Inclusion and exclusion problems in geodemographic targeting

Jakob Petersen[12], Pablo Mateos[1], Paul Longley[1], Phil Atkinson[12]

[1]Centre for Advanced Spatial Analysis, Department of Geography, University College London, Gower Street, WC1E 6BT, Tel. +44 20 7525 0267, Email j.petersen@ucl.ac.uk
[2]Southwark Primary Care Trust, Public Health, Neckinger, London SE16 3QN

KEYWORDS: Targeting, Resource Allocation, Incomplete Knowledge, Hospital Episode Statistics, Geodemographics

Introduction

Most developed countries have experienced a gradual rise in health care expenditure since the 1960s (Organisation for Economic Cooperation and Development 2005). With ageing populations, this 'health care inflation' is predicted to worsen with fewer of working age to support the funding of health care systems that at the same time are expected to meet a growing demand for their services (Gray 2005). The UK National Health Service (NHS) is principally funded by the government and as such is central to the political debate not only about taxation, but also in questions of its share of the state budget in competition with other public sectors like education, policing, social security, defence, etc. (Hsiao and Heller 2007). Within the health care system itself, there are complicated trade-offs between equity and efficiency objectives with respect to improving population health, reducing risks and inequalities, and a need to ration services balanced with maintaining a certain level of user satisfaction (Musgrove 2003). There is consequently increasing pressure to reform health care systems and the NHS is undergoing reforms to make its organisation more cost-effective and to attract private enterprise (Talbot-Smith and Pollock 2006, Pollock et al. 2007). Unsurprisingly this development has resulted in a plethora of new health management solutions and planning tools. One planning tool that has garnered much interest is the use of geodemographic systems for health equity assessments and public health campaigns (Webber 2004, Farr and Evans 2005, Jones et al. 2005). Geodemographic systems are neighbourhood classifications that for three decades have been used for direct marketing and market research in the commercial sector (Webber and Craig 1978, Harris et al. 2005, Longley 2005). There are a number of competing systems and most now present extensive support material for demographic profiling and neighbourhood targeting across the public sector. The power of geodemographics in the commercial sector clearly lies in estimating demand under *incomplete* knowledge of the market, e.g. by coding the 25,000 annual responses to the Target Group Index surveys with a geodemographic system, it is possible to create maps of demand for a wide range of consumer goods and services across the whole country. Similar ideas have been proposed for the mapping of health care demand, and several systems offer health need indices based on Hospital Episode Statistics (HES: (Webber 2004, Farr and Evans 2005).

One fundamental difference between the private and public sector applications, however, may be that the NHS has already invested considerably in HES and other electronic health care databases and arguably has *complete* knowledge of the health care demand to a very fine level of geography, i.e. the user postcode. In this paper we will investigate how well a number of geodemographic systems facilitate prediction of observed demand for hospital admissions for a number of the most common chronic diseases. This work adds to the more traditional ways of evaluating geodemographic systems with the use of Lorenz curves and Gini coefficients (Novak et al. 1992). We will discuss the findings in relation to targeting of public health campaigns.

Method

To estimate the success of a geodemographic targeting strategy we set a hypothetic goal of reaching the top 20% of admissions for a given disease and compared it to reaching the same goal by a geographic targeting strategy. To do this we used a simple 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 we used geographic targeting as gold standard and geodemographic as the alternative test.

All admissions (spells) at hospital for a number of common chronic diseases were obtained at postcode level for Greater London, for the years 2001-2004 (Table 1).

Table 1. Number of chronic disease admissions to hospital for residents of Greater London, 2001-2004

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

The 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[26], e.g. the 50 subgroups in the Output Area Classification covering Greater London. We also included Indices of Multiple Deprivation (IMD) as a potential segmentation system. To do this we have divided Greater London into a comparable number of segments (we chose 50 quantiles) according to IMD score. The

[26] Postcode level: 1. Mosaic UK Type (60 segments in Greater London); 2. Acorn Type (58); 3. Health Acorn Type (27). Output Area level: 4. Output Area Classification Sub Groups (50); 5. London Output Area Classification Groups (49). Lower Layer Super Output Area level: 6. Index of Multiple Deprivation in 50 local quantiles according to scores (50); 7. a bespoke classification based on HES diagnostic groups (48); 8. P2 Branches (38).

performance of the systems was evaluated by: (1) Gini coefficients weighted by Census 2001 population counts. The Gini coefficient is an overall mathematical measures of inequality. By weighting target frequencies with base population frequencies across all segments in a segmentation, it is intended to give higher values to a system that minimises base population to target; and (2) ranking geographic areas according to crude rate for each disease indicator and selecting those containing the top 20% of all frequencies. These sets were treated as the diagnostic gold standard and the same procedure was repeated with area types within each geodemographic system to create the alternative diagnostic sets. For each disease indicator and geodemographic system, diagnostic sensitivity was calculated as the percentage of gold standard admissions included in the geodemographic target. The base population included in the geographic target of the top 20% of admissions (the gold standard) were 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.

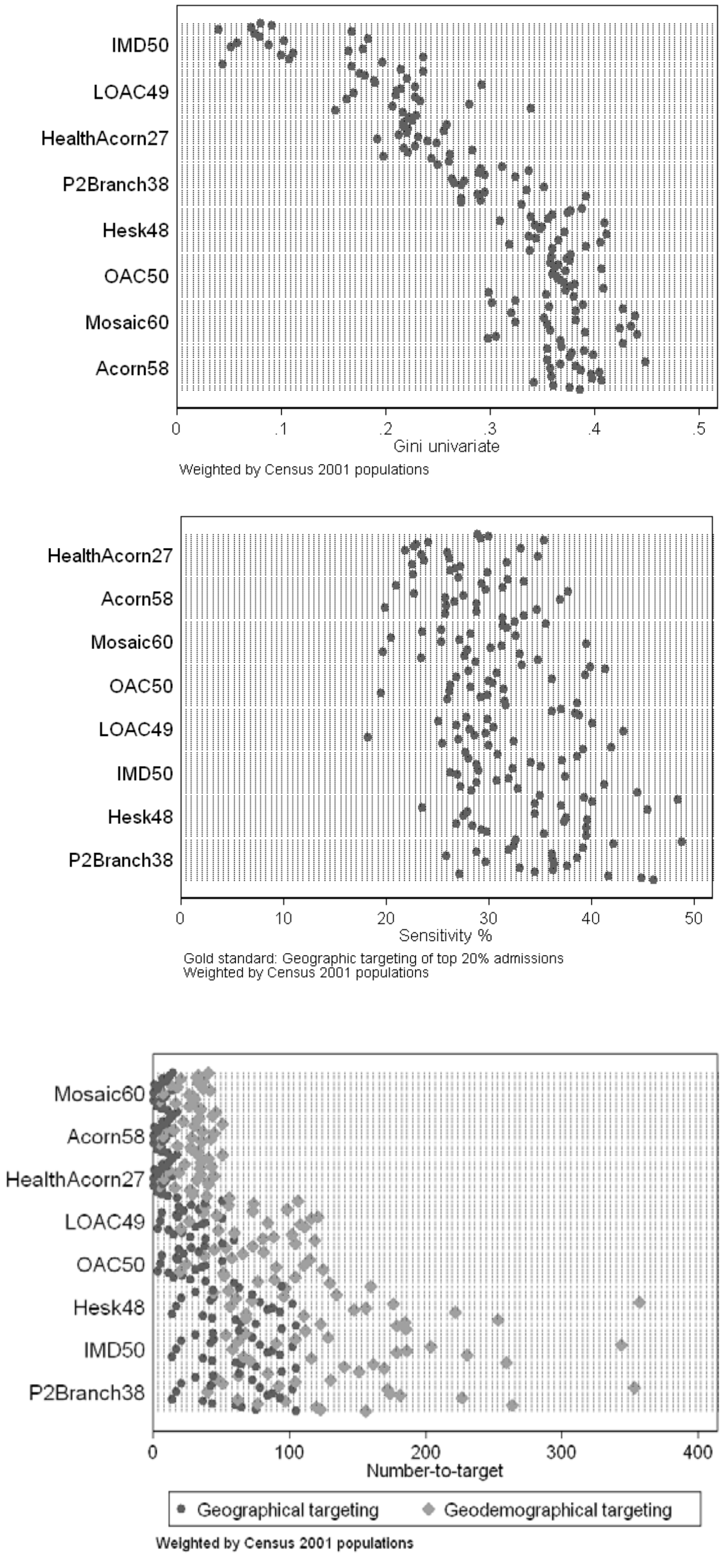

Figure 1 Top: Gini coefficients for chronic diseases across eight different geodemographic systems in horizontal panels. Middle: Sensitivity relative to geographic targeting as gold standard. Bottom: Number-to-target for same

.

Results

The Gini coefficients showed the segmentations with two general commercial systems, Acorn and Mosaic, to be best optimised relative to the base populations (Figure 1). Comparable results were obtained with OAC (output area level) and HESK[26] (super 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 overall followed by output and super output area systems. 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.

Discussion

Besley and Kanbur (1990) proposed that given scarce resources geographical targeting should favour areas in order of need until the budget is exhausted (Besley and Kanbur 1990). Any targeted strategy based on aggregated data however opens up for problems of inclusion and exclusion. A public health campaign strategy, for example, may include individuals who are not at risk for the health outcome it was designed to counter or ameliorate; i.e. 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. Both these problems pertain to efficiency and fairness considerations of a strategy and should be evaluated in line with other welfare policy interventions (Pellegrino and Thomasma 2004).

Our empirical analysis suggests that all of the geodemographic systems had a low sensitivity in comparison with geographic targeting. 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. High numbers of base population to target (indicated by number-to-target) demonstrates the *inclusion* problems in both types of targeting; although geodemographic strategies would be more expensive to deploy, e.g. in terms of mailshots or other campaign means magnified by base population numbers.

The results of the diagnostic approach deployed here also suggest that 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 popular, are too sensitive to the huge variability in base population size within the geodemographic classifications and applied in this way hence become a measure of this variability rather than of actual targeting 'efficiency'.

In evaluating geodemographic systems for the targeting of public health campaigns we need to consider two different situations. First, cases where we have data on actual demands and geographic targeting would thus be more accurate, more fair and potentially less costly for campaigning than a geodemographic alternative. Second, cases where we would like to predict lifestyle information. In theses cases geodemographic systems has potential value. Choosing a geodemographic system will rely on a number of factors, including budget and health data quality. This study suggests that postcode systems are not necessarily more accurate, but that they would be cheaper to deploy in campaigns, which takes us back to the central dilemma in health care delivery: to balance equity and efficiency.

Acknowledgements

We would like to thank the ESRC and DTI for support through Knowledge Transfer Partnership, no. 666. We are further indebted to Bromley Local Research Ethics Committee, Security and Confidentiality Advisory Group (SCAG) and Patient Information Advisory Group (PIAG) for permitting access to Hospital Episode Statistics data. We would also like to thank Experian Ltd., CACI and Beacon-Dodsworth for permission to use their geodemographic directories.

References

Besley, T., and R. Kanbur. 1990. The principles of targeting. Policy Research Working Paper Series **385**.

Farr, M., and A. Evans. 2005. Identifying 'unknown diabetics' using geodemographics and social marketing. Interactive Marketing **7**:47-58.

Gray, A. 2005. Population ageing and health care expenditure. Ageing Horizons **2**:15-20.

Harris, R., P. Sleight, and R. Webber. 2005. Geodemographics, GIS and neighbourhood targeting. John Wiley & Sons Ltd., Chichester.

Hsiao, W. C., and P. S. Heller. 2007. What macroeconomists should know about health care policy. International Monetary Fund, Washington DC.

Jones, C. E., P. Mateos, P. Longley, and R. Webber. 2005. Who does not eat their greens? Geodemographics, health promotion and neighbourhood health inequalities. Proceedings to 13th GISRUK Conference **13**.

Kirkwood, B. R., and J. A. C. Sterne. 2003. Essential Medical Statistics, 2nd edition. Blackwell Science, Malden.

Longley, P. A. 2005. A renaissance of geodemographics for public service delivery. Progress in Human Geography **29**:57-63.

Musgrove, P. 2003. Health economics in development. World Bank, Washington DC.

Novak, T. P., J. Leeuw, and B. MacEvoy. 1992. Richness curves for evaluating market segmentation. Journal of Marketing Research **19**:254-267.

Organisation for Economic Cooperation and Development. 2005. Health data. OECD, Paris.

Pellegrino, E. D., and D. C. Thomasma. 2004. The good of patients and the good of society: striking a moral balance. *in* M. Boylan, editor. Public health policy and ethics. Kluwer Academic Publishers, Dordrecht.

Pollock, A. M., D. Price, E. Viebrock, E. Miller, and G. Watt. 2007. The market in primary care. British Medical Journal **335**:475-477.

Talbot-Smith, A., and A. M. Pollock. 2006. The New NHS - a guide. Routledge, London.

Webber, R. 2004. Neighbourhood inequalities in the pattern of hospital admissions and their application to the targeting of health promotion campaigns. CASA working papers **90**.

Webber, R., and J. Craig. 1978. Socio-economic classifications of local authority areas. Studies in Medical and Population Subjects **35**.

Biographies

Jakob Petersen is working on a knowledge transfer partnership between Southwark Primary Care Trust and Centre for Advanced Spatial Analysis, Department of Geography, UCL. His research interests concerns public health targeting and the use of health care databases. He has an MSc in GISc and Biology and is studying for a PhD in health geography.

Dr. Pablo Mateos is lecturer in Human Geography at the Department of Geography in UCL. His research interests lie within Population and Urban Geography, and focus on investigating new ontologies of ethnicity, migration and mobility in today's rapidly changing cities and societies.

Professor Paul Longley, Department of Geography, UCL. Paul Longley's research interests concerns the use of GIS and quantitative methods in urban analysis. He is Editor-in-Chief of Computers, Environment and Urban Systems and reviews editor of Environment and Planning B as well as Deputy Director of Centre for Advanced Spatial Analyses at UCL.

Dr. Phil Atkinson is Public Health consultant in infectious disease epidemiology and mental health. He is head of Research & Development at Southwark Primary Care Trust and also directs the Research Development Centre for primary, community and social care research in southeast London.