported in more detail elsewhere).

This analysis of daily counts from eight consecutive years suggests that the timing of the autumn rise in asthma admissions among pre-school and school-aged children is predictable and relates closely to the commencement of the school term. This may have implications for health service planning, but argues against an important environmental component to this seasonal phenomenon.

Figure 16.5 Emergency Admissions for 0-4 year olds in England, 1987-1994

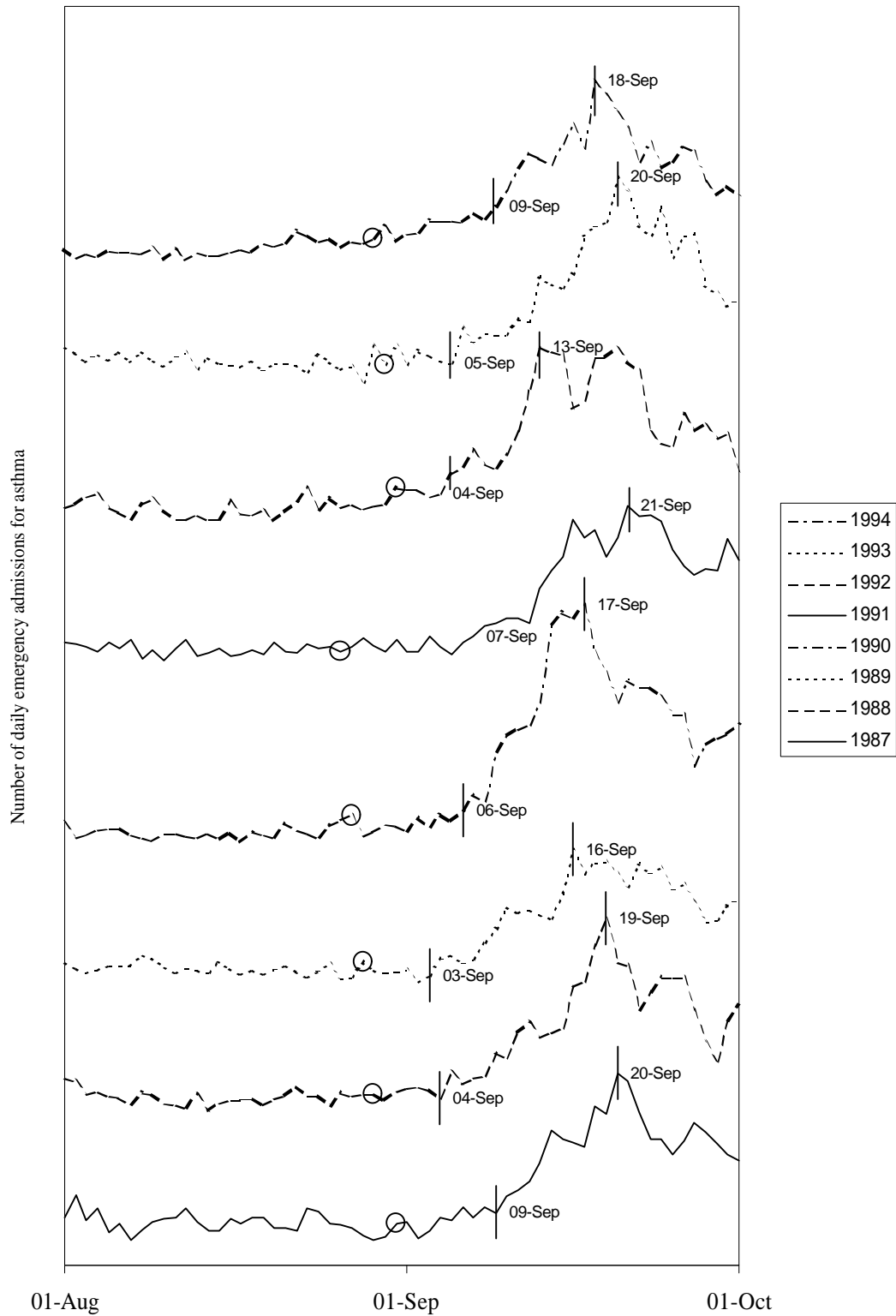
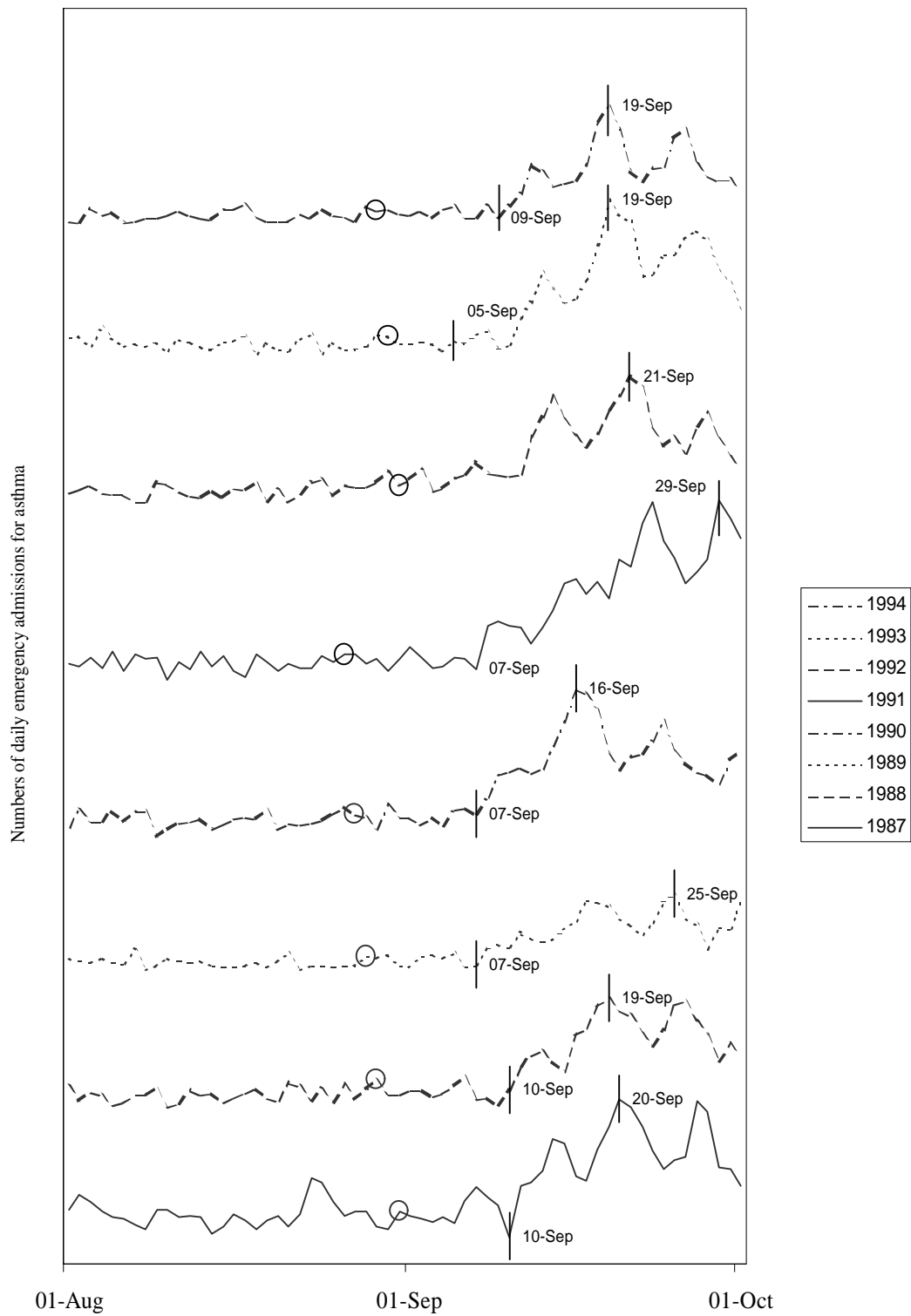


Figure 16.6 Emergency Admissions for 5-18 year olds in England, 1987-1994



Differences in uses of medical care for wheezing illness at different ages

There are differences in the relations between different measures of asthma and wheezing illness at different ages (Table 16.10). Making comparisons at the extremes of age, for children aged 0-4 symptom prevalence was numerically close to the rates of inhaler prescription for wheezing illness in general practice and of the same order of magnitude as emergency hospital admissions. For the elderly (aged 85+), with a similar prevalence of symptoms to children aged 0-4, rates of inhaler prescriptions for wheezing illness were lower. Rates of emergency hospital admissions for wheezing illness were an order of magnitude lower than wheezing symptoms, while deaths from wheezing illness were of the same order of magnitude as hospital admissions. These findings would support two opposing interpretations:

- (i) That high rates of hospital admissions in young children are preventing asthma deaths.
- (ii) That the higher rates of hospital admissions in children are inappropriate, because they are very unlikely to die from wheezing illness.

Interpreting the data to suggest that the lower use of medical care in those aged 85+ is related to the higher mortality would be prone to the ecological fallacy – individual level data would be needed to determine if those who died were more likely to not be using an inhaler or had not been admitted to hospital.

Table 16.10 Approximate prevalence or rates for different measures for asthma at extremes of age in HSE95, GPRD and mortality 1991-5 and HES 1991-4

	Age 0-4 years	Age 85+ years
Wheeze in the past year in HSE95	220 per 1000	210 per 1000
Patient inhaler prescription in general practice plus asthma diagnosis	80 per 1000 pyar	30 per 1000 pyar
<i>Patient inhaler prescription in general practice plus symptoms (no asthma or COPD diagnosis)</i>	<i>100 per 1000 pyar</i>	<i>30 per 1000 pyar</i>
<i>Patient inhaler prescription in general practice plus diagnosis of COPD</i>	<i>0.7 per 1000 pyar</i>	<i>50 per 1000 pyar</i>
Emergency hospital admissions for asthma	120 per 1000 pyar	1 per 1000 pyar
<i>Emergency hospital admissions for COPD</i>	<i>0.7 per 1000 pyar</i>	<i>5 per 1000 pyar</i>
Deaths from asthma	3 per 1,000,000	250 per 1,000,000
<i>Deaths from COPD</i>	<i>0.5 per 1,000,000</i>	<i>6,000 per 1,000,000</i>

Investigation of environmental influences on asthma

Geographical correlations between data sources were all weak, suggesting that hospital admission levels would not be a good indicator of the prevalence of disease locally. Part of this may be related to differential labelling of wheezing illness. Anecdotally, some consultant paediatricians prefer to label episodes of wheeze in children under one year of age as bronchiolitis rather than asthma, while doctors dealing with adult patients vary in their preference for an asthma or COPD diagnosis for intermittent wheeze in smokers.

Subdivision by age into children and adults did not improve asthma correlations. The exceptions were for a strong negative correlation between symptoms and hospital admissions ($r_s = -0.86$) for ages 5-14 (suggesting that regions with high prevalence of symptoms had low levels of emergency admissions) and a positive correlation between symptoms and emergency hospital admissions for adults aged 15-84. The former, raises serious questions about the role of service factors in determining local admission rates in children. To investigate this further, the symptom prevalence of schoolchildren aged 12-14 from the national ISAAC UK study in 1995 (symptoms of wheeze and speech-limiting attacks in the previous 12 months) were compared with emergency hospital admissions for asthma in 10-14 year olds in 1994 (rates shown in Table 16.11). This showed a strongly negative correlation between regional rankings for wheeze and hospital admissions ($r_s = -0.92$, $p = 0.001$) and a moderately strong but non-significant negative correlation for speech-limiting attacks ($r_s = -0.60$, $p = 0.12$). More detailed studies of asthma severity in different regions are required to explore this further.

Table 16.11 Symptom prevalence (%) from the ISAAC UK survey performed in 1995 (relating to symptoms in the previous 12 months) and crude emergency hospital admission rate for asthma per 100 population in 1994 ranked by hospital admission rate

Region	HES: Emergency hospital admissions for asthma per 1000 population in 1994	ISAAC UK, 1995: Self-reported wheeze or speech-limiting attack in the past 12 months per 1000 in children aged 12-14	
		Wheeze	Speech-limiting attack
North West	3.0	300	79
West Midlands	2.0	300	85
South Thames	2.0	313	88
North Thames	1.9	305	77
North East & Yorkshire	1.9	341	87
Trent	1.6	336	82
South West	1.5	353	96
East Anglia & Oxford	1.3	343	90

Since asthma data were clearly inconsistent, the data source most clearly related to the problem should be used to investigate environmental influences. For example, asthma severity might be better assessed using hospital admissions, while prevalence might be better assessed by survey data on symptoms.

16.7 COPD

Investigation of environmental influences

In contrast to asthma, there was good consistency between data sources in both comparisons of patterns of disease and geographical correlations. This suggests that both mortality and hospital admissions are reasonable surrogates for prevalence or severity of COPD and could be used to investigate environmental or lifestyle influences. The consistency of disease patterns across data sources, suggested that differences in provision of medical care in different regions and urban and rural areas were not responsible for the observed regional differences.

North south differences in prevalence and smoking

Approximately three-quarters of COPD mortality can be attributed to smoking,[43] so smoking might be expected to be the main confounder in geographical comparisons. However, while the north south and urban rural patterns were attenuated they were not fully removed by adjustment for smoking and social class in the HSE95. It is therefore possible that an environmental factor is responsible for these differences. Differences in diet may be a key contributor – as fruit and anti-oxidant vitamin consumption may help to preserve lung function in smokers and non-smokers.[44]

Cohort analysis

Previous analyses of mortality from chronic respiratory disease covering 1941-85[45] and 1841-1994[46] have suggested a marked cohort effect, particularly in men, with a peak in men born 1871 to 1901 and in woman born around 1925, which correspond to cohorts experienced the highest age-specific rates of lung cancer mortality and to lifetime cigarette consumption.[47] It is likely that the generations with maximal lifetime cigarette consumption at any given age differ by region and between urban rural areas, due, for example, to earlier adoption of smoking in the major cities. Although this has yet to be formally demonstrated, it potentially complicates the interpretation of both long-term and short-term time trends by region and degree of urbanisation

In our analysis of only five calendar years, there was very limited power to distinguish age effects from cohort (generation) effects. No cohort effects detectable in mortality in 1991-5 but a cohort effect was seen in hospital admissions (among those born 1910-15). Analysis of the GPRD was partially consistent with the 1900 cohort for men, but the effect was not statistically significant.

16.8 Hayfever

Regional analyses

The main comparisons were between symptoms from the HSE95 and primary care utilisation from the GPRD as the numbers of hospital admissions were small and there was only one death from hayfever in the five years studied. Symptom prevalence in the past year was higher in the South East of England, a similar finding to a national survey of 23 year olds in 1991[48], but this was not clearly seen in GP prescribing for hayfever.

Urban rural differences

There was some evidence for an urban rural gradient, with higher levels in London and in conurbations but the variation was small, allowing limited scope for significant environmental influences and consistent with similar levels of underlying prevalence of allergy in urban and rural areas.[48] Higher levels of prescribing were seen in urban areas than in rural areas, but conurbations had average levels of prescribing for hayfever. The combination of below average symptom prevalence with higher levels of prescribing in urban districts may indicate a tendency to more severe disease in urban areas, but this is not consistent with the average prescribing levels observed in conurbations, the most urban environments. Possible explanations for lower levels of prescribing in conurbations may be lower pollen levels or easier access to over the counter medication, but further work would be needed to investigate these. One study in Japan found a higher prevalence of cedar pollinosis near to motorways suggesting an interaction of pollen with vehicular air pollution.[49]

Seasonal analysis and peaks in grass and tree pollens

A summer peak in hayfever prescriptions and hospital admissions was seen in mid June to mid July which corresponds to the grass pollen season, but there was no evidence of an earlier peak corresponding to the tree pollen season. The May (weeks 18-20) peak in prescriptions, consultations and non-repeat prescriptions was later than the tree pollen peak and may represent 'stocking up' for the summer. This limits the value of GP data for time-series analysis, for instance of short-term fluctuations in hayfever incidence in relation to air pollution levels.

Investigation of environmental influences

The main data sources used for hayfever were the HSE95 and the GPRD. However, consistency between these as measured by geographical correlations was poor, suggesting that utilisation of care was influenced by factors other than the disease prevalence. The seasonal pattern in GP consultations and prescriptions for hayfever also limits the use of primary care data as an indicator of environmental influences on hayfever operating at an area level.

16.9 Pneumonia

Artefactual patterns in mortality, HES and GPRD

Some of the observed patterns were artefactual. Changes in mortality coding were seen to have a large effect on year on year mortality rates, with a doubling between 1992 and 1993.

The numbers of emergency hospital admissions for pneumonia in Oxford in 1994 were very low and in marked contrast to the previous year, when the standardised admission ratio equalled the national average. Standard quality reports suggested some under-reporting but not to the extent that might be expected from the observed figures if this was due to a global reporting problem. The number of missing primary diagnoses for patients admitted under consultants in general, thoracic, paediatric or geriatric medicine was close to the national average of 4.6% for these specialities for both 1993/4 and 1994/5 (Oxford missing 6.3% in 1994/5 and 7.2% in 1993/4), while the number of FCEs was only markedly lower for 'other medicine' which may account for small numbers of patients with pneumonia (Table 16.12).

Table 16.12 FCEs expressed as a percentage of KP70 for ordinary admissions to hospitals in Oxford in 1993/4 and 1994/5

	All specialities	General medicine	Other medicine	Paediatrics	Geriatric medicine
1993/4	97%	102%	90%	90%	116%
1994/5	not given	98%	68%	90%	111%

There were small numbers of patient consultations for pneumonia in the GPRD. It is possible that there was some misclassification of pneumonia as 'chest infection' in the GPRD which was analysed as acute bronchitis or bronchiolitis in this report. Our initial investigation of the GPRD[2] suggested that most diagnoses of 'chest infection' corresponded to a diagnosis of acute bronchitis or bronchiolitis, but there was an excess of that expected by comparison with the MSGP4 in ages 75+ (Figure 16.7), which is the age group in which pneumonia is most common.

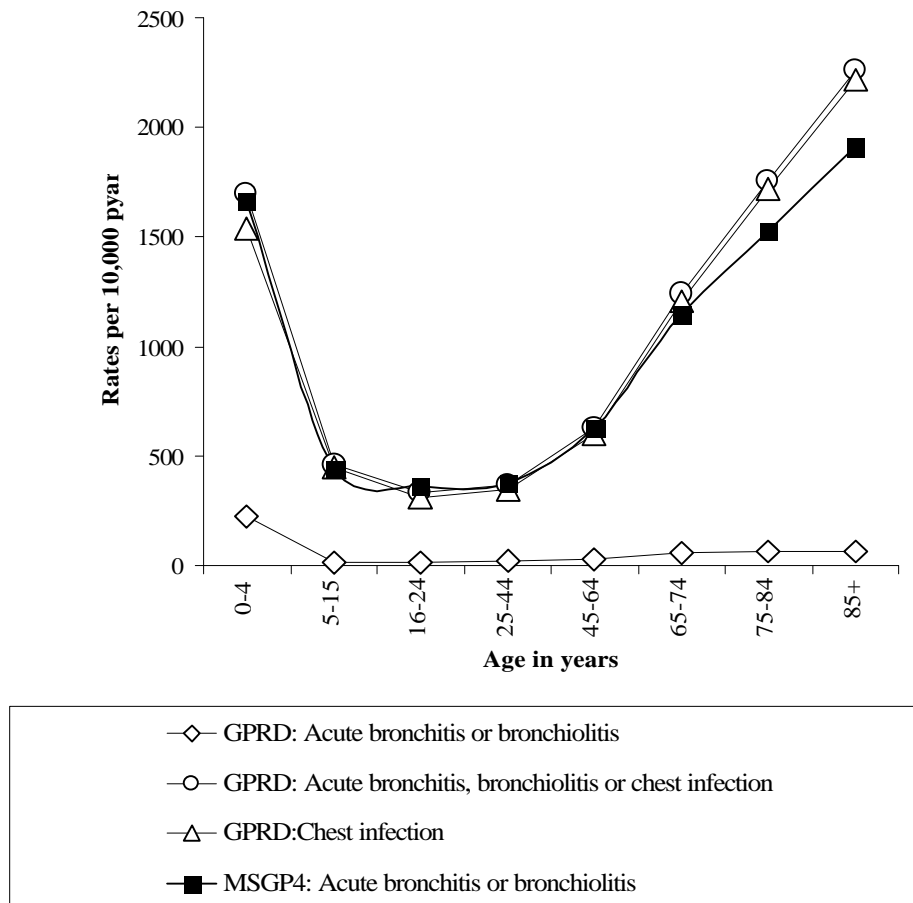
Regional and urban rural patterns

As with acute bronchitis and bronchiolitis (Section 10) there were higher rates in northern areas of England, but these was not as consistent across data sources as the findings for COPD (Section 7).

The striking urban rural gradient for pneumonia in HES and mortality, with highest levels in conurbations, was again similar to that seen for COPD (Section 7). However, this was not consistent with the GPRD, even when allowing for smaller numbers of events (illustrated by wider confidence intervals). The urban rural pattern for 'acute bronchitis, bronchiolitis or chest infection' GP patient consultations (Section 10) also did not show a gradient, but was 'U'-shaped (with higher than

average levels in rural and conurbation areas, and lower than average levels in mixed and urban areas). This may reflect differences in services, rather than a true difference in epidemiology between urban and rural areas.

Figure 16.7 The relationship between ‘chest infection’ and acute bronchitis or bronchiolitis in males from the MSGP4 and GPRD for 1991/2



Can HES be used to investigate environmental influences on pneumonia?

Geographical correlations were poor in hospital admissions by age group, suggesting a possible difference in epidemiology between pneumonia in adults and pneumonia in children. Geographical correlations were moderately good between hospital admissions and mortality, suggesting that local hospital admission rates might be a reasonable indicator of mortality in the absence of more specific data from local studies or audits.

16.10 Acute bronchitis and bronchiolitis

Seasonality

Acute bronchitis or bronchiolitis showed a marked seasonal pattern with highest rates in December and January. Small shifts in the timing of the normal winter peak lead to marked differences in yearly trends when using calendar years. Year on year analysis in diseases with marked winter peaks may be more appropriately presented by financial year.

Regional patterns

As with pneumonia (Section 9), there was an apparent excess of acute bronchitis and bronchiolitis in northern areas of England. This was similar to the regional pattern seen in COPD (Section 7), but less marked. It is possible that similar lifestyle factors such as smoking and/or environmental factors may be responsible for this excess in the north. However, as with pneumonia, urban rural patterns were inconsistent across data sources - in contrast to COPD.

Can HES data can be used to investigate environmental influences?

There was reasonable correlation between GP patient consultations and hospital admissions for acute bronchitis or bronchiolitis. However, there was a lack of consistency in the urban rural patterns and in the age sex distribution between data sources. Also, correlations between adult and childhood emergency hospital admissions were poor, while correlations between the GPRD and HES for adults separately or for children were also generally weaker than for all age groups combined. This suggests that hospital admissions could only have a limited role in environmental monitoring: they should be used for all ages combined.

As discussed in the section on asthma (16.6), there was likely to be some overlap between the diagnosis of bronchiolitis and asthma in very young children as both cause wheezing in conjunction with a respiratory infection. Similarly, there was likely to be some overlap between the diagnosis of acute bronchitis and the diagnosis of COPD in the elderly. If these variations in diagnostic labelling vary by geographical region and area then this may account for some of the weakening of correlations by age group.

16.11 Tuberculosis

Comparisons with TB notification data

An alternative source of information on tuberculosis is notifications which gives information on disease incidence. These are usually published by local authority and standard region, which were not co-terminous with health regions in the years studied in this. However, notifications by NHS regional office for the ten years 1985 to 1995 have been published[50] and are compared with other data sources in Table 16.13.

Ranking by notification rates was broadly consistent with ranking by hospital admission rates. There was wide differences in the magnitude of rates between notifications and the other data sources which can be thought of as a measures of period prevalence (for example, there was approximately a 10 fold difference between notifications and GP consultations). However, the comparative differences between data sources varied by region (for example, the notification rate expressed as a percentage of emergency admission rate was 43% in Anglia & Oxford, 38% in North Thames and 25% in West Midlands). It seems unlikely that this can be fully explained by differences in compiling the data (patients could be counted more than once in HES if they were admitted to hospital more than once in single or subsequent years and in the GPRD if they consulted for tuberculosis in more than one year). Alternative explanations include differential reporting by region, differing patterns of care (for example, West Midlands may have a lower hospital admission threshold than North Thames), or true differences in severity. The last of these may be related to differences in the age distribution of tuberculosis (e.g. due to ethnicity), which are not controlled for in these comparisons.

Table 16.13 Regional tuberculosis notifications in 1985-95[50], GP patient consultation rates 1991-5, emergency hospital admission rates 1991-4 and death rates 1991-5 ranked by notification rate* (crude rates expressed per million population per year)

Regional health office	Notifications (number, 1985-95)	Notifications (rate, 1985-95)	GPRD patient consultations (rate, 1991-5)	HES: emergency admissions (rate, 1991-4)	Deaths (rate, 1991-5)
North Thames	1,718	21.2	232.6	55.5	7.3
West Midlands	654	11.3	173.1	45.3	7.5
Trent	506	9.7	89.4	33.2	6.0
North West	633	9	120.7	36.8	6.9
South Thames	642	8.7	113.7	30.5	6.6
Northern & Yorkshire	638	8.6	159.0	26.6	6.3
Anglia & Oxford	357	7	93.1	16.3	3.2
South & West	278	4	105.8	14.1	3.8
Indeterminate				46.2	7.2

*Notification rates were calculated by dividing number of notifications by 11 and using 1991 mid year population estimate as denominator. The 'Indeterminate region' contains the 'conglomerate districts' – as detailed in the methods (section 4) – and was needed for analyses using HES and mortality. The largest constituent of the 'Indeterminate region' is the West London conglomerate, containing most of west London.

Male:female differences

Males had higher notification rates but the male:female ratio of TB notifications varied by region (Table 16.14). The two regions with lowest notification rates had the highest male:female ratio of 1.5. The two regions with the highest notification rates had different male:female ratios of 1.4 (North Thames) and 1.1 (West Midlands).

Table 16.14 TB notification rates* 1985-1995 (per million population per year) for males and females ranked by male:female ratio

Region	Males	Females	M:F ratio
South & West	4.8	3.3	1.5
Anglia & Oxford	8.3	5.7	1.5
North Thames	25.0	17.6	1.4
North West	10.4	7.5	1.4
South Thames	10.1	7.3	1.4
Northern & Yorkshire	9.9	7.3	1.4
Trent	10.7	8.8	1.2
West Midlands	11.9	10.7	1.1

*Notification rates were calculated by dividing number of notifications by 11 and using 1991 mid year population estimate as denominator.

While the rates of GP consultations for males were only slightly higher than for females, death rates and emergency hospital admissions rates were twice as high in males. A possible explanation for this is that complications for tuberculosis are more common in males than females, perhaps related to homelessness.[51] Further analysis of elective admissions and outpatient attendances would be needed to examine this.

Seasonality

Unlike a previous analysis of notifications[52] we did not find a summer peak in rates, even in hospital admissions with the largest number of events (Figure 11.2). This may be to the difference between onset (incidence) and severity as indicated by use of medical care by prevalent cases. Other factors include small numbers in our data obscuring the pattern – the analysis of notifications used 10 as opposed to five or four years of data. Also, we used emergency hospital admissions which only accounts for two-thirds of tuberculosis admissions; it may have been more appropriate to use all admissions. Alternatively it may suggest differential notification in the summer, perhaps because doctors are less busy with respiratory illnesses in the summer[50] and have more time to complete administrative tasks such as notification.

Adjustments made for ethnicity and deprivation

Our analyses did not adjust for ethnic and social class differences in tuberculosis. Patients from non-white ethnic groups contributed 56% of tuberculosis notifications in 1993, but constituted only 6.3% of the total population, with the highest rates seen persons from the Indian Subcontinent who had immigrated within the last five years.[53] However, some studies have suggested that living in a deprived, inner city area may be a more important risk factor than ethnic origin.[54,55]

Which data source for investigations of environmental influences?

Notifications are the most widely used data source for examining geographical variations in tuberculosis.[51] Undernotification of tuberculosis is acknowledged to be common and in one study investigating notifications in 1985-9 in two hospitals in East London, this varied by speciality of clinician in charge from 17% to 82%.[56] It is possible that the degree of undernotification has decreased in recent years, with the ongoing influence of consultants in communicable disease control (first created in 1988)[51] and with local audits specifically addressing this problem.[57,58]

The comparisons above (Table 16.13) and the good level of consistency between routine data sources for tuberculosis suggest that hospital admissions could be used as an alternative to notifications for simple rank comparisons between geographical areas. However, the comparisons raised the possibility of variation in undernotification at a regional level, which warrants further investigation as it raises questions about the validity of using of notifications as an indicator of regional differences in tuberculosis incidence. The comparisons of routine data also raised the possibility of differences in severity by region, which also warrant further study.

16.12 Sarcoidosis

Numbers of events in all three datasets were small and limited the scope for detailed exploration or geographical correlation.

Age and sex pattern

A recent Thorax supplement on diffuse parenchymal lung disease (predominantly sarcoidosis and idiopathic fibrosing alveolitis) commented that the epidemiological data were sparse.[59] We found that there was a predominance of GP consultations and emergency hospital admissions in mid life, while most deaths occurred in later life (ages 55-85). However, the unadjusted mortality rates in later life were numerically close to the emergency hospital admission rates. This could reflect improved diagnosis (due to increased awareness and improving access to sophisticated investigations such as wider availability of CT scanning) concentrated in younger individuals which was detecting milder disease. The most likely presenting symptoms of sarcoidosis are cough and shortness of breath[59] which may be less likely to be aggressively investigated in the elderly or, in the case of shortness of breath, less apparent as the elderly generally have lower levels of physical activity. We found no evidence of a cohort effect in hospital admissions and deaths, which would support this, but the improvement in diagnosis may have occurred gradually and we had limited power to detect an effect with only five years of data.

Male rates were higher than female rates until approximately 50 years, when female rates became higher than males until ages 80+ when rates fluctuated due to small numbers. Possible influences include the menopause and female smoking patterns (patients with sarcoidosis are less likely to be smokers than the general population[60]).

'Emergency' hospital admissions

Admission to hospital as an emergency is not a common mode of presentation for sarcoidosis, where 72% of admissions were elective (Table 5.2). However, 61% of these admissions were probable day cases (Table 5.3) which makes it likely that these were conducted for diagnostic investigations such as bronchoscopy.[59]

Regional differences

Small numbers limited the ability to detect statistically significant differences between regions. Urban rural patterns were inconsistent, with highest GP patient consultations in rural areas and highest emergency hospital admissions in conurbations, perhaps reflecting access to care and to specialist centres. In the past, space-time clustering has been suggested, but this would be at a smaller scale than district or region.[61]

Which data source for investigations of environmental influences?

Small numbers of events in all data sources limit the scope of their use as indicators of the disease burden of sarcoidosis at regional level unless several years of data are combined. Even with combined years, numbers are likely to be too small to permit stable estimates of the disease burden at district level. Emergency hospital admissions may be the most suitable indicator of the *severity* of disease at regional level (showing a relationship with mortality, but not GP consultations). However, since emergency hospital admissions constituted only 26% of all admissions for sarcoidosis, and as the bulk of the elective admissions are probably for diagnosis (as discussed above), all hospital admissions may better reflect the regional *prevalence* of the disease.

16.13 Idiopathic fibrosing alveolitis

Our findings were consistent with the known epidemiology of the disease.[62] Although sarcoidosis and idiopathic fibrosing alveolitis are both classed as diffuse parenchymal lung disease,[16] our analysis suggested that their epidemiology is different. While the highest rates for sarcoidosis were seen in mid life, the highest rates for idiopathic fibrosing alveolitis were seen in the elderly. Peak rates in men were double those in females which may reflect unrecognised occupational exposures to fibrogenic dusts or chemicals.[63]

Consistency between data sources

The close consistency between HES and mortality for 1991-5 may be a diagnostic artefact – in regions where respiratory physicians have a particular interest in the condition, it may be more likely to appear in the death certificates. GPRD had rates an order of magnitude higher than in the other two data sources. Numbers were too small to investigate regional correlations between databases or regional patterns (as confidence intervals were too wide).

It is likely that information on in-patient and out-patient contacts for idiopathic fibrosing alveolitis would be recorded in the GPRD, even if the patient's care was solely managed by a hospital physician through regular out-patient visits. GPRD might therefore be regarded as the best measure of period prevalence of the disease, but because of small numbers, stable estimates could only be calculated at a national level or with combined years of data at a regional level.

Which data source for investigations of environmental influences?

Small numbers of events limit the scope of use of routine data sources as annual indicators of the disease burden in all data sources at district level. Scope is also limited by small numbers in the GPRD and to some extent mortality at regional level, unless several years of data are combined, but numbers of events in HES would permit relatively stable annual regional rates.

The different types of analyses we performed suggested that emergency hospital admissions were more consistent with mortality than with GP consultations and therefore would be more likely to reflect severity than prevalence of the disease. All hospital admissions could be considered as an alternative measure for prevalence of the disease (emergency hospital admissions only accounted for 53% of all hospital admissions in 1991-4), but the consistency of this measure with the other data sources would need to be investigated first.

16.14 Cystic fibrosis

Consistency with presentation and natural history of cystic fibrosis

The age sex patterns were consistent with the presentation and natural history of cystic fibrosis, with early consultation and hospital admission for diagnosis and later hospital admissions and deaths related to complications. Analysis by single years of age showed a high mortality in the first year of life and a peak in deaths in ages 15-19 in girls and 20-24 in males. An analysis of 1,405 cystic fibrosis deaths in 1977-85[64] found peak rates in the first year of life and in ages 20-24 in both males and females (taking the denominator as the total number of people with cystic fibrosis).

The analysis[64] found that meconium ileus was recorded on the death certificate for just under half of those for whom the reason for death was known in the first year of life (the ICD9 code for meconium ileus is 777.0A, but these are all coded to cystic fibrosis (ICD9 277.0) by ONS).

Approximately half of all admissions for cystic fibrosis are elective and half are emergency. It may, therefore be more appropriate to analyse cystic fibrosis using all admissions. However, it is likely that a number of patients are admitted several times during the year, so the HES data may overestimate the burden of the disease in absolute terms.

Sex differences

Higher death rates were seen in females of all age groups until the age of 20 in the analysis of cystic fibrosis previously mentioned for 1977-85[64], but our analysis for 1991-5 suggested that males had higher death rates in the first year of life and similar rates for ages 1-4. Higher emergency hospital admissions were generally also seen in females (except in the first two years of life and ages 30-34), suggesting that females continue to experience more severe disease than males. It is not clear if this reflects the natural history of the disease or care factors as male GP consultations were higher than females until the mid 20s.

Cohort effect

Although examination of only five years of data offers limited scope for cohort analyses, some suggestive evidence emerged of generation effects, with higher admission and mortality rates among persons born in the 1970s than before or after. This has not previously been reported. There was also an apparent trough in emergency hospital admission rates in those born in 1985-9. Rates were higher for those born in 1990-1995 perhaps reflecting the increased morbidity in the first year of life. This could be explained by improved care factors introduced in the mid 1970s, resulting in more infants surviving to be admitted to hospital with a further therapeutic advance in the mid 1980s. Previous analyses have commented on a progressive improvement in survival which have been attributed to earlier diagnosis, improved management of meconium ileus, better dietary management and pancreatic enzyme supplement, routine physiotherapy, anti-pseudomonas antibiotics and care in

large specialised centres. [51,64-66] Further investigation into the timing of when these were introduced is needed to see which correspond best to the observed mortality and morbidity patterns.

Time trends

Time trends were consistent with improving survival in cystic fibrosis,[65] but suggested increasing morbidity over the five year period of our study. It is possible that this was due to inadequate age adjustment.

Regional differences

There were clear regional differences in SERs that were consistent across datasets and which require further investigation. This is unlikely to reflect birth prevalence – a survey of patients with cystic fibrosis in 1977-85[64] suggested “no important differences” between regions in the UK. This could be due to severity (i.e. possible environmental effects on a susceptible group) or medical care (greater proximity in urban areas).

Use of emergency hospital admissions as an indicator for environmental effects

Although the numbers involved were small, the fact that rankings in datasets were relatively consistent across regions suggests that hospital admissions could be used as an indicator of the relative burden of disease in an area.

Cystic fibrosis patients represent a vulnerable group and for this reason are of particular interest in studies, for example of the short-term effects of air pollution. However, the number of emergency hospital admissions for cystic fibrosis is too small to permit daily time series analyses except, perhaps, on a national basis (although this might create difficulties with the choice of exposure measurements).

16.15 Pneumothorax

Use of emergency hospital admissions as an indicator for environmental effects

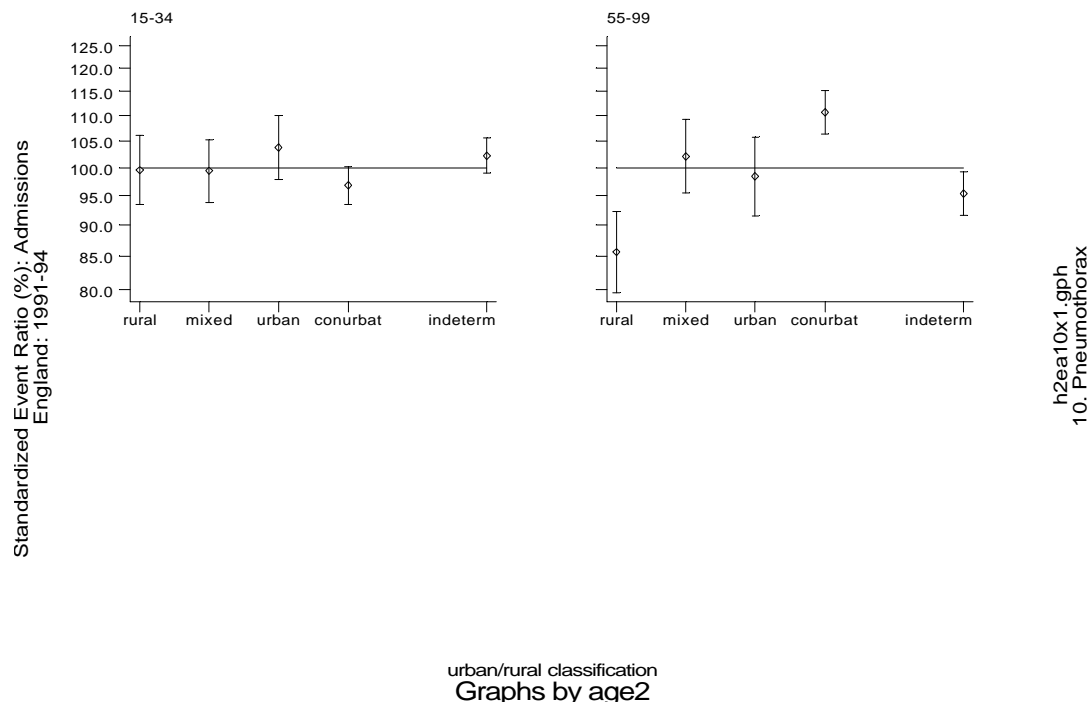
The main data source was hospital admissions as small numbers were involved in GPRD and mortality. Hospital admissions would therefore seem the best indicator for environmental effects, but may overestimate the burden of disease if the same patients are readmitted with recurrences. Readmissions for elective surgery (pleurocentesis) would not influence rates based on emergency admissions.

Age sex distribution

A bi-modal age distribution in hospital admissions was seen, with peaks in ages 20-25 and in ages 80-85. The peak in the older age group corresponds to pneumothorax as a complication of emphysema. This is supported by a similar peak at ages 80-85 in the age-sex distribution of emergency hospital admissions for COPD and a similar urban rural gradient with highest SERs in conurbations (Appendix A7). Age-specific analyses of the urban rural gradient, showed that the urban rural pattern for pneumothorax was only seen in those aged 55+ (Figure 16.8).

Pneumothorax in the younger age groups are most likely to be spontaneous. No urban rural pattern was seen in emergency hospital admissions for ages 15-34.

Figure 16.8 Urban rural pattern for emergency hospital admissions for pneumothorax for ages 15-34 and 55-99 for 1991-1994



16.16 Confounding by smoking and social class

Differences in the distribution of social class and smoking are likely to be the major confounders of any geographical comparisons after adjusting for age and sex. While social class and smoking may affect the regional distributions of disease, they would only affect correlations between data sources if they had a differential effect on severity (if, for example, there were more GP consultations per hospital admission in socially deprived areas).

Smoking and social class effects can be explored using individual level data or aggregated data at the level of postcode, GP surgery or health authority:

- Individual level data on smoking but not social class are available from the Health Survey for England and these could be linked to symptoms of asthma, COPD and hayfever or tabulated by district health authority. Individual level data on smoking and social class but not respiratory health are also available from two other surveys to standard regions subdivided into metropolitan and non-metropolitan counties (the General Household Survey) and to local authority level (the British Household Panel Study).[67]
- Individual level data on both social class and smoking[†] are available from the Fourth Morbidity Survey in General Practice.[68] The GPRD holds as yet unvalidated individual level smoking data and geographical location of practice, but not social class data. Some other GP databases such as Meditel have socio-economic data relating to the practice location, which is stored as ACORN codes, linked to each patient.
- Individual level data on smoking is not available for hospital admissions or for mortality, but ecological data on social class could be constructed from Census variables using the postcode if available, while ecological data on smoking could be constructed using one or more of the surveys mentioned above.

Adjusting for social class and smoking using individual level data

This study investigated the influence of social class and smoking on symptoms using individual level data from the HSE95. We found there was little impact on asthma symptoms, which was not surprising as there is little variation in asthma prevalence by social class,[69] although there may be a social class impact on severity as evidenced by mortality.[70] Adjustment for social class and smoking did impact on regional distribution of COPD symptoms, but did not account for all the variation.

Smoking data on patients are available from the General Practice Research Database which could be used to explore the effect of smoking on GP consultation rates for respiratory disease, but these have not been validated in published studies. A comparison with different non-smoking categories in the Health Survey for England 1995 is shown in Figures 16.9 and 16.10. Both graphs suggested under-recording of smoking habits in general practice which was most marked in teenagers. The pattern of males and females recorded as non or ex-smokers from age 20 onward in the

[†] Patients over 16 were asked “Did you smoke at all in the last seven days?” which applied to smoking any tobacco product.

GPRD was generally similar to that of never or ex-smokers from the HSE95, but approximately 15-20% lower. There was an apparent close correspondence between the age sex patterns for females recorded as non-smokers in general practice and never or never regular smokers in the HSE95. However, it is possible that the GPRD group included some ex-smokers, so the two categories were not strictly comparable. There is scope for further work in the GPRD restricted to the 'non-smoker' group to clarify non-lifestyle influences.

Figure 16.9 Non smoking categories in the GPRD in 1995 compared with the HSE95 – males, ages 15-84

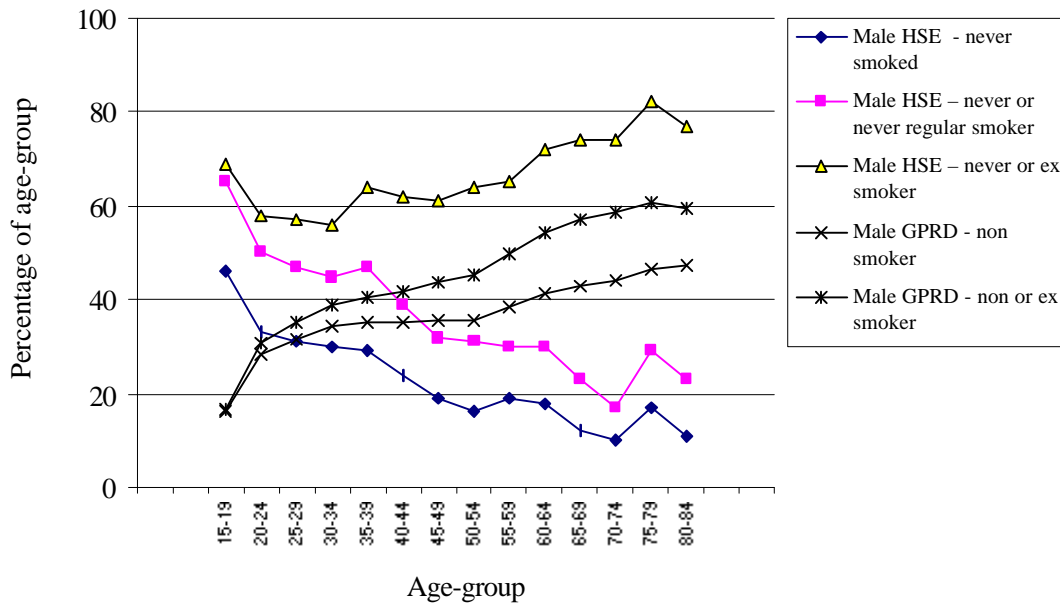
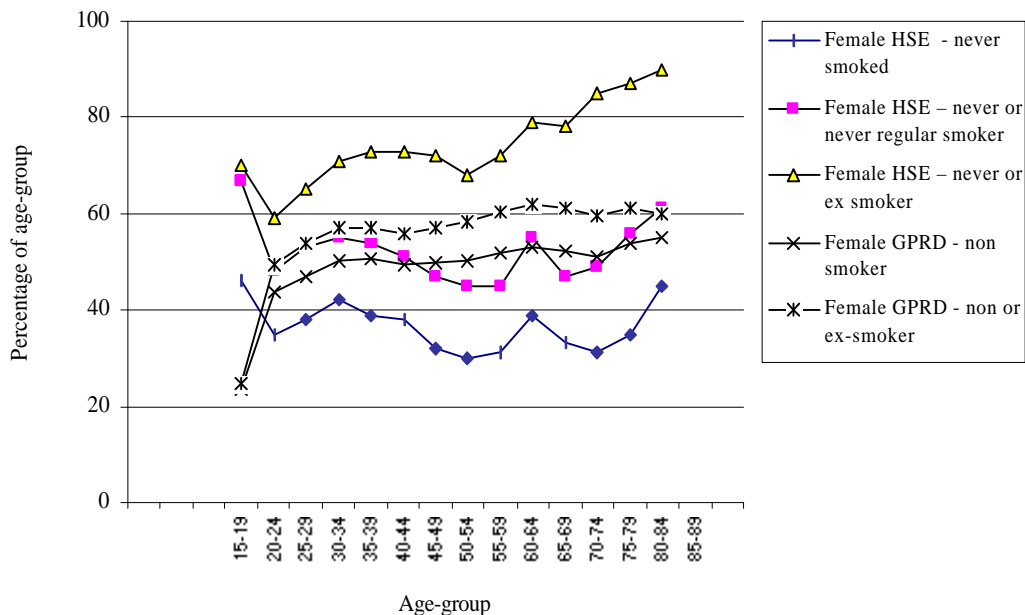


Figure 16.10 Non smoking categories in the GPRD in 1995 compared with the HSE95 – females, ages 15-84



Social class data are available for 83% of patients in the Morbidity Survey in General Practice for 1991/2. McNiece and Majeed[71] used these data to investigate the impact of social class on GP contact rates in the elderly for all conditions and found that in 65-74 years olds contact rates showed a class gradient with 23% higher contact rates in social class V than class I, in 75-84 year olds there was no clear association and in 85+ year olds contact rates were highest in social class I. Further work is needed to see if this is applicable to respiratory conditions.

Adjusting for social class and smoking using grouped data

A number of recently published ecological studies have investigated the effect of social class on GP consultations, hospital admissions and mortality, which suggest a substantial impact of social class.

Giuffrida et al[72] examined the relation at a district health authority level between age and sex standardised admission rates for asthma, diabetes and epilepsy and socio-economic characteristics from the Census together with supply of secondary care (such as number of hospital beds) and morbidity for 90 authorities over the period 1989-90 to 1994-5. They found that the 14 variables retained in the final regression analyses explained 45% of the variation in admission rates for asthma at a district health authority level. No comments were made about the quality of the hospital admissions data such as missing diagnoses, which although unlikely to vary with the explanatory variables, may have diminished the size of the association through random variations.

Reid et al[73] examined the relationship between hospital admission rates for all diseases and general practice characteristics in Merton, Sutton and Wandsworth Health Authority in South West London. They found that census derived sociodemographic variables accounted for 42% of the overall variation in admission rates between practices (45% for emergency admissions, 25% for elective admissions). They also found a strong positive correlation between deprivation and emergency but not elective admission rates, which was interpreted as raising questions about the quality of care. This is consistent with work by Griffiths et al[37] in East London, who found census derived socio-economic variables were significantly associated with admission rates for asthma in univariate analyses but not in multivariate analyses after inclusion of the size of the GP practice.

A number of studies have shown a relationship between area level deprivation based census characteristics and mortality for both all cause and disease specific mortality, especially in those aged under 65 years.[74-76] Similar area level associations have also been demonstrated with income inequality.[77]

What does adjustment for socio-economic factors adjust for?

It is difficult to know exactly what is included in socio-economic adjustments. It may include access to services, use of services, lifestyle factors such as smoking or diet. However, it is possible that some environmental factors are included, as poorer groups may be concentrated in areas with a poorer quality of environment, such as close to heavy industry.

16.17 Scope of routine data to explore environmental influences on disease

Practical problems

There are a number of practical problems in using routine data to explore the geographical distribution of disease particularly over time, which can be summarised as:

- (i) **Boundary changes.** Information is often required by administrative boundaries, but these are prone to frequent change. Postcoded or small area data are needed to aggregate data to the required boundary, but this is rarely released to individual researchers because of concerns about patient confidentiality, which have intensified in recent years. [78] Facilities need to be in place to allow researchers access to data aggregated to specified boundaries (for example, five years of data to 1994 boundaries).
- (ii) **Data quality.** Particularly in HES, this was seen to be variable by region and by district health authority, which may be important in certain types of studies.
- (iii) **Regional differences in clinical practice and clinical coding, not captured on routine quality reports.** There may be regional differences in willingness to diagnose, recording and coding clinical.
- (iv) **Comparability of data sources.** Different data sources use different clinical coding system or versions. HES is collected by finished consultant episodes rather than by admissions.
- (v) **Small numbers.** These may limit analyses, particularly if age-specific rates are examined using single years of data or small geographical areas such as district health authority.

Circumstances in which investigation of environmental influences on respiratory disease might be conducted

Generally, routine data might be sought to investigate environmental influences on disease at a district health authority in two circumstances:

- (i) **Rates are higher than average:** Information might come from surveillance or performance indicators based on routine data which show statistically significant higher rates, or from anecdotal reports
- (ii) **A suspected or known environmental hazard exists locally**

The course of the investigation will be determined by the data sources appropriate and the level of consistency between them (as detailed in Table 16.15).

Table 16.15 Suggested routine data sources and degree of consistency between them by disease

Disease	Sufficient nos* for district rankings	Sufficient nos† for regional rankings	Consistency of regional rankings‡
<i>Common diseases</i>			
Asthma	HES GPRD	Mortality HES GPRD HSE95	Weak geographical correlations across data sources
Acute bronchitis or bronchiolitis	HES GPRD	HES GPRD	Moderately good geographical correlation between GPRD and HES
COPD	Mortality HES GPRD	Mortality HES GPRD HSE95	Good geographical correlations between data sources
Hayfever	GPRD	GPRD HSE95	Weak geographical correlation between symptoms and GP prescriptions for hayfever
Pneumonia	Mortality HES	Mortality HES	Moderately good positive correlations between HES and mortality
Rarer diseases			
Cystic fibrosis	-	HES	Good consistency of regional rankings across data sources
Idiopathic fibrosing alveolitis	-	Mortality HES	Moderate consistency across data sources of regional rankings
Pneumothorax	-	HES	Could not be assessed due to small numbers even in combined years
Sarcoidosis	-	-	Moderate consistency across data sources of regional rankings
Tuberculosis	-	HES Notifications	Good consistency of rankings between HES and mortality. Moderate consistency of GPRD with HES and mortality Good consistency of notifications with HES and mortality. Moderate consistency with GPRD.

* at least 100 events per average district, based on observed number of events in 1994

† total of at least 800 events, based on observed number of events in 1994

‡ based on one year of data for common diseases, based on several years data for rarer diseases

(i) Rates from a data source are higher than average

Where there is marked inconsistency between routine data sources, such as for asthma, high rates of hospital admissions locally cannot be assumed to correspond to high prevalence or high mortality rates. Factors such as the threshold for hospital admission, geographical proximity to hospital and quality of and access to primary care are more likely to explain the inconsistency between data sources. Adjustment for local levels of smoking and socio-economic factors may be conducted in further analyses, but as discussed above, adjusting for socio-economic factors may also partially adjust for environmental influences or for lifestyle factors such as diet.

For diseases where routine data sources are consistent across data sources, it is reasonable to assume that a common factor is involved. For example, COPD has higher rates in Northern areas of England. Known confounders such as smoking need to be adjusted for. If differences persist, this may be due to environmental and/or social class factors. Adjusting for social class can be performed, but may partially adjust for environmental influences. If differences remain, it would be reasonable to suggest that environmental influences may be responsible. However, it may not be clear whether current environmental exposures or prior exposures (such as in childhood or previous years) are more important and further work is needed to clarify this.

(ii) There is a known or suspected environmental hazard

Where data are clearly inconsistent, such as asthma, the data source most clearly related to question posed needs to be used. For example, asthma severity might be better assessed using hospital admissions, while prevalence might be better assessed by survey data on symptoms. Where data are clearly consistent, such as COPD, any data source could be used to estimate the impact of environmental influences.

16.12.3 Conclusions

Routine data can certainly be used to give information about the patterns of disease at a regional level. However, there are a number of practical problems which limit their use at the level of individual districts. Having allowed for these problems, the extent to which routine data can be used to investigate environmental effects depends on the disease concerned. In particular, asthma shows striking inconsistency between routine data sources, which questions the appropriateness of using routine data for asthma as a marker for environmental effects. In contrast, routine data for COPD show consistently higher rates in the North of England and in urban areas, which persist after adjusting for smoking and social class and which warrant further investigation in relation to current and past environmental exposures.

References

1. Toren K, Brisman J, Jarvholm B. Asthma and asthma-like symptoms in adults assessed by questionnaires. A literature review. *Chest* 1994;104(2):600-8.
2. Hansell A, Hollowell J, Nichols T, McNiece R, Strachan DP. Use of the General Practice Research Database (GPRD) for respiratory epidemiology: a comparison with the 4th Morbidity Survey in General Practice (MSGP4). *Thorax* 1999;54(5):413-9.
3. Beardon PHG, McGilchrist MM, McKendrick AD, McDevitt DG, MacDonald TM. Primary non-compliance with prescribed medication in primary care. *BMJ* 1993;307:846-8.
4. Majeed A, Moser K. Prescribing for patients with asthma by general practitioners in England and Wales 1994-96. *Health Statistics Quarterly* 1999;01(Spring):16-9.
5. Moser K, Majeed A. Prevalence of treated chronic diseases in general practice in England and Wales - trends over time and variations by the ONS areas classification. *Health Statistics Quarterly* 1999;02:25-32.
6. Key health statistics in General Practice. London, UK: The Stationery Office; 1996.
7. Hollowell J. General Practice Research Database (GPRD) scope and quality of data. London, UK: OPCS; 1994.
8. Roberts SJ, Bateman DN. Which patients are prescribed inhaled anti-asthma drugs? *Thorax* 1994;49(1090):1095
9. Ponce de Leon A, Anderson HR, Bland JM, Strachan DP, Bower J. Effects of air pollution on daily hospital admissions for respiratory disease in London between 1987-88 and 1991-92. *Journal of Epidemiology & Community Health* 1996;50(Suppl 1):S63-S70
10. Newson R, Strachan DP, Archibald E, Emberlin J, Hardaker P, Collier C. Effect of thunderstorms and airborne grass pollen on the incidence of acute asthma in England, 1990-94. *Thorax* 1997;52:680-5.
11. Elliott P, Martuzzi M, Shaddick G. Spatial statistical methods in environmental epidemiology: a critique. *Statistical methods in Medical Research* 1995;4:137-59.
12. Jarman B, Gault S, Alves B, Hider A, Dolan S, Cook A, Hurwitz B, Iezzoni LI. Explaining differences in English hospital death rates using routinely collected data. *BMJ* 1999;318(7197):1515-20.
13. Use of hospital admissions data as a measure of morbidity for residents near cokeworks. *International Society for Environmental Epidemiology and International Society for Exposure Assessment*. 99 A.D. Sep; 1999;
14. Wilkinson P, Elliott P, Grundy C, et al. Case-control study of hospital admission with asthma in children aged 5-14 years: relation with road traffic in North West London. [In Press] *Thorax* 1999;
15. Strachan DP, Anderson HR, Limb ES, O'Neill A, Wells N. A national survey of asthma prevalence, severity and treatment in Great Britain. *Archives of Disease in Childhood* 1994;70:174-8.
16. Guite HF, Burney PGJ. Accuracy of recording of deaths from asthma in the UK: the false negative rate. *Thorax* 1996;51:924-8.
17. A subcommittee of the BTA research committee. Accuracy of death certificates in bronchial asthma. *Thorax* 1984;39:505-9.
18. Smyth ET, Wright SC, Evans AE, Sinnamon DG, MacMahon J. Death from airways obstruction: accuracy of certification in Northern Ireland. *Thorax* 1995;51:293-7.
19. Sears MR, Rea HR, Rothwell RPG, O'Donnell TV, Holst PE, Gillies AJD, Beaglehole R. Asthma mortality: comparison between New Zealand and England. *BMJ* 1986;293:1342-5.
20. Hunt LW, Silverstein MD, Reed CE, O'Connell EJ, O'Fallon WM, Yunginger JW. Accuracy of the death certificate in a population-based study of asthmatic patients. *JAMA* 1993;269(15):1947-52.

21. Johnston I, Britton J, Kinnear W, Logan R. Rising mortality from cryptogenic fibrosing alveolitis. *BMJ* 1990;301(6759):1017-21.
22. Devis T, Rooney C. Death certification and the epidemiologist. *Health Statistics Quarterly* 1999;01(Spring):25-33.
23. NHS Executive. Health Authority and Local Government Boundaries. *Health Services Guidelines HSG(96)* 1996;50(6 September)
24. Carr-Hill RA, Sheldon TA, Smith P, Martin S, Peacock S, Hardman G. Allocating resources to health authorities: development of method for small area analysis of use of inpatient services. *BMJ* 1994;309:1046-9.
25. Denham C, White I. Differences in urban and rural Britain. *Population Trends* 1998;91(Spring):23-34.
26. Wallace M; Denham C. The ONS classification of local and health authorities of Great Britain. London, UK: HMSO; 1996.
27. Social and Community Planning Research (SCPR), editor. *Health Survey for England 1994-1996*. London, UK: Department of Health; 1999.
28. von Mutius E. Progression of allergy and asthma through childhood to adolescence. *Thorax* 1996;51(Suppl 1):S3-S6
29. Martinez FD, Wright AL, Taussig LM, Holberg CJ, Halonen M, Morgan WJ, The Group Health Medical Associates. Asthma and Wheezing in the First Six Years of Life. *New England Journal of Medicine* 1995;332:133-1381111.
30. Hill J, Woodfolk JA, Chapman MD, Heymann PW. Changing concepts of allergic disease: the attempt to keep up with real changes in lifestyles. *Journal of Allergy & Clinical Immunology* 1996;98:S297-S306
31. Roorda RJ. Prognostic factors for the outcome of childhood asthma in adolescence. *Thorax* 1996;51(Suppl 1):S7-S12
32. Anderson HR, Pottier AC, Strachan DP. Asthma from birth to age 23: incidence and relation to prior and concurrent atopic disease. *Thorax* 1992;47(7):537-42.
33. Venn A, Lewis S, Cooper M, Hill J, Britton J. Questionnaire study of effect of sex and age on the prevalence of wheeze and asthma in adolescence. *BMJ* 1998;316:1945-6.
34. Kaur B, Anderson HR, Austin J, Burr M, Harkins LS, Strachan DP, Warner JO. Prevalence of asthma symptoms, diagnosis and treatment in 12-14 children across Great Britain (international study of asthma and allergies in childhood, ISAAC UK). *BMJ* 1998;316:118-24.
35. von Mutius E, Martinez FD, Fritzsche C, Nicolai T, Reitmeir P, Thiemann H-H. Skin test reactivity and number of siblings. *BMJ* 1994;308(12 March):692-5.
36. Sibbald B, Kerry S, Strachan DP, Anderson HR. Patient characteristics associated with the labelling of asthma. *Family Practice* 1994;11(2):127-32.
37. Griffiths C, Sturdy P, Naish J, Omar R, Dolan S, Feder G. Hospital admissions for asthma in east London: associations with characteristics of local general practices, prescribing and population. *BMJ* 1997;314:482-6.
38. Fleming DM. Weekly Returns Service of the Royal College of General Practitioners. *Communicable Disease and Public Health* 1999;2(2):96-100.
39. LAIA (Lung & Asthma Information Agency). Seasonal variations in asthma. *LAIA Factsheet* 1993;93/4
40. Woodward RH; Goldsmith PL. Cumulative sum techniques. Edinburgh, Scotland: Oliver and Boyd; 1964.
41. Storr J, Lenney W. School holidays and admissions with asthma. *Archives of Disease in Childhood* 1989;64:103-7.
42. Johnston SL, Pattemore PK, Sanderson G, et al. Community study of the role of viral infections in exacerbations of asthma in 9-11 year old children. *BMJ* 1995;310:1225-9.
43. Department of Health. Report of the Scientific Committee on Tobacco and Health. London, UK: The Stationery Office; 1998; Part one, The Scale of the Smoking Problem. p. 17-26.

44. Strachan DP, Cox BD, Erzinclioglu SW, Walters DE, Whichelow MJ. Ventilatory function and winter fresh fruit consumption in a random sample of British adults. *Thorax* 1991;46(9):624-9.
45. Lee PN, Fry JS, Forey BA. Trends in lung cancer, chronic obstructive lung disease and emphysema death rates for England and Wales 1941-85 and their relation to trends in cigarette smoking. *Thorax* 1990;45(9):657-65.
46. Marks G, Burney PGJ, Charlton J, Murphy M, editors. *The Health of Adult Britain 1841-1994*. London, UK: The Stationery Office; 1997; 20, Diseases of the respiratory system.
47. Strachan DP, Calverley P, Pride N, editors. *Chronic Obstructive Pulmonary Disease*. 1 ed. London, UK: Chapman & Hall; 1995; 4, Epidemiology: A British perspective. p. 47-67.
48. Strachan DP. Epidemiology of hay fever: towards a community diagnosis. *Clinical & Experimental Allergy* 1995;25:296-303.
49. Ishizaki T, Koizumi K, Ikemori R, Ishiyama Y, Kushibiki E. Studies of prevalence of Japanese cedar pollinosis among the residents in a densely cultivated area. *Annals of Allergy* 1987;58(April):265-70.
50. Office for National Statistics. *Infectious disease monitor. Series MB2 Number 22*. ONS. 1996.
51. Darbyshire JH. Tuberculosis: old reasons for a new increase. *BMJ* 1995;310:954-5.
52. Douglas AS, Strachan DP, Maxwell JD. Seasonality of tuberculosis: the reverse of other respiratory diseases in the UK. *Thorax* 1996;51:944-6.
53. Omerod LP, Charlett A, Gilham C, Darbyshire JH, Watson J. Geographical distribution of tuberculosis notifications in national surveys of England and Wales in 1988 and 1993: report of the Public Health Laboratory Service/British Thoracic Society/Department of Health Collaborative Group. *Thorax* 1998;53:176-81.
54. Lavender M, Black NMI. Residence in an inner city is more important than ethnic origin. *BMJ* 1995;311(6998):187-8.
55. Bhatti N, Law MR, Morris JK, Halliday R, Moore-Gillon J. Increasing incidence of tuberculosis in England and Wales: a study of the likely causes. *BMJ* 1995;310:967-9.
56. Sheldon CD, King K, Cock H, Wilkinson P, Barnes NC. Notification of tuberculosis: how many cases are never reported? *Thorax* 1992;47:1015-8.
57. Pym AS, Churchill DR, Coker RJ, Gleissberg V. Reasons for increased incidence of tuberculosis. *BMJ* 1995;311:570-.
58. Brown JS, Wells F, Duckworth G, Paul EA, Barnes NC. Improving notification rates for tuberculosis. *BMJ* 1995;310:974-.
59. British Thoracic Society SoCC. The diagnosis, assessment and treatment of diffuse parenchymal lung disease in adults. *Thorax* 1999;54 (Supplement):S1-S30
60. Valeyre D, Soler P, Clerici C, Pré J, Battesti J, Georges R, Hance AJ. Smoking and pulmonary sarcoidosis: effect of cigarette smoking on prevalence, clinical manifestations, alveolitis and evolution of the disease. *Thorax* 1988;43:516-24.
61. Hills SE, Parkes SA, Baker SBd. Epidemiology of sarcoidosis in the Isle of Man - 2:- Evidence for space-time clustering. *Thorax* 1987;42(427):430
62. LAIA (Lung & Asthma Information Agency). *Cryptogenic fibrosing alveolitis*. LAIA Factsheet 1997;97/1
63. Hubbard R, Lewis S, Richards K, Johnston I, Britton J. Occupational exposure to metal or wood dust and aetiology of cryptogenic fibrosing alveolitis. *Lancet* 1996;347:284-9.
64. British Paediatric Association Working Party on Cystic Fibrosis. Cystic fibrosis in the United Kingdom 1977-85: an improving picture. *BMJ* 1988;297(17 December):1599-602.
65. Elborn JS, Shale DJ, Britton JR. Cystic fibrosis: current survival and population estimates to the year 2000. *Thorax* 1991;46:881-5.
66. Phelan P, Hey E. Cystic fibrosis mortality in England and Wales and in Victoria, Australia 1976-80. *Archives of Disease in Childhood* 1984;59:71-83.

67. Twigg L. Choosing a national survey to investigate smoking behaviour: making comparisons between the general Household Survey, the British Household Panel Survey and the Health Survey for England. *Journal of Public Health Medicine* 1999;(21):1-14.
68. McCormick A; Fleming DM; Charlton J. Morbidity Statistics from General Practice. Fourth national study 1991-1992. London, UK: HMSO; 1995.
69. Anderson HR, Britton J, Esmail A, et al. Botting B, editors. The health of our children Decennial supplement. HMSO; 1995; 9, Respiratory disease and sudden infant death syndrome. p. 113-29.
70. Department of Health. Asthma An Epidemiological Overview. London, UK: HMSO; 1995.
71. McNiece R, Majeed A. Socioeconomic differences in general practice consultation rates in patients aged 65 and over: prospective cohort study. *BMJ* 1999;319(7201):26-8.
72. Giuffrida A, Gravelle H, Roland M. Measuring quality of care with routine data: avoiding confusion between performance indicators and health outcomes. *BMJ* 1999;319(7202):94-8.
73. Reid FDA, Cook DG, Majeed A. Explaining variation in hospital admission rates between general practices: cross sectional study. *BMJ* 1999;319(7202):98-103.
74. Carstairs V. Deprivation indices: their interpretation and use in relation to health. *Journal of Epidemiology & Community Health* 1995;49(Suppl 2):S3-S8
75. Eames M, Ben-Shlomo T, Marmot MG. Social deprivation and premature mortality: regional comparisons across England. *BMJ* 1993;307(6912):1097-102.
76. Drever F, Whitehead M. Mortality in regions and local authority districts in the 1990s: exploring the relationship with deprivation. *Population Trends* 1995;82(Winter):19-26.
77. Stanistreet D, Scott-Samuel A, Bellis MA. Income inequality and mortality in England. *Journal of Public Health Medicine* 1999;21(2):205-7.
78. NHS Executive. The Caldicott Committee: Report on the review of patient-identifiable information. London, UK: Department of Health; 1997.