Geodemographics and spatial microsimulation: using survey data to infer health milieu geographies

Jens Kandt*[1]

[1]Department of Geography, University College London

9 January 2015

Summary

This paper presents an approach to infer lifestyle geographies from survey microdata as building block of purpose-built geodemographics. 33,000 England and Wales residents have been clustered into nine lifestyle milieus based on a range of behavioural and attitudinal variables. The milieus strongly differ by individual social and demographic circumstances. Spatial microsimulation can be used to estimate probabilistically the geographical distribution of milieus. Preliminary results for London are presented in this abstract, demonstrating how extensive behavioural information of social surveys can be combined with the nearly complete coverage of spatial census data within a geodemographics framework to inform policy interventions.

KEYWORDS: geodemographics, spatial microsimulation, health behaviours, urban lifestyles, milieus

1 Introduction

So-called health behaviours and their spatial manifestation have long been of interest to social epidemiologists and health geographers, because while they appear to significantly affect population health, they are potentially modifiable [Blaxter (1990, 2010)]. But as much of conventional health geography and social epidemiology focusses on the role of objective measures of social similarity in shaping health, theoretical social science suggests that individual subjective orientations and experiences play an important part in shaping health behaviours [Veenstra and Burnett (2014); Baum and Fisher (2014); Williams (1995), calling for a stronger focus on lifestyles in health research.

Geodemographics - the arts of classifying local areas by the characteristics of their residents - has long been discussed as tool to infer lifestyle milieus at the ecological level and in so doing inform strategic public health interventions [Abbas et al. (2009); Openshaw and Blake (1995)]. But traditionally, geodemographic classifications have been rather generic, little conceptually targeted

*j.kandt.12@ucl.ac.uk

and strongly rely on objective population characteristics [Longley (2005); Singleton and Longley (2009); Voas and Williamson (2001)].

This paper summarises work that is underway to incorporate subjective orientations and lifestyle aspects into health geodemographics. The work involves the integration of extensive, individual-level social survey data with nearly complete-coverage census neighbourhood statistics through a combination of sample segmentation and spatial microsimulation. The output will be discussed with reference to uncertainties arising in this undertaking and scope and limits of informing public health and social policy interventions.

2 Data and methods

Lifestyle research originates with Bourdieu's work on social practice, in particular his detailed investigations of French middle class taste and cultural consumption [Bourdieu (1984, 1977, 1990)]. It is beyond the scope of this paper to discuss his theory in detail; but two conclusions are of particular relevance here. First, health behaviours do not occur in isolation but are socially situated in an often unconsciously adopted array of social practices. Second, social practices result from interactions between subjective orientations and individual (social) circumstances. These subjective orientations are expressed in actions such as leisure activities, cultural consumption, taste, social, civic and political participation, media use as well as stated values and beliefs. Health behaviours occur within these lifestyle dimensions, and it has been argued that addressing health behaviours requires an understanding of the subjective, behavioural context [Williams (1995); Veenstra and Burnett (2014); Nettleton and Green (2014)].

The UK Understanding Society longitudinal survey collects this information on a sample of more than 40,000 individuals. Waves 2 and 3, collected between 2010 and 2012, are available and provide a range of relevant information including some health behaviours themselves [Knies (2014)]. 56 relevant questions have been identified and have been combined to 31 scales after Principal Component Analysis [see Appendix for a list variables]. 33,000 respondents living in England and Wales have been clustered based on these variables through a two stage clustering procedure involving Ward's hierarchical and k means clustering. The resulting milieus were investigated with respect to behavioural patterns as well as their socio-demographic and economic profiles by means of χ^2-based tests and one-way ANOVAs.

These milieus are then geographically projected into small areas using deterministic spatial microsimulation, specifically Iterative Proportional Fitting (IPF) [more on these methods, see Harland et al. (2012); Lovelace and Ballas (2013)]. In short, IPF weighs a survey respondent's representativeness of a given spatial zone (e.g. ward) based on matching socio-demographic variables which are available for the respondent and for each zone in form of aggregate statistics. Subsequently, weighted statistics can be created for any outcome of interest that is included in the survey. In this case, IPF is being used to estimate the spatial distribution of lifestyle milieus and their associated health behaviours. This stage of the work is on-going, and some preliminary output is included in

Table 1: Summary of health milieus 1 to 4. All aspects reflect statistically significant cluster differences.

#	**cluster label and frequency**	**key characteristics**	**socio-demographic profile**
1	enduring isolation (11%)	unhealthy behaviours (smoking, low physical activity, diet), low leisure participation, low social and political participation, low news consumption	middle-aged, low income, low qualification, single or in couple with one child, in public housing
2	unconcerned starters (8%)	lower physical activity, unhealthy diet, low local attachment, low social and political participation, low news consumption, higher internet use	younger (16-34), low income, basic qualifications, early career, majority single, in private accommodation, urban, London
3	retiring generation (12%)	low physical activity, lower levels of smoking, mixed diet, low leisure participation, lower social integration, basic political participation, very low news consumption, no internet use, high TV consumption	majority over pensionable age, low income, low qualifications, married or widowed, in owned home, often providing care for other person
4	locally anchored (10%)	average health behaviours, lower levels of smoking, average leisure participation, very high local attachment, higher social integration, average political participation, average-lower news consumption	often women, approaching pensionable age, lower-medium income, basic qualifications (majority GCSE or below), couples often with children, in owned home/on mortgage

this extended abstract for illustration.[1]

3 The health milieus

Nine distinct milieus were identified and investigated with respect to socio-demographic and economic characteristics as well as measure of self-rated health. Tables 1 and 2 summarise key behavioural characteristics and the socio-demographic profiles of each milieu. All reported characteristics refer to statistically significant differences between clusters. The labels are provisional and still subject to refinements.

The size of the milieus ranges between 8 and 15 per cent. Three clusters, currently called *enduring isolation*, *unconcerned starters* and *retiring generation* represent three groups whose health behaviours would be considered unhealthy, in particular with respect to physical activity and diet. Although these clusters are of similarly low social status and incomes, they provide different demographic and behaviour contexts which are suggestive of different types of social pathways at work with differential impacts and implications for public health responses. The three wealthier groups *involved cultural consumers*, *rising extroverts* and *committed citizens* show minor differences

[1]The statistical software used has been R **?**.

Table 2: Summary of health milieus 1 to 4. All aspects reflect statistically significant cluster differences.

#	label	key characteristics	socio-demographic profile
5	established cultural consumers (13%)	high level of sports, healthy diet, lower levels of smoking, very high leisure participation, above-average local attachment, higher political participation, high news consumption, higher internet use	middle-aged, very high income, high qualification, in advanced careers, families with children, in owned home, in London and South East
6	rising extroverts (10%)	high levels of sports, lower levels of smoking, high leisure participation, very low local attachment, higher political participation, high news consumption, high internet use	younger, very high income, high qualification, young couples sometimes with children, in transition to home ownership, live in London and South, urban
7	committed citizens (8%)	medium to higher levels of sports, healthy diet, lower levels of smoking, higher leisure participation (arts and sights), higher local attachment, very high civic pariticpation (organisation member and volunteering), higher political participation, high news consumption, higher internet use	approaching retirement, very high income, high qualification, married with children, in own home, almost half live in London and South, often providing care for other person
8	laid-back detachment (13%)	very low levels of physical activity, mixed diet, lower levels of smoking, lower leisure participation, moderate local attachment, higher political participation, lower news consumption, higher internet use	younger to middle aged, lower-medium income, basic qualification (majority GCSE or lower), mixed ethnic background (10% Asian), families with children, in owned property/on mortgage
9	digital age autonomy (15%)	higher levels of sports, mixed diet, lower levels of smoking, specifically arts-related leisure activities, low local attachment, low civic and political participation, low news consumption, high internet and social media use	younger, lower-medium income, basic qualification (majority GCSE or lower), single or young couples often with children, on mortgage

in overall healthy behaviours. Yet they, too, reveal different behavioural tendencies with respect to leisure activities, local orientations, social, civic and political participation. Finally, another set of three clusters, *locally anchored*, *laid-back detachment* and *digital age autonomy*, with medium levels of incomes and basic qualifications differ with respect to levels of exercising and a range of leisure and social orientations.

Figure 1 shows how the clusters distribute across chronic disease risk and mean income. While each income level is broadly associated with a particular range of disease risks, the distribution of milieus suggests further differentiation of health-related pathways. This becomes particularly clear

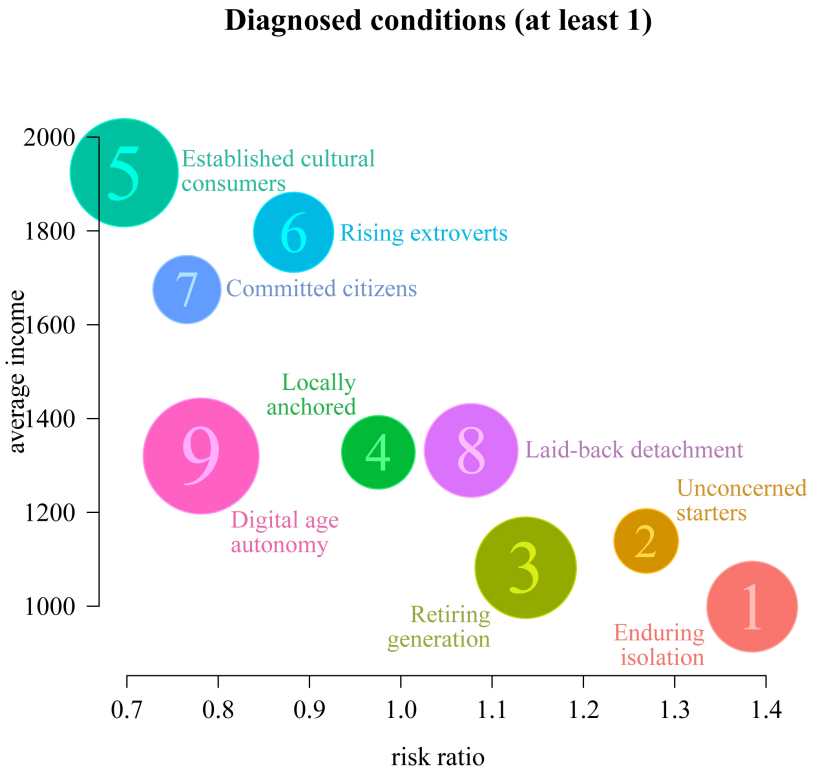

Figure 1: Milieu differentiation by income and age-and sex-standardised chronic disease risk. The size of the circles reflects the size of the clusters.

when viewing milieus 4, 8 and 9: their economic and demographic profiles are similar; yet their difference in disease risk is significant. Vice versa, larger income differences between clusters 5, 6 and 7 do not translate into corresponding proportional risk differences. In fact, contrary to our expectations, cluster 6 is worse of than clusters 7 and 9. Conventional social epidemiological studies typically assert a social gradient across status groups in health; but this uni-dimensional view of social status is likely to mask milieu-specific pathways that do not strictly reproduce the social gradient in health.

4 Adding the spatial perspective: probable prevalence of health milieus (preliminary results)

The strong socio-demographic distinctiveness of milieus offers particular opportunities to project the milieus geographically. Figure 2 shows the first experimental run of spatial microsimulation for 2011 wards in London, matching UK 2011 census neighbourhood statistics with respondent characteristics sex, age and socio-economic status (NSSEC-5). Even this very limited range of matching variables on this coarse geographical scale produces distinct spatial distributions for each

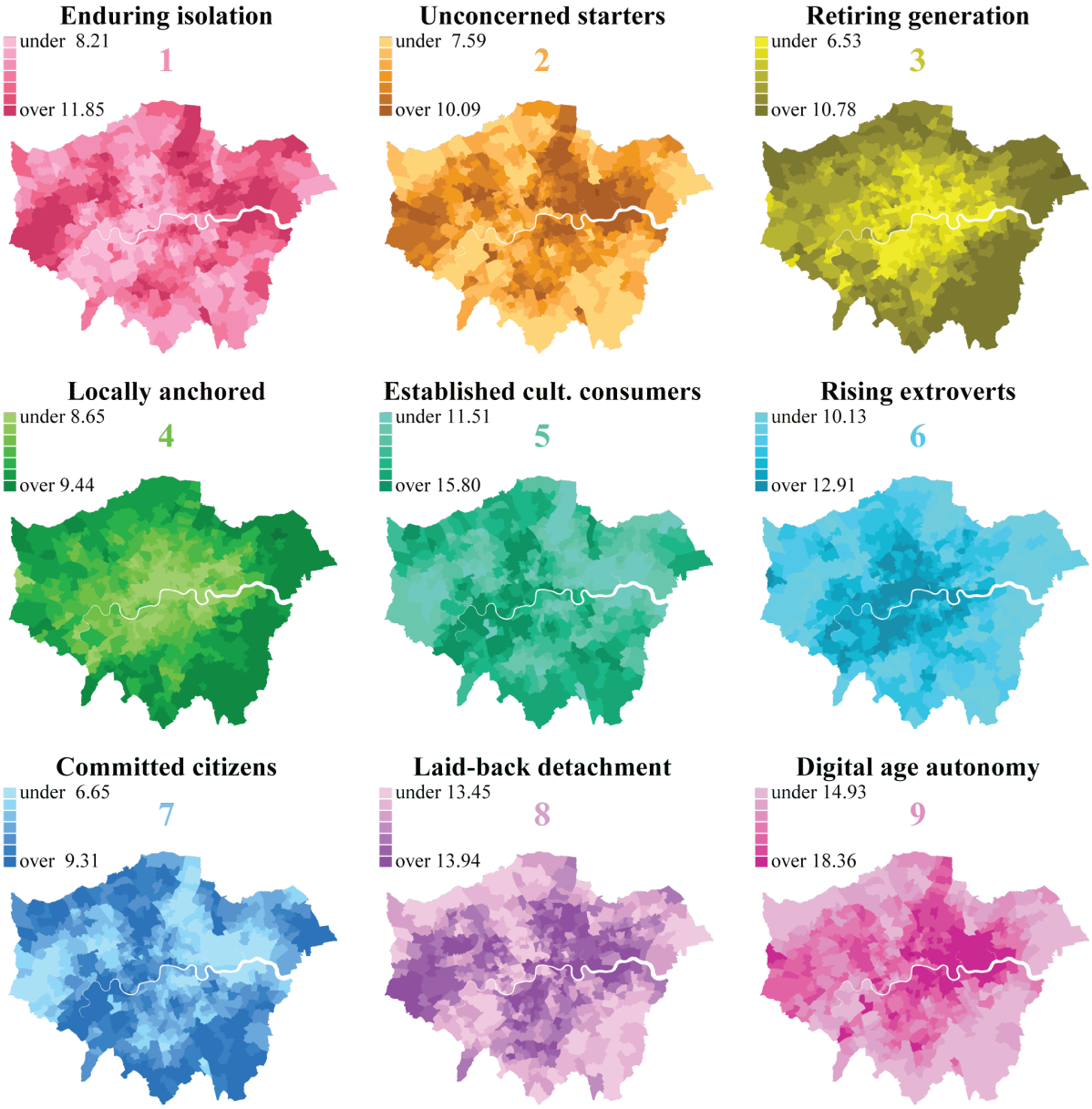

Figure 2: Milieu probabilities in London wards (2011) derived from deterministic spatial microsimultion

milieu. The maps show different relative frequencies of milieus and, given the IPF-derived weighting, can be interpreted as probabilities of the prevalence of the respective health milieu. Overall, the emerging milieu-specific geographies seem plausible in the London context. Nevertheless, it should

be remembered that this is an illustration of on-going work rather than a definitive result. It remains to be seen how the geographies change as both the spatial resolution and the number of matching variables are increased.

5 Conclusions

This paper presents a social theory-grounded approach within a geodemographic framework to combine the power of extensive social surveys with the wide coverage of an administrative data.The work will continue with refining the spatial microsimulation, specifically by extending the matching variables to those that also prove milieu-discriminant and by increasing the spatial granularity to better reflect local variations of population characteristics. Subsequently, some validation will be carried out by comparing the resulting geographies with the detailed geocoded information that is available in Understanding Society as well as other data, notably the 2011 Output Area Classification.

In summary, the findings suggests that behavioural orientations and their co-varying health behaviours vary by multiple social and demographic characteristics; they may therefore be geographically simulated under close observation of the uncertainties associated with synthetic estimates. Thus, adding a behavioural building block to health geodemographics may be a promising way forward in making the tool more relevant for strategic public health and social policy interventions.

6 Acknowledgements

This work is part of an ongoing, ESRC-funded PhD, supervised by Prof. Paul Longley and Prof. Jenny Robinson, Department of Geography, University College London, and is benefitting form their comments and feedback. I would like to extend further thanks to James Cheshire and Nicola Shelton for their reviews.

7 Biography

Jens Kandt is a PhD candidate at the Department of Geography, University College London, and a researcher at LSE Cities, London School of Economics and Political Science. His work focusses on linking spatial statistics and social theory to understand dynamics of urban environments and their implications for transport, mobility and people's health. He holds an engineering degree in planning from the German University of Dortmund and has research and work experience in the UK, India, Germany, Ghana and Hong Kong.

References

Abbas, J., Ojo, a., and Orange, S. (2009). Geodemographics–a tool for health intelligence? *Public health*, 123(1):e35–9.

Baum, F. and Fisher, M. (2014). Why behavioural health promotion endures despite its failure to reduce health inequities. *Sociology of health & illness*, 36(2):213–25.

Blaxter, M. (1990). *Health and lifestyles*. Routledge, London.

Blaxter, M. (2010). *Health*. Cambridge: Polity.

Bourdieu, P. (1977). *Outline of a Theory of Practice*. Cambridge University Press, New York.

Bourdieu, P. (1984). *Distinction: a Social Critique of the Judgment of Taste*. Harvard University Press, Cambridge.

Bourdieu, P. (1990). *In Other Words*. Stanford University Press, Stanford.

Harland, K., Heppenstall, A., Smith, D., and Birkin, M. H. (2012). Creating realistic synthetic populations at varying spatial scales: a comparative critique of population synthesis techniques. *Journal of Artifical Societies and Social Simulation*, 15:1–24.

Knies, G. (2014). *Understanding Society UK Household Longitudinal Study: Wave 1-4, 2009-2013 User Manual*. ISER Institute for Social and Economic Research, Colchester.

Longley, P. A. (2005). Geographical Information Systems: a renaissance of geodemographics for public service delivery. *Progress in Human Geography*, 29(1):57–63.

Lovelace, R. and Ballas, D. (2013). Truncate, replicate, sample: A method for creating integer weights for spatial microsimulation. *Computers, Environment and Urban Systems*, 41:1–11.

Nettleton, S. and Green, J. (2014). Thinking about changing mobility practices: how a social practice approach can help. *Sociology of health & illness*, 36(2):239–51.

Openshaw, S. and Blake, M. (1995). Geodemographic segmentation systems for screening health data. *Journal of epidemiology and community health*, 49 Suppl 2(Suppl 2):S34–8.

Singleton, A. D. and Longley, P. A. (2009). Geodemographics, visualisation, and social networks in applied geography. *Applied Geography*, 29(3):289–298.

Veenstra, G. and Burnett, P. J. (2014). A relational approach to health practices: towards transcending the agency-structure divide. *Sociology of health & illness*, 36(2):187–98.

Voas, D. and Williamson, P. (2001). The diversity of diversity: a critique of geodemographic classification. *Area*, 33(1):63–76.

Williams, S. J. (1995). Theorising class, health and lifestyles: can Bourdieu help us? *Sociology of Health and Illness*, 17(5):577–604.

Appendix - List of scales and variables

scale	variable name	description	survey module
nutrition	Usdairy	Usual type of dairy consumption	nutrition
nutrition	Usbread	Type of bread eats most frequently	nutrition
nutrition	Wkfruit	Days each week eat fruit	nutrition
nutrition	Wkvege	Days each week eat vegetables	nutrition
smoke	Ncigs	Usual no. of cigarettes smoked per day	smoking
smoke	Smcigs	Ever smoked cigarettes regularly	smoking
smoke	Smncigs	Number of cigarettes smoked in past	smoking
walk	Wlk30min	Number of days walked at least 30 minutes	physical activity
sports	Sportsfreq	Moderate intensity sports frequency	leisure, culture and sport
sports	Sports3freq	Mild intensity sports frequency	leisure, culture and sport
advice	Scopngbhc	Advice obtainable locally	neighbourhood (self-compl)
belong	Scopngbha	Belong to neighbourhood	neighbourhood (self-compl)
borrow	Scopngbhd	Can borrow things from neighbours	neighbourhood (self-compl)
dark	Crdark	Feel safe walking alone at night	local neighbourhood
friends	Scopngbhb	Local friends mean a lot	neighbourhood (self-compl)
improve	Scopngbhe	Willing to improve neighbourhood	neighbourhood (self-compl)
stay	Scopngbhf	Plan to stay in neighbourhood	neighbourhood (self-compl)
talk	Scopngbhh	Talk regularly to neighbours	neighbourhood (self-compl)
close	Closenum	How many close friends	social network
family	Simfam	Proportion of friends who are also family members	social network
local	Simarea	Proportion of friends living in local area	social network
network	Simage	Proportion of friends with similar age	social network
network	Simrace	Proportion of friends of same race	social network
network	Simeduc	Proportion of friends with similar level of education	social network
network	Simjob	Proportion of friends who have a job	social network
socnet	Netcht	Hours spent interacting with friends through social websites	social network
civic	Civicduty	Sense of civic duty	political engagement
civic	Civicduty	Sense of civic duty	political engagement
polcomp	Poleff1	Qualified to participate in politics	political self-efficacy
polcomp	Poleff2	Better informed about politics	political self-efficacy
polcost	Polcost	Cost of political engagement	political engagement
polcost	Polcost	Cost of political engagement	political engagement
polcyn	Poleff3	Public officials don't care	political self-efficacy
polcyn	Poleff4	Don't have a say in what government does	political self-efficacy
polinf	Perpolinf	Perceived political influence	political engagement
polinf	Perpolinf	Perceived political influence	political engagement
polit	Vote6	Level of interest in politics	politics
polit	Vote6	Level of interest in politics	politics
voteben	Perbfts	Personal benefit in voting	political engagement
voteben	Perbfts	Personal benefit in voting	political engagement
voteint	Voteintent	Voting intention	political engagement
voteint	Voteintent	Voting intention	political engagement

scale	variable name	description	survey module
votenorm	Votenorm	Voting as a social norm	political engagement
votenorm	Votenorm	Voting as a social norm	political engagement
org	Orgm	Which organisations member of	groups and organisations
org	Orga	Active in organisations	groups and organisations
org	Orgmt	Member of organisations NSC	groups and organisations
org	Orgat	Active in organisations NSC	groups and organisations
volun	Volfreq	Frequency of volunteering	voluntary work
arts1	Arts1freq	Arts activities frequency	leisure, culture and sport
arts2	Arts2freq	Arts events frequency	leisure, culture and sport
hist	Herfreq	Historical sites frequency	leisure, culture and sport
lib	Libfreq	Library frequency	leisure, culture and sport
musm	Musfreq	Museum frequency	leisure, culture and sport
news	Newsource	Sources of News	news and media use
tv	Tvhours	Hours of TV per weekday	news and media use