PERGAMON

Health & Place 9 (2003) 139–149

HEALTH & PLACE

www.elsevier.com/locate/healthplace

Predicting smoking behaviour in census output areas across Scotland

Jamie Pearce*, Paul Boyle, Robin Flowerdew

School of Geography and Geosciences, University of St Andrews, St Andrews, Fife KY16 9AL, Scotland UK

Received 10 April 2001; received in revised form 30 May 2002; accepted 14 July 2002

Abstract

Spatially disaggregated surveys of smoking behaviour are rare and hence estimating the geography of the incidence of smoking is difficult. The main aim of this study is to develop a technique for estimating smoking probability for different age/sex groups in small areas across the whole of Scotland using information on smoking behaviour from the Scottish Household Survey. This is useful not only in its own right, but as an aid to studies of geographical variations in diseases such as lung cancer that, as a first step, need to control for smoking behaviour. The method developed uses individual-level characteristics from the Scottish Household Survey combined with a set of output area and pseudo-postcode sector measures from the 1991 census to model the probability of smoking. The parameters from this model are then used to make smoking predictions by age and sex for output areas across Scotland. This is the first time that such geographically detailed estimates of smoking have been made available.

Keywords: Smoking; Scotland; Logit models; Scottish Household Survey; 1991 British Census

Introduction

This study is part of a larger project on the geography of lung cancer incidence across Scotland. Of the known causes of lung cancer smoking is the most significant. It is therefore important to be able to estimate the probability of smoking among different groups of the population as a first step in investigating other causes of lung cancer. For example ecological studies that consider the influence of radon upon the incidence of lung cancer must control for smoking (Stidley and Samet, 1993). This paper therefore presents a method that can be used to estimate smoking probabilities for different age and sex groups in small areas across Scotland.

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

*Corresponding author. Tel.: +44-1334-462-811; fax: +44-1334-463-949.

E-mail address: jrp6@st-andrews.ac.uk (J. Pearce).

1353-8292/03/$ - see front matter © 2003 Elsevier Science Ltd. All rights reserved.
PII: S 1 3 5 3 - 8 2 9 2 (0 2) 0 0 0 3 6 - 9

The aim of this study is to make predictions about the probability of smoking among people in different age/sex groups in small geographical areas across Scotland. Such geographical estimates are valuable, as they enable us to control for smoking behaviour in epidemiological and ecological studies of various diseases. In particular, they are useful for studies of the effects of factors other than smoking on the incidence of lung cancer. Given that these estimates need to be reasonably reliable, it is important that the model used is as accurate as possible.

In addition to data at the individual level, we have used small-area data to assist in the modelling process. A number of studies have demonstrated the importance of both individual and contextual influences on health-related behaviour (e.g. Macintyre et al., 1993; Phillimore, 1993; Jones and Duncan, 1995). Twigg et al. (2000) demonstrate the importance of using both individual and contextual factors to predict the probability of smoking in England.

Related work includes that of Barnett (2000), who focused on the contextual factors affecting smoking rates in Christchurch, New Zealand. He showed that smoking levels vary according to the level of deprivation and degree of social mixing in an area with deprived populations having the highest smoking rates. Diehr et al. (1993) found that the prevalence of smoking varied significantly between 15 different communities in the Western United States, even once they had controlled for demographic, health status, and health behaviour characteristics of the people in those communities. Duncan et al. (1999) used multi-level modelling techniques and found that neighbourhood deprivation had an independent effect on individual smoking behaviour. Each of these studies suggests that the characteristics of local areas are an important determinant of smoking behaviour.

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

We have attempted a similar analysis for Scotland, although for 1991 Census output areas (OAs)[1], rather than larger units such as postcode sectors or wards. There were 38,254 OAs in Scotland in 1991. The results are produced for 12 age/sex groups. In addition to variables at the individual level, Output Area composition variables were used, as well as two variables measured for larger areal units: ward type (as defined by the ONS ward classification) and the Carstairs index of deprivation.

[1] Output Areas in Scotland are equivalent to Enumeration Districts in England and Wales.

Data and methods

This study uses a mixture of individual-level survey and area-level census data to make predictions about the probability of smoking among 12 age/sex groups in Scotland. In an ideal situation, individual-level data would be available for small geographical areas, both for smoking behaviour and other individual characteristics as well as for those local area characteristics expected to influence smoking. However, much of the information collected in surveys is considered to be confidential and is therefore not released below broad geographical regions. In the absence of individual-level information, it is often necessary to undertake ecological analysis, involving the use of comprehensive aggregate data sets over wider geographical areas. In this study the available information included a small number of individual-level variables from the Scottish Household Survey (SHS) and a larger number of area-level variables from the 1991 British Census. Below we describe these data and the multi-level modelling approach we used to estimate smoking behaviour across Scotland.

Scottish Household Survey

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

The SHS includes individuals from a diverse set of geographical areas using the Postcode Address File (PAF) as its sampling frame. One of two sample structures was used depending on the population density of the local authority. If the population density was 500 or more persons per square kilometre, a systematic random approach was adopted, where a random sample of addresses were drawn from each stratum of a geodemographic classification system called the Scottish MOSAIC. If the population of the local authority was less than 500 per square kilometre then the interviews were clustered in primary sampling units (PSUs). The

SHS aimed to conduct eleven interviews per PSU and to have a minimum of 50 PSUs within each of the local authorities (Anderson and Hope, 2000).

Our study used data from the first year of the survey, and includes records on 13,784 individuals over the age of 16, distributed across Scotland. Usually this information is disseminated for broad geographical areas such as local authorities or for population subgroups, but we required more geographically detailed information at the OA level. In order to maintain the confidentiality of each survey respondent, only a few individual-level variables were provided. These were age, sex, whether the person smoked and the OA in which he or she lived. The ages of the individuals were recoded into six categories that are commonly used in reporting results from the 1991 census (16–24, 25–34, 35–44, 45–54, 55–64 and 65 and above) providing 12 age/sex groups for which the probabilities of smoking could be calculated (Table 1).

These probabilities are consistent with trends that have been observed in other studies (e.g. Uitenbroek and McQueen, 1993). The probabilities are higher for younger women (especially ages 16–24) than for younger men. In contrast, the rates are lower among older women compared to men in the same age group, reflecting historical gender differences in smoking behaviour. It was not possible to obtain any other variables at the individual level. However, because we know the OA in which each individual lived, it was possible to add area-based information from the census to improve our model of geographical variations in smoking behaviour.

Census data

A number of 1991 census variables were extracted at the OA level and were used as additional explanatory variables in order to predict the probability of a respondent being a smoker. The 13,784 individuals were distributed across 7,127 of the 38,254 Scottish OAs, which are the smallest set of units for which data are available in the Scottish census (containing on average approximately 50 households). A selection of census and census-derived variables were chosen at the OA level to represent the characteristics of either individuals or households resident in the OA. In addition, two variables were extracted for pseudo-postcode sectors (PPSs) (Table 2). PPSs are neat aggregations of OAs and there are 1,003 PPSs in Scotland. Because of their size

Table 1
Probability of smoking using the raw data from the Scottish Household Survey

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

Table 2
Census variables used in the logit model

Person variables (*OA level*)
% unemployed[a]
% of males unemployed[b]
% of females unemployed[b]
% of unemployed individuals in 9 age groups[b]

% of individuals working in managerial and administrative occupations[a]
% of individuals working in professional occupations[a]
% of individuals working in associated professional occupations[a]
% of individuals working in clerical and secretarial occupations[a]
% of individuals working in craft and related occupations[a]
% of individuals working in personal and protective occupations[a]
% of individuals working in sales occupations[a]
% of individuals working in plant and machine operative occupations[a]
% of individuals working in other occupations[a]

% of individuals with further education qualifications[a]

% of individuals who were widowed or divorced[a]

% of residents in households with an economically active head of household in a lower social class group (IV or V)[a]

Household variables (*OA level*)
% owner occupied households[a]
% privately rented households[a]
% of households renting from a housing association or local authority[a] (HA/LA)

% of one person households[a]
% of households with dependent children[a]
% of households with a lone parent[a]
% of households without a car[a]
% of overcrowded households (1 or more persons per room)[a]

Population density (a surrogate urban/rural indicator)[a]

Place variables (*PPS level*)
ONS ward classification
Carstairs index of deprivation[a]

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

these areas may better reflect the likely 'daily activity spaces' of people providing an indication of the 'place' in which people live. OAs, on the other hand, are much smaller and homogeneous, providing a good indication of the type of person who lives in a place.

The variables used in the modelling procedure were chosen to provide the best possible estimates of smoking. Here we control for the individual's age and sex, and the characteristics of individuals and households of those living in the same OA, before testing the two PPS-level variables for significance.

The characteristics of people within OAs used in the model included percentage information on unemployment, occupation, qualifications, marital status and social class. Where a percentage variable had a skewed distribution, logarithms were used. The variable was then used either in logged or unlogged form, according to which was most significant in the model. In addition to these variables both age-specific and sex-specific alternatives for each of the variables were also tested for the relevant age/sex group. For example, instead of using total unemployment the percentage of unemployed males or females in the relevant age group was attached to the individual data. Household characteristics were also included, reflecting tenure, the number of people living in the household, car ownership and overcrowding. Again, if the variables had a skewed distribution, logs were taken and compared. Population density was also included at the OA level as a proxy for the urban/rural character of a place; it has been shown elsewhere that there are urban/rural differences in smoking behaviour (e.g. Twigg et al., 1998).

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

The ONS ward classification divides all wards (or PPSs in Scotland) into a set of geodemographic classes using a number of census variables (Wallace and Denham, 1996). The classification has 14 groups, further subdivided into 43 clusters, which are intended to identify wards with similar housing and population characteristics. Here we used the 14 groups which are: suburbia; rural areas; rural fringe; industrial areas; middling Britain; prosperous areas; inner city estates; established owner-occupier; transient populations; metropolitan professionals; deprived city areas; lower status owner occupied; mature populations; and deprived industrial areas.

The study therefore utilises a wide range of potential explanatory variables. Table 3 shows that the correlation between the explanatory variables is relatively low in most cases, and so multicollinearity is not a major issue. This is even the case for the Carstairs score and the four variables that are used to construct the score.

Modelling smoking

In Scotland 30.9 percent of people aged over 16 declared themselves as smokers (Scottish Household Survey, 2000). The modelling procedure was designed to determine which individual and area level variables could best predict who was a smoker and who was not. There were two stages in the modelling procedure. First, using a multi-level model we estimated the probability of smoking for the 13,784 individuals extracted from the SHS. A number of models were examined and the optimum model was identified. The explanatory variables included those measured at the individual, OA and PPS levels. In the second stage, we used the parameter estimates from the optimum model to estimate the probability of smoking for the twelve age/sex groups in all 38,254 OAs in Scotland.

In stage one, a logistic regression model was used. This is appropriate where one wants to predict the presence or absence of an outcome (in this case smoking or non-smoking) using a set of predictor variables. A multi-level model (Goldstein, 1995) was used to allow us to examine whether variables measured at the PPS level (contextual effects) influenced smoking behaviour above and beyond the effects measured for individuals and OAs. Given our data, a three-level model would be preferable, where individuals were nested within OAs, which themselves nest within PPSs. However, the number of individuals per OA (an average of less than two) meant that it was not possible to treat OAs as the second level. Consequently, we used the individual-level variables (age and sex) along with the OA variables at level one and two PPS variables at level two.

Initially, we used a simple model that included only age and sex as explanatory variables. The aim of this model was to test whether there was significant random variation at the PPS level (level 2) before the census-based explanatory variables were included. Then, each of the remaining potential explanatory variables were tested individually, both in their raw and logged form, where applicable, and the more significant of the two was retained in the model. Interactions between age and sex and between age, sex and all other census-based variables were tested. If any of the interactions were found to be significant they were retained in the model, along with the variable involved. A stepwise approach was used to fit the model and the results are discussed in Section 3 below.

Table 3
Pearson correlations between variables

	% Unemp.	% Manage. & admin.	% Profess.	% Assoc. Profess.	% Clerical & secret.	% Craft & related	% Personal & protective	% Sales	% Plant & machine op.	% Other occup.	% Further educ. qual's	% Widowed or divorced
% Unemployed	1.00											
% Managerial & admin.	−0.23	1.00										
% Professional	−0.22	0.02	1.00									
% Associated profess.	−0.12	−0.06	0.02	1.00								
% Clerical & secretarial	−0.04	−0.13	−0.11	−0.10	1.00							
% Craft & related	0.11	−0.22	−0.21	−0.16	−0.14	1.00						
% Personal & protective	0.07	−0.15	−0.13	−0.11	−0.15	−0.12	1.00					
% Sales	0.01	−0.11	−0.10	−0.09	−0.11	−0.07	−0.09	1.00				
% Plant & machine op.	0.15	−0.20	−0.19	−0.15	−0.14	−0.10	−0.10	−0.07	1.00			
% Other occupations	0.23	−0.20	−0.20	−0.17	−0.19	−0.09	−0.09	−0.10	−0.05	1.00		
% Further educ. qual's	−0.33	0.21	0.58	0.31	−0.05	−0.24	−0.13	−0.08	−0.23	−0.26	1.00	
% Widowed or divorced	0.54	−0.18	−0.09	−0.01	−0.03	0.02	0.07	0.00	0.07	0.14	−0.15	1.00
% Social class (IV or V)	0.23	−0.25	−0.23	−0.17	−0.08	−0.01	0.14	−0.05	0.16	0.44	−0.29	0.17
% Owner occupied	−0.64	0.30	0.33	0.18	0.08	−0.15	−0.11	−0.02	−0.22	−0.32	0.50	−0.51
% Privately rented	−0.13	0.12	0.11	0.09	−0.03	−0.09	0.01	−0.05	−0.10	−0.05	0.22	0.17
% Renting from HA/LA	0.65	−0.33	−0.33	−0.19	−0.05	0.18	0.08	0.04	0.25	0.30	−0.53	0.45
% One person households	0.22	−0.08	0.00	0.02	0.03	−0.02	0.02	−0.02	0.00	0.05	−0.08	0.51
% With depend. children	0.05	0.00	−0.06	−0.04	−0.06	0.04	0.03	0.03	0.04	0.02	0.00	−0.05
% Lone parent	0.49	−0.18	−0.19	−0.10	−0.05	0.08	0.06	0.04	0.13	0.18	−0.24	0.44
% Without car	0.72	−0.32	−0.27	−0.12	0.01	0.12	0.09	0.03	0.17	0.24	−0.43	0.67
% Overcrowding	0.46	−0.19	−0.20	−0.08	−0.02	0.10	0.05	0.02	0.12	0.17	−0.25	0.39
Population density	0.28	−0.16	−0.09	−0.02	0.06	0.05	0.03	0.03	0.06	0.05	−0.11	0.32
Carstairs index	0.40	−0.22	−0.25	−0.11	−0.03	0.13	0.06	0.02	0.17	0.20	−0.35	0.29

	% Social class (IV or V)	% Owner occupied	% Privately rented	% Renting from HA/LA	% One person h'holds	% With depend. children	% Lone Parent	% Without car	% Over - crowding	Pop density	ONS ward class	Carstairs index
% Social class (IV or V)	1.00											
% Owner occupied	−0.34	1.00										
% Privately rented	−0.06	0.09	1.00									
% Renting from HA/LA	0.33	−0.93	−0.40	1.00								
% One person households	0.07	−0.28	0.21	0.23	1.00							
% With depend. children	0.01	0.02	−0.20	0.01	−0.70	1.00						
% Lone parent	0.17	−0.46	−0.12	0.47	−0.06	0.38	1.00					
% Without car	0.28	−0.78	−0.09	0.78	0.51	−0.20	0.47	1.00				
% Overcrowding	0.18	−0.43	−0.04	0.42	−0.03	0.24	0.37	0.45	1.00			
Population density	0.08	−0.26	−0.05	0.28	0.19	−0.02	0.21	0.42	0.23	1.00		
Carstairs index	0.22	−0.48	−0.10	0.48	0.10	0.03	0.32	0.49	0.33	0.16	−0.02	1.00

In the second stage of the analysis, the probability of smoking was calculated from the multi-level model parameters for each age/sex group. These probabilities could then be put together with the age/sex composition of each OA to calculate the probability of smoking for all OAs in Scotland. The model results are based on data for individuals distributed across only 7,127 OAs, but the model parameters for each age/sex group were applied to all 38,254 OAs in Scotland.

Results and discussion

The parameters of the first multi-level logistic model are shown in Table 4. The results show that the probability of smoking was lower for females compared to males, and lowest for those older than 65. The highest probability of smoking was for those aged 25–34. These results correspond with what we would expect from the raw probabilities in Table 1. The random variation at level 2 is significant, suggesting that the probability of smoking does vary contextually.

The model was then extended to include the census-based explanatory variables and a total of ten variables were significant in the final model, along with one significant interaction (Table 5). Both of the individual-level variables (age and sex) were highly significant, which is consistent with a number of previous studies (e.g. Amos et al., 1990; Uitenbroek and McQueen, 1993; Twigg et al., 2000). It was perhaps surprising that the interaction between age and sex was insignificant in this model in light of the raw probabilities in Table 1 and this suggests that the observed age/sex differences are explained by other characteristics that are accounted for in this model. Note that during the stepwise modelling procedure the interaction between age group and sex was significant initially, in line with the results in Table 1, but it became insignificant when other variables were introduced.

Table 4
Logit model with smoking as the dependent variable and age and sex as independent variables

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

Table 5
Final logit model with smoking as the dependent variable

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

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

It was perhaps surprising that some of the other variables were not significant in light of previous studies. For example, a number of studies have shown that educational attainment is associated with smoking (e.g. King et al., 1990; Shewry et al., 1992; Escobedo and Pedicord, 1996; Graham and Der, 1999) whereas in this study the percentage of individuals with further education qualifications was insignificant. Marital status has also been shown to be associated with smoking (Amos et al., 1990; Shewry et al., 1992) but the percentage of one-person households and the percentage of individuals who were widowed or divorced were insignificant here. Studies such as those by Amos et al. (1990) and Shewry et al. (1992) have also shown smoking to be related to unemployment but in this analysis the percentage unemployed was insignificant. The age-specific and sex-specific variables (for example, the percentage of unemployed males and unemployed females) were not found to be as significant as the non-specific variables (such as the overall percentage unemployed).

Note, however, that the studies cited above used individual-level data and the variables we have used are measured at the OA level, raising questions relating to the ecological fallacy. For example, we do not know whether the individuals in the SHS sample have further qualifications or not; we only know the percentage of people living in the same area who had further qualifications according to the 1991 census. It is not possible, therefore, for us to state that having further qualifications is unrelated to the probability of smoking. We can only report that the percentage of people with further qualifications living in the same OA does not have a significant effect on smoking probabilities, controlling for the other variables in the model.

The Carstairs index, measured for PPSs, was also significant in the final model. It had a positive effect upon the probability of smoking, suggesting that even when other characteristics had been controlled for, an increased level of deprivation in an area increased the probability of smoking. The ONS ward classification was not significant. The level 2 random variation that was significant in the model that included only age and sex as explanatory variables (Table 4) is now insignificant. The inclusion of the area-based variables in the model (Table 5) appears to have accounted for the random variation.

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

The results for each age/sex group from calculating the probabilities using the final multi-level logistic model are shown in columns two and three of Table 5. These values represent the probability of smoking in different age/sex groups using the information from the SHS and the associated census data and we would expect them to be slightly different to the raw probabilities in Table 1.

The next step involved the calculation of the probabilities of smoking for each age/sex group at the OA level across Scotland; this was possible for 35,994 out of the total of 38,254 OAs (94.1%). It was not possible to calculate probabilities where one or more of the explanatory variables was missing in a particular OA (such data were not provided by the Census Offices because of confidentiality constraints). The 'percentage of individuals classified as working in clerical and secretarial occupations', which is derived from the 10 percent sample from the census, was missing in 1,699 OAs, for example.

The means of the resulting calculated probabilities for each age/sex group are shown in Table 7. These can be compared to the probabilities from the raw data (Table 1) and the probabilities derived from the multi-level logistic model (Table 6). In general, the calculated

Table 6
Logit model probabilities of smoking

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

Table 7
Summary statistics for the predicted probabilities for output areas in Scotland

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

probabilities of smoking were higher in the younger age groups and among men. Because the age/sex interaction was insignificant in the model the predicted male smoking rate will be higher than the female smoking probability for all age groups, unlike the results in the raw data (Table 1). This suggests that the difference

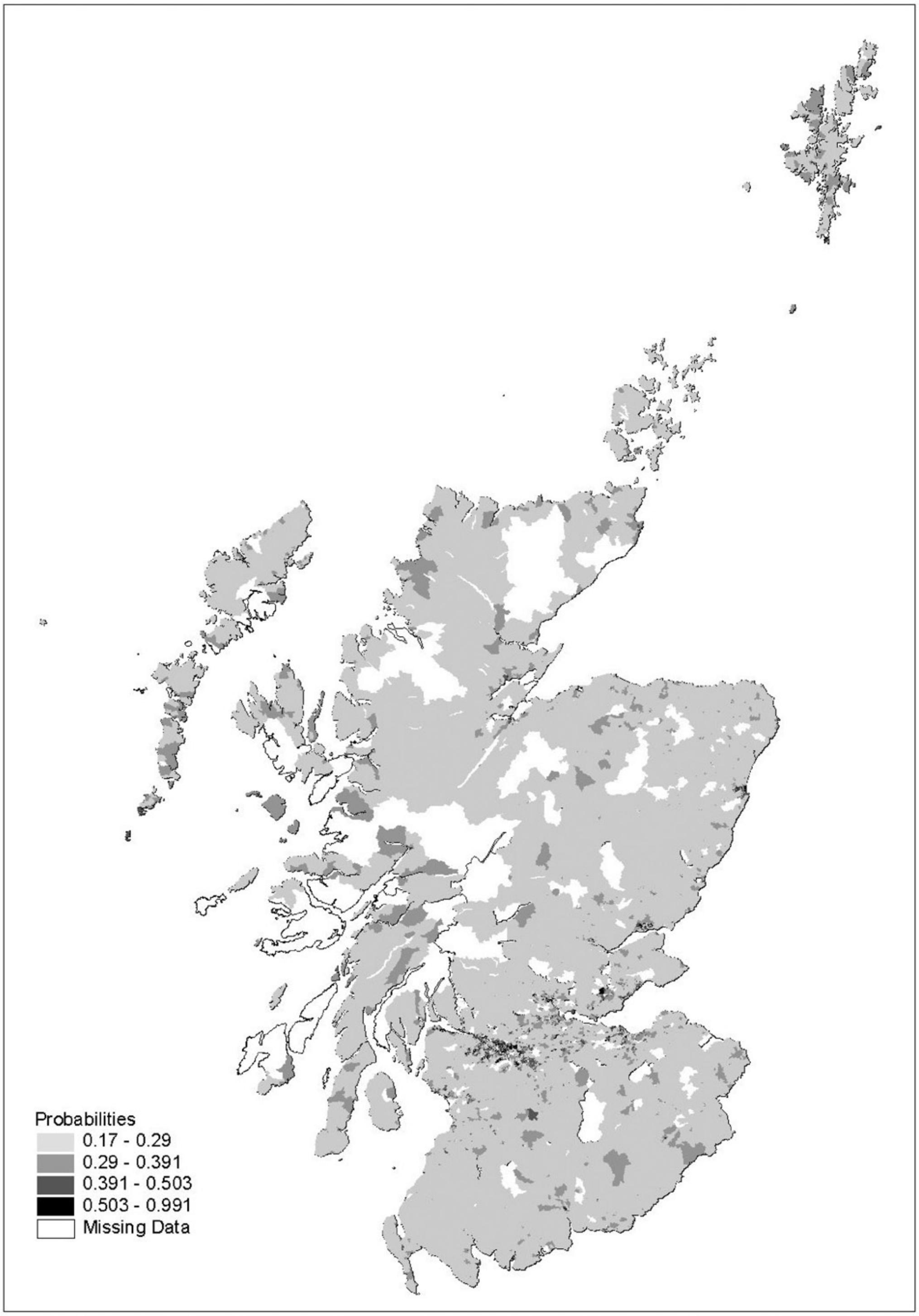

Fig. 1. Probability of smoking among males aged 16–24 in Scotland.

between the men and women in Table 1 is related to other characteristics of men and women that are controlled for in the model (Table 5).

Table 7 also summarises the range of predictions for each age/sex group across the Scottish OAs. A different probability of smoking has been calculated for each group in each OA, based on the characteristics of the OA. The range of probabilities for each of the groups is quite large. For example, the predicted smoking probability for males aged 16–24 has a maximum value of 0.792, suggesting that approximately 79 percent of men aged 16–24 smoke in that particular OA. This OA is in Edinburgh City and has particularly high values for: population density (144,000 per km^2), Carstairs Score (4.74), households with no car (86.2%), households renting from a housing association or local authority (100%), and lone parent households (22.7%). The lowest probability that was calculated in an OA for males aged 16–24 was 0.17. This OA is in Renfrew on the western edge of Glasgow. It had low values for population density (1562.5 per km^2), Carstairs score (−7.37), households with no car (0.74%), percentage households renting from a housing association or local authority (0%), percentage of privately renting households (0%), percentage of individuals classified as having worked in clerical and secretarial occupations (0%) and percentage of lone parent households (0%).

Finally, the probabilities of smoking for each age/sex group can be mapped. Fig. 1 shows the probability of smoking for people aged 16–24 for the whole of Scotland. There is a tendency for smoking rates to be higher in urban areas than rural areas; there were particularly high values in more deprived urban areas. Figs. 2 and 3 show the smoking probabilities for women aged 16–24 and over 65, respectively, in an area around Dundee including parts of Fife and Perthshire. The geography of smoking is broadly similar for these two age groups, even though the levels of smoking are much lower among the elderly. This is because the only interaction term that was significant in the model was a fairly small effect relating age and public housing; otherwise the effects of other variables were the same regardless of age group. Essentially this model suggests that, apart from areas of housing association and local authority housing, there is little reason to expect smoking by different age groups to have different geographical patterns.

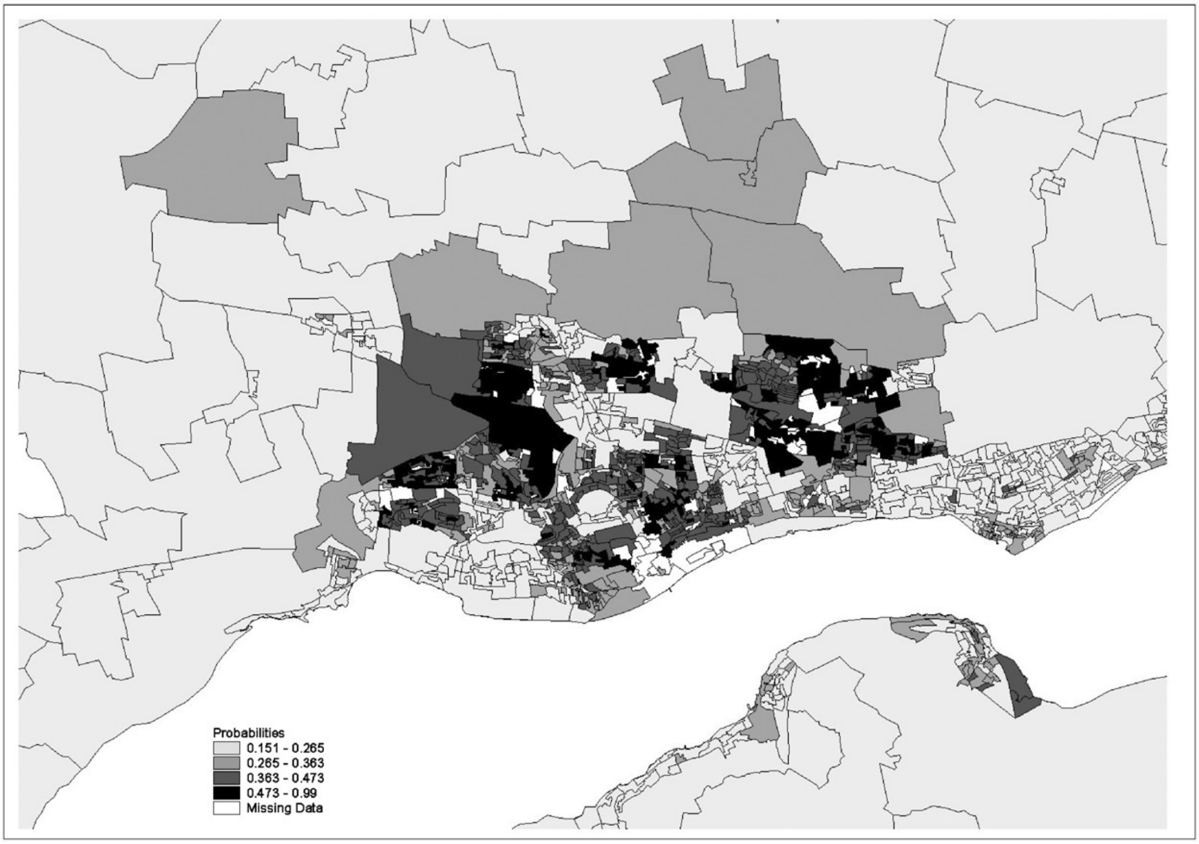

Fig. 2. Probability of smoking among females aged 16–24 in and around Dundee.

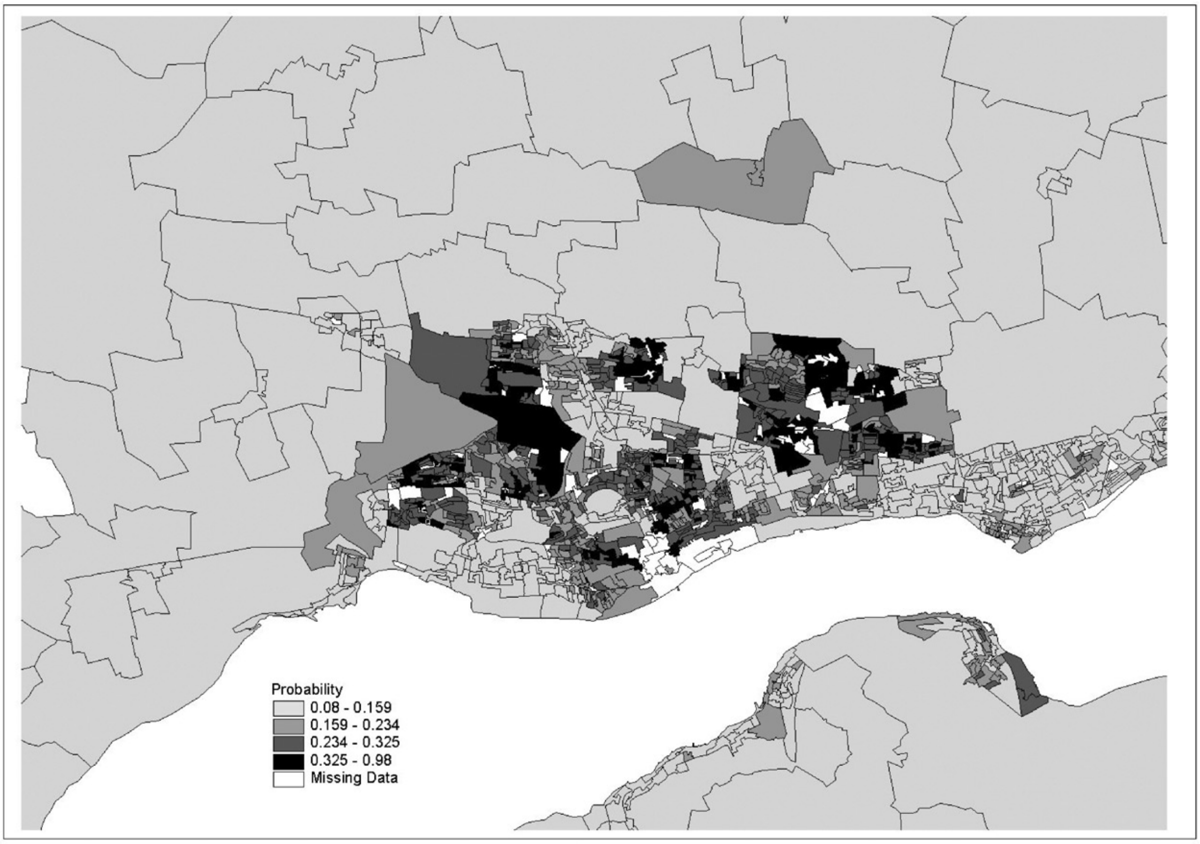

Fig. 3. Probability of smoking among females aged 65+ in and around Dundee.

Conclusion

This study is the first attempt to estimate smoking probabilities for OAs across the whole of Scotland using a multi-level logit modelling approach. These estimates will be helpful for use in our wider study of the incidence of lung cancer, and we hope for other researchers in related areas. The estimates have been calculated by using individual-level data from the SHS in tandem with area-based census and census-derived variables. The mapped results from this model demonstrate the higher probability of smoking in urban areas with particularly high values in more deprived areas. Lower probabilities are found in more suburban and rural areas. The results accord with those from earlier studies that have attempted to address this problem at coarser geographical scales. However, although the results are broadly in accord with expectations there are a number of data issues that should be borne in mind.

First, we used 1991 census data to model individual level survey data from 2000. Even though the demographic and economic makeup of most OAs change only very gradually, some OAs may have changed markedly in 9 years between the 1991 census and the SHS. It will be interesting to run a similar analysis using the results of the 2001 census, when they become available, to see how the results from the two analyses differ.

Second, there are confidentiality issues relevant to the release of the 1991 census data that may have had a small influence on our analysis. To maintain confidentiality, OA level counts have 1, 0 or -1 added randomly to them (Marsh, 1993). Clearly this may influence the percentage values of the explanatory variables, especially where the counts were small. In a number of OAs with small populations, data suppression for confidentiality reasons made it impossible to calculate a probability.

Third, we are interested in using these estimates of smoking behaviour to help explain lung cancer across Scotland for the period around 1991. Of course, it may be argued that it would be preferable to have smoking estimates for a period some time ago, because we would anticipate that there should be a lagged relationship between environmental risk factors, such as smoking, and lung cancer. This is difficult to achieve, as historical data on smoking are not generally available for small geographical areas. Also, given that our study is ecological, even were lagged data available there would be no guarantee that those who developed lung cancer resided in the same place during the lag period. On the other hand, it is also possible that the relative geographical differences in smoking behaviour today are likely to resemble historical patterns of smoking quite strongly. These combined concerns make it difficult to judge whether crude lagged smoking data would be better than contemporary estimates, which are likely to be more accurate for small areas.

Acknowledgements

Thanks to Louise Finlayson at the Scottish Executive for help with obtaining data from the Scottish Household Survey. We acknowledge use of Crown Copyright data from the 1991 Census, including digital boundary data, purchased for the academic community by ESRC and JISC. Robin Rice calculated the Carstairs scores at the University of Edinburgh. Thanks also to Liz Twigg for discussions about the modelling methodology.

References

Amos, A., Currie, C., Hunt, S., Martin, C.J., 1990. Health related behaviour in a small Scottish community. Public Health 104, 131–140.

Anderson, S., Hope, S., 2000. Scotland's people: results from the 1999 Scottish Household Survey. Volume 2: Technical Report. National Statistics Publication.

Balarajan, R., Yuen, P., 1986. British smoking and drinking habits: regional variations. Community Medicine 8, 131–137.

Barnett, J.R., 2000. Does place of residence matter? Contextual effects and smoking in Christchurch. New Zealand Journal of Medicine 113, 433–435.

Blaxter, M., 1990. Health and Lifestyles. Routledge, London.

Boyle, P.J., Gatrell, A.C., Duke-Williams, O., 1999. The effect of variability in deprivation and population stability in England and Wales: an investigation at small-area level. Social Science and Medicine 49, 791–799.

Diehr, P., Koespell, T., Cheadle, A., Psaty, B.M., Wagner, E., Curry, S., 1993. Do communities differ in health behaviours? Journal of Clinical Epidemiology 46, 1141–1149.

Dixon, F., Finlayson, L., 2000. Research Note: the Scottish Household Survey. Scottish Geographical Journal 116, 143–148.

Duncan, C., Jones, K., Moon, G., 1999. Smoking and deprivation: are there neighbourhood effects? Social Science and Medicine 48, 497–505.

Escobedo, L., Pedicord, M., 1996. Smoking prevalence in US birth cohorts: the influence of gender and education. American Journal of Public Health 86, 231–236.

Goldstein, H., 1995. Multi-level Statistical Models. Edward Arnold, London.

Graham, H., Der, G., 1999. Influences on women's smoking status—The contribution of socio-economic status in adolescence and adulthood. European Journal of Public Health 9, 137–141.

Jones, K., Duncan, C., 1995. Individuals and their ecologies: analyzing the geography of chronic illness within a multilevel modelling framework. Health and Place 1, 27–40.

King, A.C., Taylor, C.B., Haskell, W., 1990. Smoking in older women: is being a female a risk factor for continued cigarette use? Archives of Internal Medicine 150, 1841–1846.

Law, M.R., Morris, J.K., 1998. Why is mortality higher in poorer areas and in more northern areas of England and Wales? Journal of Epidemiology and Community Health 52, 344–352.

Macintyre, S., MacIver, S., Sooman, A., 1993. Area, class and health: should we be focusing on places or people? Journal of Social Policy 22, 213–234.

Marsh, C., 1993. Privacy, confidentiality and anonymity in the 1991 census. In: Dale, A., Marsh, C. (Eds.), The 1991 Census User's Guide. HMSO, London, pp. 111–128.

Peto, R., 1994. Smoking and death: the past 40 years and the next 40. British Medical Journal 309, 937–939.

Phillimore, P., 1993. How do places shape health? Rethinking locality and lifestyle in northeast England. In: Platt, S., Thomas, H., Scott, S., Williams, G. (Eds.), Locating Health: Sociological and Historical Explanations. Avebury, Aldershot.

Saul, C., Payne, N., 1999. How does the prevalence of specific morbidities compare with measures of socio-economic status at small area level? Journal of Public Health Medicine 21, 340–347.

Shewry, M.C., Smith, W.C.S., Woodward, M., Tunstall-Pedoe, H., 1992. Variations in coronary risk factors by social status: results from the Scottish heart health study. British Journal of General Practice 42, 406–410.

Stidley, C.A., Samet, J.M., 1993. A review of ecologic studies of lung cancer and indoor radon. Health Physics 65, 234–251.

Twigg, L., 1999. 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 21, 14–21.

Twigg, L., Moon, G., Jones, K., 1998. Predicting small-area health related behaviour: a comparison of smoking and drinking indicators. In: Earickson, R., Schneider, D. (Eds.), Proceedings of the Eighth International Medical Geography Symposium. University of Maryland, Baltimore.

Twigg, L., Moon, G., Jones, K., 2000. Predicting small-area health related behaviour: a comparison of smoking and drinking indicators. Social Science and Medicine 50, 1109–1120.

Uitenbroek, D.G., McQueen, D.V., 1993. Trends in cigarette smoking by gender, age and occupational status. Scottish Medical Journal 38, 12–15.

Wallace, M., Denham, C., 1996. ONS Classification of local and health authorities of Great Britain, Studies on Medical and Population Subjects, Vol. 59, HMSO, London.