Predicting small-area health-related behaviour: a comparison of smoking and drinking indicators

Liz Twigg[a,*], Graham Moon[a], Kelvyn Jones[b]

[a]*School of Social and Historical Studies, Institute for the Geography of Health, University of Portsmouth, Milldam, Burnaby Road, Portsmouth PO1 3AS, UK*
[b]*Department of Geography, Institute for the Geography of Health, University of Portsmouth, Buckingham Building, Burnaby Road, Portsmouth PO1 3HE, UK*

Abstract

Health-related behaviours are of central importance to health promotion and to the promotion of enhanced population health. In the UK, localised knowledge of the quantitative dimensions of health-related behaviours is traditionally attained by conducting a costly sample survey. Such surveys seldom generate reliable data at scales more local than that of the health authority, they also need to be repeated regularly. This paper outlines an alternative framework for generating statistics on small-area health related behaviours using routinely available data from the annual Health Survey for England ($N = 17,000$) and the decennial Population Census. Using a multilevel modelling approach nesting individuals within postcode sectors within health authorities, and focusing on the prevalence of smoking and 'problem' drinking, the paper comprises four sections: a consideration of the modelling strategy, a comparison of the smoking and drinking models, an outline of the estimation strategy, and the presentation and discussion of ward-level estimates of smoking and drinking behaviour for England. The paper concludes that the method is better at estimating smoking than drinking but that it offers a feasible, cheap and more informative alternative to the survey approach to the generation of information on smoking and drinking behaviour. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: Multilevel models; Health survey for England; Smoking; Drinking

Introduction

In the UK, small-area data on health-related behaviours are not widely available on a routine basis. This is an important constraint on health promotion and public health activities. The absence of small area data means that monitoring and target setting are done at a relatively crude geographical scale. This is a key shortcoming because health-related behaviours such as smoking and drinking are implicated deeply in the various targets set out both internationally for Health for All, and nationally as in the UK for the Health of the Nation/Our Healthier Nation initiative (WHO, 1981; DoH, 1992, 1998). The specific case for small-area data reflects the need to respond to local variations in the prevalence of behaviours and develop programmes for intervention which are grounded in locally-relevant evidence. Health related behaviour is not uniform across geographical space and substantial variations

---

* Corresponding author. Tel.: +44-1705-846-117; fax: +44-1705-842-174.

E-mail address: liz.twigg@port.ac.uk (L. Twigg).

may exist within large spatial units (Duncan et al., 1993, 1996), yet current health promotion and public health activity, of necessity, generally proceeds in ignorance of this situation.

This paper addresses the need for small-area data on health-related behaviours. It outlines a technically-robust and innovative framework for generating such data for each of the 8288 local government wards in England. The approach represents a major advance on previous strategies. The structure of the paper comprises two substantive sections. First, a short critique of previous approaches to small-area data provision is presented. Second, attention turns to an outline of a new approach based on multilevel models. Discussion focuses on the individualistic and ecological aspects of smoking and drinking behaviour and the need for a modelling approach, which takes account of this multilevel aetiological basis. Consideration is also given to data sources for model generation and the limitations posed for modelling by the chosen sources. Attention then shifts to the results of the modelling exercise. The coefficients derived from the final multilevel models of smoking and drinking behaviour are briefly discussed before moving to a detailed examination of the predicted small area estimates for smoking and drinking behaviour. The conclusion to the paper reflects on the utility of the approach.

Small-area estimates of health-related behaviour: existing approaches

Health promotion or public health specialists seeking fine-scale locally-sensitive information on health-related behaviours have traditionally pursued two tactics. The first, and most straightforward option has been direct measurement by conducting a local survey. This approach is costly and generates information which dates rapidly (HEA, 1990). Well-structured surveys with sound sampling design capable of generating representative results at the subdistrict level are estimated to cost at least £50,000 for a single district health authority. Design and sampling shortcomings are frequent and most surveys achieve little beyond a crude subdistrict partitioning. New building and urban redevelopment alone suggest that replication at five-yearly intervals is sensible. Local surveys do however provide a high profile manifestation of health promotion and public health activity within an area and may, if well done, generate high quality data. 'Oversampling' in national surveys might provide another approach but this too would be costly. In view of the fact that other contextual information would also be collected it would also have a degree of sensitivity which would be likely to preclude wide release.

A second strategy is simply to assume that regional or national data are representative of a local area. This assumption is simply untenable as it denies the possibility of local health variations. Were regional averages to apply uniformly across a region, there would be no need to seek data on local-scale variations. While the evidence regarding such internal differentiation is indeed mixed (Humphreys and Carr-Hill, 1991; Diehr et al., 1993), few would suggest that there is no variation within the relatively crude spatial reporting units for which large routine government surveys publish their results. As a consequence there is a temptation to seek fine disaggregation of such data. In the main however the sampling design of national surveys is insufficiently robust to permit disaggregation below the scale of 14–25 regions — and, paradoxically, any attempts at such disaggregation could only be validated with local survey data. As a result only crude disaggregations are published using national survey data (Blaxter, 1990).

In the late 1980s, geodemographics emerged as a potential third strategy for identifying local variations in health-related behaviour (Speller and Hale, 1985). Health-related behaviours were treated in the same way as other forms of consumer behaviour. It was assumed that they reflected the type of area in which a person was resident. Geo-referenced responses to national surveys of the relevant health-related behaviour were tagged using a national classification of small areas generated using a cluster analysis of data drawn from the national census. Response levels were then constructed for each of the area types and cross-area comparisons made. The approach, though sensitive to variations in health-related behaviour according to the socio-economic status of an area, did not accommodate the possibility that the behaviours might vary within an area, or that the relationship between area type and behaviour might not be constant in all parts of the country. In short, geodemographics worked on ecological principles and suffered from the pitfalls of aggregation bias (Schwartz, 1994; Susser, 1994a,b).

The application of geodemographics to health-related behaviour was also limited by three further factors. First, though specialist geodemographic systems began to evolve in the early 1990s, none were specifically developed using input data known to be associated with health-related behaviour. Second, though again exceptions could be noted, most geodemographic systems were based on measures of distinctiveness. The areas were identified using those indicators, which maximised their distinctiveness from other areas. The phenomenon, which made an area distinctive might be relatively rare within that area. Consequently, any regularity of association between a geodemographic area type and a particular health-related behaviour might not necessarily reflect a regular association

Table 1
The factors associated with smoking

Factor	Study examples
Social class	Golding (1987)[a]; Public Health Service, 1989; Amos et al. (1990)[a]; OPCS (1991)[a]; Campion et al. (1994)[a]; Courti (1994)[a]; Shewry et al. (1992)[a]; Glendinning et al. (1994)[a]
Gender	Golding (1987)[a]; Amos et al. (1990)[a]; Oakley et al. (1992)[a]; Uitenbroek and McQueen (1993)[a]; Galt et al. (1994)[a]; Escobedo and Pedicord (1996)
Age	Public Health Service (1989); Amos et al. (1990)[a]; Taioli and Wynder (1991); Uitenbroek and McQueen (1993)[a]; Courti (1994)[a]
Marital status	Amos et al. (1990)[a]; King et al. (1990)
Employment status	Amos et al. (1990)[a]; Shewry et al. (1992)[a]; Campion et al. (1994)[a]
Income	Oakley et al. (1992)[a]
Smoking status of family members	Golding (1987)[a]; Lau et al. (1990); Campion et al. (1994)[a]; Wang et al. (1995)
Smoking status of friends	Wang et al. (1995)
Educational attainment	King et al. (1990); Shewry et al. (1992)[a]; Campion et al. (1994)[a]; Escobedo and Pedicord (1996)
Tenure	Oakley et al. (1992)[a]; Shewry et al. (1992)[a]
Ethnicity	Public Health Service (1989); MMWR (1992)
Use of alcohol and drugs	Blaxter (1990)[a]; Oakley et al. (1992)[a]

[a] These studies are drawn from UK data and are more relevant for the purpose of covariate identification for the project outlined in this paper. The importance and effects of individual covariates may vary from nation to nation.

between the behaviour and the dominant characteristics of the residents of an area. Finally, through their focus on area-based associations, geodemographic perspectives on health-related behaviour emphasise the context within which the behaviour takes place. Moreover, they ignore the many individual factors, which may impact upon health-related behaviour.

A summary verdict on existing approaches to generating small-area statistics on health-related behaviour would be that none are adequate. Surveys are costly and may suffer design deficiencies, which preclude local sensitivity. Available routine data certainly lack such local sensitivity. Geodemographic approaches are methodologically unsatisfactory. Given the policy-driven need for small-area data, the need for an alternative approach to data generation is clear. We now turn to an outline of such an approach.

A new approach to the generation of small-area data on health-related behaviour

Multilevel perspectives on health-related behaviour

It is the simultaneously individual and contextual nature of the influences on health-related behaviour, which provides the basis for the development of a new approach to the generation of small-area data on health-related behaviours. The distinction of the individual and contextual components of health-related behaviour attempts to capture the interplay of, on the one hand, autonomously-made personal decisions and individual genetic inheritance, and, on the other hand the diverse collective influences stemming from the settings within which an individual's behaviour takes place (Macintyre et al., 1993; Phillimore, 1993; Diez-Roux, 1998). As an example, the case of smoking is illustrative. Table 1 summarises a range of variables associated with smoking prevalence.

Many of these factors may be considered from either an individual or an ecological perspective. For example, it may be that an individual's social class influences their decision to smoke. Equally, areas of higher social status may provide a cultural context that influences smoking prevalence. Third, and importantly, individual and ecological aspects of a factor can interact. Thus, the likelihood of a low social class individual being a smoker may rise if that individual is resident in a lower social status area; the influence of an individual's social class on individual smoking behaviour may be critically affected by the social status of the area in which the individual lives. Finally, interactions can, of course, also take place between factors: Oakley et al. (1992) are among many who discuss the well-known interaction of age and gender with regard to the prevalence of smoking among young women.

The presence of both individual and ecological influences on a health-related behaviour indicates the applicability of a multilevel approach to the analysis of these behaviours. The characteristics of multilevel analysis and the associated statistical theory are well-documented and do not require repetition here (Bryk and Raudenbusch, 1992; Longford, 1993; Goldstein, 1995; Kreft and de Leeuw, 1998). The analytical use of

Table 2
Candidate surveys for modelling health-related behaviour

Survey	Years and population coverage	Size and geography	Agency
Health Survey for England HSE (cross-sectional) smoking, alcohol	1991–(ann.); all adults, 16+; children 2–15, 1995 (max 2 per hhold)	Individuals (16,000); hholds (+10,000); PSUs (720, unnamed); DHAs (170 named); RHAs (14, named) England only	SCPR/UCL (DoH)
General Household Survey GHS (cross-sectional) smoking, alcohol (biennually)	1971–(ann. until 1997); all adults (16+)	1994/5; individuals (18,000); hholds (10,000); PSUs (unnamed–576); regions (named–22) Great Britain	ONS (Office for National Statistics)
British Household Panel Survey BHPS (panel survey) smoking	1991–(ann.); all adults, 16+; 11–15 yr olds 1994 onwards	1995; individuals (10,000); hholds (5000); local authority approximation (named); regions (named) Great Britain	ESRC Centre for Microsocial Change
Health and Lifestyle Survey HALS (two-wave panel) smoking, alcohol	1984/5 & 1991; selected individuals (18+)	1984/5 (1991); individuals 9003 (5000); wards (named–396); parl. constits (named–198); regions (named–22) Great Britain	Health Promotion Research Trust/ University of Cambridge School of Clinical Medicine

multilevel models in the study of health data is also well-established. The national datasets used to calibrate these existing studies provide an added reason for the use of multilevel models: not only are the behaviours in question the subject of individual and ecological influences, the datasets are based on hierarchical sampling structures in which individuals or households are selected from higher levels of aggregation. Effective modelling requires recognition of multilevel processes and the autocorrelation consequent upon hierarchical sampling designs.

There are thus substantive and methodological reasons for the use of multilevel models in the analysis of national survey data on health-related behaviour. A comprehensive model of a particular health-related behaviour should enable the prediction of that behaviour as the outcome of processes operating at different 'levels' — individual or ecological. Importantly a multilevel approach can also take *simultaneous* account of these influences whereby the effect of a process at any one level can be assessed given the process at play at the other levels. Furthermore, this methodology allows for the accommodation of between-level interaction effects and an understanding of the nature of residual unaccounted variation at each level.

Methodology

Candidate survey data

There are several candidate data sources for the calibration of multilevel models of smoking and drinking behaviour in England and Wales. Suitable sources need to have in common a number of key characteristics. First, in order to provide a basis for addressing the charge that local surveys date rapidly, they must be repeated regularly employing standardised uniform question formats. Second, to accommodate the need to consider both individual and ecological aspects of smoking and drinking behaviour, raw individual data nested within 'higher level' areas must be available. Third, they must satisfy the statistical requirements for multilevel modelling: quantitatively large and with well-found, tested and robust hierarchical sampling designs. Fourth, to facilitate wide usage, they should be free or low-cost to academic and user communities. The notion of the ideal survey covering all the factors associated with smoking or drinking within a clearly defined and openly disclosed multilevel structure is unrealisable in practice. Nevertheless there should also be a core range of contextual variables.

Four sources satisfied these desiderata: the Health Survey for England (Colhoun and Prescott Clarke, 1996), the General Household Survey (ONS, 1996), the British Household Panel Survey (Rose et al., 1994; Taylor, 1996), and the Health and Lifestyle Survey (Cox et al., 1987; Cox, 1988). Table 2 summarises the

Table 3
Model results

	Smoking		Problem drinking		
	estimate	standard error	estimate	standard error	
Constant	−0.988	0.0676	−2.037	0.091	
Male	0.388	0.0942	1.113	0.1123	Individual terms: main effects
Single	0.751	0.1098	0.423	0.1427	
16–24	0.339	0.1414	−0.142	0.2086	
35–44	−0.097	0.0942	0.141	0.1202	
45–54	0.032	0.0969	−0.011	0.1291	
55–64	−0.337	0.1103	−0.258	0.1463	
65–74	−0.643	0.1297	−0.633	0.1833	
75+	−1.235	0.2593	−0.675	0.3097	
Male 16–24	−0.092	0.2336	0.196	0.2865	Individual terms: 2-way interactions
Male 35–44	−0.183	0.1343	−0.156	0.1541	
Male 45–54	−0.347	0.1388	−0.034	0.1628	
Male 55–64	−0.415	0.1556	−0.121	0.1814	
Male 65–74	−0.308	0.1783	−0.116	0.2209	
Male 75+	−0.476	0.3385	−0.693	0.3747	
Single 16–24	−0.809	0.1832	0.176	0.2535	
Single 35–44	0.094	0.1729	−0.161	0.2224	
Single 45–54	−0.394	0.1819	−0.107	0.2352	
Single 55–64	−0.111	0.1916	−0.822	0.3038	
Single 65–74	−0.485	0.1894	−0.552	0.2787	
Single 75+	−0.585	0.2995	−1.071	0.3881	
Single male	−0.727	0.1609	0.071	0.1846	
Single: male: 16–24	0.433	0.2881	−0.433	0.3422	Individual terms: 3-way interactions
Single: male 35–44	0.328	0.2549	−0.183	0.2922	
Single: male 45–54	0.854	0.2745	0.168	0.3111	
Single: male 55–64	0.596	0.2926	0.357	0.3811	
Single: male 65–74	1.167	0.2881	0.342	0.3601	
Single: male 75+	1.203	0.4373	0.863	0.5101	
%No car	0.009	0.0015	NS	NS	Ecological effects: level-2
%2+cars	−0.007	0.0017	0.005	0.0017	
%Private rent	0.003	0.0014	0.005	0.0029	
%Social class 1 or 2[a]	NS	NS	0.007	0.002	
Single: %2+cars	0.005	0.0022	NS	NS	Cross-level interactions
Male: %social class 1 or 2	NS	NS	−0.007	0.0022	
Male:%private rent	NS	NS	−0.008	0.0029	
Single: %Social class 1 or 2	NS	NS	−0.006	0.0023	
Single: %private rent	NS	NS	0.008	0.0029	

[a] These are the two highest social classes in the current UK classification.

key differences between these sources. There are strengths and weaknesses to each and some variations in the multilevel structures. The Health Survey for England (HSE) was devised initially as a monitoring instrument for certain Health of the Nation goals and is a large health-focussed survey. The General Household Survey (GHS) has a broadly similar sample size and design but is a multipurpose survey of which health matters are just one component. Although traditionally undertaken annually, its collection was suspended in 1997 though subsequently reinstated. The British Household Panel Survey (BHPS) is a rather different type of study. It follows the same respondents through time and is thus ideal for longitudinal work though consequently it has a smaller sample size. It does not however cover alcohol consumption. The Health and Lifestyle Survey (HALS) is included for reasons of completeness although the 1991 resurvey of the original HALS respondents saw an inevitable reduction on the original sample size.

A series of exploratory multilevel models was run using comparable variables drawn from each of the candidate surveys. These revealed broadly similar results for the two large routine governmental surveys

(the HSE and the GHS). The HSE was chosen for further analysis in view of its greater health focus, its more extensive coverage of detailed aspects of smoking and drinking, its importance in the context of official health policy, and, at the time of research, its more secure future. Both the BHPS and HALS were rejected because of their smaller sample sizes. In the case of the BHPS this led to exploratory multilevel models with factors unexpectedly failing to reach statistical significance. For HALS there were arguably too few observations for effective multilevel analysis in the 1991 dataset and the 1985/6 set was thought too dated. The BHPS was also rejected because it did not cover alcohol consumption.

Developing models for predicting smoking and problem drinking prevalence

The HSE, as suggested, offers a number of ways of measuring smoking and alcohol consumption behaviour. As the aim of the research was the generation of small-area estimates of the prevalence of particular behaviours, straightforward definitions were adopted. The response variables were defined in terms of a binomial healthy/unhealthy dichotomy. For smoking, the response variable was simply whether or not a person smoked tobacco. For drinking, men were classified as problem drinkers if their consumption equalled or exceeded an average of 21 units per week over the past year, while the cutoff for women was 14 units[1] (see Table 3).

The multilevel structures adopted for both smoking and drinking attempted to model individual behaviour in terms of factors operating at three levels: the individual, the postcode sector unit (PSU) and the district health authority. The rationale for the inclusion of individual-level variables has been outlined above. PSUs were used as crude analogues of the local community, the areas providing the context for individual behaviour. Health districts offered a relatively disaggregate level at which to take account of broader 'regional' variation while simultaneously benefiting from a resonance with NHS policy-making and avoiding the overly-heterogeneous diversity of regional health authorities. Though potentially of considerable sociological significance, households were not modelled as an explicit level in the analysis. First, there remain technical difficulties in estimating the random part for a binomial model with household as level-2. When health-related behaviour is measured on a safe/unsafe dichotomy, the assumption of normality is unlikely to

exist at the household level. The problem is further complicated by the fact that there is a high proportion of one person households in the HSE. It therefore becomes very difficult to separate out between- and within-household variation because of confounding across level-1 and -2. Additionally there are also substantive reasons for not including the household level: in a correct model, the overall effect of including the level of household would be to reduce variation at the higher levels. As the objective of the paper is to predict variation at the (higher) level of the PSU, it may be appropriate to allow these higher level differences to exist even if some of their variation can be explained by household differences.

The multilevel software package MLn was employed to develop the models (Rasbash and Woodhouse, 1995). The choice of MLn reflected user expertise and the availability and flexibility of the software. Other multilevel software could accommodate the relatively straightforward hierarchical structures described in this paper. A review of multilevel software is provided by Kreft and de Leeuw (1998). MLn uses precision-weighted estimation in model fitting. This means that the approach is relatively robust to variations in the number of observations in each sampling unit. For the dichotomous response variables used in the analysis a binomial error assumption was needed and the appropriate MLn macro was employed (Goldstein, 1991). A second order penalised quasi-likelihood (PQL) approximation, using a Taylor series expansion, was employed to estimate the coefficients (Goldstein and Rasbash, 1996).

The form and content of the explanatory part of the final multilevel models of smoking and drinking behaviour were determined both by the nature of the HSE and by the availability of relevant information for the local areas on which the predictions were ultimately to be based. The chosen spatial unit for the generation of the predictions was the local government ward. This is a spatial area for which mortality data are routinely released and the possibilities for associative analyses linking mortality and health-related behaviour are thus considerable. The source of choice for detailed contextual information about factors likely to be linked to health-related behaviours is the UK Population Census. Unfortunately the census does not, in England, routinely provide data for postcode sector units, nor does the HSE identify the ward of residence of its respondents. Two consequences arise. First, the PSU and the local government ward have, perforce, to assume a degree of analytical equivalence. Both are of roughly similar size and both can be taken to offer some reflection of the local contextual setting for health-related behaviour. Second, the ecological variables for each PSU must be estimated by using the characteristics of the individual HSE respondents in

---

[1] These are accepted UK and international definitions of problem drinking. In the UK there have been recent attempts to raise these limits. In this study the more conservative figures have been used.

For every ward
derive the ecological variables (from Census)
deviate these % from national HSE mean
multiply by their estimated logits
sum these ecological logits
add in the logit for the L3 differential (DHA)
(EcoLog)

⇓

For each age/sex/marital status group (28 in total)
derive the estimated logit for each group
(GroLog)

⇓

Add the total ecological logit (and cross-level logits) to the estimated logit for each group
(EcoLog + CLevLog + GroLog)

⇓

Untransform this logit to get the proportion of each group in each ward who smoke/drink
(PropSmok)

⇓

Derive count of the number of people in each ward in each group
(NumRes)

⇓

Multiply the count of people in each group in each ward by the proportion of each group in
each ward who smoke/drink to get the estimate of the number of people who drink/smoke.
(PropSmok * NumRes)

⇓

Total the smokers/drinkers for all groups in each ward

⇓

Total the number of people for all groups in each ward

⇓

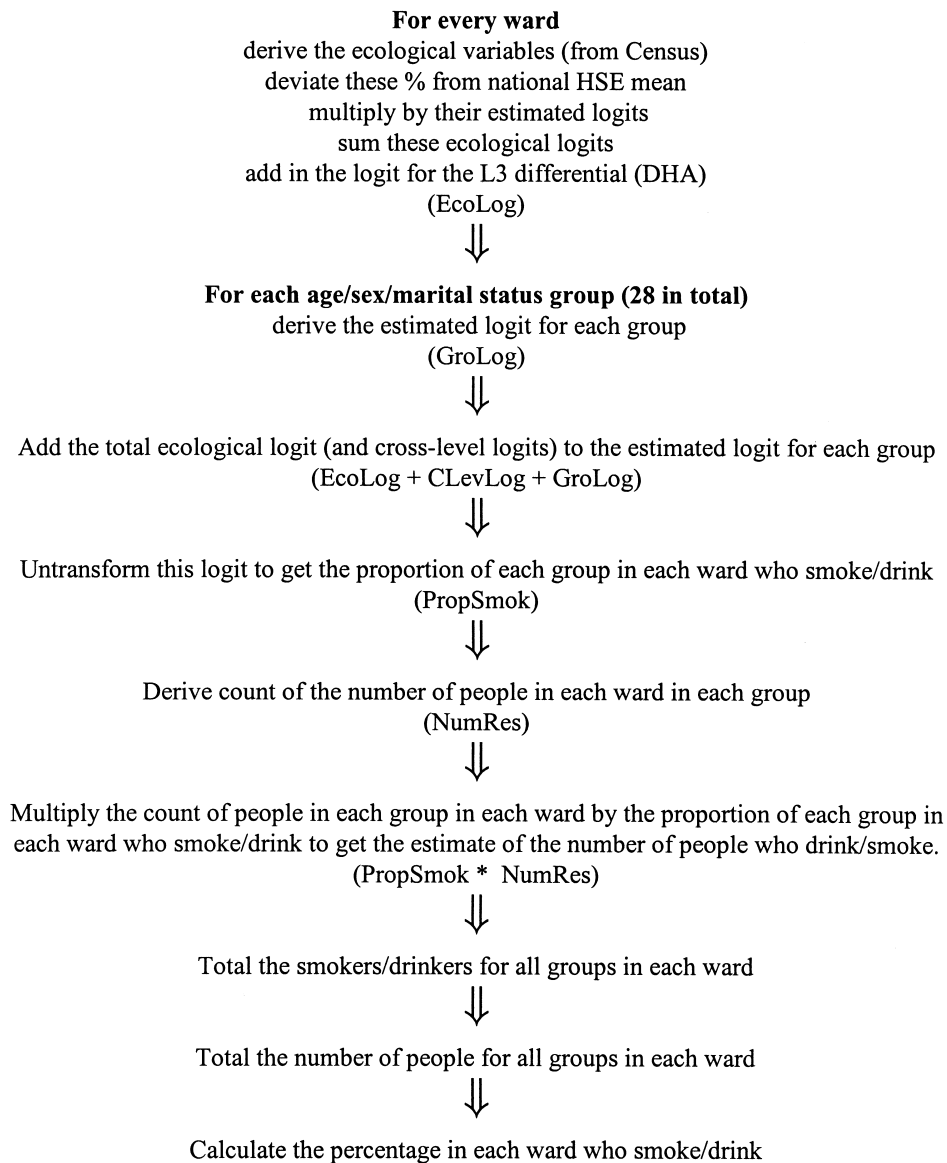Calculate the percentage in each ward who smoke/drink

Fig. 1. Prediction process.

the relevant PSU; this is possible because, although no PSU is named, the HSE discloses whether individuals live in the same PSU.

The importance of the census as the basis for local predictions of health-related behaviour also constrained the individual-level explanatory variables. The HSE survey data generated a model of individual smoking or drinking behaviour. The requirement for prediction purposes was thus individual geo-located data. The UK Census Sample of Anonymised Records is not suitable for this purpose as its level of geocoding is too coarse. It is however possible to use complex cross-tabulations of routine local base statistics from the UK Census to provide counts of the numbers of individuals in each ward who fall in particular socio-demographic categories. The most detailed such cross-tabulation available at the census ward level and relevant to health-related behaviour is age (grouped into several age bands), marital status and gender. Other important contextual variables such as tenure and
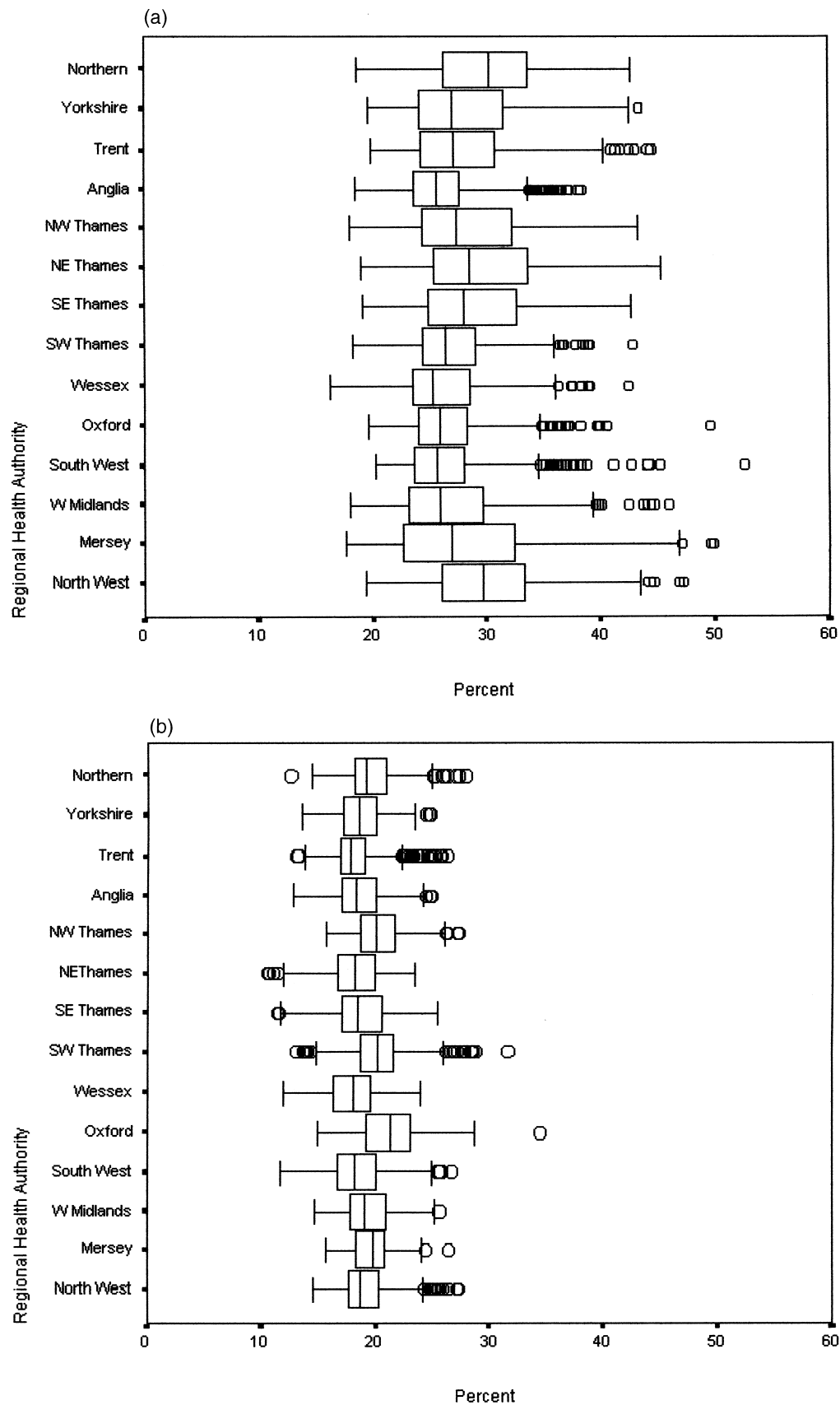
Fig. 2. (a) Predicted ward smoking by English Regional Health Authority. (b) Predicted problem drinking by English Regional Health Authority.

Table 4
Highest and lowest ward predictions

Smoking		Problem drinking	
ward	DHA	ward	DHA
Five highest predictions			
Tresco[a]	Cornwall & Scilly	Central Oxford	Oxfordshire
Central Oxford	Oxfordshire	Walbrook[a]	City & Hackney
Abercromby	Liverpool	Farringdon within[a]	City & Hackney
Granby	Liverpool	Queens Aldershot	W Surrey & NE Hampshire
Hulme	Central Manchester	Stamfordham	Northumberland
Five lowest predictions			
St Leonards & St Ives East	East Dorset	Sackville	Hastings
Little Aston	SE Staffordshire	Clyst Valley	Exeter
West Chillington	Worthing	Haven	NE Essex
Stourton	Mid Staffordshire	Partsoken[2]	City & Hackney
St Leonards & St Ives West	East Dorset	Bartholemews	NE Essex

[a] Predictions should be viewed as unstable as the denominator ward populations are extremely small in comparison to ward mean populations.

social class are not available in a small area cross-tabulation, which also contains age and gender thus losing the basis for age-sex standardisation. Table 3 summarises the final models and the results of the modelling process[2]. The estimates are based on contrast coding referenced to the modal HSE respondent (a married or cohabiting woman aged between 25 and 34 years of age) and are expressed as logits. The ecological variables are deviated around the HSE mean for the variable in question.

The resulting models uphold conventional wisdoms concerning both smoking and problem drinking. Male gender and being single are particularly important factors in both cases, the former notably so in the case of problem drinking. Age is only really of significance in the case of smoking where there are also effects stemming from the interaction of age and marital status and sex and marital status. The three-way interaction of age, sex and marital status is of some relevance in understanding the smoking behaviour of older people. Ecological variables appear to be somewhat more important in understanding problem drinking but although statistical significance is achieved the magnitude of the effects is small. Interestingly, the affluence of an area, as measured by the surrogate of dual car ownership, equates with a reduced likelihood of smoking but an increased likelihood of problem drinking. Cross-level interactions suggest that single women living in areas with high percentages of private rented households have a raised likelihood of problem drink-

ing while single women living in affluent areas are more likely to smoke.

Generating predictions

Modelling health-related behaviour using individual-based survey instruments inevitably generates estimates of the probability of an individual undertaking the type of health-related behaviour selected as the response variable. Moving from a position of knowledge about the probability of individual behaviour to predictions of small-area prevalence involves linking local census data to the estimates derived for the explanatory part of the model. In essence this requires a recognition that individual probabilities are for particular types of individual in areas with particular characteristics. With census data on the number of individuals of each type in each area, and the ecological characteristics of each area, it is possible to rework the multilevel equations for each behaviour to generate area-specific predictions.

Figure 1 summarises the processes involved. These essentially revolve around generating logit estimates of the proportion of people in particular age-sex-marital status groups who smoke, untransforming the logit form to a more familiar proportion and applying that proportion to census data. Working principally with the individual and level-2 components of the multilevel model, the approach also incorporates use of the level-3 residual. These figures, derived from the level-3 (DHA) random part of the model, indicate for each DHA the specific variance unaccounted for by the fixed part of the model. They allow adjustments to be made to the estimates to take account of the specific

---

[2] A listing of the log file used in the analysis reported in the paper is available from the authors upon request.

circumstances in each of the 177 district health authorities identified by the HSE.

The importance and application of this method lies in its generation of ward predictions for smoking and drinking. The summary geography of these predictions indicates that, on average around 28% of a ward population will be smokers and around 19% will be problem drinkers. There is little variation across England at the district health authority level for problem drinking (range 15–25%) with higher levels of problem drinking being concentrated in district health authorities with younger affluent populations and high levels of army personnel. Smoking exhibits slightly more geographical variation at the district health authority scale with predicted district health authority means ranging from 23% in parts of East Anglia and the Welsh Marches to around 40% in Liverpool and Central Manchester and the northeast.

Figures 2(a) and (b) focus down to ward-level geographical variation through boxplots for, respectively, smoking and problem drinking at the ward level for each of the 14 former regional health authorities for England. Wards in the north and northwest RHAs have, on average higher predicted levels of smoking whilst the mean of the ward values for problem drinking is highest in the Oxford RHA. The range of predicted ward values for smoking is far greater than that for drinking and outliers and extreme values suggest that there may be four wards where around 50% of the post-16 population are smokers. Equally there are two wards where around one-third of the post-16 population may be problem drinkers.

Table 4 extends this consideration of the extreme values from the predictive models. Four wards are highlighted with an asterisk, three in the City & Hackney DHA in Central London and one on the Isles of Scilly. The predictions for these wards should be viewed as unstable as the denominator ward populations are extremely small in comparison to ward mean populations. The Central ward in the City of Oxford yields high predictions of both smoking and problem drinking prevalence reflecting in large part its peculiar demographic mix and blend of high and low social status. Other high smoking wards are in inner city Liverpool and Manchester. Low levels of smoking and problem drinking are predicted in rural areas and the urban fringes of resort towns.

---

[3] One such limitation has some relevance for the predictions as generated. The random components of the problem drinking model indicated significant remaining variation at the PSU level. As the location of PSUs is unknown this variation cannot, unlike DHA-level variation, be accommodated by an adjustment factor. Consequently, rather less reliance should be placed on the problem drinking predictions.

Conclusions

This paper has outlined a new approach to generating small-area predictions of health-related behaviours, in this case smoking and problem drinking. The feasibility of an approach based on calibrating a multilevel model using national survey data and then using the resulting model in conjunction with local Census data has been confirmed. It is possible to generate small-area predictions of health-related behaviours, which conform to expected patterns and vary in an expected way around established national and regional means. Further work has explored the relationship between these 'synthetic estimations' and the results of local surveys and the impact of using a multilevel approach (Moon et al., 1998). For the purposes of this methodological paper however, two key concluding points can be made.

First, the models on which the predictions outlined in this paper are based are deliberately simple. Following Kreft and de Leeuw (1998), the models have been developed on the basis of theoretical justification and with the objective of prediction and the avoidance of over-parameterisation. Thus, the chosen parameters were theoretically justified as appropriate measures to achieve standardisation of individual responses and capture the impact of deprivation ecologies and, while more complex models might have been developed had the objective of the study been the description of health-related behaviours, few parameters were used because of the limits imposed by the need subsequently to use the Census to generate predictions. The form of the models was also, of necessity kept relatively simple. Although within-level interactions and cross-level interactions were explored at both the individual and the ecological levels and all individual within-level interactions were retained in the final models, cross-level interactions and interactions at the ecological level were only retained if significant. Furthermore, a simple random intercepts form of model was used. This enabled DHA-level adjustment factors to be identified. More sophisticated fully-random models would have introduced terms which would not have assisted the final generation of predictions. The net result of this modelling strategy is that the models, as presented, are 'fit-for-purpose' but in other senses limited[3].

Despite these matters, the second conclusion must be that the approach presented in this paper represents an advance on previous practice regarding the generation of small-area data on health-related behaviours. The survey with which the multilevel model basis to prediction is calibrated, the HSE, is repeated on an annual basis and, though the Census itself is decennial, the most important variables in the multilevel models, those at the individual level, are frequently collected or themselves estimated for ward areas at more frequent

intervals. The approach therefore offers the prospect of more timely information on health-related behaviours. Given public availability of the component datasets it is also considerably cheaper than the more traditional local survey.

Acknowledgements

References

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

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

Bryk, A., Raudenbusch, S., 1992. Hierarchical Linear Models: Applications and Data Analysis Methods. Sage, London.

Campion, P., Owen, L., McNeill, A., 1994. Smoking before, during and after pregnancy in England. Health Education Journal 53, 163–173.

Colhoun, H., Prescott Clarke, P., 1996. Survey Methodology and Documentation, Health Survey for England 1994, vol. 2. HMSO, London.

Courti, L., 1994. For better of worse? Annual change in smoking, self-assessed health and subjective well-being. In: Buck, N., Gershuny, J., Rose, D., Scott, J. (Eds.), Changing Households. The British Household Panel Survey 1990–1992. University of Essex: ESRC Research Centre on Micro-Social Change, pp. 199–219 (chapter 8).

Cox, B.D., 1988. Health and Lifestyle Survey, 1984-5 (computer file). ESRC Data Archive: Colchester.

Cox, B.D., Golding, J.F., Wadsworth, M.E.J. (Eds.), 1987. Preliminary report of a nationwide survey of the physical and mental health, attitudes and lifestyle of a random sample of 9003 British adults. Health Promotion Research Trust, London.

Department of Health, 1992. The Health of the Nation: a Strategy for Health in England. HMSO, London.

Department of Health, 1998. Our Healthier Nation. The Stationery Office, London.

Diehr, P., Koepsell, T., Cheadle, A., Psaty, B., Wagner, E., Curry, C., 1993. Do communities differ in health behaviours? Journal of Clinical Epidemiology 46 (10), 1141–1149.

Diez-Roux, A., 1998. Bringing context back into epidemiology: variables and fallacies in multilevel analysis. American Journal of Public Health 88, 216–222.

Duncan, C., Jones, K., Moon, G., 1993. Do places matter? A multilevel analysis of regional variations in health-related behaviour in Britain. Social Science & Medicine 37, 725–733.

Duncan, C., Jones, K., Moon, G., 1996. Health-related behaviour in context: a multilevel modelling approach. Social Science & Medicine 42, 817–830.

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.

Galt, M., Gillies, P., Scott, A., 1994. Smoking prevalence among 15–16 year olds in Doncaster, England. Journal of Public Health Medicine 16, 172–178.

Glendinning, A., Shucksmith, J., Hendry, L., 1994. Social class and adolescent behaviour. Social Science & Medicine 38, 1449–1460.

Golding, J., 1987. Smoking. In: Cox, B.D., Golding, J.F., Wadsworth, M.E.J. (Eds.), Preliminary Report of a Nationwide Survey of the Physical and Mental Health, Attitudes and Lifestyle of a Random Sample of 9003 British Adults. Health Promotion Research Trust, London (chapter 10).

Goldstein, H., 1991. Nonlinear multilevel models, with an application to discrete response data. Biometrika 78, 45–51.

Goldstein, H., 1995. Multilevel Statistical Models. Edward Arnold, London.

Goldstein, H., Rasbash, J., 1996. Improved approximations for multilevel models with binary responses. Journal of the Royal Statistical Society A 159, 505–513.

Health Education Authority, 1990. Health and Lifestyle Surveys: Towards a Common Approach. HEA, London.

Humphreys, K., Carr-Hill, R., 1991. Area variations in health outcomes: artefact or ecology. International Journal of Epidemiology 20 (1), 251–258.

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

Kreft, I., de Leeuw, J., 1998. Introducing Multilevel Modelling. Sage, London.

Lau, R.R., Quadrel, M.J., Hartman, K.A., 1990. Development and change of young adults preventive health beliefs and behaviour: influence from parents and peers. Journal of Health and Social Behavior 31, 240.

Longford, N., 1993. Random Coefficients Models. OUP, Oxford.

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.

Moon, G., Twigg, L., Roberts, C., 1998. Predicting small-area health-related behaviour: a comparison of synthetic estimation and survey results. Institute for the Geography of Health, University of Portsmouth, Portsmouth.

MMWR, 1992. Cigarette smoking among adults, United States, Morbidity and Mortality Weekly Report, 41, pp. 354–355, pp. 361–362.

Oakley, A., Brannen, J., Dodd, K., 1992. Young people, gender and smoking in the United Kingdom. Health Promotion International 7 (2), 75–88.

ONS, 1996. General Household Survey, 1994 (computer file and user guide). Data Archive: Colchester.

OPCS 1991. General Household Survey: Cigarette Smoking 1972 to 1990, OPCS Monitor SS 91/3. HMSO: London.

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.

Public Health Service, 1989. Reducing the Health Consequences of Smoking: 25 Years of Progress, a Report of the Surgeon General, DHHS publication, No. (CDC) 89-8411. US Government Printing Office, Washington DC.

Rasbash, J., Woodhouse, G., 1995. MLn Command Reference. Multilevel Models Project, Institute of Education, University of London, London.

Rose, D., Buck, N., Johnston, R., 1994. The British Household Panel Survey: a valuable new resource for geographical research. Area 26 (4), 368–376.

Schwartz, S., 1994. The fallacy of the ecological fallacy: the potential misuse of a concept and the consequences. American Journal of Public Health 84, 819–824.

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

Speller, V., Hale, D., 1985. Making the most of your postcode. Health Services Journal 28, 53–53.

Susser, M., 1994a. The logic of ecological. I. The logic of analysis. American Journal of Public Health 84, 825–829.

Susser, M., 1994b. The logic of ecological. II. The logic of design. American Journal of Public Health 84, 830–835.

Taioli, E., Wynder, E., 1991. Effect of the age at which smoking begins on frequency of smoking in adulthood. New England Journal of Medicine 32, 968–969.

Taylor, M.F. (Ed.) with Brice, J., Buck, N., Prentice, E., 1996. British Household Panel Survey user manual, vol. B: codebook. University of Essex, Colchester.

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

Wang, M.Q., Fitzhugh, E., Westerfield, C., Eddy, J., 1995. Family and peer influences on smoking behaviour among American adolescents: an age trend. Journal of Adolescent Health 16, 200–203.

WHO, 1981. Global Strategy for Health for All by the Year 2000. In: Health for All Series, No. 3. WHO, Geneva.