J Geogr Syst (2011) 13:173–192
DOI 10.1007/s10109-010-0113-9

ORIGINAL ARTICLE

Geodemographics as a tool for targeting neighbourhoods in public health campaigns

Jakob Petersen · Maurizio Gibin · Paul Longley · Pablo Mateos · Philip Atkinson · David Ashby

Received: 11 April 2009 / Accepted: 1 April 2010 / Published online: 5 May 2010

Abstract Geodemographics offers the prospects of integrating, modelling and mapping health care needs and other health indicators that are useful for targeting neighbourhoods in public health campaigns. Yet reports about this application domain has to date been sporadic. The purpose of this paper is to examine the potential of a bespoke geodemographic system for neighbourhood targeting in an inner city public health authority, Southwark Primary Care Trust, London. This system, the London Output Area Classification (LOAC), is compared to six other geodemographic systems from both governmental and commercial sources. The paper proposes two new indicators for assessing the performance of geodemographic systems for neighbourhood targeting based on local hospital demand data. The paper also analyses and discusses the utility of age- and sex standardisation of geodemographic profiles of health care demand.

Keywords Geodemographics · Neighbourhood targeting · Public health · Hospital episode statistics

JEL Classification I18 · N30

J. Petersen · P. Longley (✉) · P. Mateos
Department of Geography and Centre for Advanced Spatial Analysis, University College London, Gower Street, Bloomsbury WC1E 6BT, UK
e-mail: plongley@geog.ucl.ac.uk

J. Petersen · M. Gibin · P. Atkinson
Southwark Primary Care Trust, 160 Tooley Street, London SE1 2TZ, UK

M. Gibin
School of Geography Birkbeck, University of London, Malet Street, London WC1E 7HX, UK

D. Ashby
Dr Foster (Research) Ltd., 12 Smithfield Street, London EC1A 9LA, UK

1 Introduction

The growing demand upon hospital and health care resources has followed a general trend across the economically developed world as a consequence of increasingly sedentary lifestyles, rising calorie supply per capita and ageing population (Wagner 1998; Bodenheimer et al. 2002; Department of Health 2004b; Saxena et al. 2006; Bleich et al. 2008). Hospital admissions are of prime concern because of their costs and social implications. Although hospital admissions only account for 19% of all contacts, they account for 58% of the NHS expenditure (Talbot-Smith and Pollock 2006). Patients with long-term or chronic diseases have the greatest needs: they represent only 5–10% of patients, but account for up to 55% of hospital bed days (Department of Health 2004b). Concerns about hospital admission rates have led to intense research into their upstream causes, and, especially, information regarding individual-level risk factors associated with long-term diseases is used for opportunistic ‘healthy lifestyle’ advice within the health care system as well as forming the basis for public health campaigns.

Public health campaigns are increasingly guided by the principles of social marketing, viz. balancing ‘product, place, price and promotion’ (Kotler and Zaltman 1971; Kotler et al. 2002). This paper focuses upon the use of geodemographics as a social marketing tool in order to create information about ‘place’ and to target local population with new public health campaigns.

The basic principles underpinning geodemographics, the analysis of people by where they live, have been comprehensively covered in recent publications (Sleight 2004; Harris et al. 2005), and this paper focuses on evaluating the utility of geodemographics in order to plan targeted public health campaigns within the jurisdiction of a local health authority, Southwark Primary Care Trust, London (UK). Geodemographics offers the prospects of integrating, modelling and mapping health care needs and other health indicators that are useful for the geographic targeting of such campaigns. Yet, the technology itself is not new, and sporadic reports about its potential remain just as boosterist in the recent past (Department of Health 2004a) as they were in 1985 (Speller and Hale 1985). One of the major obstacles has presumably been the lack of validation beyond the observation that it works sufficiently to be of interest for commercially driven, direct marketing (Vickers and Birkin 2007). The Office for National Statistics (ONS) has built publicly available geodemographic systems from Census data since the mid-1990s, but again reports of health applications have been sporadic (Openshaw 1995). The latest Census classification, the 2001 Census Output Area Classification (OAC), has been promoted in a more sustained manner, not least by appending it to the UK’s geographical look-up table, National Statistics Postcode Directory, as well as several national surveys. The ‘open source’ approach of the OAC system also offers opportunities for evaluating the system or modifying it for new purposes (Singleton and Longley 2008).

In what follows, geo-coded hospital admission data from Hospital Episode Statistics (HES) are used to map and evaluate the potential of geodemographics as means of targeting in the setting of the inner city London Borough of Southwark. The paper uses in the first instance the official Census Output Area Classification

(OAC) and compares it with a commercial geodemographic system, Mosaic UK, to illustrate strengths and limitations associated with the current geodemographic systems for public health applications. The paper then goes onto present a new open source geodemographic system for London based on Census data to compare seven competing geodemographic systems for targeting guided by hospital data and, finally, to explore the potential for age- and sex standardisation in a geodemographic analysis of hospital demand.

2 Problem definition

The ONS Output Area Classification (OAC) consists of a three-tier hierarchy with seven Supergroups subdivided into twenty-one Groups, again subdivided into a total of fifty-two Subgroups. Mapping OAC for Greater London reveals a pattern with large swathes of Inner London belonging to the single segment, the “Multicultural” Supergroup, and many other areas forming a mosaic of only sparsely represented segments (see Fig. 1). The OAC classification was devised to present a UK national geodemographic classification, yet Greater London’s special status in the settlement system is apparent in the heavy concentration of neighbourhoods into just two of the seven Supergroups.

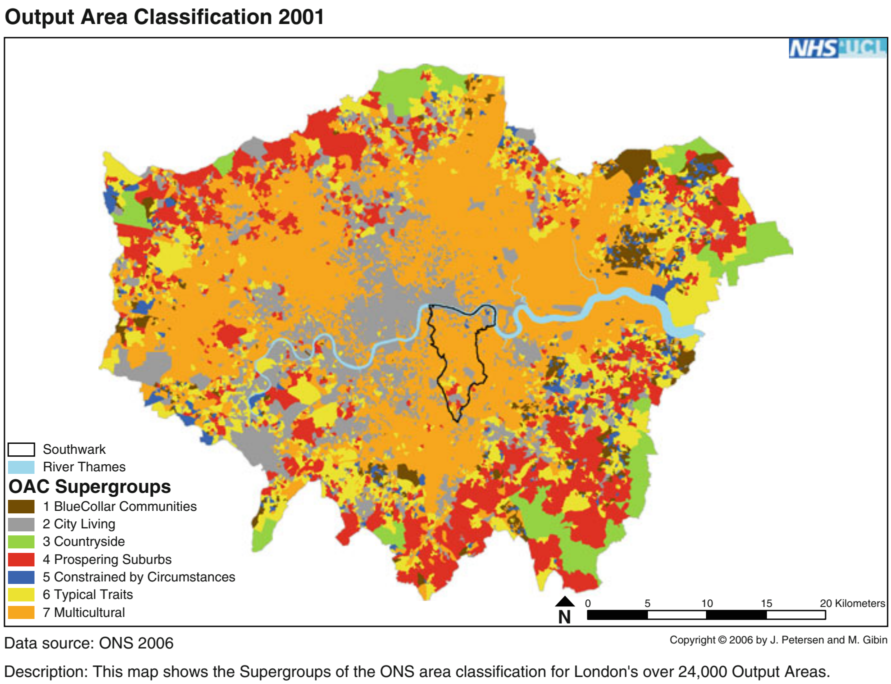

Fig. 1 Output Area Classification mapped for Greater London (OAC Supergroups)

Table 1 Output Area Classification profile of patients registered in Southwark PCT, April 2006

OAC supergroup	Freq.	Percent	Mean IMD score	Typical census attributes
1 Blue collar communities	7	<0.1	28	Terraced housing, renting publicly
2 City living	35,450	11.1	25	Higher education qualifications, single-person household (not pensioner), born abroad, renting privately, all flats
3 Countryside	0	0	–	2+ cars per household, working from home, agriculture/fishing employment, detached housing
4 Prospering suburbs	2,266	0.7	17	2+ cars per household, detached housing
5 Constrained by circumstances	2,297	0.7	43	All flats, renting publicly
6 Typical traits	3,743	1.2	17	Terraced housing
7 Multicultural	275,488	86.3	37	Renting privately or publicly, commuting to work on public transport, all flats, born abroad, South Asian or Black ethnic background
Total	319,251	100.0	36	

Sources: ONS, Indices of Multiple Deprivation (IMD) 2004 score (Noble et al. 2004)

The 'Multicultural' Supergroup is very dominant in Inner London, and indeed this neighbourhood Group accounts for some 86% of the 319,000 patients[1] registered in Southwark (Table 1). The dominance of a single segment continues in the Group classification—one tier down from Supergroups. The majority of patients in Southwark reside in "Afro-Caribbean Communities". Segmentation with the finest and third tier segments, the Subgroups, still reveals that 44% of patients reside in just one unnamed Subgroup (data not shown).

A similar picture repeats itself with commercial postcode level geodemographic systems such as Experian's (Nottingham, UK) Mosaic UK (Table 2). Although this system exhibit greater differentiation of neighbourhoods, the majority of Groups account for fewer than 1% of patients and the classification is dominated by a single very large neighbourhood segment with 49% of all patients, the "Welfare Borderline" Group (Table 2), and 48% in the second tier, "Metro-Multiculture" segment (data not shown).

The lack of specificity in these classifications makes them less suitable for targeted interventions, and this is the prime motivation for modifying the existing Census classification, OAC. The problem of specificity will in the following be discussed in a more general context in order to identify particular shortcomings that a modified version of the Census classification should address.

[1] A patient is here defined as anyone currently registered with a GP practice in the borough.

Table 2 Mosaic UK profile of patients registered in Southwark PCT, April 2006

Mosaic UK Group	Freq.	Percent	Mean IMD score	Description
A. Symbols of success	13,276	4.2	19	People with rewarding careers who live in sought after locations, affording luxuries and premium quality products
B. Happy families	1,162	0.4	39	Families with focus on career and home, mostly younger age groups now raising children
C. Suburban comfort	2,885	0.9	22	Families who are successfully established in comfortable, mature homes. Children are growing up and finances are easier
D. Ties of community	29,251	9.2	35	People living in close-knit inner city and manufacturing town communities, responsible workers with unsophisticated tastes
E. Urban intelligence	104,026	32.6	30	Young, single and mostly well educated, these people are cosmopolitan in tastes and liberal in attitudes
F. Welfare borderline	156,554	49.0	42	People who are struggling to achieve rewards and are mostly reliant on the council for accommodation and benefits
G. Municipal dependency	372	0.1	44	Families on lower incomes who often live in large council estates where there is little owner occupation
H. Blue collar enterprise	2,327	0.7	39	People who though not well educated are practical and enterprising and may well have exercised their right to buy
I. Twilight subsistence	3,593	1.1	39	Elderly people subsisting on meagre incomes in council accommodation
J. Grey perspectives	2,416	0.8	21	Independent pensioners living in their own homes who are relatively active in their lifestyles
K. Rural isolation	0	0.0	–	People living deep in the countryside in small communities little influenced by influx of urban commuters
L. Unclassified	3,389	1.1	34	Unclassified localities
Total	319,251	100.0	36	

Sources: ©Experian Ltd (Nottingham, UK), Indices of Multiple Deprivation (IMD) 2004 score (Noble et al. 2004)

First, most geodemographic systems are based on clustering of attribute variables with no acknowledgement of the relative geographical location of localities. As a consequence, they contain residual regional, or local, variation that may erode the usefulness of the construct per se, i.e. the *regionality problem* (Sleight 2004). Second, all national classifications tend to be represented regionally by one dominant neighbourhood type and a number of other types of exponentially falling representation, the *fragment problem* (Batty 2006). This causes problems for the robustness of the imputation of attribute values, where base population numbers are low. The fragment problem is exacerbated in health care applications because of the strong and ubiquitous age gradient generally observed to characterise health-associated conditions. Place effects may hence be confounded by differences in population age structure alone. To avoid this classic problem in health geography, health outcomes are typically standardised by age and sex. These standardisations, however, require further stratification of base population numbers that in many cases are already very small (see e.g. Table 1).

The regionality cum fragment problem is in other words manifested in attempts to generalise attribute space in a national classification, whereas Inner London Boroughs are distinctly different from most other locations in the United Kingdom on a variety of demographic and socio-economic factors. Clustering techniques such as *k*-means search at random for the optimal position of a predetermined number of 'seeds', *k*, that minimises the distance each neighbourhood has to the nearest seed when ordinated into a multi-dimensional attribute space spanned by the clustering variables. This process is sensitive to two 'forces'; (a) proximity in attribute space; and (b) outliers. In the case of Inner London, localities appear to cluster together more by being different from everywhere else (outliers) than by being uniquely similar (proximity).

There seems to be two immediate solutions to this problem. First, to segment clusters further recursively. This is notably the most common route of creating differentiation in geodemographic systems (Sleight 2004; Harris et al. 2005). The problem with this type of procedure is that it, if anything exacerbates the fragment problem, i.e. it creates even greater heterogeneity in base population sizes. Another solution, which will be pursued here, is to create a regional classification. Hypothetically, a regional classification would achieve greater differentiation *and* more evenness in base population structure simply by narrowing the attribute space to contain data from this region only; in this case Greater London.

3 Methods

The Output Area Classification (OAC) is based on a selection of forty-one Census 2001 variables ranging from age to ethnicity, family structure, tenure, education, occupation, transportation, and health (Vickers and Birkin 2007; Table 3). Localities (Census Output Areas in Great Britain) were first divided into seven OAC Supergroups using the *k*-means clustering algorithm. The stopping rule for the generation of these Supergroups was guided by mean centroid distance, which is a statistic that decreases with increasing number of seeds, *k*. Abrupt changes in this statistic are taken to indicate a parsimonious number of clusters. The final number of clusters is also guided by a more pragmatic consideration, i.e. the fact that it is inherently difficult for end users to differentiate between more than about seven classes when the classification is mapped (Callingham 2006).

In order to explore whether greater differentiation could be obtained for London neighbourhoods, the same Census variables were re-clustered according to the OAC methodology, but with three modifications to form an alternative regional classification, London Output Area Classification (LOAC hereinafter);

i. Data pertaining to the Greater London area only

Table 3 Census clustering variables

Clustering variables from Census 2001 used in the OAC and LOAC classifications	
Age 0–4	No central heating
Age 5–14	Rooms per household
Age 25–44	People per room
Age 45–64	HE qualifications
Age 65+	Routine/semi-routine occupation
Indian/Pakistani/Bangladeshi	2+ Car household
Black African	Public transport to work
Black Caribbean or Black Other	Work from home
Born outside UK	Standardised long-term illness ratio
Population density	Provide unpaid care
Divorced	Students (full time)
Single-person household (not pensioner)	Unemployed
Single-pensioner household (pensioner)	Working part-time
Lone parent household	Economically inactive looking after family
Two adult no children	Agriculture/fishing employment
Households with non-dependent children	Mining/quarrying/construction employment
Rent (public)	Manufacturing employment
Rent (private)	Hotel and catering employment
Terraced housing	Health/social work employment
Detached housing	Financial intermediation employment
All flats	Wholesale/retail employment

Source: Vickers and Birkin 2007

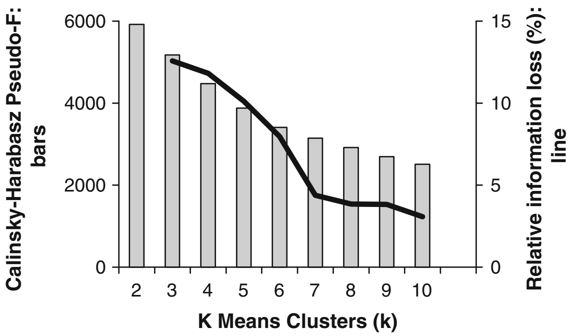

Fig. 2 Stopping rule applied in the creation of London Output Area Classification (LOAC). *Bars*: show Callinsky-Harabaz pseudo-F values on the left axis. *Line*: relative information loss (%)

ii. All variables were logarithmically transformed and their range was standardised according to the Greater London distribution

iii. The number of clusters in each group and their subdivision into lower tiers was guided by the appearance of distinct thresholds in the ratio of within- versus between-cluster variability, in accordance with the Callinsky-Harabaz pseudo-F (Rabe-Hesketh and Everitt 2004).

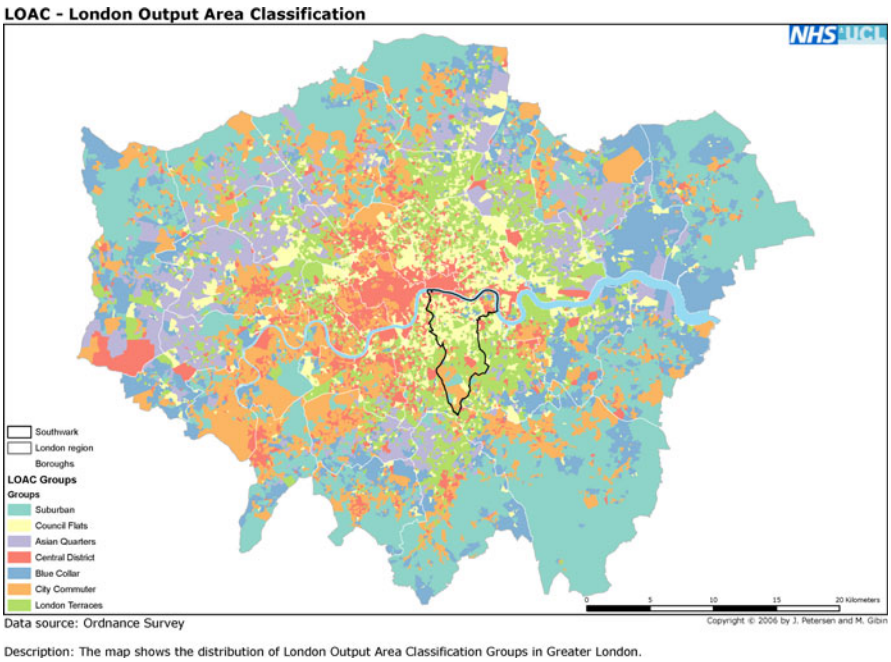

Fig. 3 London Output Area Classification (LOAC Groups)

Table 4 Geodemographic systems and neighbourhood differentiation

Aggregation level	Neighbourhood differentiation	Number of segments (Greater London)	Provider
Postcode	Mosaic UK Type	60	Experian Ltd
	Acorn Type	58	CACI
	Health Acorn Type	27	CACI
Output area	OAC subgroups	50	ONS
	LOAC groups	49	Internal
Super output area	Index of multiple deprivation	50	Department of Communities and Local Government
	P2 branches	38	Beacon-Dodsworth

NB only segments represented in Greater London are included; this therefore excludes, for example, the Mosaic UK segment 'Rural isolation'

Permission to use the commercial systems were kindly obtained through academic licence agreements between Centre for Advanced Spatial Analysis, UCL, and Experian Ltd. (Nottingham), CACI (London) and Beacon-Dodsworth (Bishopthorpe)

Seven Supergroups were formed according to a distinct threshold in the information structure (Figs. 2, 3). Each LOAC Supergroup was further subdivided into forty-nine Groups following the same procedure. This makes the LOAC Group tier comparable to the OAC Subgroup tier with fifty subgroups represented in Greater London (Table 4).

Table 5 Number of chronic disease admissions to hospital for residents of Greater London, Hospital Episode Statistics 2001–2004

Chronic disease indicator (ICD10)	Admissions
Angina pectoris (I20,I25)	119,538
Breast cancer (C50,D05)	91,026
Colorectal cancer (C17–C21)	87,988
All chest pain (R073–074,R101)	85,861
Back pain (M50–M54)	55,713
Mental health (F20–F48)	54,180
All arthroses (M15–M19)	51,332
Leukaemia (C91–C95)	49,963
COPD (J40–J44)	49,751
Stroke (I60–I69)	43,930
Lung cancer (C33–C34)	43,774
Asthma (J45–J46)	36,573
Congestive heart failure (I50)	32,759
Acute myocardial infarction (I21–I24)	31,458
Cholelithiasis (K80)	29,145
Traumatic brain injury (TBI)	26,333
Diabetes (E10–E14)	25,508
Skin cancer (C43–C44)	23,663
Epilepsy (G40–G41)	19,087
Prostate cancer (C61)	16,669
Cervical cancer (C53,D06)	10,620
Total	984,871

Source: Department of Health

Hospital episodes data for Greater London, between 2001 and 2004, were obtained with age, sex, residential postcode, and primary diagnosis. Ethical approval was obtained through Bromley Local Research Ethical Committee and Patient Information Advisory Group (PIAG).

The diagnoses (ICD10 four-character system) were categorised into indicators according to the most commonly occurring chronic diseases (Office for National Statistics 2000; Department of Health 2004b, 2005); see Table 5.

In order to estimate the success of a geodemographic targeting strategy, a goal of reaching the top 20% of admissions was set for each of the diseases and compared to reaching the same goal by a geographic targeting strategy of using the data themselves to identify the minimum number of neighbourhoods that supply at least 20% of patients for each of the diseases. This follows the logic of a design used in medical diagnostics, i.e. comparing the overlap in frequencies of patients with a positive diagnosis using a high intervention test (the gold standard) with an alternative and usually cheaper low intervention test (Kirkwood and Sterne 2003). In this study, geographic targeting was used as gold standard and geodemographics, targeting based on 'area-type', as the alternative test.

The hospital admissions data were aggregated at three geographical levels (unit postcode, output area, and super output area) and labelled with the finest level codes of eight different geodemographic systems, e.g. the fifty subgroups in the Output

Area Classification covering Greater London. The composite Index of Multiple Deprivation (IMD) was also included as a potential segmentation system (fifty quantiles according to IMD score in Greater London). Base population counts per unit postcode were derived from the 2001 Census and aggregated up to the different levels of geography using the National Statistics Postcode Directory (Office for National Statistics 2005).

The performance of the systems was evaluated by:

(1) Gini coefficients weighted by Census 2001 population counts. The Gini coefficient is an overall mathematical measure of heterogeneity (or inequality). Higher values of the Gini coefficient are associated with classifications that minimise the respective base population to target ratios. The Gini coefficient is, purportedly, the 'unofficial' industry standard for evaluating geodemographic systems (Callingham 2006).

(2) As an additional measure, two new performance indicators of targeting efficiency were created. They are both based on a ranking of geographic areas according to the crude rates for each disease indicator. The areas containing the top 20% of all frequencies for a given disease were flagged as 'targets'. These sets were treated as the diagnostic gold standard, and the same procedure was repeated with area types within each geodemographic system in order to create the alternative diagnostic sets.

    (a) For each disease indicator and geodemographic system, diagnostic *sensitivity* was calculated as the percentage of gold standard admissions (admissions targeted in the geographic set) included in the alternative geodemographic target set. Each system is hence evaluated against other systems for geographic targeting at the same level of aggregation, e.g. a unit postcode system such as Mosaic UK is compared to a geographic targeting scheme ordering postcodes according to their crude disease rates and likewise with a system coded at Census Output Area level, and so on.

    (b) The base population included in the geographic target of the top 20% of admissions (the gold standard) was divided by the number of admissions in the target for each disease to produce a *numbers-to-target* ratio. This was likewise repeated for the geodemographic target sets (the alternative standard).

3.1 Age-standardised geodemographic profile

Because LOAC as a regional classification has more even-sized segments than any of the other geodemographic systems, including OAC, it enables age standardisation of geodemographic risk profiles of, for example, London hospital admissions. Age standardisation is dependent upon detailed age-banded population estimates and many geodemographic systems are currently only delivered with total headcounts. LOAC is based on the Census geography and it is relatively simple to impute age-banded estimates for men and women from the Census itself or from the population estimates released by ONS annually at the Lower level Super Output Area (LSOA, a

LSOA typically contains five Output Areas) level. In this study, ONS age-banded population estimates for the same years as the HES extract (2001–2004) were used with the twist of assigning populations from the coarser lower layer super output area (LSOA) level data to the finer output area level (OA) proportionally to the population share of each OA in their respective LSOAs in the 2001 Census. ONS use different age bands for women and men in this data product: 0–15 years, 16–29 years, 30–44 years all for both sexes; and 45–59 years and 60+ years for women, and 45–64 year and 65+ year for men (for simplicity; only results for the analysis of the male population are reported here). Age-specific population-years-at-risk denominators were obtained by summing the population estimates over the 4 years of observation.

The age standardisation of hospital admission ratios, SARs, is derived following Kirkwood and Sterne (2003) with i as age band, p_i as population-at-risk and λ_i as the age-specific rate per population-at-risk;

$$\mathrm{SAR} = \frac{\text{Observed number of admissions}}{\begin{array}{r}\text{Expected number of admissions}\\ \text{were the age-specific rates the same}\\ \text{as in the reference population}\end{array}} = \frac{\sum a_i}{\sum \lambda_i p_i} \times 100$$

The data were analysed with an internal reference, i.e. the age-banded population of Greater London as reference. The equivalent crude admission ratios, CAR, were calculated following the same formula as for SAR only with a net crude admission rate based on headcounts instead. In order to highlight the potential effect of age standardising the geodemographic risk profile, crude and age-standardised ratios were compared in an attenuation index;

$$\text{Attenuation index} = \frac{\mathrm{SAR}}{\mathrm{CAR}} \times 100$$

If the age-standardised ratio is lower than the crude ratio, the effect of residing in a given neighbourhood type is—all other things being equal—explained by a higher proportion of elderly in that population (also known as effect attenuation). The opposite case would be expected where crude rates are lower due to a lower proportion of elderly—all else equal (also known as effect deattenuation).

4 Results

The regional Census neighbourhood classification had more evenly sized segments that were better differentiated on distinctive regional characteristics subsumed within the broader brush national classifications. Inspection of the mapped results of the OAC versus the LOAC clearly reveals that the inner city neighbourhoods are better differentiated using the LOAC (see Figs. 1, 3). In broad brush terms, this comparison revealed that the umbrella OAC “Multicultural” segment is replaced by three alternative LOAC segments, “2 Council Flats”, “3 Asian Quarters,” and “7 London Terraces”. The main discriminatory variables were those related to tenure and ethnicity (Table 6). Comparing two LOAC Supergroups, “2 Council Flats” versus "4 Central District" across all clustering variables revealed marked differences in age structure, family structure, ethnic composition, education, and occupational variables (Fig. 4).

Using LOAC, patients in Southwark would be differentiated into neighbourhood groups with 52% of patients in the dominant group, "2 Council Flats" in the first tier of divisions (Table 6) and only 14% in the second unnamed Group in the second tier. Within OAC, the largest fragment occupied as already mentioned 86% in first tier ("Multicultural") (Table 1) and 80% in second tier. More importantly, the regional classification avoided the fragment problem experienced with other systems (Tables 1, 2).

The Gini coefficients showed the segmentations with two general commercial systems, Acorn and Mosaic, to be best optimised relative to the base populations

Table 6 London Output Area Classification profile of patients registered with Southwark PCT, April 2006

LOAC supergroup	Freq.	Per cent	Mean IMD score	Typical census attributes
1 Suburban	3,346	1.1	16	Working age, white ethnic background, two-adult households, large houses, higher education, 2+ cars, routine jobs and part-time employment
2 Council flats	166,635	52.2	42	Children and young adults, black ethnic minorities and born abroad, divorcees, single non-pensioner households, lone parents, publicly rented accommodation, apartment blocks, routine jobs, long-term illnesses, unemployed, part-time or economically inactive looking after family
3 Asian quarters	227	0.1	21	Families, South Asian communities and people born abroad, two-adult households, terraced housing, routine jobs and part-time employment
4 Central district	28,297	8.9	25	Young adults, born abroad, singles and two-adult households, renting privately, apartment blocks, higher education
5 Blue collar	6,721	2.1	38	Families, white ethnic background, divorcees, lone parents, two-adult households, renting publicly, terraced housing, routine jobs and part-time employment, economically inactive looking after family
6 City commuter	19,889	6.2	20	Working age, born abroad, single and two-adult households, apartment blocks, terraced housing, renting privately, large houses, higher education, 2+ cars, part-time employment or economically inactive looking after family
7 London terraces	94,136	29.5	33	Young adults, black ethnic background and born abroad, single and two-adult households, renting publicly, apartment blocks, terraced housing, higher education, routine jobs, long-term illnesses, part-time employment or economically inactive looking after family
Total	319,251	100.0	36	

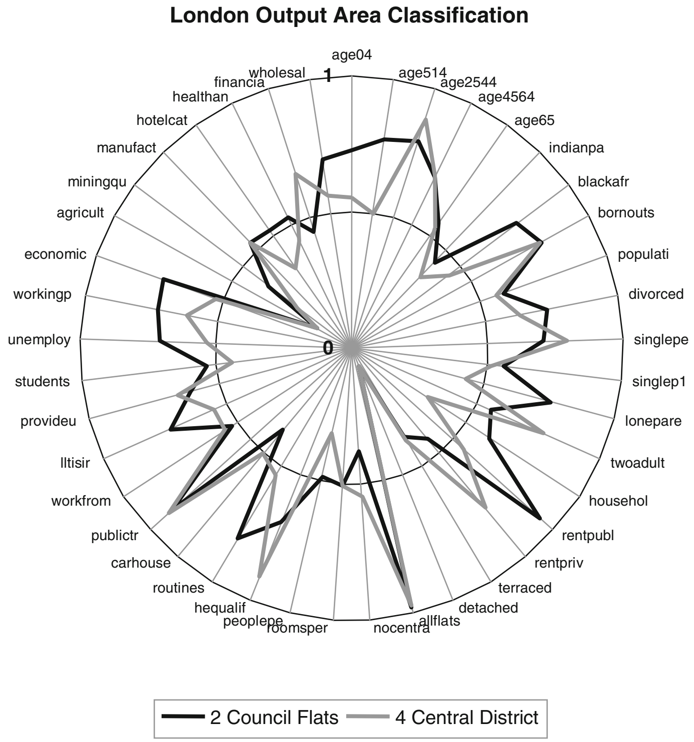

Fig. 4 Median Census attributes for two LOAC first tier clusters. Each attribute has been standardised to unity. For the full variable labels see Table 3

(Fig. 5). Comparable results was obtained with OAC (output area level); whilst the segmentation using the IMD performed the least well. Evaluating the systems relative to geographical targeting showed very low sensitivity overall. Postcode systems had the lowest sensitivity followed by output and super output area systems, respectively. The geodemographic strategies would in this study reach 20% of admissions, albeit not the same 20% as determined by the geographic targeting strategies. In fact, the proportional overlap, i.e. the sensitivity, could be as low as 20% and never exceeded 50%. Strategies using geodemographic systems at postcode level would potentially provide a cheaper means of reaching the target population because of the relatively low base populations indicated by the lower number-to-target indicator values.

The potential confounding of age and sex in geodemographic risk profiles of health outcomes was demonstrated in different ways. Detailed age-banded denominators were created using ONS population estimates for the same time period. The differences in age structure were clearly demonstrated across the different LOAC Groups, as evidenced by the occurrence of greater numbers of elderly people in suburban neighbourhood types (Fig. 6). The age structure of patients of key long-term diseases was, as expected, strongly associated with age. Congestive heart failure was, for instance, almost exclusively associated with the

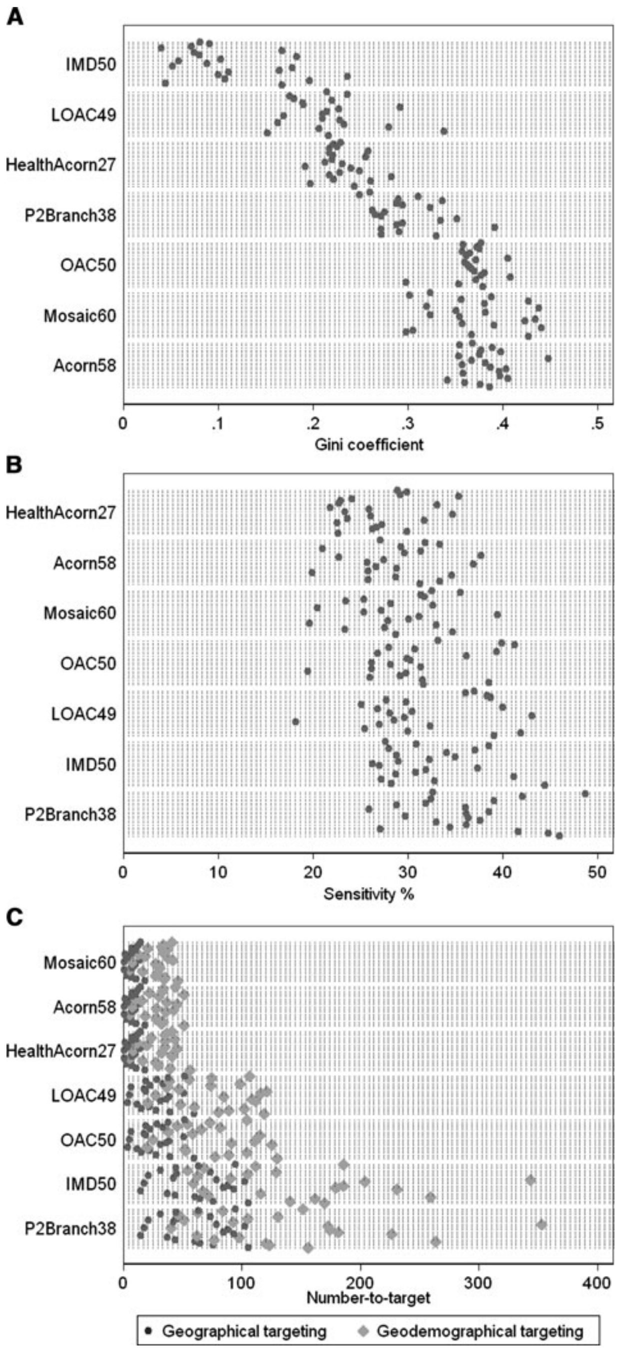

Fig. 5 **a** Gini coefficients for chronic diseases across seven different geodemographic systems in horizontal panels. **b** Sensitivity relative to geographic targeting as gold standard. **c** Number-to-target of same

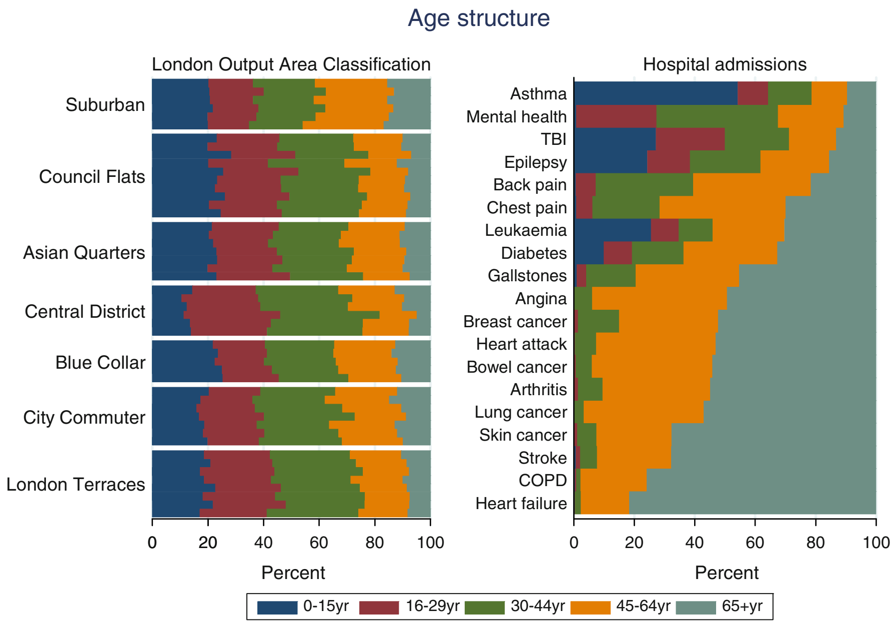

Fig. 6 Age structure in LOAC Groups and hospital admission data. Male population of Greater London 2001–2004

oldest age groups(only results for men are shown). Conversely, asthma predominantly affected children (Fig. 6).

In campaigns that are more concerned with upstream causes, age standardisation may thus be an important addition to geodemographic profiling. The usual crude profiles of angina, for example, showed that the "1 Suburban" Supergroup neighbourhood types had crude admission ratios above the regional average (CAR = 100), although when standardised for age, they clearly fell below average (SAR = 100) (Fig. 7). The effect of a particular geodemographic neighbourhood type, as a container for lifestyle factors, was in this case confounded by a higher proportion of elderly; the geodemographic 'lifestyle effect' was in other words attenuated by age. Attenuation as well as de-attenuation effects were demonstrated across key long-term diseases for both women and men (Fig. 8). These effects were, as expected, strongest with diseases associated with either many or few elderly relative to other age groups.

5 Discussion

Besley and Kanbur (1990) propose that given scarce resources, geographical targeting should favour areas in order of need until the available budget is exhausted. Any targeted strategy based on aggregated data, however, opens up issues of inclusion and exclusion. A public health campaign strategy, for example,

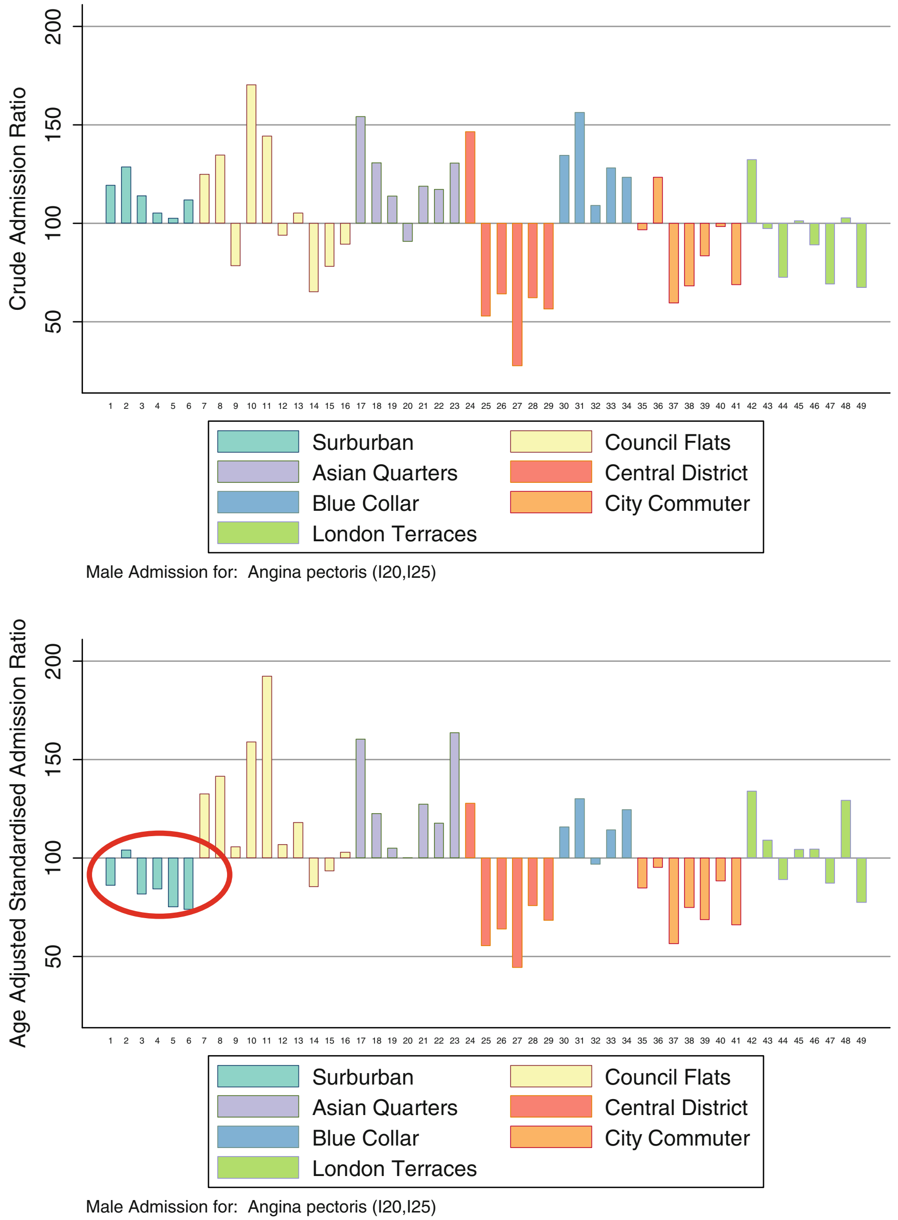

Fig. 7 Geodemographic profile (LOAC) for Angina pectoris hospital admissions. *Top*: Crude admission ratios. *Bottom*: Age-standardised ratios. NB Shift in the Suburban category with age standardisation

may include individuals who are not at risk for the health outcome it was designed to counter or ameliorate; i.e. the problem of inclusion. Conversely, there may be citizens with those exact needs that are excluded by the strategy simply by having the 'wrong' postcode, i.e. the exclusion problem. There seems no immediate

LOAC Supergroup	LOAC Group	Diabetes (E10-E14)	Mental health (F20-F48)	Epilepsy (G40-G41)	Angina pectoris (I20,I25)	Stroke (I60-I69)	Congestive heart failure (I50)	Acute myocardial infarction (I21-I24)	All arthroses (M15-M19)	Asthma (J45-J46)	COPD (J40-J44)	All chest pain (R073-074,R101)	Back pain (M50-M54)	Cholelithiasis (K80)	Colorectal cancer (C17-C21)	Leukaemia (C91-C95)	Lung cancer (C33-C34)	Prostate cancer (C61)	Skin cancer (C43-C44)	Traumatic Brain Injury
1	1	83	103	94	72	71	68	72	73	97	69	81	86	77	72	84	71	69	71	99
	2	89	103	96	81	81	80	81	82	98	80	88	92	85	81	90	80	79	81	99
	3	82	104	93	72	71	68	72	72	96	69	81	86	77	71	84	70	68	71	99
	4	88	104	95	80	80	78	80	81	95	79	87	91	84	80	89	79	78	80	99
	5	84	103	94	73	72	70	73	74	98	70	82	87	78	73	85	72	70	73	99
	6	78	104	92	66	65	63	66	67	97	63	76	83	72	66	81	65	63	65	98
2	7	103	104	100	106	105	104	106	105	94	104	106	106	105	106	100	106	105	105	99
	8	103	99	101	105	102	101	104	104	101	101	104	104	103	104	102	104	102	102	99
	9	116	110	101	135	137	143	135	135	86	141	124	120	129	136	105	137	142	138	97
	10	96	100	99	93	90	88	93	92	100	89	97	98	94	92	96	92	89	90	99
	11	115	106	103	133	129	130	132	130	91	131	124	121	126	132	107	133	132	130	97
	12	107	103	101	114	112	113	113	113	95	113	110	108	111	114	103	114	114	113	99
	13	106	102	101	112	110	110	112	111	96	111	109	108	109	112	103	112	111	110	99
	14	114	106	101	131	133	137	131	130	90	136	120	116	125	132	106	133	137	133	98
	15	110	99	103	119	120	122	119	118	101	122	111	107	114	120	108	121	123	120	100
	16	107	105	100	115	115	116	115	115	92	116	111	109	112	115	102	116	116	115	98
3	17	102	102	100	104	107	108	105	105	98	107	103	103	105	105	101	105	107	107	99
	18	97	101	99	94	94	93	94	94	99	93	97	99	96	94	97	93	93	94	99
	19	96	103	98	92	93	93	93	93	96	93	95	97	95	93	95	92	93	93	99
	20	105	103	100	110	113	116	111	111	95	115	107	105	109	111	102	112	115	113	99
	21	104	104	100	107	108	109	107	107	94	108	106	106	107	107	100	107	108	108	98
	22	100	100	100	100	101	102	101	101	101	102	100	100	101	101	101	101	102	101	100
	23	113	103	103	125	128	132	126	125	96	131	117	113	121	126	107	127	131	128	99
4	24	93	94	100	87	84	82	87	86	113	83	90	92	88	86	98	86	83	84	102
	25	103	89	106	105	105	107	105	104	128	107	99	95	101	105	112	106	107	105	105
	26	100	91	104	100	100	101	100	99	121	101	97	95	98	100	108	100	101	100	104
	27	127	88	113	161	169	188	161	158	129	184	125	112	139	163	135	170	186	168	105
	28	111	91	107	122	124	129	122	121	119	129	109	103	115	123	117	125	129	124	104
	29	110	91	106	121	123	128	120	120	118	127	108	102	113	121	116	123	128	123	104
5	30	92	103	96	86	85	83	86	86	95	83	92	95	89	86	92	85	83	85	99
	31	90	106	95	83	80	77	83	83	92	78	91	95	86	82	88	81	78	80	97
	32	93	103	97	89	89	88	89	89	95	88	93	95	91	89	93	88	88	89	99
	33	94	108	96	89	87	85	89	89	90	85	95	99	92	89	90	88	85	87	97
	34	100	107	98	101	100	99	101	101	90	99	103	104	102	101	96	100	99	100	98
6	35	93	101	97	88	87	86	88	88	99	86	92	93	90	88	94	87	86	87	100
	36	86	98	96	77	75	72	77	77	104	73	84	88	80	76	89	76	73	75	100
	37	97	94	101	95	96	97	95	95	110	97	94	93	95	95	102	95	97	96	103
	38	105	95	103	110	112	115	110	110	109	114	103	99	106	110	108	111	114	112	103
	39	90	99	97	82	82	81	83	83	102	81	88	91	85	82	92	82	81	82	100
	40	94	98	99	90	89	87	90	90	104	88	93	95	92	90	96	89	88	89	100
	41	98	99	99	96	99	101	97	97	100	100	95	95	97	97	99	97	100	99	101
7	42	101	98	101	101	100	99	101	100	104	100	101	100	100	101	102	101	100	100	101
	43	106	100	101	112	115	117	113	112	99	117	106	104	109	113	105	114	117	115	101
	44	111	95	105	123	125	129	122	122	109	128	111	106	116	123	113	125	129	124	102
	45	102	98	101	103	103	103	103	103	104	103	102	101	102	103	103	103	103	103	101
	46	109	102	102	118	118	120	117	117	96	120	111	108	114	118	105	119	120	118	99
	47	113	96	105	126	128	134	126	125	107	133	113	107	118	127	114	129	134	128	102
	48	113	101	103	126	128	132	126	125	98	131	116	112	120	127	109	128	131	128	100
	49	108	95	104	115	118	121	115	115	109	121	106	101	110	116	111	117	121	117	103

≤ 80 Attenuation - Age-standardisation decreases admission ratio
≥ 120 Deattenuation - Age-standardisation increases admission ratio

Fig. 8 Effect of age standardisation of hospital admission ratios for key long-term diseases. Male patients in Greater London 2001–2004. *Shading* illustrates effects 20 percent below (*light grey*) or above the crude ratios (*dark grey*)

solution to this problem. Two different geodemographic systems would not suggest the same prediction for a local area simply due to differences in the methodology used to generalise complex, multivariable data. Different zonations as well as systems will invariably lead to different local estimations, as manifest in the Modifiable Area Unit Problem (Openshaw 1984). As a consequence, different areas and populations are selected and excluded in targeted campaigns.

The empirical analysis presented here suggested that, with respect to London, all of the geodemographic systems were relatively poor discriminators in comparison with geographic targeting (in this case: targeting areas rather than area types with the highest crude admission rates). This exemplifies exclusion problems: geodemographic allocation strategies would still reach 20% of admissions, albeit not the same 20% displaying the highest needs as determined by geographic targeting. The relatively high number-to-target ratios demonstrate the inclusion problems in both types of targeting, although geodemographic strategies would be more expensive to deploy, i.e. in terms of mail shots or other campaign means magnified by base population numbers.

The results of the diagnostic approach deployed here also suggest that, for these health indicators, it is the geographic order of aggregation (unit postcode, Output Area, or Super Output Area), more than the geodemographic classifications themselves, that is critical for the accuracy of targeting. This also questions whether Gini coefficients, however widely used, are in fact too sensitive to the huge within-region variability in base population sizes. If this is indeed the case, their use becomes little more than a measure of population size heterogeneity rather than a measure of actual targeting 'efficiency'.

In this paper, two new performance indicators of targeting efficiency have been proposed in order to quantify these problems: sensitivity (the degree of overlap with a geographic targeting alternative) and the number-to-target (the sum of target and non-target population included in a campaign relative to the target). The (commercial and government) geodemographic systems evaluated in our London case study of hospital admissions proved to be rather insensitive to local conditions, because they were created using UK-wide Census and other data. In this way, there was only modest evidence of zone effects (in effect, the choice between classifications is not an important issue), whilst scale effects were clearly evident for the number-to-target criteria. Thus, our results confirm that scale of analysis does matter! In conclusion, the more fine-scale geodemographic systems were superior, not in sensitivity which in general was lower, but by including a lower number of non-target population (people not eligible for the campaign) in their target. The number-to-target is thus of relevance for the evaluation of campaigns intensified by mail shots, interviews or other methods involving direct contact to either households or individuals.

Geodemographic profiles of health outcomes revealed stark differences in the apparent health care needs of different populations. Geodemographics can in this way be said to project the same kinds of health inequality problems exposed by area deprivation scores or social class measures. Yet, compared to deprivation measures, geodemographics present a richer tapestry of potential factors that at least hypothetically can explain the emergence of health problems. Labels such as "social class IV" or "IMD score 27.9", on the other hand, seems less relevant to the interpretation of complex health information (Longley 2005).

The techniques presented here demonstrate how geodemographics can be used to differentiate neighbourhoods on an array of demographic and socio-economic variables. It is important to note that the strength of geodemographics is to *explore* or describe, but not necessarily to *explain* particular health outcomes. Geodemographics does in other words nor replace carefully designed epidemiological studies aimed at oncovering individual-level risk factors. Geodemographics does, however, offer the social marketer a wealth of demographic information that can help define 'product, place, price, and promotion' for a given health campaign.

Age standardisation of geodemographic profiles is rare, possibly because it requires age-banded denominator data and robust estimates of risk across the different geodemographic segments, which again is hindered by small-number problems experienced when the current national geodemographic systems are applied locally, that is the fragment problem. The empirical example with geodemographic profiles of hospital admission data demonstrated that age-banded populations-at-risk denominators were achievable using current official population estimates. It was furthermore demonstrated that robust estimates of hospital admission risk were achievable when using a new, regional geodemographic system with more evenly sized base populations. The value of age standardisation was evident especially from geodemographic profiles of long-term diseases associated with elderly patients. The elderly with poorer health would for instance be 'hidden' if residing in areas with a younger-than-average age structure. Suburban neighbourhoods would by the same token 'qualify' for targeting, although the implied association with 'upstream' policy variables would be confounded by differences in the base population age structure alone.

This paper is not written with the aim of boosting geodemographics as a single solution, but to present it as a technique that can give valuable demographic context to many public sector applications, where social class and deprivation indices project similar inequalities, yet typically remain harder to interpret and operationalise in local planning. The modification of the national Output Area Classification for Greater London illustrates how bespoke geodemographic systems can assist in creating better differentiation on regional neighbourhood characteristics and, as of special interest to the health geography domain, enable analyses that are sensitive to the 'worst disease of all', age.

References

Batty M (2006) Rank clocks. Nature 444:592–596

Bleich S, Cutler D, Murray C, Adams A (2008) Why Is the developed world obese? Ann Rev Public Health 29:273–295

Bodenheimer T, Wagner H, Grumbach K (2002) Improving primary care for patients with chronic illness: the chronic care model, Part 2. JAMA 288(15):1909–1914

Callingham M (2006) Exploring the use and value of the ONS output area classification. http://www.rss.org.uk/rssadmin/uploads/577303_Callingham%20Descriptions%203%20Oct%202006.pdf. Accessed 3 April 2009

Department of Health (2004a) Choosing health: making healthier choices easier. Department of Health, London

Department of Health (2004b) Improving chronic disease management. Department of Health, London
Department of Health (2005) National service framework for long-term conditions. Department of Health, London
Harris R, Sleight P, Webber R (2005) Geodemographics, GIS and neighbourhood targeting. Wiley, Chichester
Besley T, Kanbur, R (1990) The principles of targeting. Policy Res Working Paper Series 385
Kirkwood BR, Sterne JAC (2003) Essential medical statistics, 2nd edn. Blackwell, Malden
Kotler P, Zaltman G (1971) Social marketing: an approach to planned social change. J Mark 35(3):3–12
Kotler P, Roberto N, Lee N (2002) Social marketing: improving the quality of life. SAGE, London
Longley P (2005) A renaissance of geodemographics for public service delivery. Prog Hum Geogr 29(1):57–63
Noble M, Wright G, Dibben C, Smith G, McLennan D, Anttila C, Barnes H (2004) Indices of Deprivation 2004. Office of the Deputy Prime Minister, London
Office for National Statistics (2005) All fields postcode directory. London
Office for National Statistics (1998) Key health statistics from general practice 1998. Analyses of morbidity and treatment data, including time trends, England and Wales. Office for National Statistics, London
Openshaw S (1984) The modifiable areal unit problem. Concepts and techniques in modern geography. Geo Books, London
Openshaw S (1995) Geodemographic segmentation systems for screening health data. J Epidemiol Community Health 49:S34–S38
Rabe-Hesketh S, Everitt B (2004) A handbook of statistical analyses using Stata. Chapman & Hall, CRC, Boca Raton
Saxena S, George J, Barber J, Fitzpatrick J, Majeed A (2006) Association of population and practice factors with potentially avoidable admission rates for chronic diseases in London: cross sectional analysis. J R Soc Med 99(2):81–89
Singleton A, Longley P (2008) Creating open source geodemographics–refining a national classification of Census Output Areas for applications in higher education. Pap Reg Sci 88(3):643–666
Sleight P (2004) Targeting customers. How to use geodemographics and lifestyle data in your business. WARC, Henley-on-Thames
Speller V, Hale D (1985) Making the most of your postcode. Health Serv J 28:53
Talbot-Smith A, Pollock AM (2006) The new NHS—a guide. Routledge, London
Vickers D, Birkin M (2007) Creating the UK National Statistics 2001 output areas classification. J R Stat Soc 170(2):379–403
Wagner H (1998) Chronic disease management: what will it take to improve care for chronic illness? Eff Clin Pract 1(1):2–4