Health & Place 17 (2011) 691–695

Contents lists available at ScienceDirect

Health & Place

journal homepage: www.elsevier.com/locate/healthplace

Short Report

An index of unhealthy lifestyle is associated with coronary heart disease mortality rates for small areas in England after adjustment for deprivation

P. Scarborough *, S. Allender, M. Rayner, M. Goldacre

Department of Public Health, University of Oxford, Old Road Campus, Headington, Oxford, OX3 7LF, UK

ARTICLE INFO

Article history:
Received 13 April 2010
Received in revised form
31 August 2010
Accepted 7 December 2010
Available online 14 December 2010

Keywords:
Coronary disease
Synthetic estimates
Deprivation
Smoking
Fruit and vegetable consumption
Obesity
Blood pressure
Cholesterol
Small area analysis

ABSTRACT

Indices of socio-economic deprivation are often used as a proxy for differences in the health behaviours of populations within small areas, but these indices are a measure of the economic environment rather than the health environment. Sets of synthetic estimates of the ward-level prevalence of low fruit and vegetable consumption, obesity, raised blood pressure, raised cholesterol and smoking were combined to develop an index of unhealthy lifestyle. Multi-level regression models showed that this index described about 50% of the large-scale geographic variation in CHD mortality rates in England, and substantially adds to the ability of an index of deprivation to explain geographic variations in CHD mortality rates.

1. Introduction

The relationships between coronary heart disease (CHD) and smoking, poor diet, excessive alcohol consumption, physical inactivity, obesity, raised cholesterol, raised blood pressure and diabetes are well-established at the individual-level (Stamler, 2005; Yusuf et al., 2004; WHO, 2003). It does not necessarily follow that these established risk factors are powerful predictors of geographic variation of CHD rates: for example, the British Regional Heart Study found that CHD incidence rates in men were *negatively* associated with the prevalence of raised cholesterol, after other individual-level risk factors for CHD had been taken into account (Morris et al., 2001). The fact that the association between cholesterol levels and CHD has been shown to be different at the individual-level and the area-level illustrates the danger of interpreting results regarding this association using data collected at only one level—there is the potential for either 'ecological fallacy' or 'individualistic fallacy' (Robinson, 1950; Subramanian et al., 2009).

This paper explores ecological, compositional aspects of geographic variation in CHD mortality rates in England, specifically the variation that is due to differences in the behaviour of the populations in different areas. The paper uses model-based estimates (referred to here as *synthetic estimates*) of the prevalence of individual-level risk factors for CHD for all wards in England. The synthetic estimation is a technique that has been developed to allow for small-area estimation of phenomena (in the absence of direct small-area measurements of the phenomena), including health indicators, principally to refine resource allocation to the areas (Heady et al., 2003). The technique involves using data collected for a national survey to generate a logistic regression of the health indicator of interest (e.g. smoking) with both individual-level (e.g. age; sex) and area-level covariates (e.g. percentage of privately rented accommodation; geographic location). The parameters predicted by these logistic regression models are used to generate small area prevalence estimates using national census data. The technique is described in detail elsewhere (Heady et al., 2003; Bajekal et al., 2004; EURAREA Consortium, 2004; Twigg et al., 2000; Twigg and Moon, 2002). This paper introduces an index of unhealthy lifestyle, developed using synthetic estimates of the prevalence of several cardiovascular risk factors, and assesses whether this index can explain geographic variations in CHD mortality rates over and above those predicted by an index of socio-economic deprivation. The models developed for this paper use ecological data. It follows that the results can only provide information about the geographic variation in CHD rates in England, and not about the relationship between risk factors and CHD in individuals.

2. Methods

The units of analysis used for this paper are Standard Table Wards—a statistical set of boundaries based on the electoral ward

* Corresponding author. Tel.: +44 1865 289248; fax: +44 1865 617790.
E-mail address: peter.scarborough@dphpc.ox.ac.uk (P. Scarborough).

1353-8292/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.healthplace.2010.12.007

boundaries as of 1st January 2003. Henceforth these areas are referred to simply as 'wards'. There are 7,929 wards in England, which can be grouped into 355 local authorities (LAs).

Data on the number of CHD (ICD-9 410-414; ICD-10 I20-I25) deaths by sex, five year age group and ward of residence for the calendar years 1999–2004 inclusive were combined with population estimates from the 2001 census by sex, five year age group and ward of residence to derive sex-specific mortality rates that were age-standardised to the European Standard Population (West Midlands Public Health Observatory, 2009) for each ward in England. The mortality rates were reasonably normally distributed, and hence suited to regression analysis. Data on the ward-level prevalence of individual-level risk factors for CHD were provided by synthetic estimates. A previous analysis has assessed the validity of 16 different sets of synthetic estimates of the prevalence of individual-level risk factors for CHD for all wards in England (Scarborough et al., 2009). It assessed face, construct and convergent validity of the estimates by exploration of the supporting regression equations, and by comparison of the synthetic estimates with small area survey findings on related health behaviours, national survey findings on health behaviours and CHD mortality rates. Only the five sets of synthetic estimates that displayed reasonable validity were included as explanatory variables in the analyses reported here:

- prevalence of consuming less than five portions of fruit and vegetables per day;
- prevalence of obesity (body mass index greater than or equal to 30 kg/m^2);
- prevalence of raised blood pressure (systolic blood pressure greater than or equal to 160 mmHg, or diastolic blood pressure greater than or equal to 95 mmHg);
- prevalence of raised cholesterol (total blood cholesterol greater than or equal to 6.5 mmol/l);
- prevalence of current smoking.

The first four sets of synthetic estimates were developed for the Health Poverty Index website, by researchers from the universities of St Andrews and Oxford, funded by the Department of Health (Dibben et al., 2004). The estimates used for the website are weighted combinations of LA-level national survey estimates and a modelled element. The synthetic estimates used for this paper, and that were used in the previous exploration of validity (Scarborough et al., 2009), are based only on the modelled element and have been produced at ward-level. The smoking synthetic estimates were developed for the Health Development Agency as part of an investigation of the impact of smoking on mortality in England (Twigg et al., 2004). All of the sets of synthetic estimates are derived from the Health Surveys for England conducted between 1998 and 2001 and a table outlining the coefficients for individual-level and area-level covariates in their supporting regression equations can be found elsewhere (Scarborough et al., 2009). The synthetic estimates have been age-standardised to the European Standard Population using ten-year age bands for the analyses reported here. Prevalence rates for men and women were generated separately. The synthetic estimates are not well-suited for use as independent variables in regression analyses as they are highly correlated. Accordingly, Principal Components Analysis (PCA) was applied to z scores of the synthetic estimates to produce a set of uncorrelated explanatory variables, described as 'unhealthy lifestyle' variables.

The Carstairs index (Carstairs and Morris, 1990), constructed using data from the 2001 census (Morgan and Baker, 2006), was used as a measure of ward-level deprivation. The index is constructed by adding the z scores of the following variables: percentage of all economically active males aged 16 and over who are unemployed; percentage of households defined as overcrowded; percentage of population without access to a car; percentage of population living in households where the head of the household is defined as social class IV or V. The index has previously been shown to be highly correlated with CHD rates for wards in England (Romeri et al., 2006).

Multi-level regression models (wards nested in LAs) and spatial error regression models of the two outcome variables (male and female CHD mortality rates) were built to explore how much of the geographic variation in CHD rates can be explained by the unhealthy lifestyle variables, and by the deprivation index, in both univariate and multivariate analyses. The multi-level models allow for an investigation of both small scale and large-scale geographic variations in CHD mortality rates simultaneously, that is the differences in CHD mortality rates over a small geographic area i.e. within a local authority, and the general differences between regions of the country i.e. the North and South of England (modelled by between-wards and between-LAs variance, respectively). The spatial error regression models were built to investigate whether the results of the multi-level models may be affected by spatial autocorrelation bias, by comparing the size and sign of estimated parameters in the two sets of models. The PCA was conducted using Stata v10 (StataCorp, 2007), the spatial error regression modelling was conducted using the GeoDa software package (Anselin, 2003), and the multi-level modelling was conducted using MLwiN v2.02 (Rasbash et al., 2003).

3. Results

The results of the PCA are displayed in Table 1. Two of the transformed PCA variables were taken forward for further

Table 1
Transformation matrices calculated by principal components analysis for the sets of synthetic estimates for (1) male prevalence of risk factors for CHD and (2) female prevalence of risk factors for CHD, and amount of original variance explained by the transformed variables (wards, $n=7929$).

	Fruit & Veg	Obesity	Blood pressure	Cholesterol	Smoking	*Proportion of original variance*
(1) Synthetic estimates of male prevalence rates						
Unhealthy lifestyle 1	0.51	0.54	0.51	0.17	0.41	*0.63*
Unhealthy lifestyle 2	0.14	−0.08	−0.06	0.92	−0.36	*0.21*
PCA 3	0.17	−0.38	−0.45	0.22	0.76	*0.11*
PCA 4	−0.75	−0.10	0.50	0.26	0.33	*0.06*
PCA 5	−0.37	0.74	−0.54	0.13	0.10	*0.00*
(2) Synthetic estimates of female prevalence rates						
Unhealthy lifestyle 1	0.51	0.51	0.48	0.22	0.45	*0.66*
Unhealthy lifestyle 2	0.05	−0.31	−0.28	0.89	0.17	*0.19*
PCA 3	−0.26	0.27	0.46	0.39	−0.70	*0.11*
PCA 4	−0.70	−0.18	0.47	0.02	0.50	*0.04*
PCA 5	−0.42	0.73	−0.51	0.09	0.15	*0.00*

PCA3–PCA5 refer to the transformed variables that were not retained for further analysis.