# The social and spatial context of urban health inequalities

## Towards an interpretive geodemographic framework

**Jens Kandt**

Thesis submitted in conformity with the requirements of
**Doctor of Philosophy (Ph.D.)**

**Department of Geography**
**University College London**
September 2015

Supervisors
**Prof. Paul Longley**
**Prof. Jennifer Robinson**

### Declaration

I, Jens Kandt, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

signed: ______________________

# The social and spatial context of urban health inequalities: towards an interpretive geodemographic framework


## Abstract

Geodemographics, the technique of classifying small areas by the aggregate characteristics of their residents, is a promising method to study health inequalities and prepare the development of preventive policy. Yet, current approaches do not account for the complexity and contingency of health pathways, which are found to be differentially activated in different groups of populations at different ecological levels. Building on the strength of geodemographics to integrate diverse data and link them ecologically, I suggest an interpretive framework, which characterises population vulnerability to health disadvantage at the level of regions, neighbourhoods and individuals. At the regional level, I explore vulnerability in terms of population structure in Great Britain and UK metropolitan areas in order to assess probable geographies of specific cultural or biological factors that may shape vulnerability. Based on indicators derived from hospital admission data and the UK Census, I identify different specific expressions of vulnerability at the neighbourhood level (which I call health environments), reflecting generic social causes of health advantage and disadvantage as well as specific forms of vulnerability. A comparison of metropolitan areas further reveals specific, local guises of vulnerability across England's cities. At the individual level, I discover from social survey data different groups in society (health milieus), which are characterised by distinct activity patterns, subjective orientations, attitudes and everyday life routines. I model the geographical distribution of health milieus probabilistically for London. Geographical linkage of these layers of information results in suggestions for an alternative urban policy programme to reduce population vulnerability through an emphasis on multi-sectoral and preventive action. The thus advanced geodemographic framework provides a conceptually focussed view of health, socially and spatially contextualised at multiple ecological levels, that contributes to interpreting health inequalities in social science and addressing their root causes through strategic policy.

# Contents

## List of Acronyms

| | |
|---|---|
| BYM | Besag, York and Mollie [1991] spatial-structural model |
| CSDH | Commission on Social Determinants of Health |
| DIC | Deviance Information Criterion |
| GLM | General Linear Model |
| HCA | Hierarchical clustering procedure |
| HES | Hospital Episode Statistics for England |
| ICT | Information and Communication Technology |
| IDW | Inverse Distance Weighting |
| IMD | Index of Multiple Deprivation |
| IPF | Iterative Proportional Fitting |
| JSNA | Joint Strategic Health Assessment |
| KDE | Kernel Density Estimation |
| LHC | London Health Commission |
| LOAC | London Output Area Classification |
| LSOA | Lower Level Super Output Area (Census geography) |
| LQ | Location Quotient |
| MDS | Multi-dimensional Scaling |
| MSOA | Middle Level Super Output Area (Census geography) |
| NCC | Non-communicable conditions |
| NMI | Normalised Mutual Information |
| OAC | Output Area Classification of the United Kingdom |
| PCA | Principal Component Analysis |
| SHR | Age and sex-standardised health ratio |
| SMbRR | Age and sex-standardised morbidity risk ratio |
| SMtRR | Age and sex-standardised mortality risk ratio |
| VIF | Variance Inflation Factor |
| WHO | World Health Organisation |

# Preface

The thesis is an attempt to combine different kinds of data and make sense of the complex phenomenon of urban health inequalities. This phenomenon crosses social domains and, as will be reiterated, requires transdisciplinary interpretations. I take inspiration from sociology, urban studies, GIScience, spatial statistics, epidemiology, genetics, public health and spatial planning to develop a framework that may offer such interpretations. But, since the thesis is written by only one person, these are necessarily incomplete and require critical review by experts.

I am particularly indebted to two experts, Prof. Paul Longley and Prof. Jenny Robinson, both of whom have advised on this work from different disciplinary angles within geography. Paul has been a source of constant support in all matters academic and administrative, and his openness to lend support at a week's notice to someone he had never heard of has made it possible for me to be in the position of completing this thesis at this juncture. I am most grateful to him. Jenny has given valuable advice whenever it was needed, and while I know that she has enriched my work in important ways (not least through her regular reading meetings), I hope she will find her views well-accommodated in a thesis of a different call than those she normally takes on.

I would like to thank Dr. James Cheshire and Dr. Nicola Shelton for having taken the time to review parts of the work and for recommending me for upgrade from M.Phil. status. Their comments were critical, informative and reassuring.

Many thanks to Dr. Daniel Lewis for a number of conversations on the subject of this thesis, in the course of which he has also advised on and supported data access. Alistair Leak and Dr. Adnan Muhammad have helped me at various times with their profound knowledge of IT, which has saved me hours if not days of clumsy problem-solving. Thank you both for having been so helpful and, indeed, patient.

At the very beginning, Myfanwy Taylor and Dr. Antoine Paccoud encouraged me at LSE Cities to send out my short proposal; I am grateful to them as, but for this, I am not sure I would have done so back then. LSE Cities has been a source of inspiration for a doctoral project, notably through my work on health inequalities in Hong Kong, during which I also had the opportunity to work with the Centre of Suicide Research and Prevention at the University of Hong Kong. Many thanks to the entire teams at the LSE and HKU.

I thank Prof. Annette Spellerberg for our regular exchanges, which have always kept me mindful to employ a sociological angle to my work. I am also indebted to her and Prof. Joachim Scheiner for having provided references for my application for a doctoral project at UCL.

I am grateful to Prof. Roger Burrows and Prof. Graham Moon for being willing to take the time to read this thesis and examine its author. I hope that the thesis is sufficiently clear to make for an easy and perhaps also interesting read.

The ESRC funded this work with a grant for Advanced Quantitative Methods (AQM) after a review at the Department of Geography at University College London. I would like to thank all those who were involved in the decision; and I hope to satisfy them of the gravity with which I took my appointment in pursuit of the ambition to contribute to social science and urban geography in the future.

London, 25 September 2015

# 1 Introduction: the complex character of urban health inequalities

How can we conceptualise, assess and address health needs in a complex world? These question have long engaged epidemiologists, geographers, planners, policy makers and others. The questions imply both theoretical and practical problems, starting from what may be causal to what may be appropriate policy priorities for strategic planning, prevention and adaptation. While social epidemiology has demonstrated with certainty the link between socio-economic position and health, the precise identification and characterisation of causal pathways has remained a challenge. Causal pathways appear to be local rather than universal, contingent rather than deterministic, directional rather than reversible; it appears, they need to be framed in a broader social and geographical context, within an ecological framework that proves productive for both social science and strategic policy.

## 1.1 The complex notion of cause

Health can be and has been studied from different perspectives: biomedical, epidemiological, sociological and geographical. A paradigm shift to the study of health occurred with the introduction of the social determinants approach. Illness may not just have pathophysiological causes; pathophysiological changes may themselves be socially induced. This recognition has brought into focus socio-economic position in health as the starting point at which socially differentiated pathways become activated and produce socially patterned outcomes in health [Brunner & Marmot 2006]. Outcomes, however, do not mark the end of pathways: illness may limit life chances and the ability to improve socio-economic position.

This social epidemiological model coexists with sociological macro models of health, which focus on the causes of social inequality. Social pathways comprise ways in which societal processes distribute health-relevant assets and risks among individuals with different social attributes and contribute to the embodiment of social inequalities [Blaxter 2010; Krieger 2005; Scambler 2007; Wilkinson & Pickett 2010]. But societal processes are not just abstract macro phenomena that determine health: rather, they are made and re-made by people through their practices and choices fashioning heterogeneous modes in which pathways play out. Why do healthy versus unhealthy behaviours occur and how should they be understood in a wider context of social practice? Although not explicitly discussing health, Bourdieu's [1977] relational view of behaviour is instructive in characterising social pathways [Williams 1995].

The epidemiological and the sociological debates make clear that neither individual biology nor social and material conditions alone are sufficient to understand health. In an effort to theoretically integrate the social and the biological, there has been a recent move towards complexity theory and systems thinking in social epidemiology and the sociological study of health [Byrne 1998; Diez-Roux 2004; Rydin et al. 2012]. It is hoped that a complex systems approach can overcome primarily two theoretical and methodological problems in quantitative social science: establishing causality and addressing the spatio-temporal uncertainty of knowledge.

In quantitative social science, conventional techniques – such as general linear modelling – employed to trace the causes of health outcomes through population averages now appear inadequate to search for causal pathways, since there are no logical grounds for their detection by methods founded on correlated variables [cf Byrne 2002, 114]. From a realist point of view, socio-epidemiological pathways can be compared to Bhaskar's "transfactuals" [Bhaskar 2008 (1975)]. Transfactuals are causal forces, which exist irrespective of observable outcomes. They may give rise to observable instances under certain conditions, which constitute their empirical manifestation.

> "The world consists of things, not events. Most things are complex objects, in virtue of which they possess an ensemble of tendencies, liabilities and powers. It is by reference to the exercise of their tendencies, liabilities and powers that the phenomena of the world are explained." [ibid., 51]

Pathways are constituted by complex objects with tendencies that may or may not be manifest in discernible events. This understanding compels us to expect heterogeneity but not randomness: certain social conditions may activate pathways for certain social groups and not for others or in certain places and not in others. Group or place specificities interact with pathways and produce diverse phenomena (heterogeneity) within a spectrum of possible phenomena (non-randomness).

It follows that knowledge of associations is primarily local and cannot be generalised [Byrne 2002, 75]. The 'locality' of knowledge not only leads us to view with caution the status of current knowledge but it also translates into practical problems of strategic planning. Our ability to learn from one place about another, or from one group about another across different points of time, depends on our ability to identify transfactuals through plural interacting conditions. Robinson [2006, 2011] has shown for urban studies that one way to discover local specificities and thereby critically appraise extant knowledge is through comparative research. But she also alerts us that universalising interpretations from comparative research can result in knowledge production that is biased towards privileged places. The biased knowledge is then used to sustain inferences about other, less studied places, supporting distorted views and counter-productive urban policy. Her critique and systematisation of comparative research practices has strong implications for the quantitative research programme, not least because quantitative methods rest entirely upon comparison.

Comparative research conducted with a presumption of plurality and local specificity of pathways involves greater attention to the question of what cases we consider to be comparable in the first place. Tacit assumptions on similarity and therefore what constitutes a 'reasonable' comparison tend to reproduce rather than challenge conventional beliefs [Robinson 2011, 13]. Applied to quantitative geography, this point calls into question how we formalise comparisons, where we measure difference and where we need to hand over to qualitative interpretation. We also ought to prefer quantitative techniques that highlight difference, diversity and dissonance across places and view with care reductionist methods that focus on universal, statistical summary.

A programme of quantitative geography that accounts for complexity, therefore, focusses on diversity of people and places as well as the significance of difference. In so doing, it favours methods that acknowledge interactions and contingency and provide clues to emergent phenomena. Rather than intending abstract explanation, quantitative social science should therefore focus on exploration and systematic description to characterise the plural causalities in a complex world [cf Byrne 2002, 95; Savage & Burrows 2009, 796]. This orientation of research has implications for the systematic assessment of health needs – which is being called for in policy by the introduction of compulsory Joint Strategic Needs Assessments (JSNA) in the UK – and appears instrumental to addressing the phenomenon of persistent health inequalities in cities. In addition, trends in local climates, environmental hazards and economic developments pose the important question of how resilient urban communities are in the context of risk and what strategic solutions may support them.

## 1.2 Towards contextual and place-sensitive solutions

There are new opportunities to advance in the quantitative investigation of urban health inequalities. Rapid progress in collecting, storing and analysing routine datasets opens new ways to theorise, study and plan cities [Batty 2013; Gonzalez-Bailon 2013; Longley 2005a]. The emerging data, often called 'Big Data', comprise routinely collected government datasets, data from social media and commercial datasets. In the UK, there are at least 70 routinely collected administrative and survey datasets with relevant information on various aspects of people's health [Figure 1.1].

Assessing these datasets in terms of their volume, variety and velocity – informally, the three V's of Big Data [O'Reilly Radar Team 2011] – shows that those datasets differ in size, contents and speed of update. If these datasets are combined and managed in a coordinated environment, they may offer the prospect of a spatial health data infrastructure [Longley et al. 2011b, 480] representing different domains of health needs ranging from general population health, to the incidence of diseases, health care utilisation and social conditions of health and well-being. The linkage of administrative records pertaining to the same individual has also become common practice in health research to study problems across different information contexts [Jutte et al. 2011]. With increased computational power and new statistical approaches including complex modelling and simulation, the chances of accessing, integrating and combining the diverse information for strategic purposes including urban planning, health care and service delivery have increased, and quantitative geography may be expected to contribute significantly to solutions of a variety of practical problems.

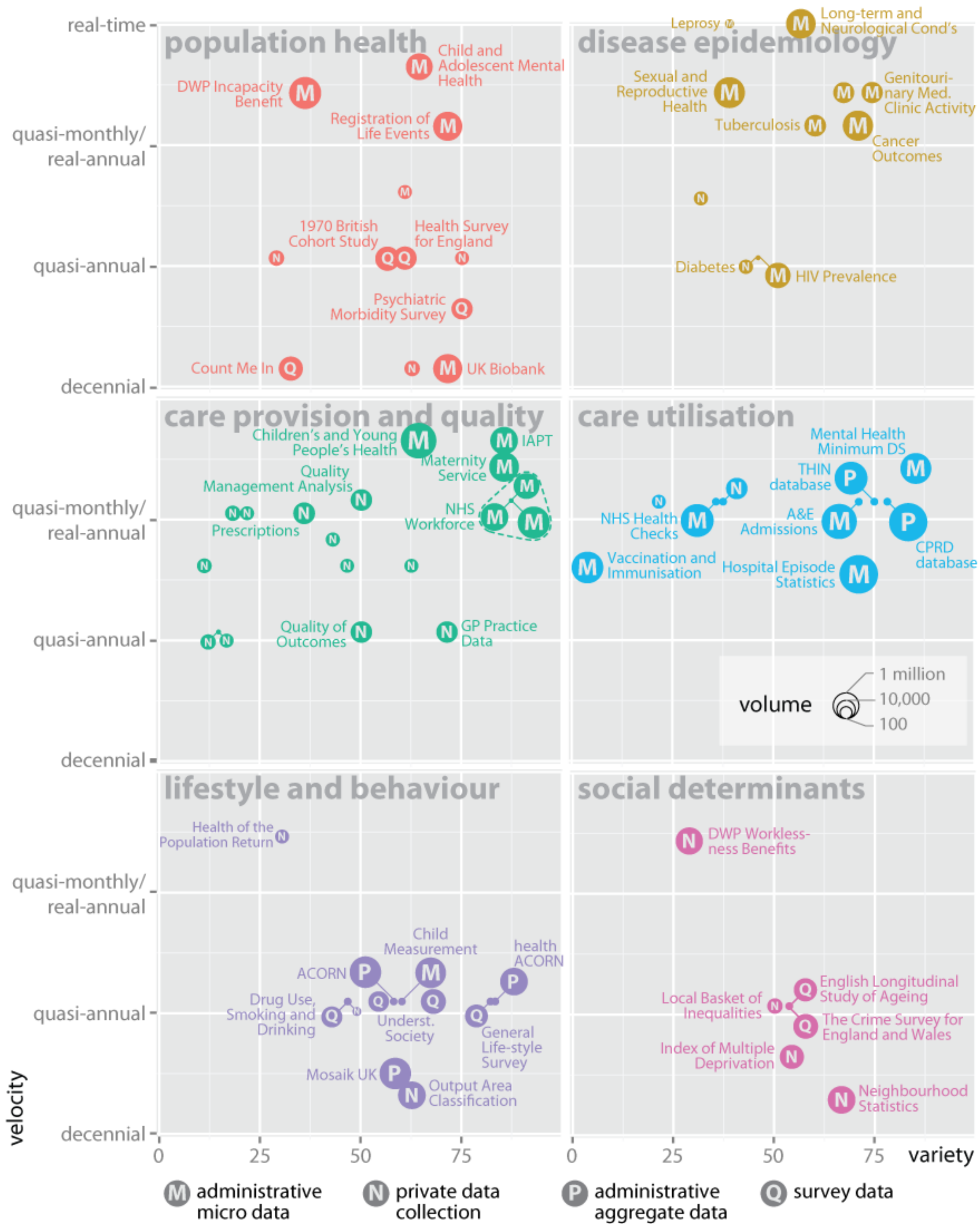



Figure 1.1: Review of routinely collected health datasets in the UK by their volume, variety and velocity (above) and more details on temporal and geographical resolution of selected datasets (righthand page).

| | population health | geographical resolution | temp. resolution |
|---|---|---|---|
| M | Child and Adolescent Mental Health Services Dataset (CAMHS) | post code | continuous |
| M | Community Mental Health Activities | Primary Care Trust | quarterly |
| M | DWP Incapacity Benefit and Severe Disablement Allowance | Super Output Area | continuous |
| M | Registration of Life Events (General Register Office/ONS) | post code | continuous |
| M | Compendium of Population Health Indicators (NHS Indicator Portal) | Primary Care Trust | annual |
| N | Health Poverty Index (HPI) | local authorities | once |
| Q | 1970 British Cohort Study | Gov't Office Region | annual |
| Q | Count Me In 2010: National Mental Health [...] Ethnicity Census | Strategic Health Authority | once |
| Q | Health Survey for England | Strategic Health Authority | annual |
| Q | Psychiatric Morbidy Survey | Strategic Health Authority | triannual |
| | **disease epidemiology** | **geographical resolution** | **temp. resolution** |
| M | Cancer Outcomes and Services Dataset (COSD) | post code | continuous |
| M | Genitourinary Medicine Clinic Activity Dataset (GUMCAD) | Lower Super Output Area | continuous |
| M | HIV and AIDS Reporting System (HARS) | Lower Super Output Area | continuous |
| M | Survey of Prevalent HIV Infections Diagnosed (SOPHID) | Primary Care Trust | annual |
| M | Leprosy Surveillance | hospitals | continuous |
| M | Sexual and Reproductive Health Activity Dataset | Lower Super Output Area | continuous |
| M | Enhanced Tuberculosis Surveillance (ETS) | Lower Super Output Area | continuous |
| N | Cancer Audits | hospitals | monthly |
| N | National Diabetes Inpatient Audit | hospitals | annual |
| | **care provision and quality** | **geographical resolution** | **temp. resolution** |
| M | National Children's and Young People's Health Services Dataset | Super Output Area | continuous |
| M | Improving Access to Psychological Therapies (IAPT) | post code | continuous |
| M | Annual Medical and Dental Workforce Census | provider-level | continuous |
| M | National Maternity Services Data Set (NMDS) | post code | continuous |
| M | General Medical Practitioners Annual Census Collection | provider-level | continuous |
| N | Bed Availability and Occupancy | hospitals | quarterly |
| N | National Cancer Waiting Times Monitoring Dataset | Strategic Health Authority | monthly |
| N | Diagnostic Waiting Times and Activity Data Collection | Primary Care Trust | monthly |
| N | GP Practice data (NHS Indicator Portal) | GP practice (some missing) | annual |
| N | Summary Hospital-level Mortality Indicator (NHS Indicator Portal) | hospitals | annual |
| N | NHS Outcomes Framework (NHS Indicator Portal) | local authorities | annual |
| N | NHS Stop Smoking Services Collection | Primary Care Trust | quarterly |
| N | Prescriptions: GP Practice Prescribing Presentation-level Data | GP practice | monthly |
| N | Quality Management Analysis System | GP practice (some missing) | continuous |
| N | Quality of Outcomes Framework | GP practice (some missing) | annual |
| | **care utilisation** | **geographical resolution** | **temp. resolution** |
| M | Hospital Episode Statistics (HES) | post code | monthly |
| M | Vaccination and Immunisation Data Return (COVER) | Primary Care Trust | quarterly |
| M | Mental Health Minimum Data Set Version 4 | post code | continuous |
| M | NHS Health Checks Data Set | Lower Super Output Area | continuous |
| N | Childhood Immunisation Programme | Primary Care Trust | continuous |
| N | Seasonal Influenza Vaccine Uptake | GP practice | continuous |
| P | CPRD database: Clinical Practice Research Datalink | GP practice | continuous |
| P | THIN database: The Health Improvement Network | Strategic Health Authority | continuous |
| | **lifestyle and behaviour** | **geographical resolution** | **temp. resolution** |
| M | National Child Measurement Programme | post code | annual |
| N | Statistics on Obesity, Physical Activity and Diet | Strategic Health Authority | annual |
| N | Output Area Classification (OAC) | Output Area | decennial |
| P | ACORN (CACI) | post code | annual |
| P | Mosaik UK | post code | quinquennial |
| Q | Drug Use, Smoking and Drinking Among Young People in England | Strategic Health Authority | annual |
| Q | General Lifestyle survey (formerly: General Household Survey) | Lower Super Output Area | annual |
| Q | Living Cost and Food Survey (formerly: National Food Survey) | local authorities | annual |
| Q | Understanding Society | Lower Super Output Area | annual |
| | **social determinants** | **geographical resolution** | **temp. resolution** |
| N | DWP Worklessness Benefits | Output Area | continuous |
| N | Index of Multiple Deprivation (IMD) | Lower Super Output Area | triannual |
| N | Local Basket of Inequalities Indicators (NHS Indicator Portal) | local authorities | annual |
| N | Neighbourhood Statistics (ONS National Census) | Output Area | decennial |
| Q | The Crime Survey for England and Wales | Lower Super Output Area | annual |

Data of this kind have been used for some time by private sector companies to describe in detail the consumer world for their own purpose of profit maximisation. Their use in academic research and by government, however, has prompted concerns over data abuse, surveillance and undermining of citizenship [Burrows & Ellison 2004; Goss 1995; Graham 2005]. Extensive data mining and processing continues to invoke the image of the transparent citizen that can be influenced, monitored and misled, if personal and confidential data is used to serve unchecked vested interests. Quantitative geography will have to strike the balance between precision and protection of the individual in an ethically sound terrain [Longley 2007]. In the UK, researchers seeking access to the datasets need to demonstrate information governance arrangements that meet constantly rising security standards. Access to high resolution datasets is granted on a *bona fide* basis to academic researchers with ongoing monitoring of its uses by data providers and ethical bodies. Record linkage has proven particularly challenging, since this method entails the need for specific consent by a potentially large number of participants as well as elaborate technical arrangements for safeguarding confidentiality. The need of quantitative geography to engage with the ethical dimension of research appears more pressing than ever as more detailed datasets become available.

The full potential of the datasets can be unlocked when combining their diverse and voluminous information within a shared, substantively meaningful context. It can be conceived that the datasets pertain to different levels of aggregation encompassing individuals, households, neighbourhoods, local authorities and regions to name a few. Some datasets are longitudinal datasets that follow cohorts of respondents over time, such as the Understanding Society government survey or the English Longitudinal Ageing Study. Other datasets contain rich, cross-sectional information, collected repeatedly from populations or samples at the time. The Health Survey of England and Wales or the Census exemplify this type. Purpose-built administrative datasets, such as Hospital Episode Statistics or General Practice patient registers, comprise all cases at a given institutions and can be very large. At a first glance, the datasets touch a variety of relevant dimensions that constitute health needs.

The multi-level nature of the different data suggests that an ecological, spatial framework may be a powerful way to combine the different datasets through common units that can be linked. Geodemographics – a technique that classifies geographical areas such as neighbourhoods by the characteristics of their residents – has proven to be a successful heuristic in making sense of diverse data [Longley 2005a]. The method offers potential in accounting for diversity of populations and thereby supporting focussed applications in urban planning or health care. A major objective of this thesis is, therefore, to review geodemographic approaches to understanding the social world and explore ways to advance them for the purposes of characterising urban health inequalities.

In terms of velocity, an important question relates to the stability of health needs over time, which may vary across contexts and for different ecological levels. In order that system characterisations of health be accurate, its spatial and temporal stability needs to be assessed. Systems thinking offers a metatheoretical framework that conceptualises change in terms of interactions between system parts and the whole and resulting emerging phenomena that are specific and historically situated [Byrne 1998].

The diversity and dynamics of health needs in urban environments pose a particular challenge for strategic health care planning. The complexity of interactions and the specifically urban types of inequality and marginality in advanced economies [Wacquant 1999] as well as the current trend of persisting and widening health inequalities in cities [GRNUHE 2010; WHO & UN-HABITAT 2010] call for special and urgent attention to health in urban environments. Urban environments are regionally embedded, but equally they are shaped by global processes, networks and circulating policies. This fact necessitates investigations to look beyond territorial definitions and take into account processes elsewhere [Robinson 2013a, 2014].

Amin and Thrift [2002] refer to future states of cities as potentialities, which can be thought of as variants of Bhaskar's tendencies of complex objects. These potentialities comprise possible phenomena that emerge from interactions of wider social processes and agency by people, communities and organisations. To which extent geodemographics represents an adequate technique to capture those potentialities shall be explored in the context of current debates in social science and political quests to address urban health inequalities.

## 1.3 Objectives and structure of the thesis

The principal objective of this thesis is to develop an advanced, interpretive geodemographic framework that supports the scientific study of health inequalities and provides utility to policy development. The central characteristics of the advanced framework are its conceptual focus or purpose orientation, its spatial explicitness – offering an adequate description of the spatial context in which health inequalities evolve –, its multi-level structure encompassing regions, cities, neighbourhoods and individuals, and its incorporation of population specificity through comparison. The thesis explores how the full evidence emanating from this framework can be used for the design of urban policy interventions – with London as a case study.

The concept of vulnerability constitutes the focus of this geodemographic framework. After a review of geodemographic research, its role in social science and opportunities for its advancement (chapter two), I introduce the concept of vulnerability and demonstrate why and how vulnerability is a useful, if not the essential, target concept in assessing and addressing health inequalities (chapter three). In brief, I choose the concept of vulnerability because it applies to all pathways thought to link socio-economic position to health, connects a multitude of policy sectors and scales to any ecological level, such as individuals, groups, neighbourhoods or regions. The components of vulnerability predispose individuals and populations towards a definite range of possible health outcomes through biology, culture, embodiment of social relations and stable tendencies to act.

Drawing on this conclusion, I first explore vulnerability in terms of regional specificity with a focus on biogenetic population structure in Great Britain (chapter four). I do so by combining a sample of DNA data with a population-wide representation of surname geographies. A distinction is made between rural and urban Great Britain in order to assess what kind of urban and rural geographies may be appropriate in offering a regional perspective of health. For rural Great Britain, I develop a regionalisation of population specificities in terms of population structure, which provides a generalised biological and cultural foundation that may shape the relationship between socio-economic position and health in regionally specific ways.

Building up on these findings with historical and contemporary surname geographies, I explore where and to what degree UK cities are inhabited by their regional populations over the long term. A series of spatial statistical heuristics suggests that although UK cities are primarily composed of regional populations, regional patterns diminish at finer granularities and sub-city parts emerge with distinctive urban populations that are common to many yet not all cities. Most probably, these represent mixed neighbourhoods with a high proportion of international migrants.

The characterisation of regions and UK cities informs the subsequent investigation of geographical disparities in health (chapter five). Interpreting data from hospital admissions in England and the UK Census, I develop a simple geodemographic classification of age and sex-standardised health outcomes in English Census wards, which incorporates explicit spatial context. This classification reveals regions with distinct challenges that reflect social causes of ill-health, general health advantage and some specialised challenges with particular profiles of disease burdens. Differences in the patterns emerge between north and south and urban and rural England. The contiguous regions of differing health challenges show a weak yet significant correspondence with a regionalisation of England based on surnames.

I subsequently compare ten metropolitan regions in England with respect to their health challenges in small areas. While many cities share common challenges, the magnitude and spatial manifestation of health inequalities differ. It appears that a societal process shaping health inequalities is attenuated or exacerbated in some cities. I develop a more granular small area classification for London, which highlights the existence of unique guises of advantage and disadvantage in London. I introduce the concept of health environments, which represent neighbourhood ensembles with empirically similar yet specific and singular health challenges.

The next step is dedicated to the individual level (chapter six). From data taken from the Understanding Society longitudinal survey, I identify ten social milieus – called health milieus – with different vulnerability profiles in the UK. These profiles can be characterised by the groups' activity patterns, subjective orientations, attitudes and everyday life routines. The milieus also differ significantly in their health and socio-demographic and economic characteristics. Using again the regionalisation based on surnames, I compare regional variants of the milieus and draw conclusion with respect to regional, cultural tendencies in the UK.

I then simulate the geographical distribution of health milieus in London (chapter seven), using a technique that belongs to the class of spatial microsimulation matching Census data with survey responses based on common information on socio-demographics. Distinct, plausible geographies emerge for each of the ten milieus; these geographies depict a contiguous geodemographic representation reflecting individual rather than aggregate level phenomena in conceptual terms.

I then integrate the different layers of vulnerability into one single perspective through ecological linkage with focus on London (chapter eight). Based on the joint evidence from the health milieus and health environments, I develop a third typology – health spaces – which indicate areas where social causes of health outcomes are likely moderated by locally specific conditions, which need to be investigated and accounted for as part of strategic efforts to address population vulnerability. Following this reasoning, I identify central elements of an urban policy programme set to address population vulnerabilities in London. Mirroring the advanced geodemographic framework, the programme speaks to different ecological levels and provides a multisectoral and context-sensitive alternative to conventional public health approaches. Finally, conclusions with respect to the significance of the advanced framework in current social science and urban policy debates complete the thesis along with recommendations for future research (chapter nine).

Internally, the empirical chapters (four to eight) are structured in almost identical ways. A first section briefly reviews research undertaken to date on the chapter's subject matter. A second section summarises the data and methods used. Subsections marked as "Technical specification" therein are intended to provide details to the reader with an interest in methods; they are not essential to the understanding of the results and may be skipped in reading. The results are then presented in the following two sections, each treating them from a particular angle. The only exception to this is the last empirical chapter (eight), as it exclusively builds on the previous chapters and develops the urban policy programme. All chapters – theoretical, conceptual and empirical – contain a final synthesising section summarising and discussing the findings and highlighting those aspects that are material to the remainder of the thesis. A glossary of technical, statistical terms is provided at the end of this document, and all terms with an entry therein appear underlined in the text.

# 2 Geodemographics and social science

Geodemographics is a branch of geographic information science concerned with the classification of local areas by the characteristics of their residents. If its construction and interpretation is based on social theory, geodemographics can be a powerful scientific device to study the nature of and processes within society [Parker et al. 2007]. Geodemographics studies society ecologically: it rests on the assumption that individuals with similar characteristics live spatially proximate to each other. Thus, small areas are considered to be sufficiently homogeneous to describe a group of individuals as a spatial unit with reasonable accuracy [Longley 2005a].

## 2.1 Theories of socio-spatial differentiation

The idea that similar residents live close to each other goes back to the socio-ecological positions of the Chicago School in the early 20th century. Some scholars, notably Ernest Burgess, Louis Wirth, Robert Park and Roderick McKenzie, observed and studied processes of socio-spatial differentiation in US-American cities. They contemplated that capitalist forces encouraged competition for residential space among different social groups, which are classifiable by their positions in the hierarchical process of production [McKenzie 1924, 290; Wirth 1938, 15]. Competition for residential space would produce socio-spatial dynamics that, at the ecological level of urban neighbourhoods, are manifest in socially differentiated zones within expanding urban systems and temporally demarcate the frontiers of distinct socio-economic communities [Burgess 1968, 54].

This socio-ecological conceptualisation of urban dynamics provided perhaps one of the earliest theories of socio-spatial differentiation and led to the creation of analytical tools, such as methods of factorial ecology, to characterise the spatial manifestation of social class [Longley 2005b, 923; Rees 1971, 222]. But studies applying these methods operated within a positivist-reductionist framework, primarily describing quantifiable phenomena that were taken to be natural. The studies became detached from ongoing, sociological debates about social structure and class; debates, which shifted emphasis away from stratification by one's position in the production process to stratification by consumption patterns [see Bourdieu 1984, 106 sqq; Beck 1992, 91 sqq; Giddens 1991, 82 sqq]. This shift commanded not just a stronger focus on lifestyles as a central driver of social difference, but, more fundamentally, the reworking of the theoretical foundations of sociological phenomena to integrate individual, subjective perceptions and objective, structural determinants of social practice [Grenfell 2012, 43; Reibel 2011, 308; Smith 2000, 1016; Harvey 1996, 358; Giddens 1982, 182].

Pierre Bourdieu was one of the influential social thinkers who advanced the theoretical debate on individual agency in the context of 'fixed' social structures. He demonstrated that the relation between a subject (actor) and social structures are manifest in distinct lifestyles discernible for different groups of individuals. In his view, objective, occupation-based measures of social class failed to adequately represent the positioning of individuals in the social order [Bourdieu 1984, 260]. His detailed consideration of the interdependent relationships between the social environment, different forms of capital and individual dispositions provided the conceptual basis for lifestyle studies as a new approach to study social class, which influenced commercial research and modern geodemographics [Maton 2012, 49; Burrows & Gane 2006, 806].

Central to Bourdieu's theory is the concept of habitus. Habitus represents a system of dispositions emanating from individuals' embodiment of social structures, lived experience of social and material circumstances (age, gender or occupational position) and subjective orientations (values, preferences or taste), all of which are in turn structured by habitus.

> "The habitus is not only a structuring structure, which organizes practices and the perception of practices, but also a structured structure: the principle of division into logical classes which organizes the perception of the social world is itself the product of internalization of the division into social classes." [Bourdieu 1984, 170]

Individuals' enduring lived experience of social structure fashions dispositions that are manifest in distinct, relatively stable tendencies to act and react in a context of choice. Social circumstances and individual properties structure subjective perceptions of behavioural possibilities and impossibilities, which produce routines that emerge as differentiated lifestyles at an aggregate (group) level [ibid., 101]. Habitus is the principle by which individuals embody social structure as it is experienced; but practice is also shaped by prevalent social rules in different domains of society and an individual's life trajectory.

Bourdieu uses the formula $([habitus][capital]) + field = practice$ to express that individuals act (practice) according to the interaction between their dispositions (habitus) and their social position (capital) in the context of the current workings of particular social domains (fields), where different forms of capital are at stake [ibid., 101; see also Maton 2012, 50]. One of the main objections raised by Bourdieu's critics is the implicit emphasis of structure over agency and a consequent danger of structural determinism [Longhurst & Savage 1996, 285; Bennett et al. 2009, 13; Jenkins 1992, 77]. But if practices are understood as the result of interactions between multiple identity-forming processes and conforming individual strategies, habitus becomes a contingent delimiter of individual intentions and actions in relation to a set of possibles rather than a definite determinant [Bourdieu 1977, 95; Bourdieu 2000, 214]. Moreover, structure and agency become intimately tied together: one constitutes the other through social practice.

In Bourdieu's thinking, social practice is recursive: practices cumulatively shape habitus and maintain or modify access to capital within a given field in a context of struggle, which in turn shapes social practice. Social practice is momentarily independent: habitus permits transposition of social actions according to the nature of the field the individual encounters in the moment of action. Practice, habitus and capital are interdependent in the long run: together, they constitute fields and direct their natures and workings. It follows that practices are contingent and themselves constitutive of social structure and its embodied variants within agents [Bourdieu 1977, 78]. Habitus is internalised through mental structures of individuals; for Bourdieu, these structures are inert, move with the body and, in this respect, resemble biological dispositions, such as genetic ones [Bourdieu 1990b, 14].

At an operational level, Bourdieu demonstrates in his empirical analyses the limited utility of unidimensional specifications of class, such as by occupation or income. Only broader, multi-dimensional specifications that alongside economic capital include other forms of capital, individual dispositions and field characteristics provide an adequate description of similarity and distance in social space [Bourdieu 1990a, 126 sqq]. Multidimensionality is an essential characteristic of geodemographics, which the technique establishes primarily through common geo-reference of different area attributes [Harris et al 2005, 82]. Bourdieu rarely discusses geographical space on its own account, but where he touches on it, he describes the role of geographical space as providing opportunities for appropriation of rare assets, resulting in socially ranked geographical spaces that are never neutral or balanced in terms of the social characteristics of their users [Bourdieu 1984, 102]. He writes elsewhere:

> "Social space tends to be translated, with more or less distortion, into physical space, in the form of certain arrangements of agents and properties. It follows that all the division and distinction of social space (high/low, left/right) are really and symbolically expressed in physical space, appropriated as reified social space." [Bourdieu 2000, 134-135]

Certain groups of agents execute their dispositions in decisions of location, whereby space is viewed as space of opportunities to cultivate, intensify, adjust or transpose one's practices [cf. Bourdieu 1984, 105]. Social distance translates into spatial distance; social position translates into ways in and modes by which physical space is occupied and appropriated. These spatially inscribed social structures are partly embodied in mental structures through experience and perception, which in turn tend to reproduce those spatial structures [Bourdieu 1999, 126]. Geographical space thus influences habitus as a physical manifestation of social distance and proximity and the associated distance (more specifically, travel time) to goods and services. [ibid., 127].

Savage et al [2005], who apply Bourdieu's theory in their study of social cohesion in selected Manchester neighbourboods, hold that habitus operate primarily in the residential environment. The authors conclude

> "One's residence is a crucial, possibly the crucial identifier of who you are. The sorting processes by which people chose to live in certain places and others leave is at the heart of contemporary battles over social distinction. Rather than seeing wider social identities as arising out of the field of employ-

> ment it would be more promising to examine their relationship to residential location." [Savage et al. 2005, 207]

## 2.2 Geodemographics – a geography of distinction?

If, then, residential location reflects agents' positions in social space, it is possible to detect different social milieus spatially through geodemographics. One way to do this is by examining nighttime geographies in terms of their measurable socio-spatial characteristics.

> "Crucially, however, 'classes' in these classifications are clustered around precise geographical coordinates, so that 'class identification' becomes a spatial practice, and hence can no longer be separated from the identification and sorting of places. This spatialization of class binds populations together in imaginary ways according to the micro-territories they inhabit." [Burrows & Gane 2006, 808]

This position suggests a spatial codification of Bourdieu's habitus, and, indeed, researchers who sought to embark on what has been called 'habitus mapping' find evidence of a generative and reproductive role of space for certain social classes [cf. Butler & Robson 2003; Savage et al. 2005]. These accounts suggest a co-existence of two processes involved in shaping socio-spatial differentiation: *sorting* as consequence of individual dispositions and *shaping* of practices through the lived experience of social position in residential space. It can be argued that whether the latter occurs within a conscious sense of community or by much more subtle forces and to weaker effects may depend on the type of neighbourhood. The extent to which habitus and its expressions are spatial may itself be unique to places rather than universal. Certainly, both types of processes are consistent with Bourdieu's theory of social distinction.

Since the relationship between what is measured as capital or demographic properties and sets of dispositions is unknown, though not random, the researcher needs to make assumptions about how the measured variables relate to capitals and habitus in the context of the prevailing workings of social environments. Those assumptions often tacitly inform cluster labelling and descriptions, which complete the process of geodemographic classification.

Harris et al [2005, 60 sqq] provide an overview of existing geodemographic classifications in the UK. Researchers run a clustering algorithm on relative frequencies of educational qualifications, occupations, housing, car ownership etc and produce a taxonomy of small areas. For example, the UK Output Area Classification 2001 identifies seven main groups (super groups) of areas, 21 groups and 52 sub-groups [ONS 2014] of Census Output Areas based on 41 census variables. CACI ACORN segments postcodes into six categories, 18 groups and 62 types [CACI 2013], and MOSAIC produces 15 groups, 66 types and 243 segments of UK post codes [Experian 2014]. Data sources used for geodemographic classification can be diverse: commercial classifications such as ACORN or MOSAIC include census data, routinely collected administrative data and surveys [Abbas et al. 2009, e37].

Although geodemographic classifications have been widely used in marketing, their application in social research is less common, partly due to a process of alienation between academia and market research since the 1970s [Reibel 2011, 310]. Broadly speaking, three types of research applications of geodemographics can be identified in the literature: explanation, study design and reference, and interpretation of the social world.

Studies seeking to explain an outcome of interest use geodemographics to describe exposure of subjects, often to augment available information on socio-economics with lifestyle context. For example, some authors use geodemographics to characterise health inequalities. Kimura et al [2011] use MOSAIC Japan groups to explain inequalities in influenza incidence in Japan. Dedman et al [2006] use the P2 geodemographic classification to characterise the most vulnerable groups to ill health as 'urban challenges', 'disadvantaged households' or 'multicultural centres'. Norman and Fraser [2014] find that the prevalence of life-limiting conditions is higher in children of specific ONS geodemographic groups in England. In another research field, Anderson [2010] studies road accidents in London and finds that drivers with certain geodemographic backgrounds are more often involved in accidents than other groups, suggesting potential, "micro-cultural" causes for road accidents. As an example from crime research, Breetzke and Horn [2009] develop their own geodemographic classification of neighbourhoods in Tshwane, South Africa, to characterise neighbourhoods with high concentrations of offenders. Harris et al [2007] study parental school choice in Birmingham and find significant interactions between ethnicity, neighbourhood and ONS neighbourhood type.

Geodeomographics is also used to review uptake of health screening. Nnoaham et al [2010] run a multilevel regression analysis of non-response to a screening invitation of colorectal cancer in South England and find that specific segments of the P2 (People and Places) classifications show low propensity to uptake. Sheringham et al [2009] employ the ACORN health classification to monitor differences in uptake of chlamydia screening in England. Muggli et al [2006] carry out a descriptive investigation of different frequencies of prenatal screening for Down's syndrome in mothers in Victoria, Australia, by geodemographic group. Their work suggest that rather than social status alone, it is the interaction of socio-economic and geographical characteristics (accessibility) that influences screening uptake. The authors testify significant potential of geodemographics for future studies of this kind.

In the second type of application, geodemographics informs either study design or serves as an analytical framework to prepare geographically targetted policy interventions. Birkin and Clarke [2011] use the 2001 UK Output Area Classification (OAC) in spatial microsimulation to synthetically estimate birth rates and car ownership in small areas in Leeds. Here geodemographics does not serve as explanation but as means of stratification to improve spatial estimates of social outcomes for the purpose of urban

Table 2.1: Review of applications of geodemographics in research

| study | domain | object | unit of study | c'ry |
|---|---|---|---|---|
| | | **explanatory** | | |
| Anderson 2010 | road accidents | people involved in road accidents | accidents in London | UK |
| Breetzke & Horn 2009 | crime | crime offender profiles | inmates in Tshwane | ZA |
| Dedman et al. 2006 | health | health care demand | hospital patients in NW England | UK |
| Harris et al. 2007 | education | parental school choice | parents in Birmingham (sample) | UK |
| Kimura et al. 2011 | health | influenza cases | influenza patients in Isahaya | JP |
| Muggli et al. 2006 | health | uptake of down syndrome screening | screened women in Victoria | AU |
| Nnoaham et al. 2010 | health | response to cancer screening | screening invitees in England | UK |
| Norman & Fraser 2014 | health | life-limiting conditions in children | hospital patients in England | UK |
| Sheringham et al. 2009 | health | chlamydia risk and screening uptake | persons screened for chlamydia | UK |
| | | **study design** | | |
| Batey et al. 2008 | urban policy | deprivation risk | small areas in England | UK |
| Brown & McCulloch 2001; McCulloch et al. 2003 | health | sample design for evaluations | GP patients in Mereyside | UK |
| Farr & Evans 2005 | health | risk of diabetes | small areas in Slough | UK |
| Petersen et al. 2010 | health | likelihood of hospital admissions | small areas in London | UK |
| | | **interpretive** | | |
| Debenham et al. 2003a,b | employment | economic vulnerability | small areas in Yorkshire | UK |
| Green et al. 2014 | health | mortality causes | small areas (MSOA) in England | UK |
| Longley et al. 2008 | ICT | ICT user types | households in Great Britain | UK |
| Ojo et al. 2013 | social geography | deprivation | small areas in Philippines | PH |
| Shelton et al. 2006 | health | mortality causes | parliamentary constituencies in Great Britain | UK |
| Spielman & Thill 2008 | social geography | segregation | small areas in New York | US |
| Webber 2007 | social geography | metropolitan communities | small areas in the UK | UK |

policy. The authors find that OAC groups indicate the geographical distribution of social phenomena better than conventional variables. Batey et al [2008] use the 2001 ONS classification in conjunction with the IMD to identify areas where deprived households are concentrated. Similarly, Petersen et al [2010] evaluate the use of geodemographic groups to capture actual health care utilisation and ascertain some potential of geodemographics to enable robust inference of local health needs. Farr and Evans [2005] demonstrate that there is potential of geodemographics in locating unknown diabetics. In an earlier study, Brown, McCulloch and colleagues employ the 1991 ONS classification to develop a novel research design that can be compared to an area-matched control trial evaluating of the success of a health screening campaign [Brown & McCulloch 2001; McCulloch et al. 2003]. Here, geodemographics is proposed as a means to inform sample design and support methods of intervention appraisal.

The third type of study comprises those that develop a bespoke geodemographic classification to describe the social system and interpret its parameters with reference to a particular problem. Shelton et al [2006] create mortality profiles for 76 regions in England and Wales in order to identify areas with similar health challenges. They discover distinct clusters of spatially co-occurring mortality causes, which remain stable over time. Green et al [2014] conduct a similar study in England classifying 7,194 small areas into eight groups of different mortality profiles. Debenham et al [2003a,b] build a new census-based classification to assess economic dependency of neighbourhoods on specific industrial sectors in North England. Longley et al [2008] use the UK Enhanced Electoral Roll to create a typology of households, correlate those with Information and Communication Technology (ICT) use and add Experian's Mosaic UK to produce a new area classification indicating engagement with ICT. They demonstrate how geodemographic classifications can reveal distinct behavioural phenomena and their complex geographies beyond simpler notions of social exclusion and digital divide. As an example from urban and regional studies, Webber [2007] investigates the relative location of Experian's Mosaic milieus and is able to identify distinct metropolitan groups. Although he uses an existing classification, his work provides an example of the use of geodemographic groups to refine descriptions of social geographies.

To summarise, geodemographics has been applied under various research designs and in various domains of social science. The majority of applications have been in data-rich and English-speaking contexts. All authors assume that geodemographic segments reflect real differences between areas, that they describe a geography of really distinct groups; but how distinction is interpreted and applied varies by type of study. Most studies of the first 'explanation' group ascertain a potential of geodemographics in accurately characterising if not explaining or predicting an outcome of interest. Sometimes geodemographic categories are treated as control variables in linear models, or are directly interpreted nominally by their labels to characterise exposures in the absence of direct data on individuals, lifestyles or ethnicity.

Some researchers compare the classification they use to the English Index of Multiple Deprivation (IMD) and find that the geodemographic classification performs better in identifying group differences. It should be noted, however, that the IMD and the geodemographic systems used operate at different geographical scales: the IMD describes deprivation at the coarser scale of Lower Super Output Areas; so the better performance of the more fine-grained geodemographic systems may be due to higher apparent geographical precision. Discussion of geographic scale, neighbourhood size or geographical extent are rare in this and the second type of studies.

The second type places greater emphasis on validating the geographical distribution of phenomena inferred from geodemographic characteristics. While they report some success in this undertaking, the lack of predictive power of generic geodemographic groups suggests that segmentations need to be more purpose-focussed if they are to inform area-based initiatives. Bespoke classifications were built in the third, interpretive type of studies, where geodemographics acts as a hermeneutic to characterise distinct system parameters and their distribution in space.

This review demonstrates that geodemographics can be applied widely for research purposes. Before specifying how geodemographics can be used to study health inequalities, it is necessary to gauge the limits of the different uses of geodemographics.

## 2.3 Geodemographics as a hermeneutic

In his extensive critique of positivist-reductionist approaches to scientific inquiry, Byrne endorses a realist research programme that 'interprets the real' and 'describes the complex' [cf. Byrne 2002, 12]. Byrne considers classification to be a 'case centred' and 'case driven' research practice [ibid, 101], by which the researcher groups cases (or systems) according to similar properties (phase states) and trajectories. Classification accounts for the case as a whole with all the information available including interactions of multiple case properties.

> "That is to say, systems whose trajectories belong in the same ensemble at any given time point are members of the same class." [ibid, 35].

Unidimensional (monothetic) measurement therefore seems inappropriate to describe things; only classifications based on multiple dimensions (polythetic) can take account of the multiplicity of tendencies that complex objects possess, which Bourdieu also pointed out with respect to social studies. Byrne considers multidimensional classification methods, such as cluster analysis, to be robust in relation to the real:

> "Basically, if there is a real underlying taxonomy to be found then different clustering methods will produce essentially similar classification account [sic] when applied to the same data set." [ibid, 100].

Uprichard describes the classifying research strategy as method to study a case in relation to other cases, that is "based on the knowledge of the *whole* case, and not on knowledge of one or more *aspects* of the case" [2009, 134, original emphasis]. Applied to geodemographics, the researcher studies the nature of neighbourhoods by viewing the neighbourhood in relation to all other neighbourhoods. The fact that there are similar neighbourhoods that fall into one class may tell us something about that underlying structure which generates differences among neighbourhoods. In Byrne's terms, geodemographics focusses on ecological system characteristics [Byrne 1998, 99] and not directly on individuals or households.

Desrosières [1998] discusses extensively the reality of social aggregates with their apparent epistemic contradiction. On the one hand, certain social phenomena show remarkable regularity and stability at aggregate population levels: birth, unemployment or crime rates do not fluctuate significantly over time. On the other hand, the underlying instances of having children, being unemployed, being criminal or the like appear too diverse for reliable predictions at the individual level [ibid., 75]. This duality has led to two opposing views in statistics: the frequentist view, whereby tendencies manifest themselves randomly around an 'average' individual or resident, and the epistemic view, in which there is a common cause that leads to aggregate outcomes and therefore lend consistency to individuals being grouped into one class [ibid., 78].

The frequentist assumption of an average individual – or average resident in case of geodemographics – in whom some or all group 'averages' are unified entails ecological fallacy. From a realist perspective, the epistemic view, is to be preferred: here "aggregates do exist" [ibid., 101] at their level and cannot be directly conferred to individuals. The stability of aggregates depends on the strength of causal factors at a given system level; those factors may themselves not operate homogeneously in space or time. Our knowledge of causal factors generating group differences is incomplete and can only be improved gradually over time by repeated measurements at multiple system levels, observation of system changes and triangulation [Byrne 1998, 97 sqq].

In this realist view, geodemographics becomes interpretive practice that characterises social systems primarily at the level of small areas at given points in time. It offers the prospect of characterising the system, its parameters and trajectories with reference to a given problem of interest. But neighbourhood classifications have their limits and geodemographics has been criticized on a number of counts. The critique can be divided into three strands: one relating to the contents of geodemographics, another relating to the method, and another to the interpretation and potential social consequences of geodemographic systems.

As for the first strand of critique, a number of authors note the substantive shortcomings of geodemographics and propose extensions through additional information to capture dynamics that have hitherto thought to be outside the remit of neighbourhood classification. First, geodemographics represents a taxonomy at one particular scale, but in reality, that scale is nested in higher level systems: neighbourhoods are affected by processes at higher and lower levels [ibid., 100]. Neighbourhood change is a contingent outcome of complex interactions at all of these levels. Higher-level events create the potential for similar phenomena at different locations; a comparison of different locations may reveal those higher level processes. But by largely relying on neighbourhood level variables, geodemographic systems alone do not contribute to analysing those processes in a meaningful way [Debenham et al. 2003b; Voas & Williamson 2001]. More utility of geodemographics could be derived by incorporating data pertaining to different levels, but little progress has been made in this respect.

Second, each neighbourhood is composed of different individuals, networks and connections with distinct physical and topographic conditions. They can be outliers compared to other neighbourhoods and possess their own unique dynamics. Geodeomographics has been critcized for averaging out or masking local specificities through simple group membership based on abstract, geographically uniform measurements [see Singleton & Longley 2009; Voas & Williamson 2001]. Failure to account for local specificity may lead to geographically naive and ineffective applications, and some critics emphasise the need to systematically consider regional patterning of area attributes.

Third, geodemographics are typically static classifications that describe neighbourhoods at one point in time with little information about the stability of either the membership of one entity or the classification system itself. Incorporating those dynamics would be important to assess the spatio-temporal validity of taxonomies and to assign entities to classes [Byrne 2002, 108].

The second strand of critique stems from doubts about the actual distinctiveness of geodemographic groups in the real world. Voas and Williamson [2001] test the spatial interaction of 1991 UK census variables and show that geodemographic classification reduces only marginally the variation of variables. They discover that any one geodemographic class often encompasses two thirds of the entire variation of any variable; thus, clusters overlap widely in multivariate, statistical space, which renders generalisations about neighbourhoods imprecise. The authors conclude that a general purpose classification is probably impossible to create and suggest purpose-built and parsimonious classifications.

Graham [2005], Pickles [1994] and Goss [1994, 1995] are more pessimistic and argue that geodemographics is after all reductionist: classification reduces the initial number of cases to a lower number of second order observations, an abstract data structure.

> "The abstract data structure is then anchored to a direct representation of reality, which leads to the conceit that the world of the GDIS [Geodemographic Information System] is itself another reality" [Goss 1994, 143]

The thus created 'second order reality' is incomplete and misleading, since not only the groups of cases are statistical abstractions, but also the cases themselves – areas as ensembles of diverse individuals – are fictitious entities. Consequently, pen profiles of neighbourhoods are based on imaginary, falsely territorialised and distorted representations of social life with questionable value for policy [cf. Goss 1995, 182]. This misconception not only gives rise to ecological fallacy, further compounded by the omnipresent problem of the modifiable area unit problem [Openshaw 1983], it also encourages analysts to misconceive social identity (or consumer identity) as coherent at best and stereotypical at worst [cf. Goss 1995, 187].

These concerns directly lead to the third strand of critique: the potential social consequences of geodemographic systems. According to the critics, geodemographics tends to take consumption as the defining feature of people's identities, implying a normalised view of social life that disregards and penalises those unmeasured deviations that contradict reductionist expectations; a similar error to committing the naturalistic fallacy. The reification of the abstract data structure into socio-ecological imaginaries may entail dangerous socio-political consequences, particularly reproduction of unequal social geographies and stigmatisation [cf. ibid., 191].

Burrows and Ellison [2004; Ellison & Burrows 2007] note, for example, that online geodemographics supports households' relocation decisions and thus enables powerful groups to sort themselves into the best matching, best supplied and most privileged neighbourhoods, leaving marginalised groups trapped in socially disorganised environments. Although this process has always existed, the authors suggest that this way of sorting is accelerated such that it culminates into a new, emergent form of neighbourhood politics with advanced polarisations of engagement and disengagement, urban citizenship, fragmentation and unequal appropriation of public space [Burrows & Ellison 2004, 326]. Here, geodemographics reproduces and instrumentalises new guises of digital divides [Burrows & Gane 2006, 805]: the differential, reflexive use of digital technologies, perhaps by mobilising a new type of informational, cultural capital, and the invisible, software-sorting that algorithmically classifies urban space to control access and resource distribution. Graham [2005] finds this possibility particularly alarming when automated software heuristics and classificatory taxonomies are produced under neoliberal power regimes that, in their instrumentalist motivation, seek to render complex and unpredictable consumer decisions predictable, manageable and controllable.

Besides calls by some authors [Goss 1995; Graham 2005] for sabotage if not the boycott of geodemographic classifications, this strand of critique raises the important question of how and in what context geodemographics should be produced, verified and ultimately used in research and policy. A carefully designed, purpose-built and validated classification may be useful to inform the design of studies that investigate a specific problem. The explanatory use of geodemographics, however, appears problematic. Indeed some of those studies come very close to an act of "moral labelling" [Dowling 2009, 834]. Breetzke and Horn [2009], based on their generic geodemographic classification, concluded that black and deprived neighbourhoods are home to most crime offenders in Tshwane, South Africa. Anderson [2010], by profiling drivers involved in car accidents based on their residences, suggests that certain groups are more likely to cause accidents and inadvertently loads normative contents into geodemographic labels. Extended to location decision-support, uncritical application of labels can reinforce stigmatisation, sorting tendencies and segregation.

Yet, one may also argue that well-balanced descriptions can challenge unquestioned neighbourhood images and, if transparent, can offer a more differentiated picture of neighbourhoods and their social compositions. Accurate pen profiles of neighbourhoods may well speak to a multitude of criteria in location decisions and serve to de-stigmatise neighbourhoods, provided the power of the classifier is mediated through transparency and consultation [cf. Thrift & French 2002, 331]. Longley and Singleton [2009] argue that user engagement and openness to scrutiny need to be central in classificatory research practices, particularly when applied to high resolution data, whose usage currently prompts growing concerns over information governance, data security and confidentiality.

In urban research, studies that use geodemographics as structural element in their research design show that software-sorting can highlight needs of otherwise invisible groups, reveal complexities beyond simplistic perceptions of the social world and may unmask unexpected aspects of community needs [e.g. Brown & McCulloch 2001; Farr & Evans 2005; Longley et al. 2008]. Applications by those who follow critical agendas and dedicate themselves to equality and justice are just as well imaginable as misuse to malign ends. As Goodchild [1994, 32] notes, "a technology that can be used to promote democracy can also be used to deny it". Residents of marginalised neighbourhoods, social activists, third sector organisations, critical researchers can use geodemographic information system to point to social issues, traces of negligence, ongoing marginalisation and the need for agendas of improvement, in particular where conventional survey data fails to capture those phenomena, and, perhaps through alternative and bespoke classifications and descriptions, to expose established ones [Burrows 2013; Savage & Burrows 2007].

The key barrier in so doing, however, is that the data to produce classifications are still largely gathered in secluded spheres by companies and governments. But in movements such as UK Open Data, we also witness improvements of data dissemination and transparency with an ambition to build open licences and what Thrift [2006, 301] calls a "new kind of creative common" to proactively put in place freedom of information and open up conventional data property regimes. This trend may create better prospects for transparency, validation and verification of geodemographic classifications, including the monitoring of their use.

From the critique it becomes clear that geodemographics not only describes but intervenes in and modifies the social world, and that with online dissemination, the "double hermeneutic" of social science [Giddens 1982] acquires a new spatial and temporal imminence. This point, however, may not be specific to software-sorting alone, but the information society as a whole, whose collective habitus produces an intensified "urge to classify" [Burrows & Gane 2006, 803] and thus reproduces the classes that themselves constitute the information society and their power relations defined by exclusion [Lash 2002, 75].

## 2.4 Synthesis: possibilities to advance geodemographics in research

Geodemographics classifies neighbourhoods in their entirety and permits the study of their observable socio-spatial attributes. Bourdieu's theory offers a conceptual framework that can help us build and interpret classifications, not just because geodemographics to date has been particularly concerned with lifestyles, but also because his theory offers sociological explanations for socio-spatial differentiation. Grounding geodemogrpaphics in Bourdieu's theory entails commitment to scientific realism in accounts of causality and by extension recognising that the social world is complex. In this understanding, geodemographics has to aim at capturing interactions as a way to study and perhaps explain diverse social phenomena at multiple levels, notably neighbourhoods.

From a technical point of view, critics note that geodemographic group membership is uncertain, and taxonomies are fuzzy rather than clear-cut. Desrosières understands the aim of statistical work as making "a priori separate things hold together, thus lending reality and consistency to larger, more complex objects" [Desrosières 1998, 236]. Robust classification is one way to make separate things hold together and, in addition, is particularly apt to reveal previously invisible differences by emphasising specific aspects of distinct classes [Bowker & Star 1999, 29]. At the same time, variable selection, standardisation and choice of algorithm may easily favour some pre-conception of the social world and weaken others – therefore, there is an acknowledged need for sensitivity testing and verification. But clusters should not simply be reified, as some studies do where geodemographic labels are taken nominalistically as control variables or explanatory categories. Brown and McCulloch describe the interpretive use of geodemographics in health research:

> "The approach provides a means of capturing the relationship between individual patients and the neighbourhoods and social space in which they live, and between the places in which people live and the incidence of disease or illness in such localities. It implies that those living in different area types are likely to have different patterns of social interaction, different degrees of exposure to risk and different levels of access to services." [Brown & McCulloch 2001, 112]

Thus it is the act of typifying that can be used to offer clues about causal pathways in health [cf. Byrne 2002, 105 sqq], although classification alone does not reveal what these different patterns of social interaction, exposures and access levels are, and if and how they are causal. Geodemographics provides a basis to explore social system characteristics through case groups as second order objects, their tendencies, stability and common trajectories, which may reveal the properties and dynamics of causal processes.

Some modifications would enhance the utility of geodemographics for social research and policy. The first is to build geodemographics for a particular purpose, to explore social phenomena of interest. Second, classifications should be parsimonious yet composed of diverse types of relevant data sources, such as supply-side information (infrastructure, services) or survey data. Third, given that neighbourhoods are surrounded by other neighbourhoods and nested in higher level systems, information that accounts for connections to other areas and influences from higher levels should be incorporated. Fourth, classifications should capture neighbourhood change to identify common trajectories of areas and consider area stability beyond snapshots at one point in time. Finally, a consciously developed comparative framework in which classifications can be contrasted to other possibles may reveal unique and locally specific patterning of classifications. Part of this step is the incorporation of information that points to the uniqueness of people and areas alongside generic measures of social distance.

Finally, it must be acknowledged that researchers or interpreters are themselves socially situated and apply extant categories in describing the social world. On the one hand, this lends aptitude and credibility to the interpreter, since personal experience with and knowledge of social reality share a degree of consistency. On the other hand, the social situatedness makes the interpreter necessarily partial. Reflection including the questioning of one's own stereotypes, transparency, verification, sharing and user engagement need to be essential parts of the research undertaking to meet the risk of subjective distortion in interpretations [cf. Bowker & Star 1999, 325; Desrosières 1998, 277; Longley 2007, 621; Uprichard et al. 2009, 2833] and to maintain critical consciousness about potential social consequences of geodemographics.

# 3 Characterising urban vulnerability

One of the central recommendations to improve geodemographics is to move away from generic towards conceptually focussed classifications. A target concept that seems instructive in designing neighbourhood classifications to characterise urban health inequalities is the concept of vulnerability. Vulnerability is a key theme in many debates held in urban research and policy. Often the term is used implicitly and resonates with resilience, exposure, deprivation or risk. Yet, there seems to be some transcending notion of vulnerability: groups of people are vulnerable to negative impacts of climate change [e.g. Hahn et al. 2009; Luers 2005; Parry 2007], to stress and health risks [e.g. Gee & Payne-Sturges 2004; Turner 2013; Zannas & West 2014], to hazards [e.g. Bankoff et al. 2004; Cutter et al. 2008; Hogan & Marandola 2007] or to displacement due to gentrification or coercive development schemes [e.g. Cao et al. 2012; Colburn & Jepson 2012; Pearsall 2012]. Geodemographics that focusses on vulnerability may therefore respond to a range of urban policy issues that have been identified as relevant in the social determinants of health debate. Capturing those may not only support the development of scenarios but also prepare strategic policy interventions.

## 3.1 Vulnerability as target concept

Vulnerability is an important concept in hazard research. Cutter and colleagues [1996; 2003] define the concept very broadly as 'potential for loss', which can acquire specific connotations depending on the problem of interest [Cutter et al. 2003, 242]. Cordona [2004] specifies this notion further:

> "Vulnerability may be defined as an internal risk factor of the subject or system that is exposed to a hazard and corresponds to its intrinsic predisposition to be affected, or to be susceptible to damage. In other words, vulnerability represents the physical, economic, political or social susceptibility or predisposition of a community to damage in the case of a destabilizing phenomenon of natural or anthropogenic origin." [ibid., 37]

Vulnerability is an intrinsic property of a subject or system, influencing the chance of survival or continuity in potentially harmful, external events. Cordona regards 'intrinsic predisposition' as the starting point of vulnerability, but the nature of this disposition remains to be specified. Hilhorst and Bankoff [2004] discuss their take on vulnerability and conclude:

> "Critical to discerning the nature of disasters, then, is an appreciation of the ways in which human systems place people at risk in relation to each other and to their environment – a relationship that can best be understood in terms of an individual's, a household's, a community's or a society's *vulnerability*." [ibid., 2]

The intrinsic predisposition results from structural processes that puts people into different risk contexts. Within a Bourdieusian reading of predisposition as habitus, vulnerability emerges from embodiment of social relations that structures perceptions, aspirations and actions, and limits their expressions to a set of possibles. But Bourdieu's view also demands that disasters – or risk – need to not only be investigated in relation to structure, but also in relation to people's agency. Awareness of vulnerability may prompt attempts to mediate it, if the necessary actions lie within the possible remit of group habitus. Or the interaction between real vulnerability and institutional response causes recognition of vulnerability and therefore action. Agency has an important place in the actualisation of vulnerability as contingent consequence of risk. Hilhorst and Bankoff [2004] maintain therefore that "vulnerability is a much more precise measurement of exposure to risk" than, for example, poverty or deprivation, because potential agency and coping capabilities modify vulnerability such that not all deprived communities can be said to be vulnerable and, vice versa, not all vulnerable communities are deprived [ibid., 2].

Cutter et al [2008] discuss how the term resilience – often understood as communities' coping capabilities – relates to vulnerability. They describe different perspectives and decide that the terms overlap while having distinct meanings. Intuitively, the term can be treated as the antonym of vulnerability: the more vulnerable communities are, the less resilient and vice versa. This notion seems consistent with Cutter and colleagues' review as well as other authors [Hogan & Marandola 2007; Pendall et al. 2012] and shall be applied in this research.

That vulnerability is an outcome of social relations has come to be a central tenet in public health within the social determinants of health paradigm. Though less explicitly defined, vulnerability is broadly understood as the inability of a community to cope with disease burden [Blas et al. 2011, 2; Koot et al. 2011, 164]. This notion originates from the work of the Commission on Social Determinants of Health [CSDH 2008], whose members postulate a causal chain starting from socioeconomic position, which causes differential exposure to risks and differential vulnerability, leading to differential health outcomes and long-term consequences at the individual level [Blas & Kurup 2010, 6 sqq].

The causal chain of the CSDH model runs through different ecological levels at which distinct causal factors operate [Figure 3.1]. Socio-economic forces operate at the societal level as macro forces that place different individuals in different positions and therefore structurally pre-confine possible exposures, vulnerabilities and outcomes. Scambler [2007] situates here the logic of capital accumulation and regulation regimes as guiding modes of reproducing health inequalities. Differential exposure refers to risk factors in the social and physical environment, to which different socio-economic groups are exposed differently. Disadvantaged groups are often exposed to poor environmental conditions, poor housing, poor working conditions and so forth. But, consistent with Hilhorst and Bankoff's reasoning described above, disadvantage due to socio-economic position and differential exposure does not necessarily lead to worse health outcomes.

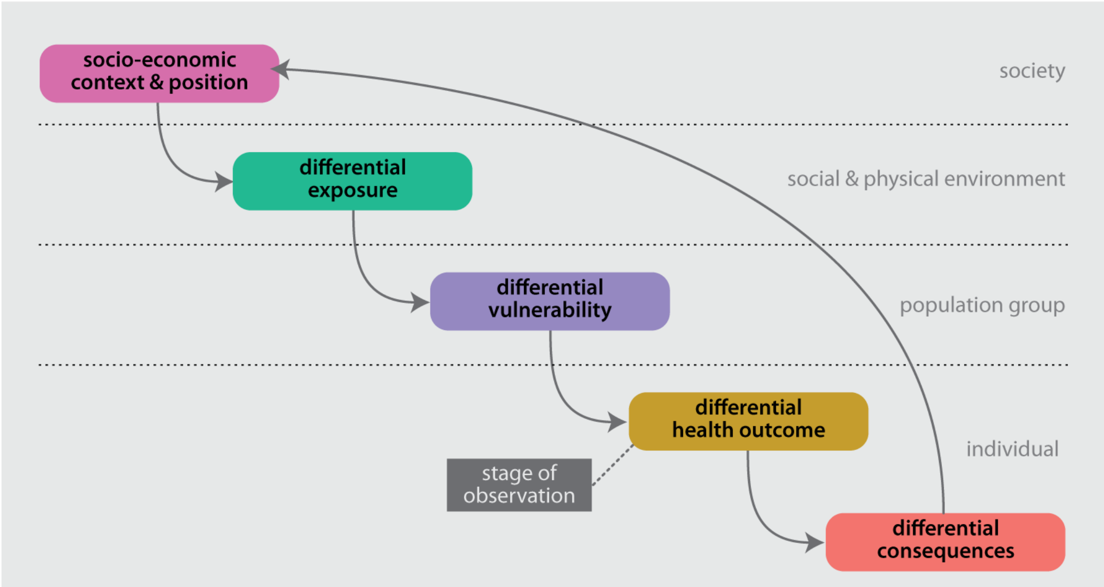



Figure 3.1: The causal pathways for health action by the Commission of Social Determinants of Health [adapted from CSDH 2008, 175]

Health is mediated by vulnerability at the group level: for example, disadvantaged communities that have access to improved preventive health care may not experience the consequence of disadvantage. According to the CSDH framework, vulnerability is a characteristic of a population group, which then determines individual outcomes and consequences [Blas & Kurup 2010]. Neighbourhood vulnerability could be understood as a special case of the CSDH's notion of population group vulnerability; in fact, it may be asked to what extent the CSDH framework tacitly assumes spatial clustering of differentially vulnerable population groups. As a consequence of group vulnerability and consequent impacts on individuals, we perceive social patterning of health outcomes – the essence of health inequalities.

The CSDH model recognises an epidemiological notion of embodiment: "historically contingent, spatial, temporal, and multilevel processes become embodied and generate population patterns of health, disease, and wellbeing, including social inequalities in health" [Krieger 2005, 350]. Brunner and Marmot [2006] identify three pathways through which socio-economic position connects to pathophysiological response [Figure 3.2]: material, behavioural and psycho-social pathways. The material pathway focusses on unequal exposure of socioeconomic groups to environmental hazards, the behavioural relates to differential behavioural tendencies of socio-economic groups, and the psycho-social presumes exposure to stress that is specific to socioeconomcic status, including low self-esteem due to experience of relative social disadvantage. Stress affects the way the brain regulates the nervous system and consequent physiological responses [ibid., 11]. In addition, biogenetic characteristics and the current life stage of an individual modify these pathways and influence the way in which they play out in terms of health. Again, conceptual resemblance with Bourdieu's habitus can be observed: socio-economic

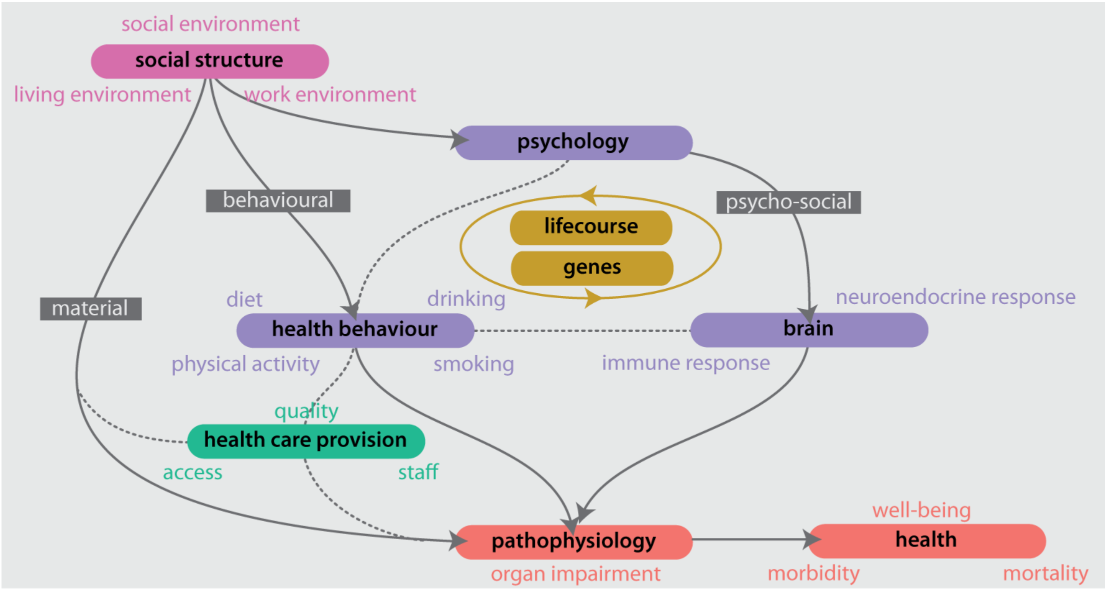



Figure 3.2: Causal pathways linking socio-economic status to health [adapted from Brunner & Marmot 2006, 10]

context, lived experience and imparted values become embodied and tend to reproduce themselves. Brunner and Marmot's model renders plausible potential effects of habitus on health and well-being in the long run [see also Blaxter 2010, 92].

Causal factors operating at multiple levels with interactions between geo-physical, social and political spheres mirror characteristics of complex systems. The language of systems thinking is also entrenched in the writings of disaster and hazard researchers [see Bankoff et al. 2004; Cardona 2004; Cutter et al. 2008; Oliver-Smith 2004]. The CSDH, too, recognises complexity and therefore recommends a health programme that acts at multiple levels [cf Blas & Kurup 2010, 38; CSDH 2008, 186]. These programmes ideally combine preventative and restorative interventions, but in practice expressions of vulnerability are only recorded at the stage when restorative action is taken. In contrast, preventative programmes focus on determinants of health and require information on the vulnerability of populations beyond individual health in order to be successful [cf Hawe et al. 2009, 268]. For "upstream" interventions at higher levels, other policy sectors need to be "joined up", including economic development, welfare, transport, housing and so forth [Moon 2009, 36].

The phenomenon of persistent health inequalities everywhere [GRNUHE 2010; WHO & UN-HABITAT 2010] has long driven a search for policy responses applicable within a social determinants approach to health. Health inequalities are stable, observed differential health outcomes across groups and are held to be unjust when they are socially induced and potentially avoidable [Asada 2005; Starfield 2001]. Health inequalities are typically identified when life expectancy, disease prevalence, premature mortality and the like fall or rise with some measure of social class [e.g. Graham & Kelly 2004; Marmot

2005, 2010]. The social gradient of health, however, does not explain the full variation of health, since any kind of illness occurs in any given class. Interventions acting upon the social gradient seem incomplete, if they are based on the 'average' deprived individual and are scaled up to higher level aggregates [e.g. Graham & Kelly 2004; Hawe et al. 2009]. Bourdieu's work suggests that heterogeneity of pathways exists at a much finer categorisation than social class. Health inequalities are complex and persistent phenomena, and the failure of strategic attempts to reduce health inequalities in the UK, for example, demonstrates that complex, long-term policy programmes responses are necessary [Mackenbach 2011].

In the UK, Joint Strategic Needs Assessments (JSNAs) were introduced as an instrument to take stock of health needs and inform so-called Joint Health and Well-being Strategies (JHWS). In 2010, these two instruments became statutory requirements for local authorities; the Health and Social Care Act 2006 sets out their responsibility to bring together stakeholders from multiple policy and social arenas relevant to health. The Department of Health defined JSNAs in 2007 as a "systematic method for reviewing the health and wellbeing needs of a population, leading to agreed commissioning priorities that will improve health and wellbeing outcomes and reduce inequalities" [UKDH 2007, 7]. The same document highlights the reliance of JSNAs on good quality data and proposes core datasets for the identification of local health needs [ibid., 13]. Needs identification and characterisation are part of a multi-step process of stakeholder and community engagement, for which disease monitoring systems, modelling tools, in-depth qualitative investigations and neighbourhood profiles are considered to be good practice [NHS 2011].

The practical problem arising from the JSNA requirement is the identification and interpretation of diverse social, environmental and clinical data and their integration into an overarching strategic needs assessment. Local authorities are expected to tackle this problem in order to deliver public services effectively. Geodemographics has long been identified as a potential tool of support here [Abbas et al. 2009; Longley 2005a; Openshaw & Blake 1995], if it overcomes the shortcomings outlined earlier and succeeds in capturing the spatial dimension of vulnerability.

## 3.2 The spatial dimension of vulnerability

If vulnerability is to be taken as a target concept for geodemographics, it is necessary that vulnerability is spatially representable. Cutter et al [2003] develop the notion of *place vulnerability*, which broadly describes vulnerability at the community level and is equated with small areas. Akin to the "spatialisation of class" [Burrows & Gane 2006], the approach presupposes a common social process that translates differential vulnerability among communities directly into spatial differentiation, which in itself constitutes a new, emergent dimension of vulnerability.

### Hazard of place studies

Oliver-Smith [2004] refers to the work of Wilches-Chaux [1989], who identifies eleven dimensions of vulnerability: natural, physical, economic, social, political, technical, ideological, cultural, educational, ecological and institutional vulnerability. Cutter and colleagues attempt to turn these dimensions into a spatially explicit model [1996; 2003]: the "hazard of place model". The model takes into account different kinds of vulnerability within a dynamic process of community adaptation and learning during and after the event of a disaster.

A number of researchers apply Cutter's model in order to create a spatial index of vulnerability. For example, Azar and Rain [2007] use the model to study social and biophysical vulnerability in Puerto Rico. Collins et al [Collins et al. 2009] apply the model within a comparative framework to assess vulnerability in Ciudad Juárez, Mexico, and El Paso, United States. Kumpulainen [2006] measures regional vulnerability of European NUTS3 regions in a study that is informed by the hazard of place model. There are other applications in diverse contexts, such as United States, Norway or Taiwan [Chang & Hsiao 2007; Cutter et al. 2000; Lujala et al. 2014; Maantay & Maroko 2009].

A number of observations – relevant to the geodemographic study of health inequalities – can be made about these studies. Hazard of place studies focus on measuring different degrees of vulnerability by means of continuous or ordinal composite indices. Unlike geodemographics, they do not distinguish different types of vulnerable areas, reflecting populations with different needs. Nevertheless, the studies reveal the specific information needs that arise in measuring multiple dimensions of vulnerability. Geographic Information Science plays a central role in managing diverse information pertaining to social characteristics, biophysical layers and the built environment. The hazard of place model provides a framework of measurement that specifies dimensions of vulnerability and allows application in different settings, different countries, including data-rich and data-poor contexts. Although most of the studies focus on environmental hazards as external events, the framework of measurement provides a useful starting point to extend the model to characterise vulnerability to ill-health. Cutter et al [2003, 246 sqq] list different indicators of social vulnerability; they all represent social determinants of health.

Exposure to environmental hazards is an important element in the study of vulnerability. In health research, air pollution, noise and crime are most obvious examples as corollaries to risks of flooding or earthquakes. From the social determinants of health perspective, the exposure layer is situated at the level of the social and physical environment and shapes vulnerability of populations. The studies also highlight the need to consider services that help communities cope with disasters. Access to health care would fall into this category, too, as well as access to different types of jobs, reflecting economic vulnerability and dependency on single industrial sectors.

Hazard of Place models conceptualise vulnerability as dynamic. Although the studies to date provide snapshots of one point in time, they acknowledge that vulnerability changes and monitoring changes is important to understand adaptation, recovery or deprivation of communities [Azar & Rain 2007; Cutter et al. 2008]. Cutter and Emerich [2006] monitor their social vulnerability index in four US states over time. They assume that changes in composite vulnerability indices reflect changes in actual vulnerability. Perhaps except for Cutter and Emrich's study, there are none that have systematically monitored change of vulnerability over time. One of the major challenges in formally measuring and characterising change consists in the multi-dimensionality of the problem. Hazard of place studies use factor analysis to construct vulnerability indices, which makes change over time difficult, since the composition of factors changes over time, too.

Finally, the hazard of place model proves robust in a multitude of contexts. Collins et al [2009] compare vulnerability within and across two different cities in two different contexts: Ciudad Juárez, Mexico and El Paso, United States. They use different data sources for the same dimensions of vulnerability and are able to demonstrate that vulnerability is three times higher in Ciudad Juárez than in El Paso. The contextualisation of local vulnerability within a comparative framework enabled the researchers to identify institutional capacity as the main driver of the increased vulnerability in Ciudad Juárez. As far as methods are concerned, they recommend that, in transnational vulnerability assessments, indicators be so selected as to "balance the need to incorporate contextual specificity with general comparability" [ibid., 459]. Thus, comparative research helps to not only assess local specificity of results but also to estimate the scale of their significance.

### Neighbourhood effects research

Neighbourhood effects research is another strand of literature that implicitly applies spatial notions of vulnerability. In contrast to hazard literature, this strand focusses on exposure in relation to outcomes in order to establish causality [cf Van Ham & Manley 2012]. Neighbourhood effects research studies potential impacts of neighbourhood characteristics on social outcomes, such as employment, educational performance or health [Van Ham et al. 2012]. In studies that take an interest in social epidemiology, the focus has primarily been on whether area characteristics influence health beyond individual ones [Cummins et al. 2007; Diez-Roux 2004; Macintyre et al. 1993; Macintyre et al. 2002]. Pearce et al [2011] discuss potential, behavioural pathways linking place to health. They distinguish between place-specific 'practices' and place-specific 'regulation and policy', which each unfold higher, neighbourhood level dynamics, affecting health-relevant behaviours of their residents. In spite of the plausibility of these pathways, evidence of neighbourhood effects remains mixed, and new research designs and techniques have been called for to improve the inquiry [Dietz 2002; Diez-Roux 2007; Galster 2012; Oakes 2004; Van Ham & Manley 2012; Van Ham et al. 2012].

Even if an independent effect of neighbourhoods is identified, one of the main issues is reversed causality. DeVerteuil et al [2007] illustrate the problem with the question of whether deprived and disorganised neighbourhoods cause mental illness – the social causation hypothesis – or whether mentally ill people drift into socially disorganised neighbourhoods over time – the social selection hypothesis. In both scenarios, neighbourhood context is significant, but the policy implications are very different. Neighbourhood effects researchers therefore call for longitudinal studies, but their sample size as well as confidentiality concerns pose a challenge for the detailed study of neighbourhood effects [Hedman et al. 2013; Manley & Van Ham 2012; Small & Feldman 2012].

It can be noted that much of the neighbourhood effects literature (including some social epidemiological studies) has focussed on the quantification of the relative contribution of areas, often with the goal to inform resource allocation [Van Ham et al. 2012, 2790]. Galster [2012, 28] uses the analogy of "dosage-response": if neighbourhood aspects are found to influence residents' well-being to a certain degree, a dose of improvement should increase the well-being to the expected degree. This idea presupposes linearity of causal mechanisms, thus failing to account for emergent properties: outcomes of interactions that bring about a qualitative change of neighbourhoods (such as 'ghettoisation'), may transform the workings of neighbourhoods and be causal to changes in well-being beyond linear relations [cf Byrne 1998, 100]. In reality, increasing advantage in neighbourhoods implies to change many 'variables' simultaneously. This is the core of Byrne's critique, when he complains that linear modelling tends to view exogenous variables (unemployment, crime, deprivation etc) as independent although they are traces of the same neighbourhood [cf Byrne 2002, 30]; in other words, the positivist programme tends to reify variables and lose track of the things being studied. It can be speculated that the inconsistency of findings in neighbourhood effects studies may be due to analytical attempts to force linearity and independence of variables on a non-linear world – a world made of things, rather than variables.

A second limitation of neighbourhood effects studies relates to the notion of neighbourhood or place. The terms are used in the sense of abstract statistically homogenous zones formally described by median income, proportion of commuters by public transport or other summary statistics. Cognitive aspects of place are left out, relegated to what epidemiologists call "residual confounding"; measurable outcomes and statistical attributes are given precedence [Macintyre et al. 2002]. Kearns and Moon [2002, 611] argue that the way place is measured in quantitative studies is more revealing about data collection practices rather than a potential social significance of place. Moon refers to this conceptualisation of space as "formal space" [Moon 1990, 166], used to organise statistical data for the purpose of efficient resource allocation without direct involvement of human experience. Formal space is zonal; based on proximity, it groups people together to fictional, arbitrary communities. This understanding is far removed from 'sense of place' in everyday life and generally not suited to deal with direct experiential aspects of health and well-being.

To overcome this limitation, Galster [2012] calls for the integration of quantitative and qualitative methods in one coherent research design. Qualitative research permits scholars to work with experiential aspects of place, but, from a statistical point of view, inevitably suffers from limited sample size and lack of confidence with regards to generalisations. Social media, such as Twitter data or Mappiness may offer an important opportunity for the enhancement of neighbourhood effects research with qualitative user data at a large scale, if the relationship between sample and the general population can be clarified.

#### Implications for geodemographics

The hazard of place and the neighbourhood effects literature usefully flag opportunities and challenges in extending geodemographics to the study of vulnerability. First, hazard of place literature suggests that combined information pertaining to a variety of small area characteristics can be valuable in creating a focussed description of an overarching aspect of interest. Yet, unlike neighbourhood effects research, hazard of place studies focus on exposure and vulnerability without considering real world outcomes.

Second, although the literatures highlight the importance of higher-level social and political root causes placing people at risk, this aspect has only been considered in some neighbourhood effects studies. The incorporation of wider systemic causes calls for deliberate comparative study designs.

Third, there is a tendency in both literatures to tacitly assume a sense of community in the terms place or neighbourhood. In hazard research, places, neighbourhoods and communities are in fact often used interchangeably. But whether or not a small area represents a coherent community, will itself determine aspects of social vulnerability and should be captured in any assessment of it.

Fourth, the cultural dimensions of risk have not been considered, except qualitatively in the study by Collins et al [2009]. But local specificities in culture and even biogenetic characteristics of populations may be important additional aspects of vulnerability, including varying perceptions of risk [see Hogan & Marandola 2007]. Subjective components, such as lifestyles and behaviours, have so far been neglected by those literatures and ways to incorporate information on those aspects may advance the inquiry.

## 3.3 Systemic processes and local specificity

Despite shortcomings, neighbourhood effects research has been useful in generating hypotheses about pathways linking exposure to outcomes. Many studies identify relevant domains of neighbourhoods as being connected to social determinants, such as neighbourhood deprivation and disorganisation [e.g. Congdon 2012; Jackson et al. 2008; Propper et al. 2005], housing [e.g. Graham et al. 2009; Hiscock et al. 2003; Lawder et al. 2013], land use [e.g. Ewing et al. 2013; Factor et al. 2013; Turrell et al. 2013] or accessibility [e.g. Olabarria et al. 2014]. The results are mixed, evidence for both causation and selection can be found, and it emerges clearly that associations between area exposure and outcome differ for distinct groups of people.

Small and Feldman [2012], for instance, summarise findings from a comparative study that investigates changes in well-being of residents who moved from deprived to wealthier neighbourhoods in some US cities. They report wide-ranging heterogeneity: girls improve on health, behavioural and educational measures, while boys fare worse on all counts; effects are found in Baltimore and Chicago, but not in Boston, Los Angeles or New York. They conclude that "neighbourhood effects do not operate homogeneously across subpopulations and across 'treatment settings'" [ibid., 63]. In other words, whether neighbourhoods matter is conditional on individuals, household types, neighbourhoods and cities – a classic problem of complexity that draws attention to interactions and contingency and suggests that heterogeneity of pathways is the norm and not anomaly.

The heterogeneity ascertained indicates that research findings are spatially and temporally specific. Local specificity has been of explicit interest in studies that seek to create regional geographies of populations based on their unique attributes. Cheshire et al [2013] determine the regional distribution of surnames and delimit subnational territories of cultural similarity within Japan. They argue that while social similarity based on ubiquitous categories such as income or social class helps characterise populations, they are incomplete without the locally specific patterning that is given by cultural and geographical micro context [ibid., 4]. Surnames have been used in other national studies to investigate regions of population similarity based on unique descriptors [e.g. Boattini et al. 2012; Longley et al. 2011a; Novotný & Cheshire 2012]. These studies offer a promising method to consider local specificity of populations.

But besides discrete indicators of specificity, comparative research designs are necessary to detect locally specific guises of pathways. Robinson [2011] distinguishes different comparative strategies in urban studies: individualising, universalising, encompassing and variation-finding [ibid., 5]. Some studies assume that the very idea of specificity renders meaningful comparison impossible, but this view is typically not held in quantitative geography. As already critiqued by neighbourhood effects researchers, many studies focus on 'average effects' within one study population, refrain from relating their findings to an international, comparative discourse and hence avoid the question of whether their findings can be generalised. An easy extension of this case study strategy is to critically review findings in the context of findings by others and discuss commonalities in methods and results as well as differences. This is what Robinson calls 'individualising', in which the study population (or case study) is brought into conversation with other cases and commonalities and differences are asserted qualitatively [cf ibid., 6]. The focus of this strategy would be, however, to highlight and explain local, specific trajectories rather than universal laws, although the processes and patterns revealed in these studies can be of wider relevance to other cases. All other comparative strategies rely on some discrete information that encompasses local specificity, such as distinct groups, cities or regions. The strategies differ in the way this information is formalised.

Universalising comparisons are employed to find universal, causal laws through formal comparison of two or more cases. They can be understood as an extension of case studies in numbers. Quantitative geographical models are typically formalised as general linear models, although interpretations of these models tend to abstain from positing universal laws. The comparative element is introduced as binary variables assigning categorical membership to a place. Hill et al [2005], for instance, compare the association between neighbourhood disorder, psychological distress and self-rated health in Chicago, Boston and San Antonio. City residence is introduced as a binary control variable for each, and the impact of residence in Boston and San Antonio on the neighbourhood-health association is compared relative to Chicago residence. There are many variants of the general linear model – ordinary least square regression for continuous variables, logistic regression for binary variables, multinomial regressions for nominal variables. What is common to all of them is the assumption that some quasi-linear mathematical form characterises all relationships between outcomes to be explained and control variables including regional membership. Byrne [2002, 113 sqq] argues that assumption of linearity as well as the conceptualisation of variables as independent, separate 'traits' of study units render these models inappropriate to study a complex, non-linear world and discover real, global processes that result in qualitative changes of systems as a whole. But some attempts to consider local context can be made through the analysis of residuals or outliers that do not correspond to the observed regularity. Fotheringham [1997] proposes a geographical extension of the general linear model that systematically incorporates sensitivity to local context to improve model fit and facilitate an investigation of scale for geographically correlated phenomena.

The encompassing strategy starts from the opposite position: it presupposes some global process that connects all comparative cases. Each case (or city or population) is understood as an instance of that global process or global system, of which the system can be understood as a whole and individual cases (or cities or populations) as parts [cf Robinson 2011, 8]. Yet, Robinson points out that the parts may not form anything that constitutes a whole system, such as is supposed by world systems hypothesis. Cases might also be "encompassed" by a variety of processes; indeed, a process might constitute quite different entities as cases: a city, a region, a neighbourhood ensemble, spatial practices, to name a few. Hence, the strategy should be used in an open-ended fashion to empirically infer a whole from a selection of cases, or reject a consistent whole, if none connecting the cases can be found.

In quantitative geography, this strategy is best exemplified in multilevel models. Specificity is treated as a nominal, categorical variable of group membership, for example an individual assigned to an area, city or region. Separate associations are estimated for each area (or city or region) and compared formally within a single model to determine to which extent outcomes of interest vary across individuals within each area as well as across areas. The areas can further be nested in higher levels, such as area types, cities, groups of cities and countries – in theory, possible extensions are infinite. Multilevel models can then be used to detect patterns that are specific to each level within the nested structure of the model. The appeal of these models lies in the fact that they can reflect real, hierarchical structures in the world and are sensitive to local specificities while highlighting more global processes. For that reason, Byrne [2002, 122] regards multi-level models as useful exploratory tools to derive clues about locally specific pathways, although the approach, too, attempts to impose linearity on a non-linear world. Additionally, each level is in some way a class or category taken as *a priori*, although the taxonomies of these categories may change. Encompassing comparisons as practiced so far seem ill-suited to account for changes in taxonomies [cf Robinson 2011, 8; Byrne 1998, 81].

Finally, there is what Robinson calls variation-finding used to discover similar or different causal pathways of a problem of interest [cf Robinson 2011, 10]. Variation-finding or the "comparable cases strategy" [Perry & Roberson 2002, 35] can be divided into two designs: a "most similar systems" and a "most different systems design". By means of the former, knowledge is derived by comparing cases with as many structural similarities as possible and different observed outcomes. For example, a set of cities may have similar levels of pollution but the incidence rates of respiratory diseases are very different. The researcher is then interested in the factor that accounts for the differences. The most different systems design begins with the recognition of a common outcome observed across a set of cases that are very diverse in their structural characteristics. The researcher intends here to find the unobserved common variable accounting for the pattern.

In quantitative research, this approach involves the creation of separate, matched models for different cases. Stephens et al [1997], for instance, estimate relative risk of mortality adjusted by socio-economic variables in two urban settings: Accra, Ghana, and São Paulo, Brazil. They estimate separate models for each city, the category of local specificity, and are able to discover common yet locally specific pathways. The comparative work consists of an informal and continuous engagement with comparison from variable selection to qualitative appraisal of results. This approach is more flexible than encompassing approaches, as it does not require use of the same variables across different settings. But, as Robinson [2011, 10] points out, unquestioned assumptions of what makes cases comparable tends to reproduce existing knowledge.

All in all, many researchers engaged in comparative urban research argue that comparative studies should be so exercised as to account for diversity of pathways, reveal plural causality as opposed to universal causality and characterise specific guises of causal interactions [Robinson 2011; Small & Feldman 2012; Wacquant 2008]. These recommendations have important implications for the quantitative research programme and resonate with similar calls made by social scientists advocating systems thinking.

## 3.4 Synthesis: focussing geodemographics on vulnerability

The antecedent review shows that geodemographics, vulnerability, and social determinants of health share theoretical and conceptual ground. Although none of these approaches refer explicitly to Bourdieu's theory of social practice, his theory is consistent with all of them. Embodiment and predisposition are central concepts in these literatures, without which social pathways in health as well as durable, multi-level manifestations of vulnerability would not be conceivable. In addition, the acknowledgement of complex interactions and system dynamics that is at least implicit in all approaches suggests that concepts and tools can be shared across these strands of research.

In substantive terms, it can be concluded that vulnerability is directly connected to health. Health is always an expression of some form of vulnerability; it is in this respect a sufficient indicator. The reverse, however, does not apply, in that vulnerability always results from ill-health; vulnerability is thus a necessary indicator of health but not a sufficient one at the time of measurement. Hence vulnerability is a potential on which ill-health is contingent in interaction with other conditions. Health needs can therefore not be narrowly defined in terms of diagnosis but need to be viewed in terms of vulnerability. The essential target of preventive health care and policy programmes must therefore be to reduce vulnerability (or increase resilience) rather than increase health.

Drawing on Bourdieu and the work of political ecology, vulnerability is a relational concept. The precise significance of vulnerability in shaping health can only be determined in the specific social (and geographical) context of an individual or population group. Access to neighbourhood social capital, for example, may not be relevant for everybody and only for some, depending on how an individual or group relates to neighbourhood, social networks or society in general. Vulnerability has to be measured in context; an absolute, substantialist reading of vulnerability is not likely to be productive in research or policy.

Another property of the concept of vulnerability is that it can be meaningfully applied to different ecological levels. Health, strictly speaking, is an individual-level outcome. Although it can be transferred metaphorically to other units ("community health", "healthy city" etc), it really refers to a community or city with a high number of healthy individuals. It denotes an aggregation. In contrast, vulnerability acquires different meanings, 'realities', at different ecological levels. Even in its broader definition as potential for loss, the concept describes different phenomena at different levels. Vulnerable individuals may be liable to lose out on health, vulnerable communities may lose common resources, shared social capital or degree of cohesion. Vulnerable communities are conceptually distinct from vulnerable individuals; and yet, the vulnerability of a community can be an important component of an individual's vulnerability and vice versa; it constitutes more than just aggregation.

This ecological notion of vulnerability invites spatial heuristics that are designed to make generalisations at the group level while accounting for local specificities. Addressing existing critiques of geodemographics [see chapter 2.3] may render the tool useful not only to study vulnerability in society but also to support preventive health programmes at a strategic level. A systematic assessment and characterisation of vulnerability at different levels can thus inform strategic planning for preventive health care and support the development of interventions. In health care policy, this seems to be intended in recently introduced policy instruments, Joint Strategic Needs Assessments (JSNA) and Joint Health and Well-being Strategies (JHWB).

A geodemographic framework of vulnerability provides a multi-dimensional, multi-level characterisation of vulnerability with suggestions why some units appear to be vulnerable in certain respects, which events may further increase or decrease their vulnerabilities and what interventions may be expedient to address their causes. When it is viewed within an encompassing, comparative research design, generic and local processes in shaping vulnerability and their expressions can be distinguished.

# 4 Encompassing specificity: geography and dynamics of population structure

To date, geodemographic studies have not accounted for local specificity, although all populations have unique cultural and biological characteristics. Social pathways to health may be structured by these unique characteristics, which operate outside the sphere of social distance, are often compounded into aspects of place and produce singular local phenomena [Moon 1995]. In order to formally incorporate heterogeneity and specificity into geodemographics, it needs to be determined, how specificity can be measured, which geographical scale may be appropriate to represent it, and how temporally stable specificities are likely to be, in particular in dynamic, urban environments.

## 4.1 Determining the locally specific

### Specificity and vulnerability

Some attempts to account for population specificity in geodemographics have been made. Debenham [2003a,b] and Petersen et al [2010] produce regionally adjusted classifications out of an existing framework of variables. The latter authors present the so-called LOAC (London Output Area Classification) and demonstrate that the classification reflect intra-London neighbourhood differences better than the UK-wide Output Area Classification (OAC).

Useful though social similarity-based classifications may be, from a vulnerability perspective they fail to capture important aspects, notably regionally specific, micro-cultural and biological characteristics of populations. Micro-cultural aspects can include social networks, historical influences of local economy and society, modes of decision-making and other socio-cultural tendencies including transnational bonds linking multiple places. Biological characteristics include specific biogenetic profiles of populations that may make them subject to health disadvantage. This latter aspect of vulnerability currently receives particular attention as research councils like ESRC and MRC as well as Wellcome Trust and other research organisations make substantial investments in biogenetic data resources – with the UK Biobank [www.ukbiobank.ac.uk] or inclusion of biomarkers in the Understanding society longitudinal survey [Benzeval et al. 2014] being perhaps the most prominent examples. Biosocial relations are relevant in the study of vulnerability, since the social and physical environment have been found to modify phenotypical expression of genes, which includes pathophysiological responses [Guthman & Mansfield 2012].

How can local specificity be captured or measured? First, specificity shall be understood as some relevant aspect that is unique to populations. Conventional quantitative models, which focus on statistical summary, treat the unique typically as residual variation.

But if specificity is of explicit interest, data describing and methods highlighting unique aspects of populations or places are needed.

Second, the spatial dimension of specificity requires a consideration of scale. Each neighbourhood, for example, is unique and therefore produces unique phenomena. Micro-cultural and biogenetic distinctiveness may already be found between neighbourhoods but are also affected by wider spatial interaction: in everyday life, cultural mix bears on experience at a wider spatial extent than neighbourhoods and biogenetic characteristics are derived from wider, regional gene pools.

Third, temporal stability is an important aspect of local specificity and depends on spatial scale. At a regional level, socio-cultural and biogenetic factors tend to remain durably distinct over time, but neighbourhood turnover may change local conditions quickly, especially in dynamic urban environments. This has implications for the ecological study of vulnerability in a corresponding geodemographic framework.

### Surnames and population structure

Surnames offer potential to capture local specificity. Local surname compositions have been shown to reflect regionally specific cultural and biogenetic traits [Darlu et al. 2012; Degioanni et al. 2003; Jobling 2001; King & Jobling 2009] and may offer potential to delineate geographies of local specificity at a variety of scales [Cheshire & Longley 2013; Longley et al. 2011a]. Surnames differ from measures of social similarity because they describe people in literally nominal categories. Since they have been geographically concentrated over centuries [Cheshire et al. 2009], they inevitably bear some cultural and biogenetic contents beyond the name itself. In particular, a direct link between genetic and surname inheritance translates into corresponding geographies of gene frequency and local surname mixes; this could be verified in a variety of national contexts [e.g. Boattini et al. 2012; Dipierri et al. 2011; Herrera Paz et al. 2014; Rodríguez-Larralde et al. 1998, 2000].

The correspondence between genes, culture and surnames can be explained by the fact that each is – at least in parts – hereditary. But although they are hereditary, genetic profiles and cultures do not remain static. Geneticists discern four evolutionary forces that produce and reproduce varying genetic profiles across populations: natural selection, mutation, genetic drift and migration [Cavalli-Sforza 2001; Cavalli-Sforza et al. 1994].

Natural selection occurs when certain genetic variants are more likely to spread or survive than other genetic variants. Today, medical innovations and widespread health care has rendered natural selection an insignificant force in human populations because they ensure that life expectancy is greater than the typical age of procreation [Cavalli-Sforza 2001, 205]. Nevertheless, other forms of selection may occur: in populations, certain genetic or social traits may be preferred in the choice of partners, resulting in what is known as "assortative mating". This results in socially induced, unequal survival chances among the genes associated with the traits.

The second type of change – mutation – occurs when genes mutate during the transmission of DNA from one generation to another. This is a random process mainly owed to copying 'errors' and can have a large-scale impact on populations if the mutated gene has gained survival advantage or is part of a small founding population that has migrated. Nowadays, mutation is likely to be insignificant in driving genetic differences between populations.

The third type – genetic drift – is genetic change encompassing an entire population. Genetic drift occurs when a population lives relatively isolated and is not much affected by migration or natural selection. Certain genes gain in frequency and slowly crowd out others over time. This type of change is based on random gene selection out of the local gene pool; thus, it is path-dependent and contingent on the profile of the founding population. Genetic drift is among the major forces shaping today's regional genetic differences.

Migration is the fourth driver of genetic change. In the event of migration, genes spread at a new location, either in the newly arrived group or through interaction with a local, earlier settled population. This type of change is only significant if the scale of migration is transformative, for example in case of large scale population movements, as witnessed during colonialisations. Sometimes, the members of the newly arrived are ill-adapted to the new local conditions and thus vulnerable to certain diseases to which the earlier settled population has become resistant. Biogenetic vulnerability or sociocultural factors innate to migrant populations can produce significant population patterning of disease, and the strong research interest in associations between ethnicity and health nowadays results from this potential pathway.

The way genes spread corresponds to surname diffusion to some extent. In the UK, surnames were created under regionally varying naming conventions, for example by profession, by lineage or by place names [Cheshire et al. 2009]. Naming conventions constitute one of the causes why surnames tend to concentrate geographically within a culture. Yet, while genes do not change during the life course, an individual may adopt a different surname notably after marriage. In the majority of Western countries, women adopt the name of their husbands [Goldin & Shim 2004; Valetas 2001], and thereby delete the link between surname and gene at the level of the individual. But if the majority of people remain in their birth region and marry locally, surnames remain indicators of gene pools at the level of regions. Cheshire et al [2009] suggest that this is the case in Great Britain.

The mix of surnames may drift, too, either when the founding population has a low surname diversity or families intermarry more often than others. This kind of drift would coincide with genetic drift – here surnames and genes spread in parallel and rarer variants may be crowded out on varying time scales depending on their initial frequencies. In consequence, the specific mix of surnames corresponds to the resulting regional gene pool and surnames remain informative of genetic population characteristics, if the population remains relatively isolated.

Finally, migration affects local surname compositions, too. Since migrating individuals carry their surname and genes, the resulting genetic profile of populations is reflected in surnames. In this case, surnames are not only informative of local population mixes but can also be used to trace short and long-term migration movements in and out of local areas [Jobling 2001; Longley et al. 2007].

### Surnames and culture

The evolutionary forces in human populations can also be related to culture. Cavalli-Sforza [2001] observes modes of cultural transmission that resemble genetic evolution.

> "Culture resembles the genome in the sense that each one accumulates useful information from generation to generation. The genome increases adaptation to the world by the automatic choice of fitter genetic types under natural selection, while cultural information accumulates in a person's nerve cells, being received from another person and selectively retained. Cultural transmission occurs in a variety of ways: by the traditional path (observation, teaching, conversation), through books, computers, or other media developed by modern technology." [ibid., 176]

Cavalli-Sforza distinguishes between cultural transmissions from generation to generation – vertical transmission – and those between individuals of the same generation – horizontal transmission [ibid., 180]. The latter can occur rapidly, sometimes with transformative consequences. Bourdieu's theory of social practice offers a more sociological explanation to Cavalli-Sforza's reasoning. The through habitus embodied cultural dispositions stay with the individual with little change throughout the life course. Bourdieu identifies the early life environment as crucial determinant of habitus; rearing practices by families primarily shape our modes and abilities to appropriate culture throughout life, while the influence of education is secondary and confined to the transmission of cultural knowledge [Bourdieu 1990a, 60; Johnson 2012, 32]. It is more difficult to relate the four forms of genetic change to Bourdieu's micro-cultural context, but similar processes can be imagined.

Dawkins [2006(1989), 69 sqq], for example, simulates the spread of what he calls 'egoistic' and 'altruistic' genes in a population and arrives at the conclusion that due to natural selection mechanism, there would be a natural equilibrium of frequencies of those two types of genes in a world without any intervening higher level structures (social hierarchies, governance etc.). An equilibrium is a situation in which nobody in a population can improve one's position through alternative "strategies"; the strategies maintaining those equilibria are evolutionary stable. But applied to culture, this is a very unrealistic situation. Bourdieu's theory suggests that socio-cultural dispositions are actually never in equilibrium, even in a society without institutions [Bourdieu 1990a, 118]. Here, natural selection may occur, if certain practices are more successful in improving a group's position in society – forms of symbolic capital may be a result of greater success and power. There are possible genetic analogies to Bourdieu's theory which could be discussed, for example the transposability of habitus; but this may be a transdisciplinary thought experiment for the future.

Cultural drift as an analogy to genetic drift seems plausible, if habitus is viewed within a complex social world. Specific interactions may cause tensions or oppositions that lead to durable transformations of habitus, including conscious attempts to change social structures or reflexive adaptations of usual practices [ibid., 118]. One may add that habitus is constantly drifting, never in equilibrium and by definition path dependent – just like genes. Yet, these drifts are not fast-paced; they remain durable and stable with changes occurring over generations. In terms of space, the geneticist's notion of "isolation by distance" suggests that both genetic and micro-cultural trajectories have their own distinct and corresponding geography, if potential transformative effects of cultural diffusion through electronic media are ignored for the moment.

Finally, migration is an influential process that simultaneously changes the location of genes, cultures and surnames. With Bourdieu in mind, we may therefore agree with Cavalli-Sforza, when he writes:

> "Even at a microgeographical level, the regions subject to detailed study have usually shown strong correlations between geography, genetics, linguistics, and other cultural aspects like surnames. Often the genetic-linguistic mosaic we observe clearly shows the effects of numerous expansions – some are known historically – and of their superimpositions and interactions. Perturbations do occur, but they do not manage in most cases to obscure the clarity of the correlation between genes, peoples, and languages." [Cavalli-Sforza 2001, 168].

At an individual level, habitus in form of socio-cultural disposition moves with the body and is transported – with possible modifications during the life course – through generations, manifesting in durable intergenerational signatures of cultural heritage and geographic origins [cf. Bourdieu 1977, 87 sqq].

### Conclusion: Surnames as descriptors of the specific

Theoretical contemplations in both sociology and genetics suggest that surnames plausibly carry biogenetic and cultural signals. These signals may bear information about biological vulnerability or cultural factors that modify pathways in health. From an intervention perspective, population registers can therefore be valuable resources for strategic planning; at the level of individual interventions, however, surnames are only informative under certain conditions [Jobling 2001, 355]. In geographical research, local surname compositions rather than individual names are useful, since local surname compositions likely correspond to regional gene pools and some socio-cultural characteristics. If so, surnames as delineators of socio-cultural and biogenetic geographies are relevant not just for epidemiological studies but also more widely for comparative regional and urban research.

## 4.2 Data and methods to identify population structure

In order to devise heuristics of incorporating regional specificity and assess both their relevant geographic scales and their spatio-temporal stability, different kinds of data are drawn together and analysed. A central dataset in this undertaking is obtained from a study that investigates genetic population structure in Great Britain: a sample of genetic data, which can be compared to surname registers at two points in time.

### Measuring population structure: study context and data preparation

As part of the Wellcome Trust funded study "The People of the British Isles", the DNA of 2,019 participants has been collected. The objective of the study is to characterise the fine genetic structure of the British population and investigate the regional distribution of genetic variants. Fine genetic population structure denotes the genetic composition of population sub-groups, for example populations at sub-national geographical scales [Hellenthal et al. 2015]. The investigation of fine population structure is a research agenda in genetics that has emerged after the decodification of the human genome and the evolution of computational capacity and powerful statistical algorithms. Fine structure promises to offer clues to historical inter-cultural interactions and to improve genome-wide association studies – studies that investigate associations between genotype and disease onset of an individual [Winney et al. 2012].

The study participants were recruited from rural regions in Britain. Rural populations are thought to be more homogeneous and less affected by migration than urban areas. The assumption of the project is, therefore, that a sample of rural residents may best reflect genetic profiles that is unique to the British population. In order to ensure long-term rural and local ancestry of participants, the study imposed strict inclusion criteria: all four grandparents of the participant were born in rural Britain, no more than 80 kilometres away from each other. These restrictions increase the likelihood of a volunteer representing a region's gene pool.

The sample of rural participants offers the opportunity to investigate the biogenetic contents of surnames by relating the sample to the wider population. Population registers are nearly complete micro datasets with some information on each inhabitant, including fore and surname, address or area of residence and sometimes demographic

Table 4.1: Sizes and geographies of population registers

| | 1881 | | | 2007 | | |
|---|---|---|---|---|---|---|
| | population* | surnames | parishes† | population* | surnames | wards |
| total | 29,912,298 | 518,153 | 7,203 | 45,667,321 | 825,999 | 9,434 |
| Anglo-Saxon | 27,213,993 | 61,286 | 7,203 | 40,677,037 | 524,352 | 9,434 |
| rural Anglo-Saxon | 18,692,871 | 57,511 | 6,848 | 21,770,043 | 123,876 | 6,354 |

* population registers are not complete · † parish groups [see technical note at the end of this section]

characteristics. Two points of time are relevant in this task: the approximate birth year of volunteer's grandparents (median 1885) and the year the Wellcome Trust sample was taken (2008). The closest population registers for the former date are the 1881 Censuses for England and Wales and the 1881 Census for Scotland. They contain information for each individual at the time, including name, age, sex, place of birth and residential parish. For the second point of time, the 2007 Enhanced Electoral Roll comes closest to the time of data collection. This dataset holds information on each resident registered on the UK Electoral Roll, "enhanced" with data from telephone directories and other sources, forming a nearly complete register with names and postcode of each resident.

In order to create best matching population registers, residents living in urban parishes or postcodes are excluded. Urban records are identified by overlaying their parish or postcode with geospatial boundaries of urban land. Non-Anglo-Saxon records are also excluded, as the study sample include native Britons only. In order to classify records as Non-Anglo-Saxon, a tool is used that takes fore and surnames of records to identify an individual's ethnicity (the ONOMAP tool, see Mateos 2007 and www.onomap.org). By considering the forename of an individual, the tool also succeeds in excluding Black British names. For example, while the name *James Taylor* would be classified as "English", *Jerome Taylor* would be "Jamaican". This procedure will necessarily produce errors, when the choice of forenames does not follow conventional naming practices, but it may nevertheless improve the categorisation of ethnicity by ancestry. The application of exclusion criteria results in an analytical population of 18.7 million rural British with 58,000 surnames in 1881 and 21.8 million rural British with 124,000 surnames in 2007 [see Table 4.1].

### Synopsis of research strategy

The three datasets – the sample of participants, the 1881 Census micro dataset and the 2007 Enhanced Electoral Roll – need to be transformed into identical formats in order to formally relate them to each other. The genetic sample has been made available by Hellenthal et al [2015] as a so-called "coancestry matrix" [Lawson et al. 2012], which measures the probability of each pair of participants sharing common ancestors and thus genetic similarity.

As for the population registers, surnames are counted and aggregated to an analytical spatial unit: parish groups in 1881 [see technical note at the end of this section] and wards in 2007. For each parish group or ward, the local surname composition is calculated by counting the relative frequency (percentage) of surnames occurring in each area. Subsequently, the similarity of surname compositions of each pair of areas is measured. The result of this procedure is a similarity matrix – a matrix that measures the 'isonymy' (relatedness) between each pair of areas [Lasker 1977]. The isonymy matrix is the analogous output to the coancestry matrix; both matrices can therefore be processed in the same way.

In order to process the information on population structure held in these matrices, the statistical technique of cluster analysis (hierarchical with Ward linkage) is used. For this purpose, the matrices are inverted such that they represent dissimilarities or distances between observations (sample participants or areas). The technique divides all cases (participants or areas, depending on the dataset) into relatively homogeneous groups, within which member cases are more similar to each other than cases that belong to other groups. This form of cluster analysis produces a taxonomy of cases which can be adjusted depending on the desired number of groups. Applied to the 1881 and 2007 population registers, cluster analysis produced between two and 80 groups of areas with distinct surname compositions, in the following called *isonymy groups*.

Using the same method, participants can be classified based on their genetic profiles. Yet, the resulting taxonomy differ from the one produced from specialised procedures that geneticists use [Hellenthal et al. 2015], and hence an assignment matrix provided by Hellenthal et al is used in further processing. The matrix assigns each participant to one of 53 groups, which each are hierarchically nested in higher level groupings of 52 to two groups.

The genotyped participants and classified areas are then spatially related to each other in the following way. The birthplace of each of the four participants' grandparents is assigned to the local area appearing in the population register. The grandparents' birthplaces are then assigned to the isonymy group of their local area (parishes or wards), and the most frequent area isonymy group among the four grandparents is assigned to the participant. In the rare event that two grandparents were assigned to two or four different isonymy groups, the isonymy group is randomly assigned from the possible ones.

The thus achieved genotyping and 'isonymy-typing' of participants results in two partitions of the sample: participants are classified by local surname compositions on the one hand and by their individual genetic profiles on the other hand. If the correspondence between these two partitions is high, we observe evidence in support of the hypothesis that surnames reflect local specificity (in this case biogenetic). The correspondence between these partitions can be measured by similarity indices [see e.g. Albatineh et al. 2006; Vinh & Epps 2009]. The indices measure the overall agreement of two partitions (typically on a scale from zero to one) and, at the same time, by comparing different cluster solutions, identify the most agreeing partitions.

Measuring the overall correspondence of different solutions of surname mixes and genotypes yields a global estimate of the overall biogenetic contents of surnames. The methods discussed so far do not permit a local view of agreement and disagreement between the two partitions, in other words, they do not convey where exactly in Great Britain coancestry and isonymy correspond. A local view can be achieved by processing the two matrices – isonymy and coancestry – in multi-dimensional statistical algorithms that – instead of creating a taxonomy – quantify the overall difference between entities (e.g. participants or areas).

The multi-dimensional statistical algorithms used are Principal Components Analysis (PCA) and Multi-dimensional Scaling (MDS), which translate a matrix of similarities or dissimilarities (distances) into scores on a given number of dimensions. The scores can be thought of as coordinates in a multi-dimensional statistical space with each pair of data point having a distance (or equivalent) given by the matrix. The result of this procedure are scores of coancestry for each case in the coancestry sample and scores of isonymy for each spatial unit indicating a 'coordinate' of surname composition in statistical space. Methods of spatial interpolation can now be used to map the scores and highlight spatial patterns of value similarity. This step permits an informal, visual exploration of regional trends in coancestry and isonymy.

In a next step, based on the joint evidence of the population register and the genetic sample, a regional geography of Great Britain based on population structure is estimated. This is done by measuring the number of genotypes that can be found in each isonymy group. In order to do so, two diversity indices are used, an index of Distinctiveness that measures the degree to which an isonymy group hosts a distinctive (sample) population in terms of genotypes, and Simpson's index of Dominance [Simpson 1949], which measures internal genetic diversity of an isonymy group, by determining the probability that two individuals drawn from the same isonymy group have the same genotype. An aggregate index is formed out of Distinctiveness and Dominance, which I call Regional Integrity of an isonymy group. The resulting regionalisation represents the scale of regional specificity with respect to population culture.

In order to assess population specificity and their dynamics in urban areas, the study is repeated in an encompassing comparative framework of 15 UK conurbations. In the absence of genetic data, the study is carried out with the two population registers (1881 and 2007), the full register with the entire population and a subset of records with Anglo-Saxon names only. The consideration of these two datasets at two points in time permits the investigation of long-term urban dynamics and some conclusions with respect to biogenetic and socio-cultural specificity of different urban environments.

**Technical specification: clustering applied to coancestry and isonymy matrices**

The isonymy matrix and the coancestry matrix are similar: both measure similarity between entities based on what these entities are composed of (surnames and genes respectively). The coancestry matrix measures similarity, more specifically the likelihood of any two individuals having a common ancestor. The method used to create a coancestry matrix out of DNA sequenced data is known as "chromosome painting" [Lawson et al. 2012]. The chromosomes of each individual are treated as if they were composed of the chromosomes of all other individuals in the sample. The coancestry matrix measures for each pair of individual the percentage of DNA that is most likely common between the two. This technique is novel and deemed particularly suitable to explore fine population structure [ibid.]. The coancestry matrix can be inverted to represent genetic distance rather than genetic similarity between individuals.

The afore-mentioned isonymy matrix measures the similarity of areas based on their surname composition. Lasker [1977] proposed a measure of isonymy which can be applied to the level of populations.

> "To the extent that persons having the same surname can be assumed to be descended from the same progenitor through the male line (in patrilineal societies), and assuming that relationships through female lines and mixed lines are proportional to those through the male line, then the frequency of shared surnames between two communities samples the genetic lineages and measures the degree of biological kinship between the communities." [ibid., 489]

The degree of relatedness of two populations can be calculated based on the relative frequency of common surnames in the population.

$$\eta_{ij} = \sum_s \frac{n_{s.i} \cdot n_{s.j}}{2 n_i n_j} \qquad [4.1]$$

where $n_{s.i}$ and $n_{s.j}$ are the frequencies of surname $s$ in populations $i$ and $j$ respectively. Transferred to areas, the repeated calculation for each area would result in a matrix of isonymy $M_\eta$ between each pair of areas.

Hierarchical clustering algorithm (HCA) is frequently used in genetic studies that aim to estimate the likelihood of individuals having a common ancestor, based on their genetic distance. The same clustering can be applied to isonymy matrices. HCAs require that the matrix measure distance rather than similarity between entities, so Lasker proposes a modification:

$$M_\eta^d = \log\, M_\eta = \log \begin{pmatrix} 1 & \eta_{12} & \cdots & \eta_{1a} \\ \eta_{21} & 1 & \cdots & \eta_{2a} \\ \vdots & \vdots & \ddots & \vdots \\ \eta_{a1} & \eta_{a2} & \cdots & 1 \end{pmatrix} \qquad [4.2]$$

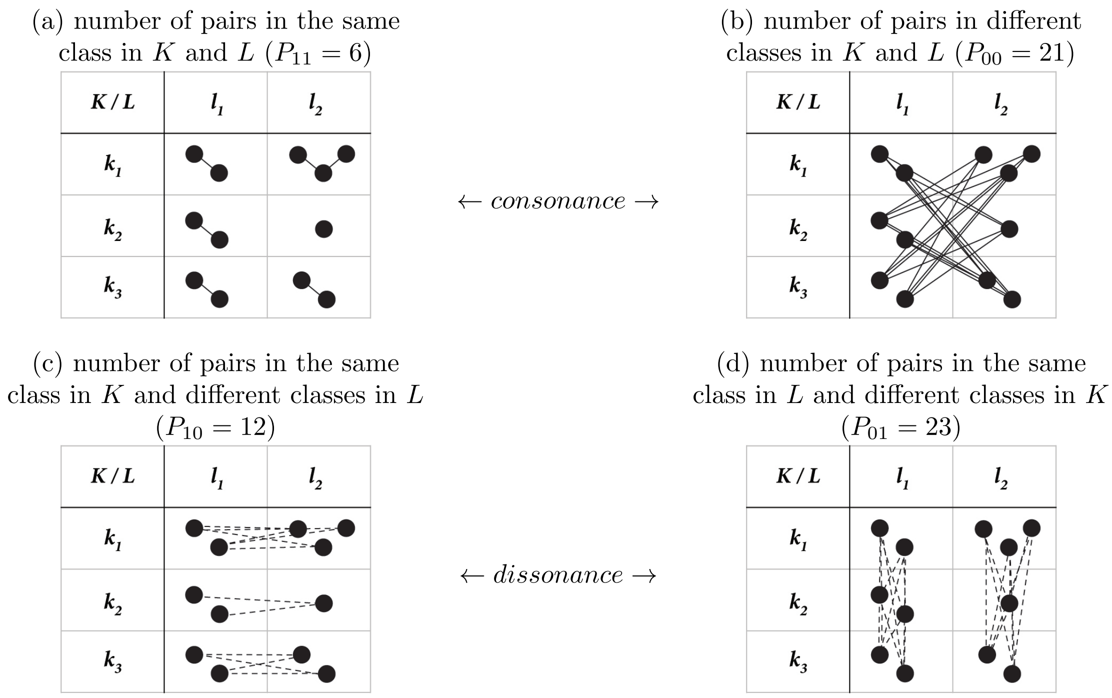



Figure 4.1: Example of two partitions – $K$ and $L$ – of $n = 12$ cases and the resulting counts of consonant and dissonant pairs.

As for the clustering itself, there are two types of procedures to group entities into classes of similars: divisive procedures and agglomerative procedures [Everitt 1974]. Divisive procedures begin with the entire sample and iteratively divide it based on distances between cases. Agglomerative procedures operate the opposite way: initially, each case is considered to be a single cluster, which are iteratively merged into larger clusters. The generic formula for any agglomerative clustering procedure is:

$$d_{(ij)k} = \alpha_i d_{ik} + \alpha_j d_{jk} + \beta d_{ij} + \gamma |d_{ik} - d_{jk}| \qquad [4.3]$$

where $d_{(ij)k}$ is the distance between the merged cluster of $i$ and $j$ and the cluster $k$, $d_{ij}$, $d_{ik}$ and $d_{jk}$ the distance between the cluster pairs $i$-$j$, $i$-$k$ and $j$-$k$ respectively. The entities $\alpha$, $\beta$ and $\gamma$ are parameters, often depending on cluster size, that further define the procedure. The Ward linkage procedure – used here – is agglomerative and defines these parameters so as to inhibit mergers of larger clusters and has proven particularly useful.

### Technical specification: comparing partitions to measure global agreement

In HCAs, any number between two and the number of cases can be chosen as the number of groups into which the sample should be partitioned. Similarity indices or indices of agreement are formal methods to compare different types of partitions. [e.g. Albatineh et al. 2006; Hubert & Arabie 1985; Vinh & Epps 2009]. The key element of comparing two partitions is a contingency table, which counts the number of cases classified alike or differently by two different partitions [see Figure 4.1].

Indices of cluster agreement can be divided into pair counting and information theoretic indices [Vinh & Epps 2009]. Pair counting indices compare the quantities of consonance against a possible maximum value. The quantity $P_{11}$ counts all pairs of cases that are assigned to the same class in both partitions; the quantity $P_{00}$ counts all pairs that are assigned to separate classes. Together $P_{11}$ and $P_{00}$ measure the consonance between the two partitions. Partition dissonance is represented by the remaining two quantities: $P_{10}$ counts the pairs which have cases in the same class in the first partition, but in a different on in the second partition; $P_{01}$ counts the reverse. These four quantities form the central elements of pair-counting indices of cluster agreement.

For example, the *Rand* index calculates the ratio of consonant pairs and all possible pairs between two partitions [Table 4.2]. The maximum value of one would be achieved if the partitions are identical, a value of zero – the minimum value – if one partition groups all cases into one class, while the other groups each case into a single class. These two extreme scenarios are not realistic in most applications, however, and thus adjustment of the index for chance has been recommended [Albatineh et al. 2006; Hubert & Arabie 1985]. The generic formula for an adjusted pair counting index would be:

$$Adjusted\_Index = \frac{Actual\_Index - Expected\_Index}{Maximum\_Index - Expected\_Index} \qquad [4.4]$$

An adjusted index takes random partitioning as a baseline of zero, with one indicating identical partitioning. Adjustment for chance allows comparison across different partitions with variable number of classes. The *diagonal* is another pair counting index that can be used to assess consonance of partitions.

The normalised mutual information (NMI) belongs to the class of information theoretic measures. Information theoretic measures focus on entropy – the amount of information a random variable adds given another covariate. Entropy is the reverse of mutual information, which can be defined as the quantity of information that is shared by two covariates or partitions. Information theoretic measures treat classifications as random variables: they show a high degree of mutual information, if they are correlated. The normalised mutual information takes the value of zero when the two partitions are independent and one when they are identical.

The five indices – diagonal, kappa, Rand, adjusted Rand and NMI [Table 4.2] – have been selected for further investigation because they represent different classes of indices with their own strengths and weaknesses as well as uncertainties.

Table 4.2: Indices of cluster agreement used

| index | description | formula |
|---|---|---|
| diagonal | Number of identically classified pairs $P_{11}$ relative to all possible pairs $\binom{n}{2}$ | $\delta = \frac{P_{11}}{\binom{n}{2}}$ |
| kappa | Number of identically classified pairs $P_{11}$ relative to number of expected identically classified pairs $\bar{P_{11}}$ | $\kappa = \frac{P_{11} - \bar{P_{11}}}{\binom{n}{2} - \bar{P_{11}}}$ |
| Rand | Number of consonant pairs $P_{11} + P_{00}$ relative to all possible pairs $\binom{n}{2}$ | $R = \frac{P_{11} + P_{00}}{\binom{n}{2}}$ |
| adjusted Rand | Number of consonant pairs $P_{11} + P_{00}$ relative to number of expected consonant pairs $\bar{P_{11}} + \bar{P_{00}}$ | $R_{adj} = \frac{P_{11} - \bar{P_{11}} + P_{00} - \bar{P_{00}}}{\binom{n}{2} - \bar{P_{11}} + \bar{P_{00}}}$ |
| norm. mutual information | The mutual Information $I$ of partitions $K$ and $L$ relative to the entropy $H$ of each partition $K$ and $L$ | $NMI(K,L) = \frac{I(K,L)}{\sqrt{H(K)H(L)}}$ |

#### Technical specification: multi-variate methods to measure local agreement

Whereas the indices of partition agreement provide a view of the overall correspondence of coancestry and isonymy groups, they do not reveal where in the population (Great Britain) isonymy and genotypes correspond much and where not. Informally, geographical agreement of two multi-dimensional datasets can be compared using methods of dimension reduction.

Principal Components Analysis (PCA) is a widely applied technique in genetic research to estimate the genetic similarity between populations and infer common ancestors as well as historic demographic events [Cavalli-Sforza et al. 1994]. If the Principal Component (PC) scores are geocoded, it is possible to estimate a spatial surface of genetic similarity of populations. PCA is a method that takes two or more variables and, by observing their covariance, reduces them to underlying components [see e.g. Hastie et al. 2009, 534 sqq]. The components can be understood as latent variables with a standardised score that follows the z score distribution (in the following PC score).

Multi-dimensional Scaling (MDS) is another form of ordination – it reduces the dimensionality of the data to a number of dimensions, this time based on a dissimilarity matrix between observations [ibid., 570 sqq]. The method translates pair-wise distances of a dataset into coordinates in a multidimensional statistical space. The coordinates, or dimension scores, can then be treated as abstract variable values of the domain in question, in this case coancestry or isonymy.

The isonymy and coancestry matrices are processed in PCA and – in inverted form – in

MDS, and a suitable number of dimensions is chosen based on the eigenvalue of each. The scores for each dimension are then translated into a spatial point distribution of population similarity, using the geo-referenced data points (sample locations and parish or ward centroids). These points are then interpolated in order to infer a continuous spatial surface of population similarity. Inverse Distance Weighting (IDW) and Kriging are among the most widely applied spatial interpolation techniques [see Longley et al. 2011a]. The choice of method depends on the nature of the data and the intended application. For visual comparison of the geographic distribution of PCA scores, the simpler method of IDW has been chosen with a value decay function over the squared distance to the most proximate observations.

### Technical specification: measuring Regional Integrity

In order to identify discrete regions of local agreement between coancestry and isonymy, three indices are developed that measure local correspondence of genotypes and isonymy groups: Distinctiveness, Dominance and Regional Integrity.

Distinctiveness is measured as the extent to which the genotypes found in areas of the same isonymy group do not also occur in other isonymy groups. If, for example, 100 per cent of observations of a genotype occur only in one isonymy group, the genotype is characteristic of that isonymy group and thus contributes to Distinctiveness. The logic is similar to that of the Location Quotient (LQ), which can be modified to range between zero and one as follows:

$$DIS_k = \sum_l \frac{n_{kl}^2}{n_k n_l} \qquad [4.5]$$

where $n_{kl}$ is the number of observations in insonymy region $k$ that belong to genotype $l$. Hence for each isonymy group, Distinctiveness is the sum of the regional share of total observations that belong to a given genotype weighted by the prevalence of the genotype within the group. In other words, Distinctiveness can be conceived of as weighted proportional LQ of all genotypes that occur in one region. Since it uses proportions, Distinctiveness can be interpreted as a probability that the regional population is distinct from the remaining regional populations.

The second component, Dominance, measures the internal homogeneity of a region's population. The index can be expressed as the Simpson Index of Dominance [Simpson 1949].

$$DOM_k = \sum_l \left(\frac{n_{kl}}{n_k}\right)^2 \qquad [4.6]$$

It describes the probability that two randomly chosen individuals from a region belong to the same genotype and thus the degree to which one population group is dominant relative to the remaining groups. Yet, while a probabilistic reading is useful, it should be noted that some values of the indices are more likely than others because of unequal sizes of isonymy groups. It is necessary to adjust for chance, as is done with the Rand index, whereby the expected index value from the actual value is subtracted and divided it by the difference between the maximum possible value and the expected value [cf Equation 4.4]. The maximum possible value for both Distinctiveness and Dominance is one in both cases. If genotypes were distributed randomly, the expected Dominance in each isonymy group would be $1/L$, where $L$ is the number of genotypes. The expected value for Distinctiveness of a group equals the share of total observations that falls into a region. Under adjustment, a value of zero indicates randomness and one perfect departure from randomness. In contrast to Dominance, adjusted Distinctiveness can fall below zero in extreme cases, for example when one region only encompasses one observation or indeed no observation at all. Negative Distinctiveness would, however, indicate poor overall partitioning.

A combined index of Regional Integrity can be calculated by multiplying both indices together.

$$RI_k = DIS_{k.adj} \cdot DOM_{k.adj} \qquad [4.7]$$

where the subscript $adj$ indicates the adjusted version of the index.

### Technical specification: assessing urban dynamics

Urban dynamics are assessed in two steps. First, the local surname compositions of parishes in 1881 and wards in 2007 are compared by calculating isonymy between the areas with the whole population and the population with Anglo-Saxon (including Celtic, Irish and Cornish) names only. This produces a measure of change for each area due to long and short term international migration. The measures correlate closely with the percentage of people with non-Anglo-Saxon names in each area.

Second, the population dynamics between 1881 and 2007 are compared by defining a common spatial unit out of 1881 parishes and 2007 and calculating isonymy between the 1881 and 2007 units. In rural areas, parishes cover larger areas than wards; and the reverse is true for urban areas. A common spatial unit corresponding to the larger aggregations are formed by a procedure as follows. The closest parish or ward is determined to the centroids of wards and parishes respectively (Spatial Join) and a lookup table is produced to relate parishes to their closest ward and vice versa. For each parish and ward, the number of wards and parishes they respectively cover is counted. Thus, for each parish-ward pair in the lookup table, the one with the higher number represents a higher level of aggregation and is hence chosen as a new area identifier. This identifier is appended to the 1881 parish and 2007 wards shapefiles and both are first intersected and then re-aggregated based on the new identifier. This results in 1,106 spatial units that approximately reflect the overlapping geographies of 1881 parishes and 2007 wards at the coarsest level.

### Technical note on digital boundaries for 1881 parishes

Since parishes vary in size and range from very small to large, an algorithm is developed to merge neighbouring parishes that belong to the same district and together form a population of at least 750 people. This results in 6,848 parish groups for Great Britain. 2007 records are aggregated to UK census wards.

Digital boundaries for Scottish parishes of 1881 are not yet available and therefore Scottish census records are geocoded based on 1951 parish boundaries in conjunction with 1881 gazetteer that was provided by www.visionofbritain.org, University of Portsmouth. This procedure resulted in an approximate jurisdiction geography of 1881 Scottish parishes.

## 4.3 A regional geography of population structure in Great Britain

### Inferring biology from surnames geographies

Classifying study participants by their genotype and by the surname mix of their ancestral places of birth produce different partitions whose agreement (partition consonance) can be measured by five indices of partition consonance: diagonal, kappa, Rand, adjusted Rand and Normalised Mutual Information (NMI). The indices of partition consonance are calculated repeatedly for all combinations of $k \in \{2..80\}$ surname clusters and $l \in \{2..53\}$ genotypes. The indices behave differently and, at first sight, highlight different combinations of isonymy groups and genotypes [Figure 4.2].

The diagonal shows the highest level of agreement (.863) between $k = 2$ isonymy groups and $l = 2$ genotypes, before it sharply drops to a level near zero. It peaks again at $k = 5$ and $l = 3$ partitions, but generally stays below a low value of .1. The chance corrected form of the diagonal – kappa – fluctuates at the beginning where partitions have few clusters. Kappa remains below a value of .2, indicating low agreement overall. Since the two indices only accept cases as consonant that are assigned to the same classes in the two partitions, they are the most restrictive and tend to favour symmetrical partitions with few clusters.

The Rand index [not illustrated] and the adjusted Rand index have different values but peak on similar solutions. The adjusted Rand index highlights combinations with $k = 3$

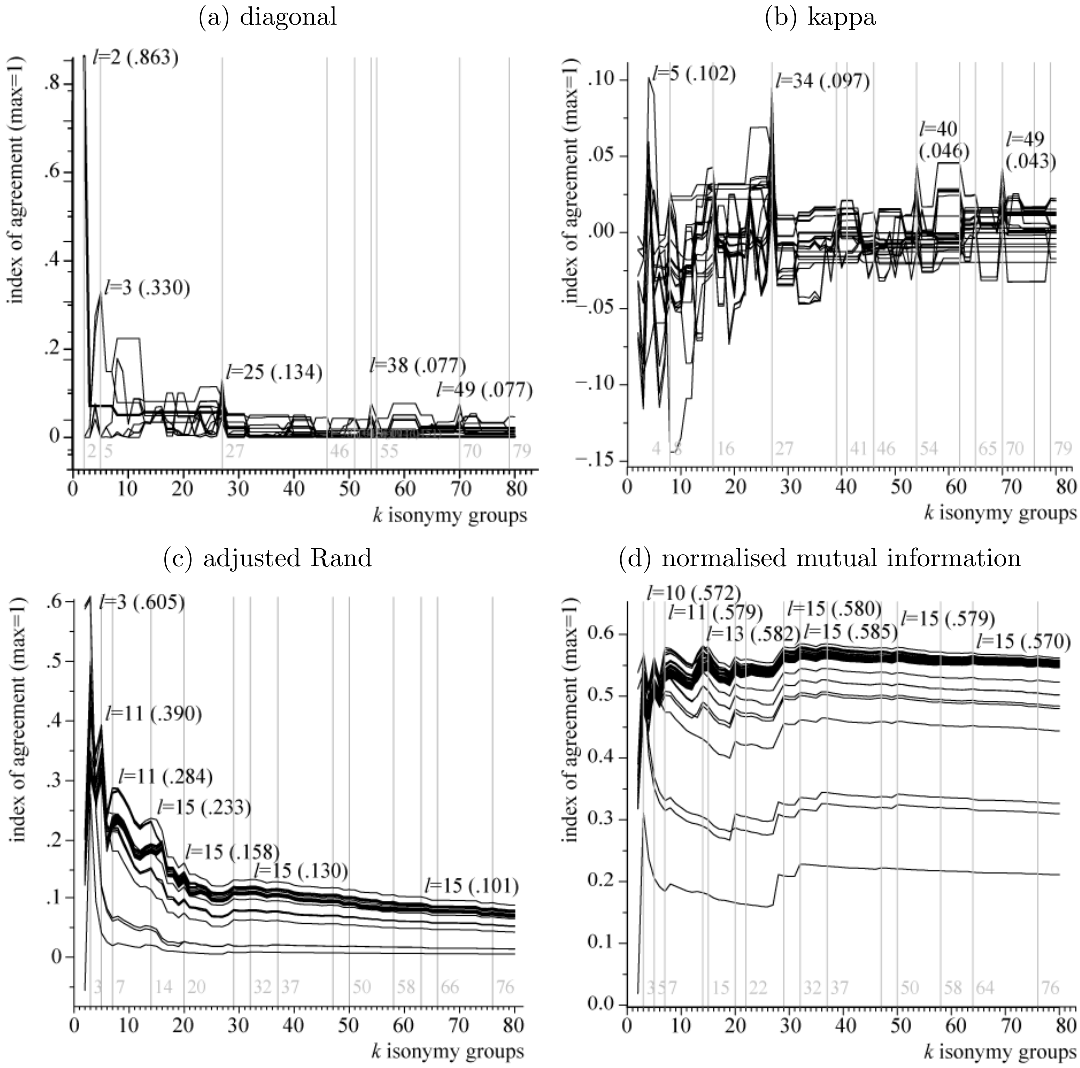



Figure 4.2: Four indices measuring partition consonance of varying $k$ surname regions and $l$ genotypes. Each line represents a different genotype solution with $l \in \{2..53\}$.

isonymy groups and $l = 3$ genotypes and successively peaks at more complex solutions. After $k = 16$ isonymy groups, any combination with $l = 15$ genotypes remains superior to any other genotype solution. The normalised mutual information peaks at similar combinations as the Rand indices. The highest peak is at $k = 14$ by $l = 13$ cluster solutions. After $k = 20$ isonymy groups, the solution of $l = 15$ genotypes consistently shows the highest level of agreement.

The review of consonance indices suggests that the adjusted Rand index provides a useful benchmark to review a succession of selected partitions that agree most. In general, the consonance levels decrease as partitions become more complex. The combination with the highest index value ($R_{adj}$ = .605) is one with $k = 3$ surname clusters and $l = 3$ genotypes.

Mapping the location of isonymy groups and genotyped individuals reveals a corresponding geographical pattern of surname clusters and genotypes [Figure 4.3(a)]. It should be noted that no geographic information is processed in the clustering; the emerging geographies are independent outcomes. The clustering of 1881 parish groups produces three regions with distinct surname mixes: a region almost corresponding to Scotland, a region nearly corresponding to England and a third, Welsh, region. The geography of genotypes marks similar regions. There is a large spread of one genotype across England and Scotland. A second genotype is concentrated in Wales, and a third genotype is concentrated in Orkney. People living in these three regions seem to be genetically most different from each other. The overall consonance between the isonymy groups and the genotype concentrations suggests with high certainty biogenetic differences between the populations of those regions.

The next partition pair, for which the adjusted Rand index peaks with .390, comprises a solution with $k = 5$ isonymy groups and $l = 11$ genotypes. This solution contains a split of England into three regions: a northern, a central region covering East Anglia and a southern region covering the Cornwall peninsula. This solution is followed by a partition of $k = 7$ and $l = 11$ and an index value of .284 [Figure 4.3(b)]. The first isonymy group still covers Scotland; it has remained unchanged. Four genotypes are now concentrated in this region, with two distinct genotypes detected in Orkney. A second isonymy group emerges in northern England, where also a number of spatially overlapping genotypes can be observed. It appears that in isonymy groups one and two, the genetic diversity of populations is higher than the mix of surnames would suggest at this level. Isonymy groups three, four and five encompass the widespread genotype number 15, which reaches into the northern England isonymy group two. A sixth isonymy group emerges in the south west, encompassing two concentrations of genotypes there: one in Cornwall (genotype three) and one in Devon (genotype ten). Isonymy group seven – the Welsh region – corresponds to the concentration of now four distinct genotypes. Although at a slightly lower level of certainty, this partition pair reveals a finer regional population structure than local from surname mixes would suggest.

The next partition pair is one with an adjusted Rand index of $R_{adj}$ = .233, $k = 14$ isonymy groups and $l = 15$ genotypes [Figure 4.3(c)]. The first change to observe is that Scotland splits now into three isonymy groups: a northern one covering the south of Orkney, an eastern one including north of Orkney and a southern Scottish region. The eastern region comprises some local concentrations of genotypes in Aberdeenshire as well as the Orkney genotype number two. The northern England isonymy groups remains unchanged, but there is now a distinct isonymy group eleven in the Liverpool-Manchester area, which in part corresponds to the geographical extent of genotype eleven. Mid and southern England is now split into five isonymy groups, across which the widespread genotype 15 can be observed, and a region at the England-Wales border

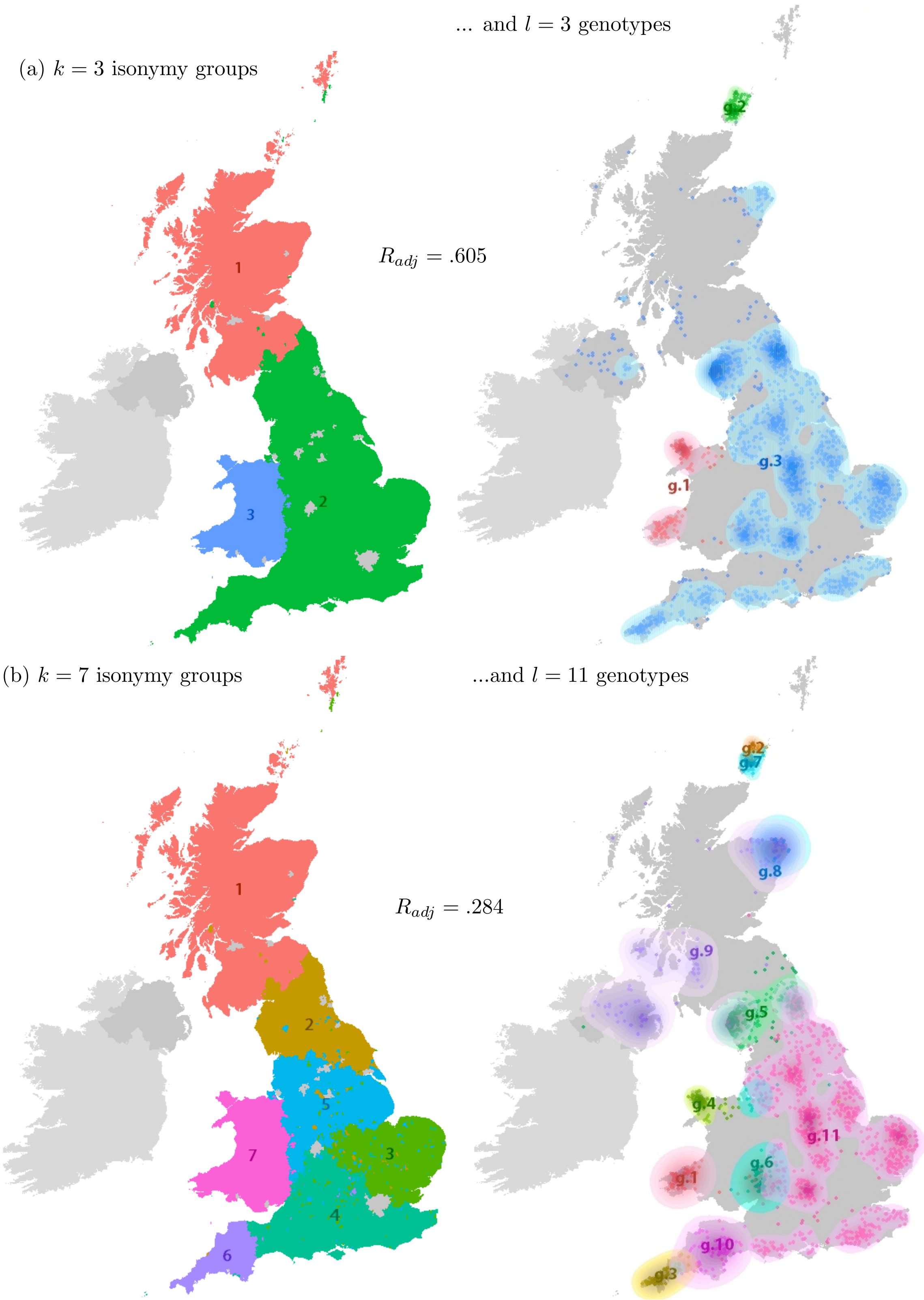
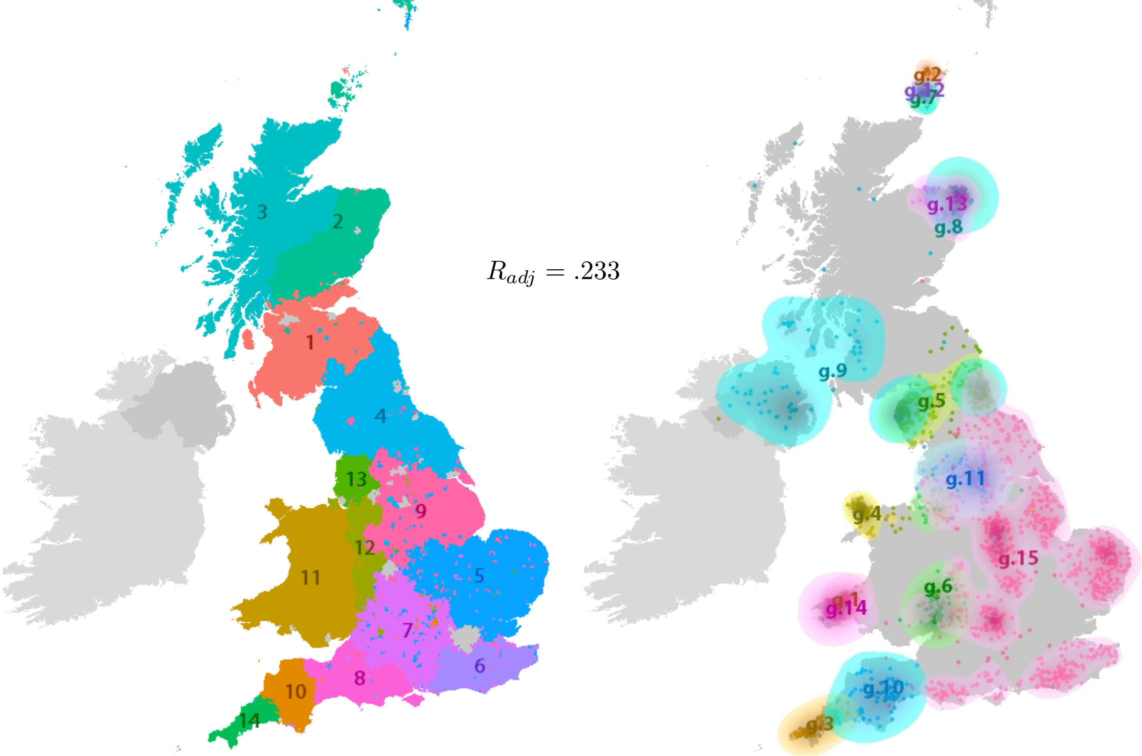



Figure 4.3: Partitions of $k$ isonymy groups and $l$ genotypes with high consonance.

(c) $k = 14$ isonymy groups ... and $l = 15$ genotypes




(d) $k = 20$ isonymy groups ...and $l = 15$ genotypes

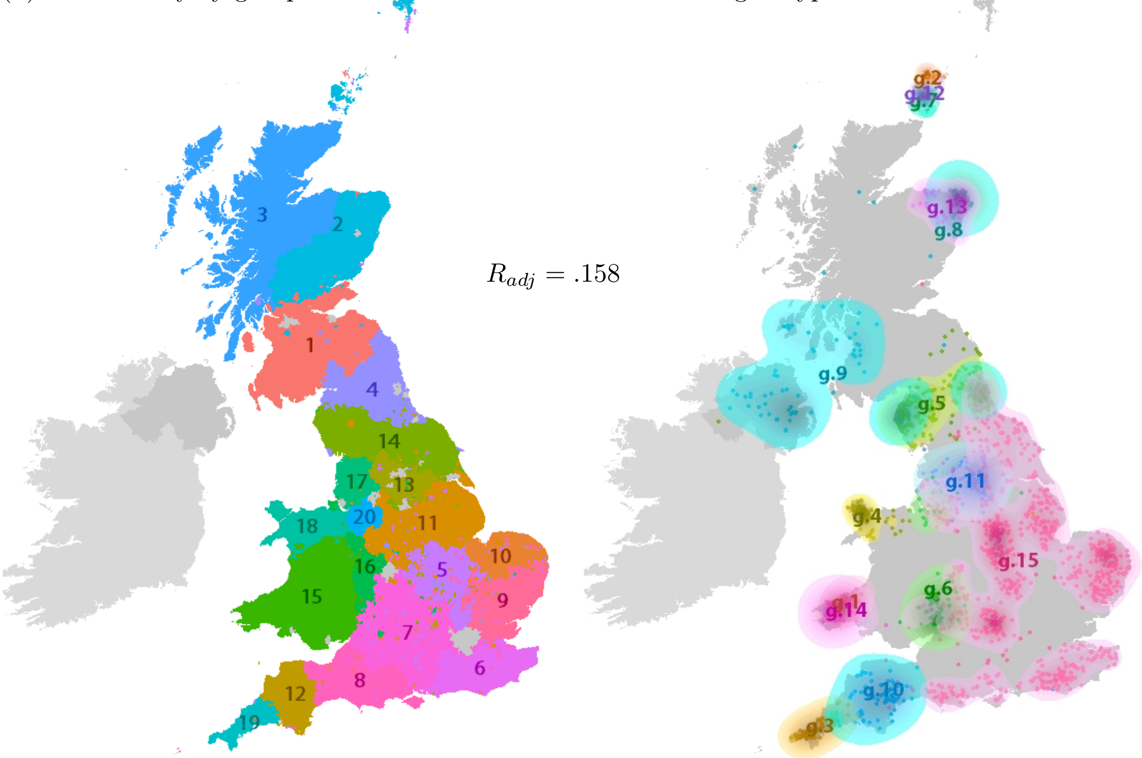



Figure 4.3 (continued)

corresponding to genotype six. The previous south western English isonymy group now splits neatly into Cornwall and Devon, corresponding to the two distinct genotypes that concentrate in each. The Welsh region remains unchanged.

The next partition pair with a peaking index shows larger complexity with $k = 20$ isonymy groups and less certainty with $R_{adj} = .158$ [Figure 4.3(d)]. The major changes that occur compared to the previous pair are further splits in the northern England region, in the Midlands and East Anglia as well as in Wales. There are now two distinct isonymy groups along the Welsh border. The surname clusters in mid and southern England suggest much more regional heterogeneity than is detected by genetic variants. Historically, this could result from a wave of early settlers which were genetically similar but spread over a large area in England before surnames were introduced. Before surnames became hereditary by the 15th century, they could change frequently between generations or individual biographical events with regionally different naming conventions. Perhaps the time that has passed since then is not sufficient to show strong signs of genetic drift, which may have further slowed down through migration in the area – but overall this result is difficult to explain and would warrant further investigation. The new northern England isonymy groups broadly correspond to different regions of overlapping genotype geographies.

So far, the results confirm that regional surname compositions may have biogenetic contents. The successive considerations of different partition pairs suggest that population structure can be asserted with increasing geographical granularity and uncertainty levels. The global estimate of cluster agreement may be a first step to assess the scale of specificity and regionalise the British population based thereupon.

### Interpolated geographies of population specificity

Applying PCA to the genetic coancestry matrix and the 1881 and 2007 isonymy surname matrices produces similar geographies of population specificity in the British 'native', rural population [Figure 4.4]. The first component of genetic coancestry identifies distinct genetic covariation in Wales, which decays by distance and then reappears in the North of Scotland towards Orkney. The first PCs of the area isonymies, too, mark Wales as different from the remainder of Britain.

The second component of genetic coancestry highlights again Wales, but this time the component shows the largest difference in northern Scotland and Orkney and along a line dividing eastern and western England. The second component of 1881 area isonymy highlights the south of Wales and southern England with contrasting localities northward from the Midlands. It reveals some similarity between southern Welsh and southern English areas, as well as Orkney and appears to partially reflect the similarity between Orkney and Wales detected by the first PC of genetic ancestry. The second PC of 2007 isonymy indicates the same pattern to the 1881 component, just in reverse form.

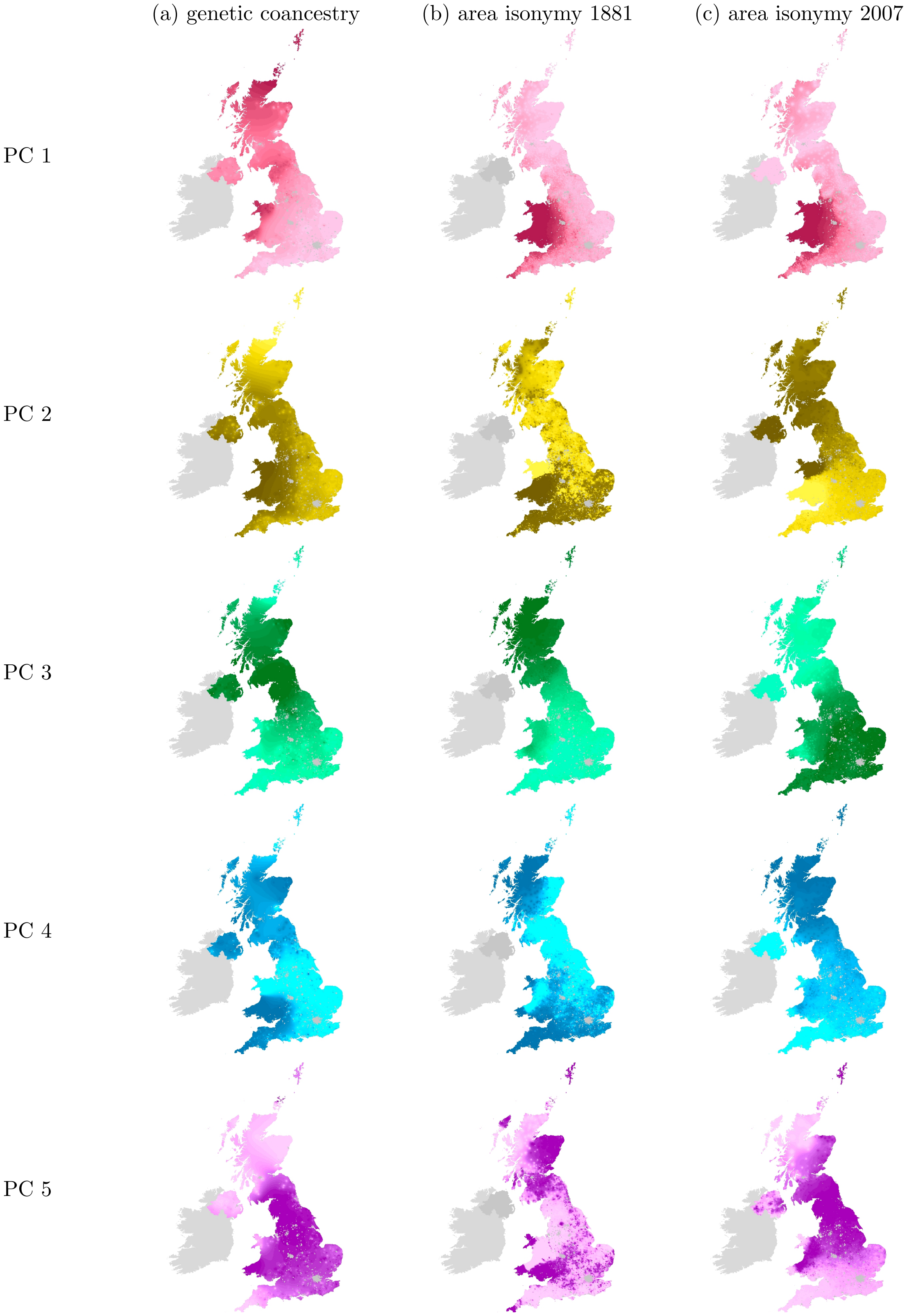



Figure 4.4: The first five principal components of genetic coancestry (a), area isonymy based on local 1881 (b) and 2007 (c) surname mixes.

The third components of all three datasets mark north-south differences between populations, with a contrasts near the England-Scotland border. Orkney again appears as a region of dissonance between the two: surnames suggest similarity and genes dissimilarity to the Scottish population. The fourth components of both genes and 1881 surnames highlights south western England, southern Wales and northern Scotland. The areas that are most different are southern Scotland, the Scottish east coast as well as northern England and south eastern England. The fifth component of 1881 isonymy resembles the fourth coancestry component; the information the two component pairs provide seem to overlap here.

Mapping the components reveals a complex picture of similarity and dissimilarity between rural British regions. Most likely, the geographies reflect the outcome of successive migrations and genetic drift over time. Both PCA solutions clearly highlight the specificity of the Welsh, northern Scottish and Orkney populations. The movement of Vikings is an example of a historic event that plausibly connects these populations: in the 9th century B.C., Norse Vikings travelled to and settled in the Shetlands, Orkney, and further down along the Scottish and English west coasts [Hellenthal et al. 2015]. This is just one example of historical population movements that may be visible to a historian in these maps. Rural British population structure appears to be a long-term outcome of historically successive admixture of both DNA and local surname compositions.

Next, MDS is applied to test whether similar patterns emerge based on dissimilarity between observations. Both the isonymy and coancestry matrices are inverted into distance matrices. The first three MDS dimensions contain – in compressed form – the information that could already be observed in the PCs [Figure 4.5]. The patterns emerging from the first two interpolated surfaces of 1881 and 2007 isonymy are nearly identical. The third dimensions each emphasise a different spatial process, however. While genetic coancestry and 1881 isonymy emphasise a north-south orientation, 2007 isonymy reveals the presence of a distinctive pattern in the centre of Great Britain. Here, it seems that spatial mobility of the Anglo-Saxon population may produce an impact on area isonymy that becomes visible at the third dimension.

### A triangulated geography of population structure in Great Britain

The succession of isonymy maps suggests different levels of regionalisation that best reflect the population in genetic terms. The interpolated surfaces further suggest that the correspondence between genes and surnames varies geographically. This may be an effect of both sampling design and divergent processes affecting the respective geographies of surnames and genes. In some parts, isonymy suggests greater distinction or greater homogeneity of the corresponding sub-sample than can actually be found in the sample.

Based on this reasoning, high local correspondence may be defined as a property of an isonymy group when it encompasses a set of observations that is distinct from other

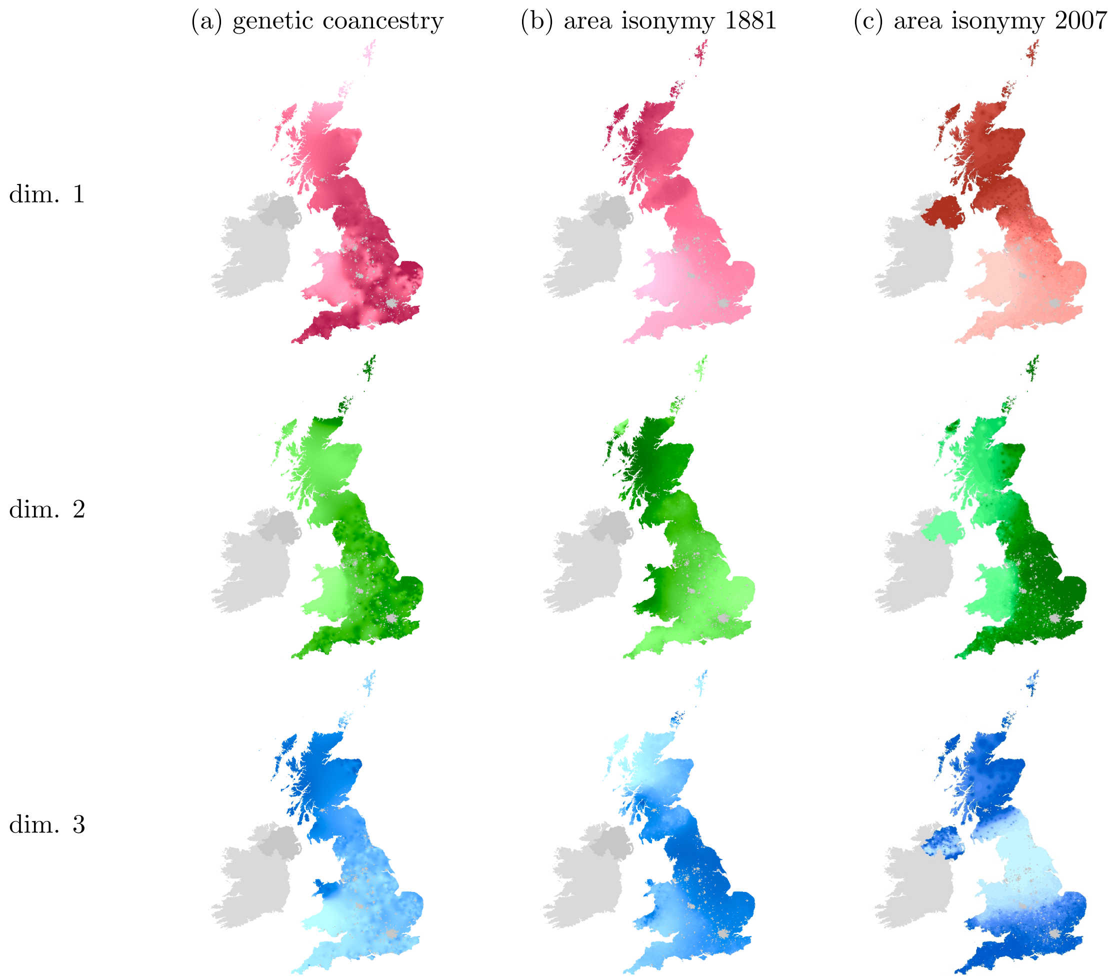



Figure 4.5: The three dimensions of multidimensional scaling of genetic coancestry (a), area isonymy based on local 1881 (b) and 2007 (c) surname compositions.

isonymy groups (Distinctiveness) and simultaneously internally homogeneous (Dominance). Multiplied together, the two components convey a sense of confidence one can ascribe to the degree to which an isonymy group truly represents population structure. This may be termed the Regional Integrity of an isonymy group in terms of the population structure found in the sample.

Distinctiveness, Dominance and Regional Integrity are calculated for all three most agreeing combinations of isonymy regions and genotypes. For the first combination [Figure 4.6(a)], the one with the highest level of global correspondence, the Welsh isonymy region has the highest level of Distinctiveness with a value of .721. The English isonymy region has the second highest level of Distinctiveness (.559). The Scottish isonymy region exhibits the lowest level of Distinctiveness (.466), which can be explained by the fact that, although Scotland hosts the distinctive population that is classified as genotype two [cf Figure 4.3(a)], these comprise just 40 per cent of the observations in Scotland. The remaining 60 per cent belong to the mostly English genotype three, which reduces Distinctiveness and even more so Dominance. In terms of Dominance, the English region is the top-scoring one because within its regional extent, only genotype three prevails. The Welsh region hosts all observations with genotype number one, but the observations in the England-Wales border region, which make up a quarter, belong to genotype number three.

These patterns attest to a higher level of Regional Integrity in the English region, followed by Wales and Scotland. Given that in the latter two regions, Distinctiveness is higher than Dominance, a more granular division may increase Regional Integrity. Conversely, it is less likely that Regional Integrity in the English regions can be improved unless further splits in terms of isonymy and coancestry are spatially congruent.

The next combination with seven isonymy regions and eleven genotypes changes some of the patterns [Figure 4.6(b)]. Now, the Scottish region exhibits a higher level of Distinctiveness than other regions because those genotypes that formerly extended across Southern Scotland and England split into more local genotypes number five, eight and nine [cf Figure 4.3(b)]. The latter two are largely distinct to southern Scotland and only a minority of the genotypes five and eleven remain in this region. Given a resulting lower expected Dominance, the more granular genotypes increase homogeneity, adding up to an overall improvement of Regional Integrity for the Scottish isonymy region. The unchanged Welsh region receives a higher Distinctiveness because of a regional subdivision of the formerly English genotype into a new genotype number six, which prevails in the England-Wales border region. Given further splits, population heterogeneity increases and therefore Dominance remains low in comparison with other regions. A highly distinct region has emerged in the south west, covering Cornwall and Devon. Given that this new region corresponds to two highly local genotypes, the level of Dominance is lower. The picture reverses for England. While most English regions lose Distinctiveness, their internal homogeneity remains high as the widespread genotype eleven is dominant among their corresponding observations in the sample.

The resulting map of Regional Integrity appears rather invariant. The most consistent region is the Cornwall-Devon cluster, while the least consistent regions are the south eastern isonymy groups. Regional Integrity has improved for the Scottish region compared to the coarser partitioning, while it has decreased for Wales. There and in Cornwall-Devon, Dominance is lower than Distinctiveness which indicates that greater integrity could be gained by further regional division.

The final combination with the highest granularity confirms the tendencies found in the previous one [Figure 4.6(c)]. 20 isonymy groups and 15 genotypes reduce the number of highly distinctive regions to one in northeastern Scotland (which includes parts of the Orkney and Shetland islands), two Welsh regions, Cornwall, Devon and to a lesser degree southern Scotland. The southern and eastern English regions are the least distinctive, which arises because the widespread English genotype number 15 does not split significantly compared to the last combination. The latter is reflected in a high level

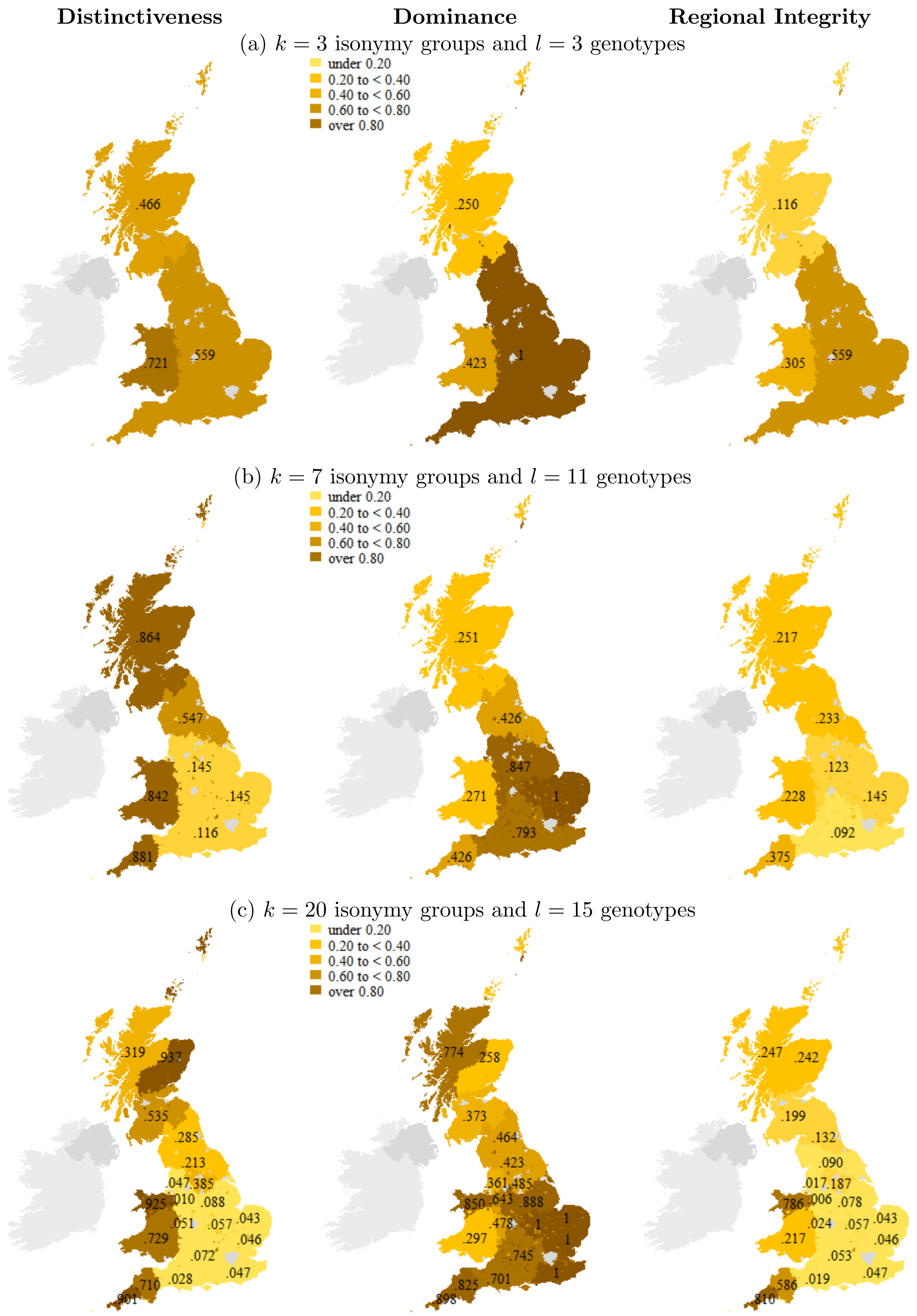



Figure 4.6: Distinctiveness, Dominance and Regional Integrity for each combination with high global correspondence.

of Dominance among the southeastern English isonymy groups. Four of these show a value of one, implying that there is only one genotype present among the corresponding observations. The absolute number of observations varies between 89 and 155 in these regions. Now, the southwestern isonymy groups exhibit higher values of homogeneity along with clusters in northern Wales and northern Scotland. At this higher granularity, it seems that the isonymy groups reflect more homogenous populations in most parts of the country based on the sample information on coancestry.

Yet, Regional Integrity suggests that the lack of Distinctiveness in some of the isonymy groups outweighs the benefits of viewing homogeneous populations. Merging some of the English regions, for example, would better reflect population structure in Great Britain than the highly granular solution. The only regions that show a high likelihood of Regional Integrity are Cornwall, Devon and northern Wales. There is a second tier of intermediate values, which can be found in Scotland and southern Wales. This is due to higher Distinctiveness in south Wales and Scotland-Orkney, suggesting that another split of this region may improve integrity values.

The Regional Integrity values at various levels of granularity can be used to derive a synthesised picture of Great Britain's geography of population structure, which assembles those isonymy groups and genotypes that lead to the highest Regional Integrity. In order to do so, one may formulate a decision algorithm about how partitions are to be constructed:

1. Identify the combination with the highest level of granularity one is comfortable with. This decision should be based on the Adjusted Rand index and the sample size.

2. Merge the regions to the next coarsest level as indicated by the next local maximum of the Adjusted Rand index.

3. If Regional Integrity of the merged regions is higher than all its sub regions, accept the merger; otherwise keep the sub-region with higher Regional Integrity and merge the remaining sub-regions.

4. Recalculate local correspondence indices for the resulting partition.

5. Repeat steps two to four until no further improvements occur or the coarsest level of regionalisation is reached.

This algorithm results in a synthesised geography of population structure in Great Britain [Figure 4.7]. Note that mergers of sub-regions that belong to different parent regions are not allowed. The resulting partition corresponds to eight isonymy regions based on the 15 coancestry groups. Scotland is now composed of far northern, northeastern and southern sub-regions. The large English region comprises all former sub-regions except Cornwall-Devon, which each remain separate. Wales is divided into a north-

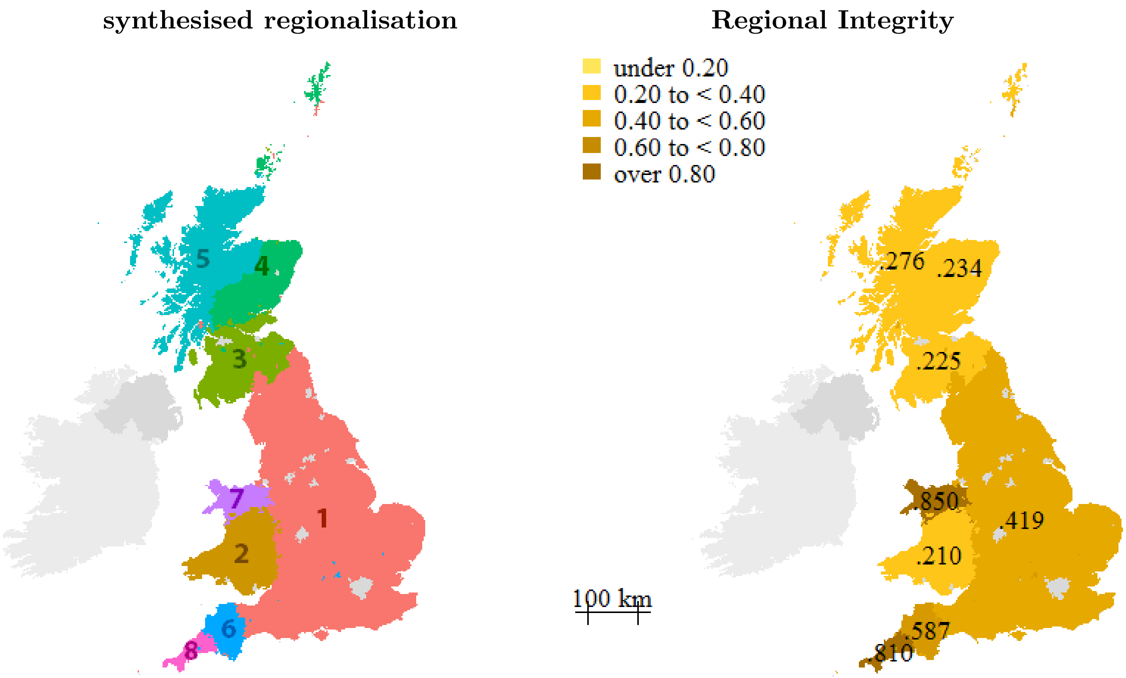



Figure 4.7: The synthesised regions of population structure in Great Britain (left) and their corresponding regional integrities (right).

ern and southern part. The resulting Regional Integrity values range between .210 in southern Wales and .850 in Cornwall.

As a whole, the map highlights those areas of Great Britain where one may place greatest confidence that they represent population structure – northern Wales and Cornwall – given the information in and geographic distribution of the sample. More caution should be employed in Scotland and southern Wales, where further investigation would be necessary to increase certainty about population structure. By viewing the components of Regional Integrity, it becomes evident that the high uncertainty in southern Wales arises because of internal heterogeneity, that is lack of Dominance among the prevalent genotypes. Further sampling in this region may help to attain more clarity about population structure in Southern Wales. This may well lead to a further split of Southern Wales or a merger with adjacent regions.

The same conclusion can be drawn for southern Scotland, while the contrary applies to northern Scotland, where the sample suggests a high level of homogeneity but a lower level of Distinctiveness. In Scotland, the sampling is particularly uneven with nearly half of all Scottish observations made in Orkney. Again, a more geographically representative sample may increase confidence about population structure in Scotland and improve the integrity of regional partitioning.

## 4.4 Local specificity and urban dynamics

How do these findings relate to diverse, dynamic urban environments? It may be expected that population specificity has a different meaning in diverse urban environments, where due to domestic and international migration the composition of neighbourhoods may change quickly, leading to more similar surname mixes across cities. The temporal stability of biogenetic and cultural phenomena is an important criterion in recognising local specificity and using surnames in health research.

The degree to which urban neighbourhoods exhibit biogenetic and cultural stability – as approximated through surnames – is now investigated by means of a quantitative, encompassing comparative study of fifteen large conurbations in the UK. According to Robinson [2011], encompassing frameworks typically presuppose connecting relations, such as hierarchical relationships among cities. Here, there is no specific hypothesis about hierarchies, but the assumption that movement of British citizens between the conurbations could occur freely for centuries while all conurbations have been subject to the same immigration regime. In theory, migration could operate on the cities in a similar fashion, with potentially similar outcomes, and yet, specific outcomes are likely on the basis of, for example, different economic structures and trajectories among cities.


Figure 4.8: The 15 selected conurbations.

Table 4.3: Sizes and geographies of population registers.

| | 1881 | | | 2007 | | |
|---|---|---|---|---|---|---|
| | population* | surnames | parishes | population* | surnames | wards |
| total | 29,912,298 | 518,153 | 15,756 | 45,667,321 | 825,999 | 9,434 |
| urban | 14,898,972 | 339,620 | 2,244 | 20,067,398 | 559,431 | 2,706 |
| Anglo-Saxon | 13,275,778 | 56,128 | 2,244 | 16,997,972 | 322,812 | 2,706 |

* population registers are not complete

The urban areas investigated here are the previously excluded cities. 15 larger conurbations in the United Kingdom are included in the subsequent study: Belfast, Birmingham, Cardiff including Newport, Edinburgh, Glasgow, Kingston-upon-Hull, Leeds including Bradford and Wakefield, Liverpool including Birkenhead, London, Manchester, Middlesborough including Darlington and Hartlepool, Newcastle including Gateshead and Sunderland, Sheffield, Swansea and York [Figure 4.8]. From each city centroid, a buffer with a radius of 20 kilometres is drawn to capture the full extent of the urban areas. Larger conurbations require increased radii, such as Manchester, Liverpool and Birmingham with 25km and London with 30km. These radii are selected because the interest here is explicitly territorial: the long-term stability of urban neighbourhoods. The radii chosen are sufficiently large to cover suburban areas as spatially proximate sites of change. Clearly, processes outside these radii may shape population dynamics in those cities, but since the focus lies on assessing rather than explaining dynamics, a territorial definition of cities is chosen.

Urban areas are processed in two stages: first, they are classified by isonymy based on the population with Anglo-Saxon (including Celtic, Cornish and Irish) names and based on the entire population in a separate step. This separate view helps to assess the impact of international migrants on similarities and differences of local areas between cities. Second, local dynamics within metropolitan areas are investigated in terms of their diversity and their long-term dynamics.

### Isonymy of urban areas

All wards of the 15 conurbations are clustered by isonymy based on, first, the population with Anglo-Saxon names only (in the following, the Anglo-Saxon population), and, second, the whole population including people with non-Anglo-Saxon names. Plotting successive cluster solutions for isonymy of the Anglo-Saxon population in urban wards side-by-side reveals a pronounced regionalisation of urban populations. First, Scottish and northern Irish, English and Welsh urban wards emerge, which then successively split into further subregions almost along the borders of the individual conurbations. At $k = 4$ clusters, northern England's cities split from the remaining English cities with a boundary between the Manchester and the Leeds and Sheffield conurbations [Figure 4.9(a)-(b)]. At the next iteration, Liverpool and Manchester split from Birmingham and

London, creating five groups of urban regions [Figure 4.9(c)]. Belfast divides from the two Scottish cities, Glasgow and Edinburgh, at the next step.

At $k = 8$ solutions, the first inner urban areas emerge as distinct from surrounding areas: outer Birmingham forms a separate cluster, while central Birmingham remains in the same cluster as London. The northern English cities closer to the east coast split from the central conurbations, except central York, which maintains greater similarity with Leeds and Sheffield. At the next iteration, Liverpool splits from Manchester, and at $k = 10$, Newcastle forms its own region of wards, distinct from the remaining cities along the northern east coast. At the next two steps, the London suburbs break away from central London, which still remains in the same cluster as central Birmingham. Belfast divides into an inner and an outer urban region. The next splits mark further conurbations as distinct and produce more inner urban areas, including in Sheffield and north London. Considering the native population alone, a clear regionalisation emerges, which emphasises the existence of regional population differences between countries and sub-regions.

Clustering urban wards by isonymy based on the entire population exhibits a few differences in regionalisation [Figure 4.10]. The first coarse cluster solution of $k = 3$ regions again highlights Scottish and Northern Irish, English and Welsh cities. But already at the subsequent iteration that divides urban wards into $k = 4$ groups, certain inner urban areas emerge across English conurbations. A few suburbs in Manchester, Leeds and Sheffield as well as central Birmingham and a larger part of north west London form a distinct, geographically non-contiguous cluster.

Only then, one can observe the same separation of Liverpool and Manchester from the remaining northern English cities as in the clustering solutions with the Anglo-Saxon population only. At $k = 6$ clusters, when Liverpool, Manchester and Birmingham split away from London, a few central areas remain part of the London cluster [Figure 4.10(d)]. Outer Birmingham wards now split from London earlier and indicate similarity with the Liverpool and Manchester conurbations. This is not the case with the Anglo-Saxon population, where Birmingham appears more similar to London. This result suggests that international migration has changed London's surname composition significantly, which sets it apart from other English cities.

The next iterations follow a similar pattern to the one observed with the Anglo-Saxon population, although a few more urban areas show intra-urban heterogeneity earlier on, notably Belfast and London. At $k = 18$ solutions, London shows the highest diversity of areas, while Liverpool and Manchester continue for longer to form one cluster with Birmingham being distinct. Liverpool and Manchester, it seems, are more similar due to international migration than they would be if one considers the Anglo-Saxon population only.

Overall, urban areas are largely composed of regional populations except a few inner urban areas that are distinct from wider local populations and shared across cities. The

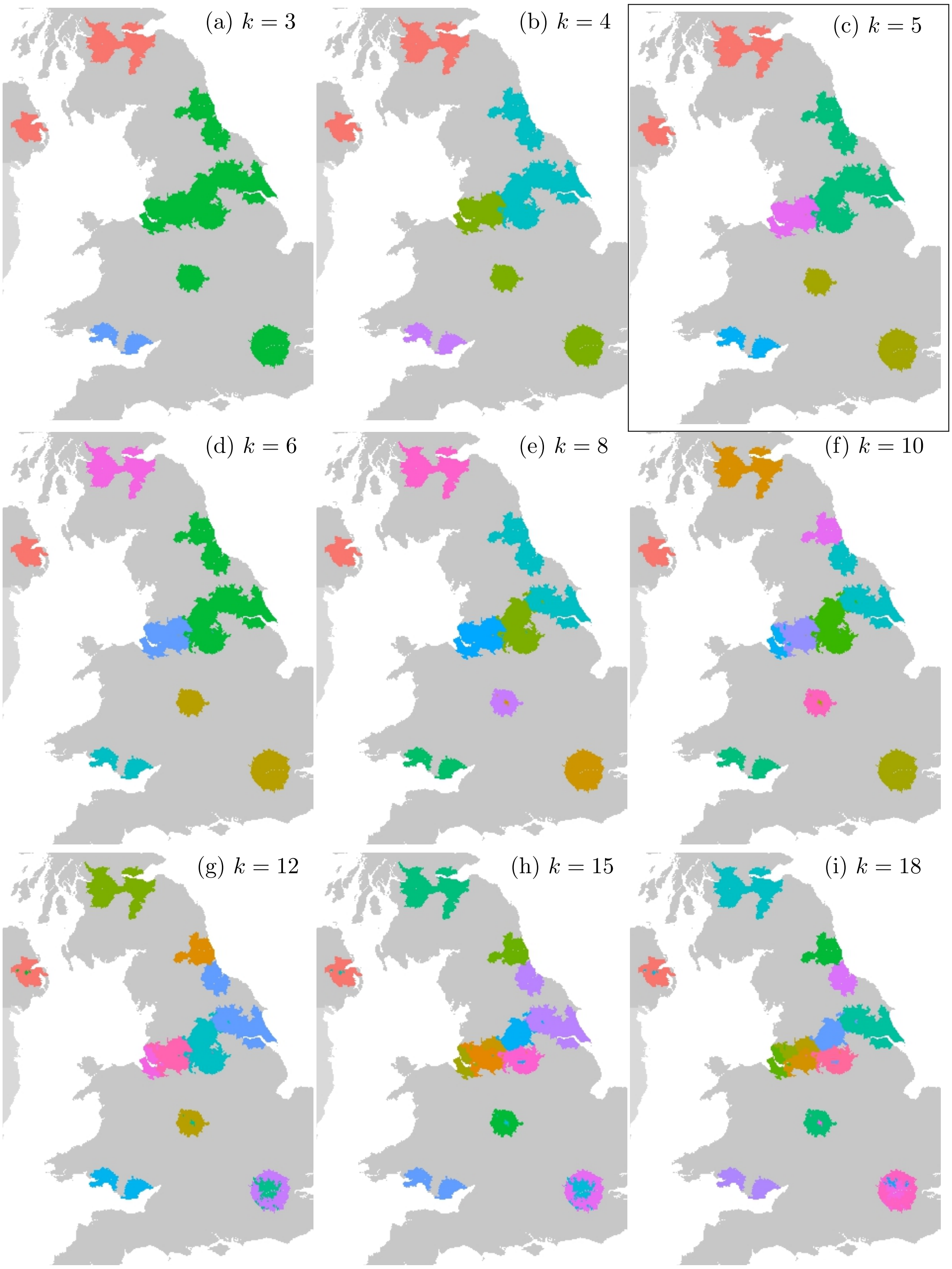



Figure 4.9: Partitions of $k$ surname regions of the urban Anglo-Saxon population.

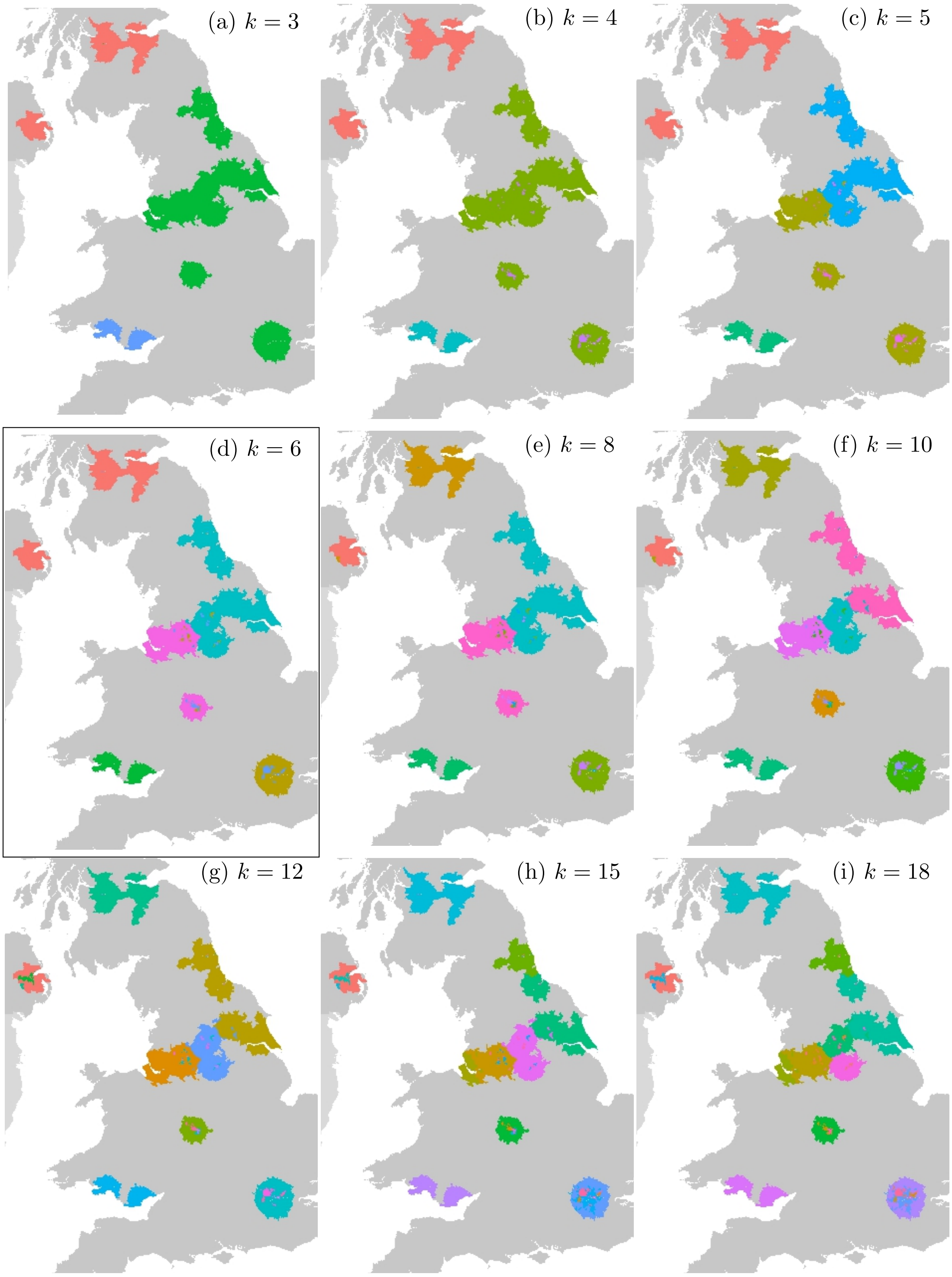



Figure 4.10: Partitions of $k$ surname regions of the entire urban population.

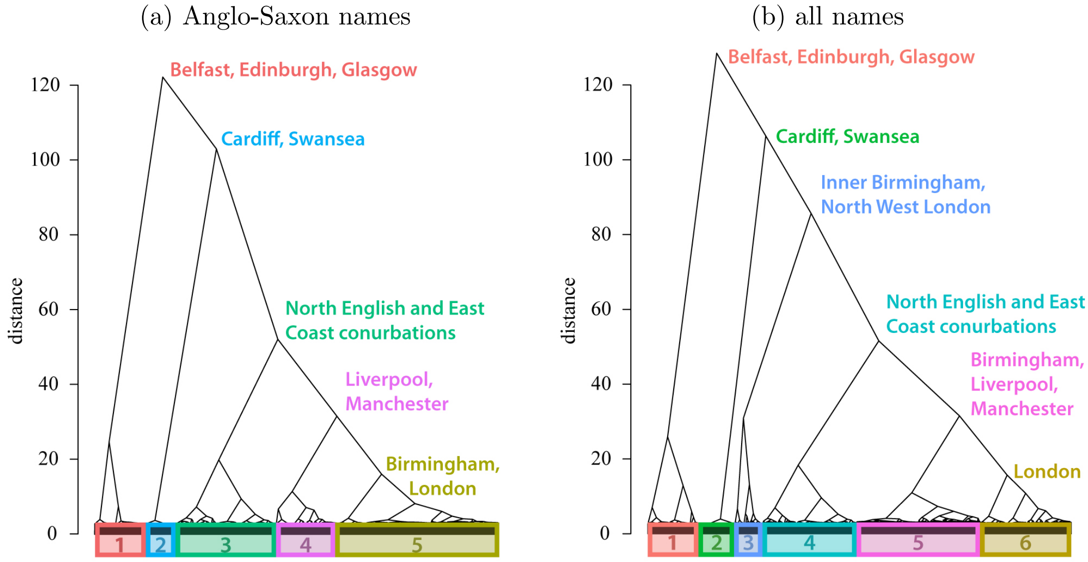



Figure 4.11: The dendrograms of clustered urban wards in 2007. At a distance coefficient of 30, five clusters prove most instructive when only population with Anglo-Saxon names and six clusters when the entire population is considered.

Adjusted Rand index of a comparison between cluster solutions of Anglo-Saxon against the entire population indicates a high level of consonance – at least 0.6 – for most partitions of similar divisions [not illustrated]. The high level of consonance suggests that, even if migrants are considered, urban populations tend to come from the nearby regional environment and migrants do not reduce significantly the inter-city geography of difference in the United Kingdom.

But how significant are these differences? The cluster dendrograms illustrate differences between urban wards as measured by a distance coefficient [Figure 4.11]. Cutting at a distance value of 30, the cluster solution with the Anglo-Saxon population splits into five, the one with the entire population splits into six groups [corresponding to figures 4.9(c) and 4.10(d) respectively]. The major difference across the two populations is the early split of inner city neighbourhoods, notably central Birmingham and north west London, when wards of the entire population are clustered. Since explicit geographic information is not part of the clustering, the emerging taxonomy illustrated in the dendrogram may be suggestive of larger-scale geographic forces that shape distinction between urban populations.

## Spatial patterns of urban diversity

In order to assess the geography of urban diversity, the whole population and the Anglo-Saxon population are compared – that is, isonymy between them is calculated – for each

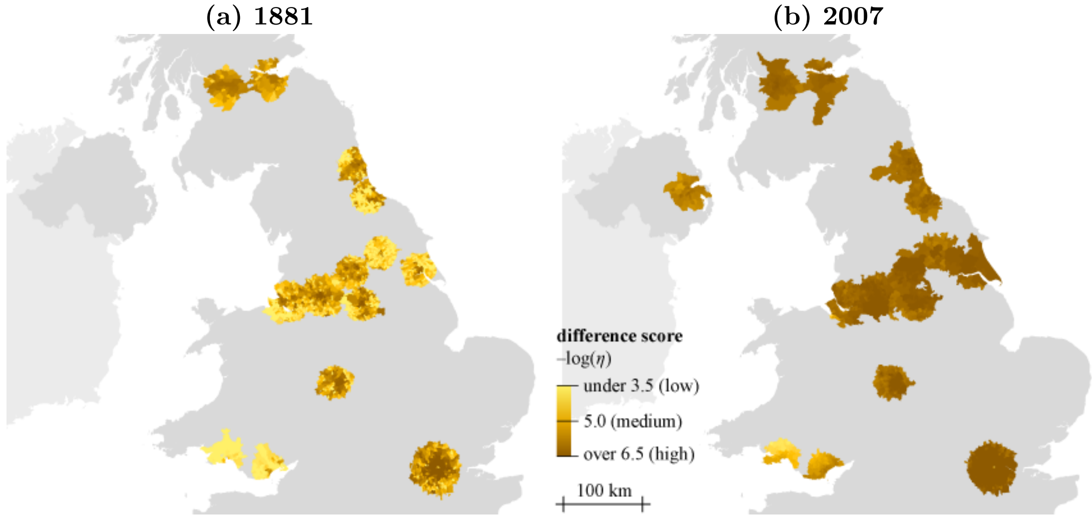



Figure 4.12: Inverted isonymy between the whole population and population with Anglo-Saxon names for (a) 1881 parishes (b) and 2007 wards. The higher the value the larger the difference.

area (1881 parishes and 2007 wards) separately for the two time periods. The resulting isonymy is inverted (negative logarithm) to represent a dissimilarity or diversity score. In this reading, diversity results from the local presence of international surnames [Figure 4.12].

A comparison between 1881 and 2007 suggests different magnitudes and geographies of long-term international migration. In 1881, people with international surnames mainly live in inner city and immediately surrounding areas across all metropolitan areas. Outer urban areas have fewer people with international names. The Welsh cities appear to be the least diverse with only a small proportion of people with non-Anglo-Saxon surnames concentrated in the centres of Cardiff and Newport. In York and Kingston, international migrants concentrate strongly in the city centre, whereas the surrounding areas show little dynamics in 1881. In Glasgow, Birmingham and London, people with international names are distributed more widely throughout the metropolitan areas.

In 2007, all metropolitan areas indicate significant dynamics due to international migration in all parts, except in the Welsh cities. The largest extent of dynamics can be found in London with high values of diversity throughout inner and outer London. They are similarly large in Leeds, Manchester and Sheffield. In Edinburgh and Birmingham, high dynamics as indicated by inverted isonymy are concentrated in more central areas. Glasgow, Liverpool and Newcastle show lower dynamics compared to other cities, which nevertheless are still significantly higher than in 1881.

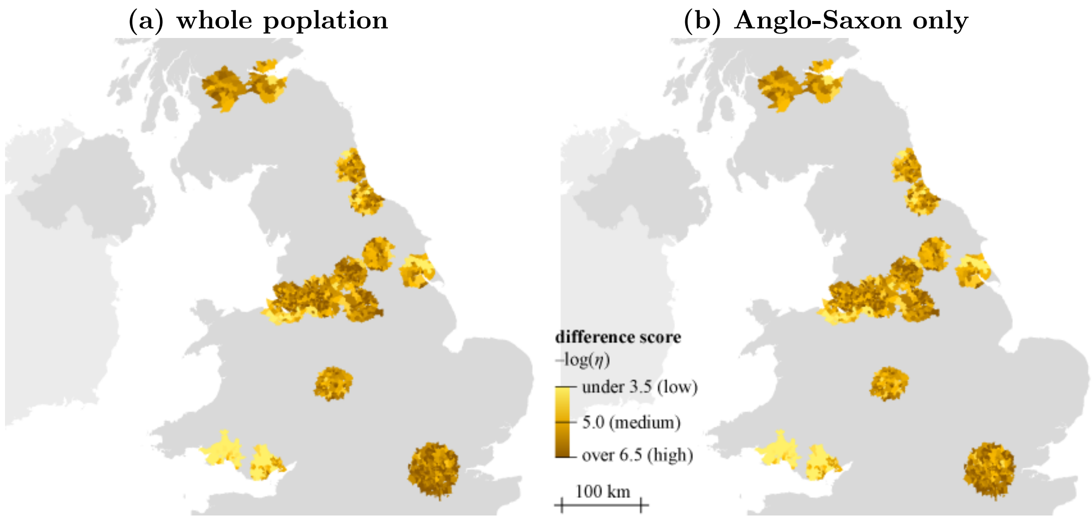



Figure 4.13: Inverted isonymy calculated for 1881 and 2007 spatial units. The higher the value, the higher the degree of change.

### Long-term change of urban areas

Changes in local surname compositions can also be compared across the two points in time for both populations, those with Anglo-Saxon names only and the entire population. Calculating again the logarithm of inverted isonymy for merged spatial units [see technical specifications on page 53] produces a geography of long-term urban change due to migration over short and long distances [Figure 4.13].

Assessing long-term dynamics for the whole population emphasises a mix of inner city neighbourhoods and suburban areas. This trend is best exemplified by the two Scottish cities, Manchester, Leeds, Birmingham and London. Swansea and Cardiff show again least population dynamics. In York and Kingston, these dynamics are concentrated in inner cities. The pattern is nearly identical for people with Anglo-Saxon surnames, indicating that the pattern of long-term dynamics is not driven by international migrants but by the residential mobility of the longer settled population. Joining evidence of the local dynamics with the antecedent clustering of urban areas leads to the conclusion that, although urban areas do not show a uniform blend of people with diverse ancestry, they nevertheless exhibit high levels of local dynamics between inner city and suburban areas. Hence, the dominant trend determining population dynamics appears to be suburbanisation in the last 120 to 130 years. The only outliers are the Welsh cities, whose patterns may be interpreted in two ways. First, surname diversity is lower in Wales and therefore dynamics in terms of population diversity is limited although there may still be high levels of neighbourhood turnover. But since the local surname 'pool' is less diverse, the dynamics are not manifest in greater population diversity. Second, the results may reflect a lower level of suburbanisation in Wales than in other parts of Great Britain, perhaps due to industrial and population decline in the two harbour cities. This hypothesis may warrant further investigation; at this point, a lower level of population diversity in Wales seems certain.

### Discussion: specificity of urban environments

The clustering of the Anglo-Saxon population compared with the entire population suggests that higher scale differences between cities remain, although some inner urban areas and suburbs appear more distinct from their neighbours than from some neighbourhoods in other conurbations. These findings may usefully inform comparative research designs. If, for example, a city-level policy intervention was to be evaluated, a city in the same cluster without the intervention may serve as an appropriate control. But depending on the research problem, studies could also compare urban settings with very different populations, or similar neighbourhoods across different urban conurbations. Generically speaking, the geography of distinct urban areas suggests regions where policies adopted at higher levels may have differential repercussions and may translate into outcomes that need to be viewed in context with local conditions. Depending on the policy sector, comparisons may be cast with different hypotheses about cultural and biogenetic population differences.

Different urban dynamics can be observed when considering differential surname compositions in 1881 and 2007 of the two populations. Local dynamics in 1881 tend to occur in inner cities, while in 2007 dynamics in the cities' outskirts predominate. The spatial trends of population change are confirmed again when considering long-term dynamics of surname compositions. Changes in the outskirts over the last century are mainly driven by suburbanisation and also by new arrivals of international migrants further away from central neighbourhoods.

While isonymy indicates differences between urban populations, the exploration of urban dynamics across cities adds important information about specificity. Some cities – notably the two Welsh cities – in the UK have changed less relative to others and in some cities population compositions of central neighbourhoods have changed, while other cities witnessed stronger dynamics in outskirts. In earlier years, international migration focussed stronger on inner city areas, whereas more recently all areas have become destinations either directly or after a succession of moves over multiple generations of migrants. Yet, higher dynamics do not necessarily imply less population turnover, since, within a framing of isonymy, dynamics are defined in terms of surname diversity. Isonymy highlights, beyond mere population turnover, a specific form of dynamics, which is linked to higher population heterogeneity in biological and cultural terms and may be associated with different conditions for social cohesion, local and global identities and neighbourhood sense of belonging; all of which are relevant aspects of vulnerability.

## 4.5 Synthesis: scale, stability and significance of regional specificity

Surname geographies can act as approximate metrics of local specificity. At a broad temporal scale, the correspondence between today's genetic variants and 1881 surnames has remained stable in rural Britain. Where the geographies of surnames and genetic variants correspond, we may be more certain as to observing real differences in population structure. Yet, inasmuch as the spatial granularity increases, so too do the uncertainty of detecting fine population structure as well as the confounding influences of gender and migration.

The most likely regionalisation of population structure in Great Britain, given the joint evidence of both datasets, confirms pronounced differences between three of the UK countries and weaker regional differences that highlight northern England, south western England, northern and southern Wales. Finer scale regionalisations are less distinctive but are nevertheless important in some parts of the country, such as Devon and Cornwall. Where isonymy and co-ancestry align, more confidence can be placed on attributing phenomena (e.g. phenotypes) to common processes that are present in the entire sub-population. Where they do not align, there is less basis to ascertain the degree to which a phenomenon is present in the sub-population. This applies particularly to Scotland and southern Wales; here, it remains uncertain whether the genotypes found are characteristics of particular, sampled individuals, sampling sites, a locality or the wider region. This limits possibilities to assess the importance of phenomena for the population at large.

Almost certainly, however, the result of combining data on surname geographies and genes is in part a manifestation of the sample design. The degree to which a region is distinct and homogeneous is still affected by the number of observations that fall within a region and how many genotypes can be identified; both characteristics are in turn a result of the sampling. It should be noted, however, that the index of Distinctiveness penalises small numbers of observations to some extent. For instance, if the sample had been more skewed in England with the same observed genotypes, some English sub-regions with more observations would have increased in Distinctiveness. Conversely, regions with fewer observations would have lost Distinctiveness and thereby Regional Integrity. Since lower values result in mergers, and since more observations tend to produce higher values of Distinctiveness, regions with few observations tend to merge to larger entities, which is a desirable effect of the decision algorithm. Spatial heuristics of this kind may be of growing importance as unstructured samples proliferate, be it as 'Big Data' or DNA samples.

The emerging geographic scale that seems most appropriate for the representation of biological and cultural specificity are coarse regions that are finer than the level of the UK countries and broader scales than neighbourhoods and often individual cities. Yet, the question remains of how significant these geographies are for health pathways. Comparative research designs can provide some clues here, and the incorporation of surname regions in multi-level models may be one extension of encompassing comparisons. The regionalisation could also be used for the design of control and contrast samples, for example in epidemiological case-control or intervention studies, thus sharpening aetiological research. Nationally representative surveys alone do not guarantee a sound basis for making inferences about pathways that are affected by factors outside the purview of social difference. It may be contemplated that the sharpness of the surname regions increases when common surnames such as Brown or Smith are removed from the cluster analysis. This modification might refine the geographical variation of correspondence between genes and surnames and the resulting validity of surnames as genetic, cultural and linguistic indicators.

With Bourdieu's theory of social practice and Cavalli-Sforza's work on cultural transmission, it may be plausible – though more speculative – that distinct regional, cultural tendencies, such as dietary habits, lifestyles, governance arrangements, may covary with the geography of surnames. As objective structures manifest in tendencies of action, the isonymy regions may constitute a foundation for place effects that limit potential actions, pathways or outcomes to sets of regionally likely ones. To assess whether this foundation of population difference is significant at all for vulnerability of populations is explored in subsequent parts of the thesis.

In the context of urban areas, however, the emergent isonymy regions may also reflect greater population diversity rather than a common cultural phenomenon. Especially where neighbourhood types are shared across cities, the greater diversity necessitates a different approach to viewing local specificity, understanding the social pathways of health and designing policy interventions. Yet, although cities have changed due to international migration, the urban isonymy regions suggest that they largely maintain their regional distinctiveness over the period of more than a century. Convergence of local surname compositions in cities towards a uniform blend of cosmopolitan surnames cannot be confirmed.

The findings suggest that surnames could inform comparative urban research by helping decide on "comparative units" [cf Robinson 2011, 15]. First, cities or groups of cities could be so selected as to represent different types of urban populations. By this means, it could be investigated whether circulation of international or national urban policy regimes translates differentially into outcomes and produces potential unintended effects, if policies are designed with a certain set of privileged cities in mind. A careful selection of comparative units in urban studies may unmask heterogeneous pathways that have their starting point outside the spatial remit of the city. For example, empirically informed comparative strategy may consider Cardiff, Glasgow and London as the largest, most different conurbations in the UK – rather than Birmingham, Leeds and London – or Cardiff, Edinburgh, Liverpool and Sheffield as the most different UK cities of similar size. Geodemographics could be used as additional controls of social similarity to decide on comparative units in urban research at multiple scales.

Second, the similarity between neighbourhoods across different cities may inform decisions of scale. Basing comparison on, for example, Central Birmingham, North West

London and some selected suburbs across England may be a useful basis to investigate phenomena, meanings and causes in view of the diversity of cities. The assignment of inner city neighbourhoods into one cluster is indicative of "topological spatialities" [Robinson 2011, 16] whereby cities, affected by the same circulations, are connected to each other. Robinson argues that assumptions about the ways in which "cities inhabit each other" is not only a subject of interest in comparative projects but should also inform the comparative design itself as to which cities are to be considered and at what scale. At an international level, surname data may support the identification of comparative units below the city scale, too, for example to delineate the impact of global mobility on inner cities or to characterise specific dimensions of globally induced neighbourhood dynamics and gentrification. If surnames are classified more broadly by their geographic origin akin to a refined measure of ethnicity, neighbourhoods may be ranked by degree of ethnic diversity or cosmopolitanism in a global hierarchy of urban areas, facilitating research of global circularities in international urban comparisons. Similarly, local surname compositions may also support the definition to control and contrast populations in cross-national urban studies.

Further research may include a dynamic perspective of specificity: with surnames classified by their geographic origin, it is possible to trace long-term migration movements in cities and view diversity beyond ethnicity in longer, historical trajectories. This aim, however, would necessitate to formalise the information a foreign surname may bear with respect to the likely time period of migration.

For health research and vulnerability studies, the findings imply that quantitative geography should apply methods to highlight difference and disagreement rather than those that focus on summary and law-like explanation. Difference-seeking techniques, with a focus of exploration and description as advocated by Byrne [2002] supports a more comprehensive and open research programme with reflexive and unstable theory-building, which Robinson [2011, 2014] also calls for.

As reviewed earlier, geodemographic classification represents such a difference-seeking technique. For an extended geodemographic system, the findings from this study imply, in brief, that a combination of an encompassing comparative research design and geodemographics could be a first move towards incorporating regional specificity in geodemographic studies. The surname geographies of cities may inform the selection of comparative regions, in which specific geodemographic classifications can be nested to characterise local pathways. The regionally nested classifications can then be formally compared against a nation-wide classification. The methods of partition consonance tested earlier can assess regional correspondence and highlight, where, whether and to which extent local specificities as approximated through surnames are at all relevant for the study of vulnerability.

# 5 Regional and urban health inequalities: a simple geodemographic classification

The previously presented regional geography of population structure and – perhaps – culture sets a high level context for further investigation of health. Hospital Episode Statistics (HES) for England and the UK Census are rich data resources, which collect extensive information on health for almost the entire population. Using the detailed data and organising it spatially at different ecological levels may reveal different types of geographically varying health advantage. A simple classification developed here contributes to existing approaches to small area classification by addressing the need to incorporate explicit spatial context and to compare across places.

## 5.1 Approaches to measuring health

There are a few studies that classify areas by health outcomes in the United Kingdom. Shelton et al [2006] and Green et al [2014] each conduct a study that develops a geodemographic classification using health data. Based on Office for National Statistics mortality data, the studies identify areas with different health challenges. Health is conceptualised in negative terms, as risk or presence of mortality due to a particular cause.

Shelton et al's investigation focusses on temporal changes of mortality in 76 parliamentary councils in England and Wales between 1981 and 2000. They identify ten clusters of different mortality profiles, which represent different types of geographical areas (London, urban England, southeast, coastal areas and former mining areas). The study by Green et al uses mortality data, too, but only for one point in time. They pool all 1.1 million death cases occurring between 2006 and 2009 and aggregate them to MSOA level (n=7,194). They identify eight groups of areas with different mortality profiles, which broadly differ in terms of the general level of mortality ("best health", "average profiles", "the middle", "mixed experiences" to name a few) with, unlike the study by Shelton et al [2006], a dispersed geography across England and Wales. In their unique ways, both studies discover a discrete ecology of mortality profiles and conclude that areas with increased mortality risk require appropriate policy focus, notably in terms of resource allocation and service provision.

Incorporating health outcomes into geodemographics implies that measuring health at an aggregate level is meaningful. The meaning can be viewed from two perspectives. Health, on the one hand, can be conceived of in terms of deficit, as the "absence of disease and infirmity" [WHO 1948]. On the other hand, health may be defined in terms of assets, whereby it implies a "state of complete physical, mental and social well-being" [ibid.]. Moon [2009] shows how the different views entail different approaches in studying the geography of health. Apart from a broader theoretical orientation towards either biomedical or social processes, each notion also places different weights on different strategies to measure, estimate and model health. Studies that adopt the deficit notion, as most conventional epidemiological studies do, focus on health measured as mortality rates, life expectancy or healthy life expectancy, disease prevalence or incidence and risk factors. The asset view emphasises more subjective information on self-rated health, physical and mental well-being, life satisfaction and lifestyles [cf Baggott 2004; Scambler 2007].

Theoretical frameworks to measure and contextualise health differ according to whether health is conceived of in asset or deficit terms. The eco-social theory of health postulates that health is an outcome of embodied social relations, contingent on individual biology. Dahlgren and Whitehead [1992] identify a causal chain that begins with general socio-economic conditions, which in turn define general living and working conditions. These result in social and community influences, within which an individual leads a lifestyle that has implications for health. This framework has been influential in informing socio-epidemiological studies: health is often studied as dependent outcome of these influences.

A more complex framework has been presented by Solar and Irwin [2007]. Health outcomes result from exposure to risks and vulnerability, both of which result from social, cultural and economic processes that put individuals with different characteristics into different contexts of risk. The nature of social determinants, mode of stratification and exposures are shaped by globalisation and socio-political processes affecting religious beliefs and practices, human rights, the workings of labour markets and educational systems. Access to the health care system modifies the relationship between vulnerability, exposure and health.

Evans and Stoddart's [1990] "health field" framework places the individual in the social and physical environment, who, given individual genetic endowment, responds in terms of health-relevant behaviour and biology. Health, bodily function and disease onset result from the duality of behaviour and biology, leading to treatment within the health care system, which feeds back to the individual, and general well-being.

Similar approaches emphasising a deficit notion of health are the fundamental cause model [Link & Phelan 1995; Phelan et al. 2004] and the stress-process model [Turner 2013]. In all these models exposure to risks triggers negative consequences for health, while exposure is defined by broader social, economic and political conditions. These conditions lie beyond the control of the individual; health nearly becomes a deterministic outcome of passive, receiving individuals with differential coping capabilities. Health is therefore the inverse of disease presence and physical or mental dysfunction.

An asset view of health, on the other hand, emphasises the distribution, access to and use of resources that are required to lead a healthy life. Scambler [2007, 2012] shows in his model how logics of capital accumulation and regulation in industrial societies produces a culture of aspirational consumerism that determines the flow of health-relevant biological, psychological, social, cultural and spatial assets. It is the distribution of these assets within society rather than negative influences in terms of exposures that enable certain individuals to live more healthily than others.

Frohlich and Abel [2014] draw on Bourdieu's work among others and view individual health in a context of biological, economic, cultural and social capital that can be accessed to ensure and increase health. These capitals partly arise from local access to community resources, which are shaped by the wider institutional environment, local sociability, the physical and the economic environment. They emphasise that local constellations of community and individual assets change over time, and are significantly shaped by individuals with unequal power, which in turn shape individuals.

Finally, Blaxter [2003] develops the notion of "health capital", by which she seeks to synthesise social inequalities, class, lifestyles and individual biology. Along with economic, cultural, social and symbolic capital, health capital can be increased or depleted over the life course. Individuals act within the scope of possibilities as they emerge within conscious and unconscious notions of health, biographical events and everyday life routines.

Asset models of health pose individual agency at the centre and focus on the socio-structural enablers and disablers of healthy practices. Deficit models begin with exposure to risk and tend to conceptualise the individual as passive recipient of externally defined influences. Whereas they emphasise social structure, associated determinants of health and absence of health, asset approaches give equal weight to agency and structure and focus on presence of health.

Both approaches to measuring health are incomplete. Although asset indicators capture presence of health, they do so in an often subjective and non-clinical way. There are limits to classifying health beyond physical or mental – a dichotomy, which in itself can be challenged [Brunner & Marmot 2006]. On the other hand, deficit indicators, such as disease diagnoses, offer a nearly unlimited possibility to classify health risk, ranging from more generic groupings, such as infectious and chronic illnesses and injuries, to individual diseases, such as tuberculosis, diabetes, asthma and so forth. Moreover, these indicators are often clinical data and therefore not based on individual judgement other than that of the professional. The downside is that deficit indicators do not provide direct information about the healthy population. Hence, both aspects are necessary to obtain a comprehensive view of health.

In grouping observations spatially, the ecological measurement of health reflects both aspects, too. In terms of assets, health of an area population relative to other area populations reflects differential access to resources that are relevant to health. An extended understanding of Blaxter's [2003] concept of health capital appears particularly useful here; it may be said that access to all health-relevant elements of economic, cultural, social and biological capital (in short health capital) produces the health outcomes that may be typical for an area given its assets. In terms of deficit, area population health reflects differential exposure to risk. Implicitly, health capital mediates the impact of exposure by defining local capabilities to cope with the exposure. Both access to assets and exposure to risk constitute vulnerability and they may or may not be connected.

It follows that geographical health disparities are rooted in societal processes that determine the spatial distribution of health-relevant assets and risks. Health geographies are thus expressions of these societal processes, associated social structures and relations of power. This view is consistent with the sociological approaches to study health and illness [Coburn 2012; Scambler 2007, 2012; Williams 1995] and politico-ecological approaches to understanding risk [Beck 1992; Hilhorst & Bankoff 2004]. In addition to global processes, health disparities reflect local conditions and population characteristics in shaping assets and risks that affect the entire area population beyond the individual. Interactions between societal processes and local conditions, including coping capabilities of populations, produce specific spatial phenomena. Area classification, then, provides a hermeneutic to study the workings of a social system in conjunction with local conditions through ecological description.

## 5.2 Health estimation and spatial structure in population datasets

The two connotations of area health – deficit and asset – imply criteria for the selection of data sources to measure health. First, the data need to capture both the asset and the deficit view for a large part of the population. Second, the information should be available at appropriate granularity in terms of geography. Ideally, neighbourhood-level information should be provided, which may be aggregated to coarser levels to match the analytical focus and also attenuate purely statistical effects, such as greater volatility typically observed for small areas. Third, a sufficient temporal resolution is required to enable sensitivity to temporal dynamics while at the same time avoiding irrelevant volatility of phenomena. No single dataset fulfils these criteria; several relevant datasets need to be combined.

### Data sources

Hospital Episode Statistics for England (HES) record each inpatient hospital admission along with patient demographics, area of residence, referral organisation, clinical data on diagnosis and treatment and the health care organisation itself. They thus provide data pertaining to the deficit view of health for the entire population. Health conditions are classified according to the 10th revision of the International Classification of Diseases (ICD-10) [WHO 2011], which is an international coding of health conditions into approximately 15,000 conditions. These may be grouped into three broad categories and 22 sub-categories [Table 5.1].

For this study, the HES dataset for the years 2008/09 is available. HES covers all of England, although some patients may reside close to the border in either Wales or

Table 5.1: The WHO Global Burden of Disease cause groups based on the International Classification of Diseases, 10th revision (ICD-10) [WHO 2004].

| level 1 | title | ICD-10 code groups |
|---|---|---|
| I | Communicable, maternal, perinatal and nutritional conditions | see below |
| II | Non-communicable diseases | see below |
| III | Injuries | see below |
| I | Infectious and parasitic diseases | A,B,G,N |
| I | Respiratory infections | B,H,J,U |
| I | Maternal conditions | O |
| I | Conditions arising during the perinatal period | P |
| I | Nutritional deficiencies | D,E |
| II | Malignant neoplasms | C |
| II | Other neoplasms | D |
| II | Diabetes mellitus | E |
| II | Endocrine disorders | D,E |
| II | Neuro-psychiatric conditions | F,G |
| II | Sense organ diseases | H |
| II | Cardiovascular diseases | I |
| II | Respiratory diseases | J |
| II | Digestive diseases | K |
| II | Genito-urinary diseases | K,N |
| II | Skin diseases | L |
| II | Musculoskeletal diseases | M |
| II | Congenital anomalies | Q |
| II | Oral conditions | K |
| III | Injuries | S,T,V-Y |

Scotland. After exclusion of records with unknown sex, Scottish or Welsh residence, a total of 11,614,937 remained for analysis. Along with patient sex, age and resident LSOA, age and sex-standardised morbidity rate ratios (SMbRRs) for each of the 22 conditions can be estimated for English LSOAs or higher geographies. SMbRRs measure the incidence risk of a given disease for a group or area, and express the risk as ratio relative to the average risk in the general population. The risk ratio is adjusted by the area's populations demographic structure in order to account for the rising risk of mortality with increased age. A risk ratio of one indicates that an area shows exactly the average risk of the population, an area risk ratio of two indicates the risk of disease incidence in the area is double that of the total population, and 0.5 indicates the risk is half.

As for the asset view, the 2011 UK Census constitutes an adequate population-wide data source. The Census asks respondents about their health on a five-point Likert scale, with a value of one indicating "excellent" and five indicating "poor". The data can be summarised in a similar way as disease data. The top two categories "excellent" and "very good" can be grouped as indicating good health, age- and sex standardised and summarised in terms of standardised health ratios (SHRs). SHRs provide a similar metric to SMbRRs with the same distributional properties.

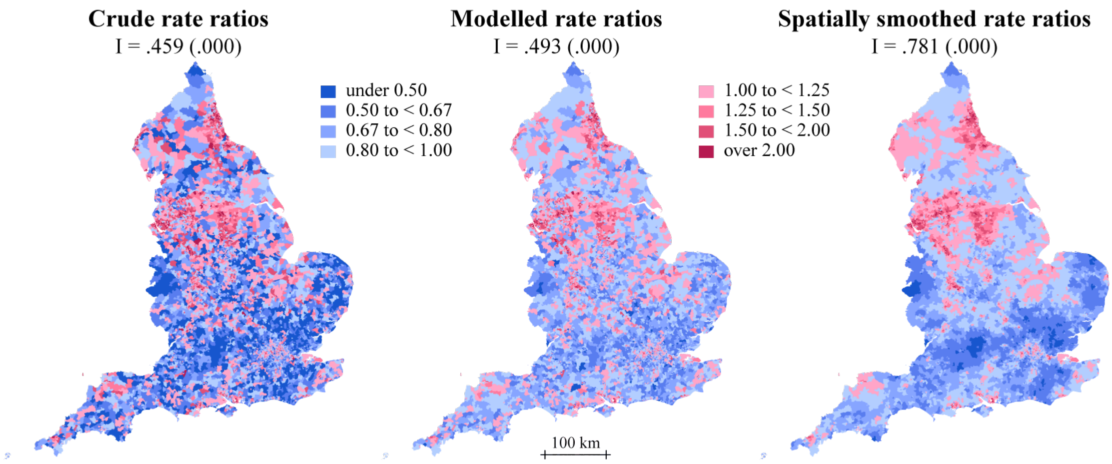



Figure 5.1: Age and sex-standardised incidence rate ratios for infectious and parasitic diseases in England's wards, including their Moran's I of spatial autocorrelation.

### Statistical framework to estimate health

Small area estimates of health, like all spatial data, are subject to spatial autocorrelation, the phenomenon by which spatial measurements are more similar in value the more proximate they are to each other in space. Spatial autocorrelation can be used as a source of information in a context of data sparseness. Datasets that allow a large number of categorisations, such as disease groups in HES, often show a low number of observations of a given condition in a given area per year, and hence estimates are subject to annual fluctuation and high level of statistical uncertainty. Considering the value of neighbouring areas, however, can be a way to reduce uncertainty and volatility of estimates. Given the principle of spatial autocorrelation, estimates of high risk areas that are surrounded by high risk areas can be assumed to be more certain and stable than high risk areas that are surrounded by low risk areas. The same applies vice versa. This strategy may be referred to as "borrowing strength" [Best et al. 2005; Elliott & Wakefield 2001] from proximate observations, whereby proximate may be defined in spatial or temporal terms.

In the context of health geography, borrowing strength provides the double advantage of producing more robust estimates of health given proximate evidence and, at the same time, reflecting the spatial context through being sensitive to the value of neighbouring areas. Besag et al [1991] propose a spatial-structural model which allows the borrowing of strength from adjacent areas. This model, informally known as the "BYM" model, is implemented under a Bayesian statistical framework and results in *spatial smoothing* of area SMbRRs [see example in Figure 5.1]. Crude estimates are simply the direct ratio between observed and expected number of cases of a given disease, which tend to be statistically uncertain in a situation of sparse data. A universal model may estimate SMbRRs based on the Poisson distribution with the result that outliers shrink towards the mean. The spatial-structural (BYM) model takes into account the value of neighbouring areas, which results in spatially smoothed estimates of SMbRRs and typically lower statistical uncertainty. The improvement of spatial smoothing can be assessed by model fit criteria, in this case the Deviance Information Criterion (DIC). If the DIC decreases by a value of three compared to the previous models, statistical uncertainty is reduced significantly and the model is to be preferred [Spiegelhalter et al. 2002].

### Technical specification: spatial-structural model

The spatial-structural model proposed by Besag et al [1991] estimates, on the basis of the Poisson distribution, the number of cases $Y_i$ that have a specified condition in an area $i$ compared to expected cases $e_i$ multiplied by the relative risk ratio (SMbRR) $\theta_i$.

$$\begin{aligned} Y_i &\sim \mathcal{P}(e_i, \theta_i) \\ \log \theta_i &= \mu_i + v_i + u_i + \mathbf{x_i}\beta \end{aligned} \tag{5.1}$$

The relative risk ratio of an area $i$ is modelled as the logarithm of the sum of the intercept $\mu_i$, a spatial component $u_i$ and a random error component $v_i$. In addition, it is possible to add area covariates $x_i$, such as deprivation, in order to model relative risk. In the BYM model, the random error component is normally distributed with mean zero and precision $\tau_v$. The precision, the inverse of the variance, is defined to be gamma distributed. The gamma distribution becomes thus the prior distribution for the unstructured random component, which is viewed jointly with the evidence from the data. The spatial-structural component is distributed according to:

$$u_i|u_j, \quad i \neq j, \quad \tau_u \sim \mathcal{N}\left(\frac{1}{n_i}\sum_{i\sim j} u_j, \frac{1}{n_i \tau_u}\right) \tag{5.2}$$

where $u_i$, the additive risk ratio of area $i$, is normally distributed around the weighted average of risk ratios of adjacent areas $j$ with the precision $\tau_u$. Adjacency is noted as $i \sim j$. $\tau_u$ is gamma distributed, the prior for the structured component. The gamma distributions for both $\tau_u$ and $\tau_v$ are defined by the hyper-parameters $\alpha$ and $\beta$ defining the shape of the gamma distributions.

$$\tau_u, \tau_v \sim \Gamma(\alpha, \beta) \tag{5.3}$$

The specification can be different for each $\tau_u$ and $\tau_v$, but often a vague distribution with a low precision is chosen ($\alpha$ = .0005, $\beta$ = .0005), which gives more weight to the evidence from the data. Other researchers have chosen the same specifications in spatial estimation of health [Chang et al. 2011; Cheung et al. 2012; Congdon 2012, 2013]; yet it should be noted that the choice of prior distribution is itself subject to debate, given its impact on the posterior distribution.

The R-package 'INLA', which is an implementation of Integrated Nested Laplace Approximation (INLA), is used. In earlier applications of Bayesian statistics, it was common to use Gibbs samplers on Monte Carlo Markov Chains (MCMCs) to derive posterior estimates of parameters and their statistical dispersion. Laplace approximation is a technique that determines the posterior distribution through deterministic integration of distributions that are part of the Gaussian family [Rue et al. 2009]. This technique can hence be used for most modelling problems; furthermore, they expend less computation time than sampling from MCMCs.

#### Technical specification: area classification

Smoothed estimates of self-rated health prevalence and disease incidence are classified using a two-stage cluster procedure of combined Ward hierarchical clustering and k means clustering. Ward clustering is an agglomerative procedure of clustering, which processes dissimilarities between observations. Statistical packages convert a matrix of $n$ observations and $p$ variables to a $n \times n$ matrix of pair-wise Euclidean distances in multi-variate statistical space. These dissimilarities produce a hierarchical taxonomy of observations, in this case areas. The taxonomy can then be split into groups and the statistical group centres in statistical space can be calculated.

The resulting partition may not always produce optimal solutions because the formation of the taxonomy is iterative and does not make adjustments to prior groupings. More robust partitions can therefore by achieved when the cluster centres are used as initialisations for the *k means* algorithm, which iteratively adjusts cluster centres and cluster assignments. The multiple potential cluster solutions with varying number of clusters are then assessed based on basic goodness of fit test, specifically the F ratio – the proportion of between-cluster sum of squares divided by within cluster sum of squares. In this way a statistically optimal number of clusters can be determined. The statistical software R [R Core Team 2014] is used to cluster the observations.

## 5.3 Regional patterning of health

Estimating SHRs of self-rated health and SMbRRs of health conditions based on spatial-structural models leads to 23 model outputs that each contain posterior estimates and credible (confidence) intervals for local areas. These estimates and their associated uncertainty provide the basis for a geographical study of health in England. These are summarised in three ways as follows. First, general measures of ecolog-

Table 5.2: Health inequalities in England per health condition.

| condition | ASR | ARD | QR | I | p(I) |
|---|---|---|---|---|---|
| **asset** | | | | | |
| Self-rated health (SHR) | 81,391 | 13,730 | 1.18 | .668 | .000 |
| **deficit** | | | | | |
| Infectious and parasitic dis's | 336 | 349 | 2.89 | .781 | .000 |
| Respiratory infections | 792 | 767 | 2.62 | .677 | .000 |
| Maternal conditions | 2,258 | 2,000 | 2.49 | .523 | .000 |
| Cond's from perinatal period | 161 | 563 | 16.61 | .275 | .000 |
| Nutritional deficiencies | 196 | 199 | 2.85 | .797 | .000 |
| Malignant neoplasms | 950 | 475 | 1.65 | .594 | .000 |
| Other neoplasms | 450 | 367 | 2.21 | .843 | .000 |
| Diabetes mellitus | 93 | 112 | 3.22 | .783 | .000 |
| Endocrine disorders | 251 | 180 | 2.07 | .812 | .000 |
| Neuro-psychiatric conditions | 711 | 575 | 2.30 | .569 | .000 |
| Sense organ diseases | 923 | 645 | 2.03 | .748 | .000 |
| Cardiovascular diseases | 1,648 | 1,079 | 1.94 | .572 | .000 |
| Respiratory diseases | 627 | 591 | 2.62 | .636 | .000 |
| Digestive diseases | 2,380 | 1,797 | 2.21 | .651 | .000 |
| Genito-urinary diseases | 1,459 | 950 | 1.94 | .642 | .000 |
| Skin diseases | 500 | 451 | 2.48 | .808 | .000 |
| Musculoskeletal diseases | 1,765 | 1,347 | 2.13 | .652 | .000 |
| Congenital anomalies | 143 | 90 | 1.87 | .637 | .000 |
| Oral conditions | 447 | 630 | 4.35 | .817 | .000 |
| Injuries | 1,590 | 1,217 | 2.14 | .569 | .000 |
| Not elsewhere classified | 2,904 | 2,859 | 2.75 | .718 | .000 |
| Health-service related | 1,290 | 1,516 | 3.28 | .708 | .000 |

**columns:** ASR = age and sex-standardised rate per 100,000; ARD = absolute rate difference; QR = quantile ratio; I = Moran's I of ward standardised rate ratios; p(i) p-value of Moran's I.

ical and spatial health inequalities among England's wards are calculated. Second, Principal Components Analysis (PCA) is used to identify health conditions that frequently occur jointly across England's wards. Finally, for each level of investigation, an area classification of health is developed and interpreted with respect to the evidence arising in all three steps.

### Health inequalities among England's wards

Most of the 23 health indicators reveal a high degree of health inequalities and spatial clustering among the spatially smoothed estimates for wards [Table 5.2]. The absolute standardised incidene rate (ASR) measures the number of cases per 100,000 people and is standardised by the demographic structure (age, sex) of the population. The absolute risk difference (ARD) measures health inequalities in absolute terms; it is the difference between the ASR of the 95th and the fifth percentile of wards ordered by their ASR. The ARDs vary considerably and are broadly related to the level of incidence rates of conditions; excluding self-rated health, they are highest for unclassified diseases, maternal conditions and digestive conditions and lowest for diabetes, congenital anomalies and endocrine disorders. The strongest absolute health inequalities are recorded for conditions arising during the perinatal period, a rarer group of diseases with only 161 cases per 100,000 people.

The quantile ratio (QR) is the ratio of the 95th compared to the fifth percentile of values and indicates the magnitude of risk dispersion among wards. Since it is a ratio, the QR is comparable across conditions with different magnitudes of ASRs, although rarer conditions tend to have inflated ratios. The QR can be understood as a relative ecological measure of health inequalities which measures the ratio of risk between a typical high risk area and a typical low risk area. The majority of quantile ratios lie between two and four, indicating different condition-wise between-ward disparities in health are two to four-fold.

The Moran's I measures the degree of spatial autocorrelation on a scale from -1 (dispersed) and +1 (clustered). It measures the degree of spatial concentration and thus constitutes an indicator of spatial health disparities . The Moran's I ranges between .55 and .80 for most conditions. The strongest spatial disparities in health, as approximated by Moran's I, can be found for other neoplasms, oral conditions and endocrine disorders.

Bayesian spatial-strutcural models provide so-called credible intervals of the local estimates, which can be compared to 95 per cent confidence intervals. These intervals are derived from the posterior marginals of local estimates; they can also be used to determine the probability of a local estimate exceeding any given value, for example one, the national average.

The SHRs vary primarily between rural and urban wards; the likelihood that people rate their health better is higher in rural areas than in urban areas [Figure 5.2]. A similar pattern can be observed for some health conditions, notably infectious and parasitic diseases, respiratory infections and nutritional deficiencies. In the northern cities, ranging from Liverpool to Leeds as well as the region around Newcastle, there appears to be increased risk of infectious and parasitic diseases. Further hotspots include east London, Southampton and Somerset. The value of Moran's I is high at .776.

The spatial pattern for malignant neoplasms (cancer) is less clustered compared to other conditions. Hot spots are in East and South London, in Norfolk and Lincolnshire, Devon, Hampshire as well as in the far North of England. Cancer is common throughout England, and relative health inequalities are lower with a QR of 1.65. Skin diseases, with a Moran's I of .761, are concentrated in many urban centres and along the coasts, particularly in the north west, East Riding, and along the south west coast.

Not elsewhere classified and health-service related conditions not only show higher prevalence but also a distinct spatial patterns. These special ICD-10 categories, which are

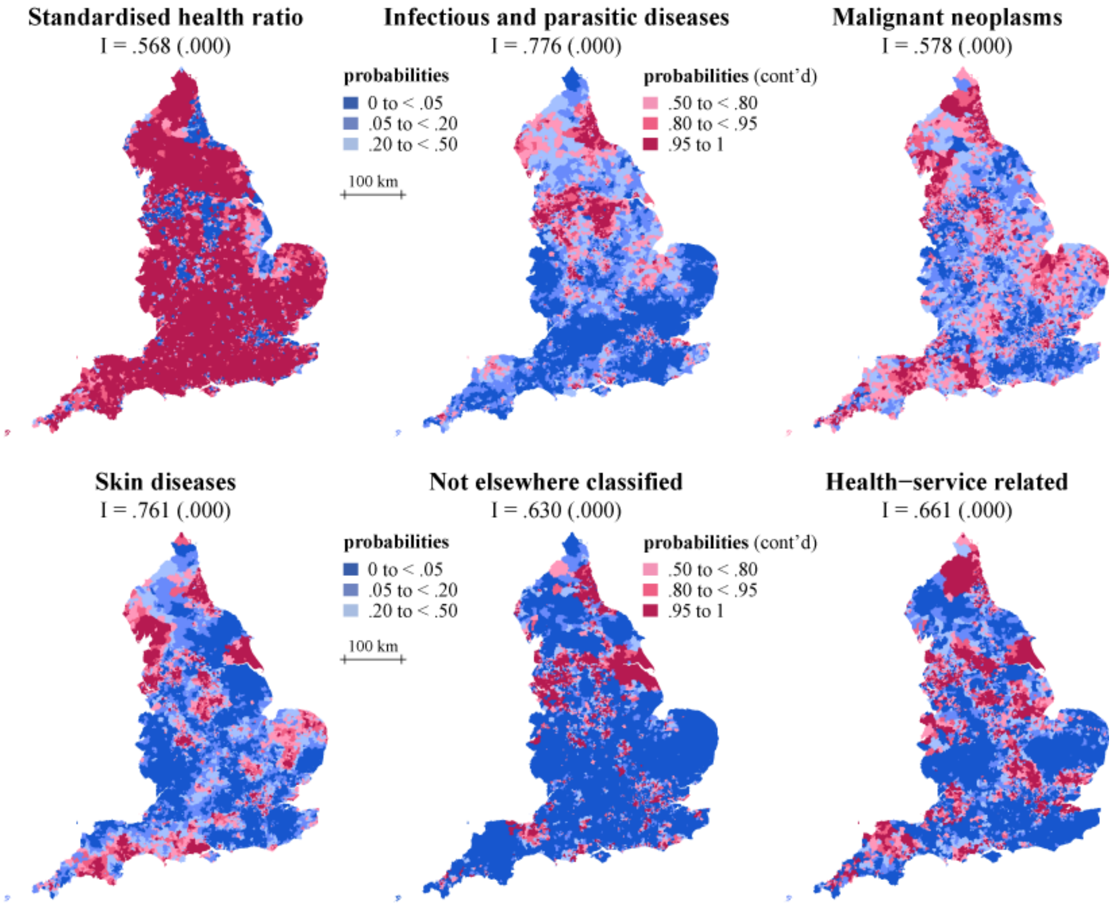



Figure 5.2: The probability of smoothed incidence rates for self-rated health and selected conditions being above average in England's wards including their Moran's I of spatial autocorrelation. Note that they may differ slightly from the spatial autocorrelation of ASRs shown in Table 5.2.

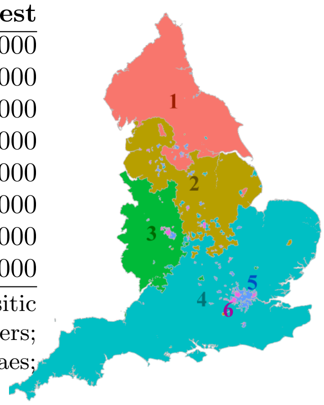
not included in the Global Burden of Disease typology of the WHO, are more revealing about the health services than direct risks of morbidity. Unclassified disease cluster in a corridor that stretches from Merseyside in the west to the river Humber in the east crossing the urban regions of Manchester and Leeds. Other clusters can be found in the areas of Newcastle and Somerset. This may either indicate that the incidence of unclassified conditions is higher in these regions or classification practices differ from other regions in England. As for conditions arising from contact with health services, the northern corridor connecting west and east coasts emerges, as well as clusters in Northumberland, Midlands towards London and Devon. The health disparities associated with the latter are significant and point towards differential quality and operation of health services.

The geographical distribution of some conditions overlap, others are more distinct in the total population. Separate estimates for women and men do not reveal significant

Table 5.3: Selected health conditions by surname cluster. Stronger patterns are underlined.

| | 1 | 2 | 3 | 4 | 5 | 6 | F test |
|---|---|---|---|---|---|---|---|
| S.H.R. | 0.99 | 0.99 | 1.01 | 1.03 | 0.98 | 0.97 | .000 |
| Inf.P.D. | 1.18 | 1.09 | 0.96 | 0.76 | 1.02 | 1.13 | .000 |
| Mal.N. | 1.08 | 1.02 | 0.94 | 0.97 | 0.96 | 0.87 | .000 |
| End.D. | 0.93 | 1.05 | 0.87 | 0.84 | 1.19 | 1.15 | .000 |
| N.Psych. | 0.89 | 1.04 | 0.96 | 0.87 | 1.05 | 1.08 | .000 |
| S.O.D. | 1.00 | 1.03 | 0.86 | 0.91 | 1.07 | 1.16 | .000 |
| C.V.D. | 1.02 | 1.03 | 0.91 | 0.89 | 1.06 | 1.17 | .000 |
| M.S.D | 1.03 | 1.06 | 1.02 | 0.95 | 0.83 | 0.89 | .000 |

S.H.R. = standardised health ratio; Inf.P.D. = infectious and parasitic diseases; Mal.N. = malicious neoplasms; End.D. = endocrine disorders; N.Psych = neuro-psychological conditions; S.O.D. = sense organ diseaes; C.V.D. = cardio-vascular diseases; M.S.D. = musculosekelatal diseases.




divergence between sexes (except for maternal conditions, which only apply to women) [not illustrated]. Only genito-urinary diseases reveal subtle differences: in addition to northern and southern urban and town centres, male cases are concentrated around the Wash and in Dorset, female cases in Somerset and East Riding. As for the rest of conditions, there is strong geographical congruence between female and male cases.

Regional patterns can also be observed for parts of England divided by surname compositions. By applying the same method as in chapter 4, England can be divided into six sub-regions based isonymy derived from the 2011 Enhanced Electoral Roll [Table 5.3]. Four large, contiguous regions and two smaller, dispersed regions emerge: one large each in north, central, west and south England, with the latter extending across the entire range from Cornwall to East Anglia, and two urban regions dividing London into an inner and an outer region and Birmingham into a southern and a northern part. The cluster of inner London and southern Birmingham also includes the northern English urban centres.

Self-rated health and disease incidence differ across the six regions, in parts significantly according to oneway ANOVA tests of standardised logarithmically transformed incidence rates. Self-rated health tends to be poorer in the two urban regions and highest in the large southern surname region. Similarly, infectious and parasitic diseases are significantly higher in the northern region and the outer urban areas; they occur less often in the southern region. There is a higher risk of endocrine disorders in the two urban surname clusters; the risk is lower in the western and southern region. A similar pattern can be observed for neuro-psychiatric conditions.

Conversely, malignant neoplasms occur less often in the urban clusters, least often in outer urban areas, while they are more common in the north. Incidence rates are also lower in the western region. A similar pattern can be observed for other neoplasms, suggesting that risk of cancer is higher in the rural north. Musculoskeletal diseases occur less often in inner urban areas and most often in the central English region.

Table 5.4: PCA solution for English wards.

| condition | PC 1 | 2 | 3 | 4 | 5 | 6 | u'ness |
|---|---|---|---|---|---|---|---|
| **asset** | | | | | | | |
| Self-rated health (SHR) | -.682 | -.130 | -.193 | -.285 | -.118 | -.252 | .322 |
| **deficit** | | | | | | | |
| Infectious and parasitic dis's | .430 | .163 | .214 | .271 | .246 | .569 | .286 |
| Respiratory infections | .450 | .102 | .423 | .293 | .177 | .589 | .144 |
| Maternal conditions | .158 | .114 | .335 | .276 | .022 | .093 | .764 |
| Cond's from perinatal period | .216 | -.111 | -.041 | .680 | .064 | .144 | .452 |
| Nutritional deficiencies | .328 | .281 | .084 | .163 | .613 | .181 | .371 |
| Malignant neoplasms | .175 | .606 | .192 | -.004 | .115 | .080 | .546 |
| Other neoplasms | .029 | .851 | .096 | .041 | .171 | .074 | .229 |
| Diabetes mellitus | .472 | .281 | .089 | .166 | .101 | .071 | .648 |
| Endocrine disorders | .595 | .132 | .000 | .314 | .211 | .044 | .484 |
| Neuro-psychiatric conditions | .629 | .180 | .269 | .084 | .080 | .176 | .455 |
| Sense organ diseases | .496 | .407 | .141 | .186 | .065 | -.030 | .529 |
| Cardiovascular diseases | .610 | .218 | .381 | .210 | .271 | .115 | .304 |
| Respiratory diseases | .717 | .135 | .320 | .227 | .131 | .228 | .245 |
| Digestive diseases | .414 | .261 | .385 | .209 | .516 | .104 | .292 |
| Genito-urinary diseases | .425 | .285 | .501 | .207 | .087 | .060 | .433 |
| Skin diseases | .346 | .667 | .264 | .095 | -.016 | .144 | .336 |
| Musculoskeletal diseases | .139 | .338 | .605 | .051 | .108 | .085 | .479 |
| Congenital anomalies | .216 | .162 | .238 | .512 | .001 | .214 | .562 |
| Oral conditions | .210 | .278 | .118 | .115 | .108 | -.096 | .831 |
| Injuries | .445 | .182 | .500 | .047 | .153 | .232 | .439 |
| Not elsewhere classified | .590 | .128 | .334 | .190 | .333 | .134 | .358 |
| Health-service related | .174 | .201 | .211 | .632 | .237 | -.084 | .422 |
| Sum of squares loadings | 4.326 | 2.511 | 2.088 | 1.882 | 1.173 | 1.089 | |
| Proportion of variance | .188 | .109 | .091 | .082 | .051 | .047 | |
| Cumulative variance | .188 | .297 | .388 | .470 | .521 | .568 | |

**interpretation:** poor health capital (1); cancer & outer organs (2); injuries & organ damage (3); birth & service-related (4); nutrition-related (5); infections (6); u'ness = variate uniqueness

In this general view, the patterns suggest a north-south and urban-rural divide rather than difference associated with biological characteristics. Further refinement of surname clusters into eleven regions confirms this, although some conditions such as infections, endocrine disorders and diseases related to sense organs show more variation that may warrant further investigation. At this point, there is not much evidence of different health risks by isonymy group, although there is certainly confounding by broader regional, social and geographical context.

### Dimensions of population health

Which conditions frequently occur together in England's wards? Principal Components Analysis (PCA) applied on the smoothed ecological estimates reveals six components with an eigenvalue larger than one accounting for 56 per cent of the variance [Table

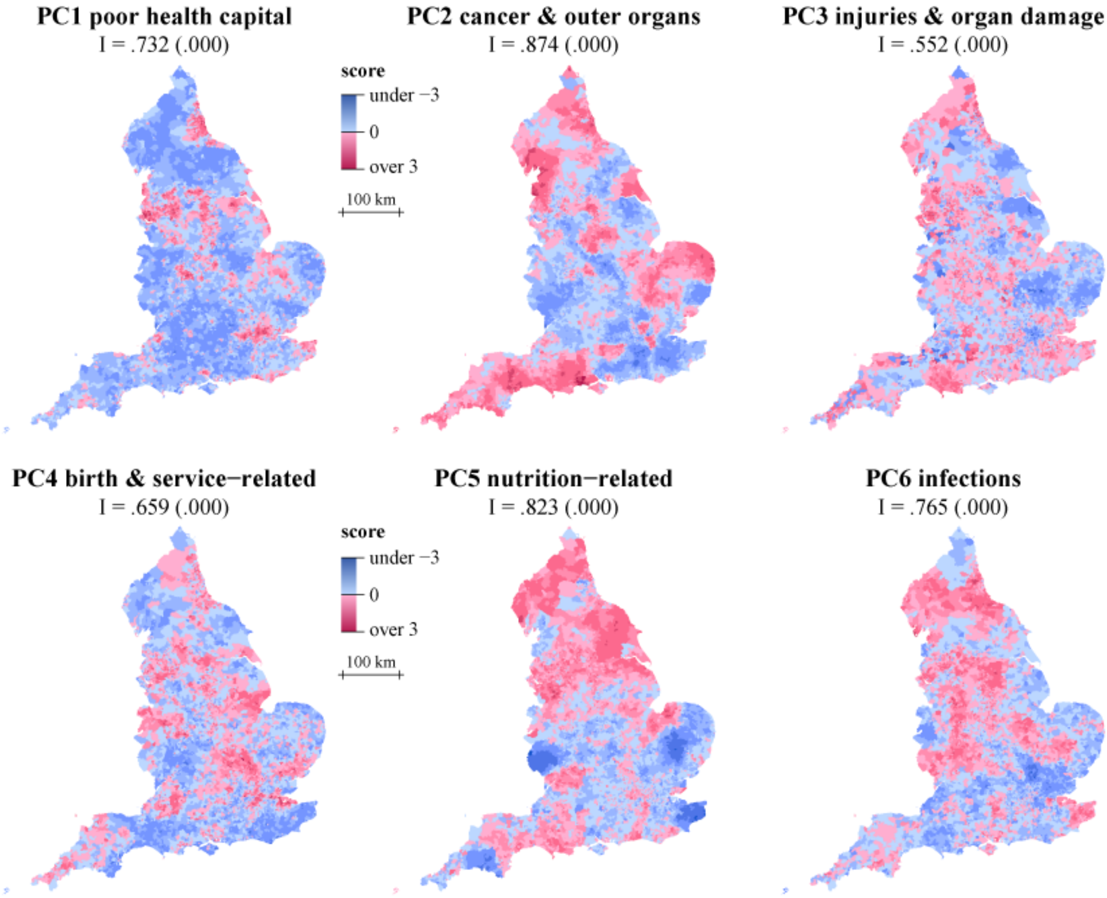



Figure 5.3: Principal Component scores in England's wards.

5.4]. Mapping the scores of each component produces distinct geographies that indicate a high degree of spatial clustering [Figure 5.3]. The largest component with 19 per cent reflects spatial co-incidence of non-cancerous chronic illnesses and low levels of good self-rated health and indicates that respiratory diseases, endocrine disorders, cardiovascular diseases, neuro-psychiatric conditions as well as unclassified conditions co-incide in English wards. In total, it may be interpreted as reflecting poor health capital of local populations. The spatial distribution of the scores suggests that the component is predominantly urban: wards with poor health capital often occur in urban centres across England. A rural exception to this is an area south of the river Humber.

The second component comprises cancers and impairment of outer organs with high loadings on malignant and other neoplasms and skin diseases. It identifies the coasts of the north, east Anglia, Devon and Dorset as high risk of cancers. In addition, south and north west London as well as some medium-sized cities Leeds, Nottingham, Leicester and Cambridge, among others, reveal high risk. The third component describes injuries and organ damage; injuries, musculoskeletal diseases and genito-urinary diseases load high on this dimension. It reveals an opposing geographical pattern to the first component, emphasising rural areas. No high scores are found in London and some other cities in England, whereas rural parts of Cumberland, Northumberland, Lancashire, Kent and Dorset emerge as high risk. The component emphasises coastal areas. Both the second and third component account for 10 per cent of the variance each. The first three components appear to reflect urban and rural dynamics in disease incidence in England.

The remaining components describe spatial coincidence of specialised health conditions with each accounting for less variance between 5.5 and 8.2 per cent: birth and service-related conditions (fourth component), nutrition-related disorders (fifth component) and infections (sixth component). The fourth component exhibits a dispersed spatial pattern in central and southwest England. It seems that near-urban areas have more risk of birth and service-related conditions than far rural or central urban areas. This may indicate specific accessibility issues in peri-urban areas in England, although this is speculative at this point. The fifth component emphasises again the north, notably Yorkshire, Cumbria and Cheshire, as well as Norfolk, Gloucestershire and Devon. Yet another geography can be observed for the sixth component: near-urban and rural areas in central and northern England show higher risk of infections in addition to a cluster in eastern Kent.

### Area classification of health

Assessing between versus within-group sum of squares of different cluster solutions suggests that six groups of areas are appropriate. Group-wise boxplots reveal distinct tendencies with respect to health in each of the groups [Figure 5.4]. Given the logarithmic nature of the value distributions, all incidence rates are transformed and subsequently standardised to form the uncentered z score. This ensures equal weights of each individual group of conditions. Boxplots are a useful tool to summarise area health profiles since they visualise the interquartile range (boxes), the median (central line within boxes), the spread of values outside the lower and upper quartiles (whiskers) and outliers (dots).

The first group comprises areas that may be assumed to have good health capital. The population in this group rates their health positively and shows lower than average incidence rates for most conditions. Only birth and service-related conditions show marginally higher incidences, although their median values still remain below average (a value of zero). The geographical pattern associated with this type of areas emphasises rural wards, particularly in the western Midlands and the south east of England. In addition, there are three agglomerations of wards of this type along the east coast: in Essex, north Lincolnshire and north Yorkshire.

The second group shows average health patterns with most of the interquartile ranges (boxes) including the mean value of zero. Just three conditions indicate lower incidences: skin diseases and cancers (malignant and other neoplasms). The geographical pattern highlights again rural areas and also some wards in and around London. There are

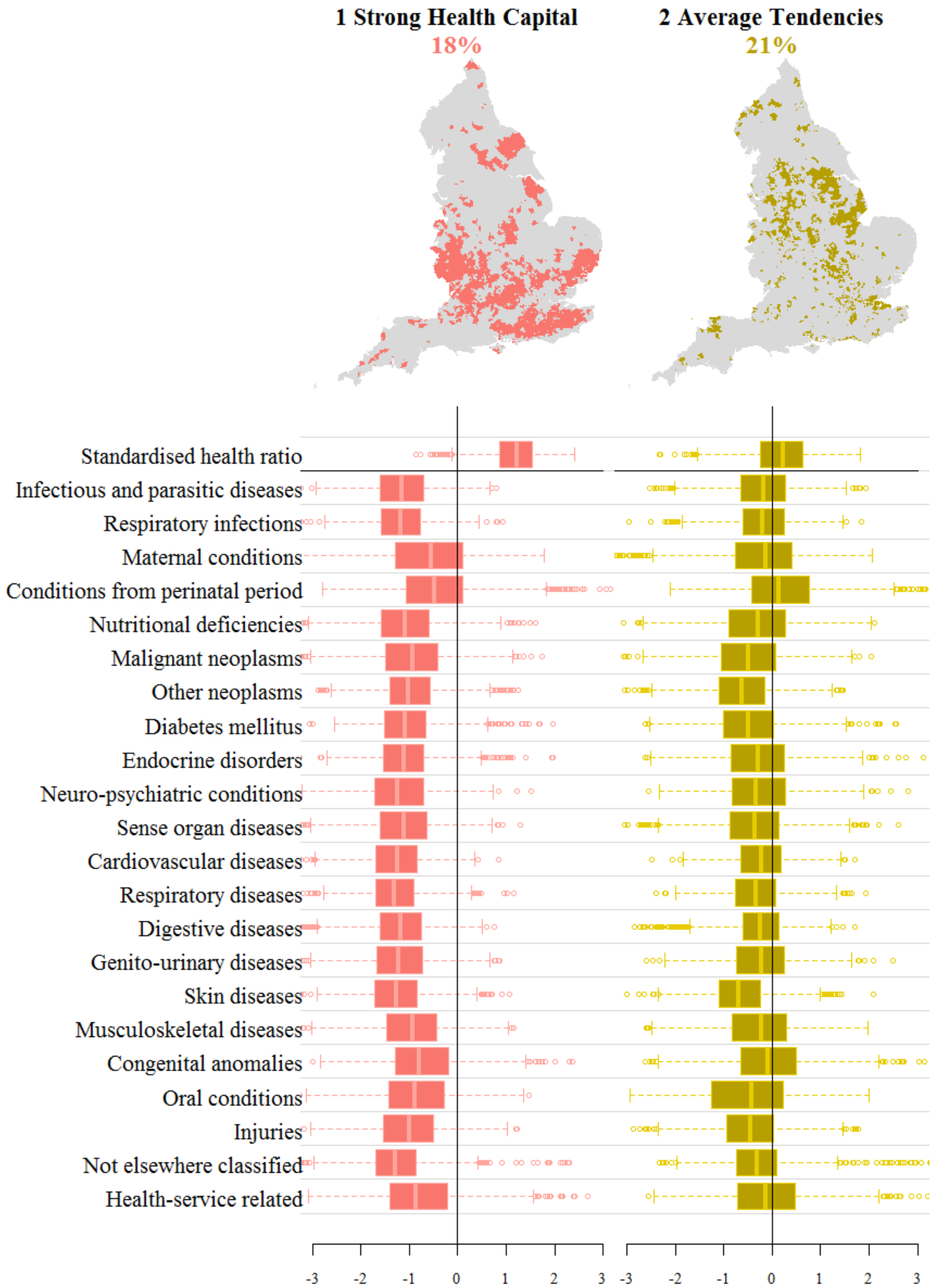



Figure 5.4: Geography and profiles of health area types in England.

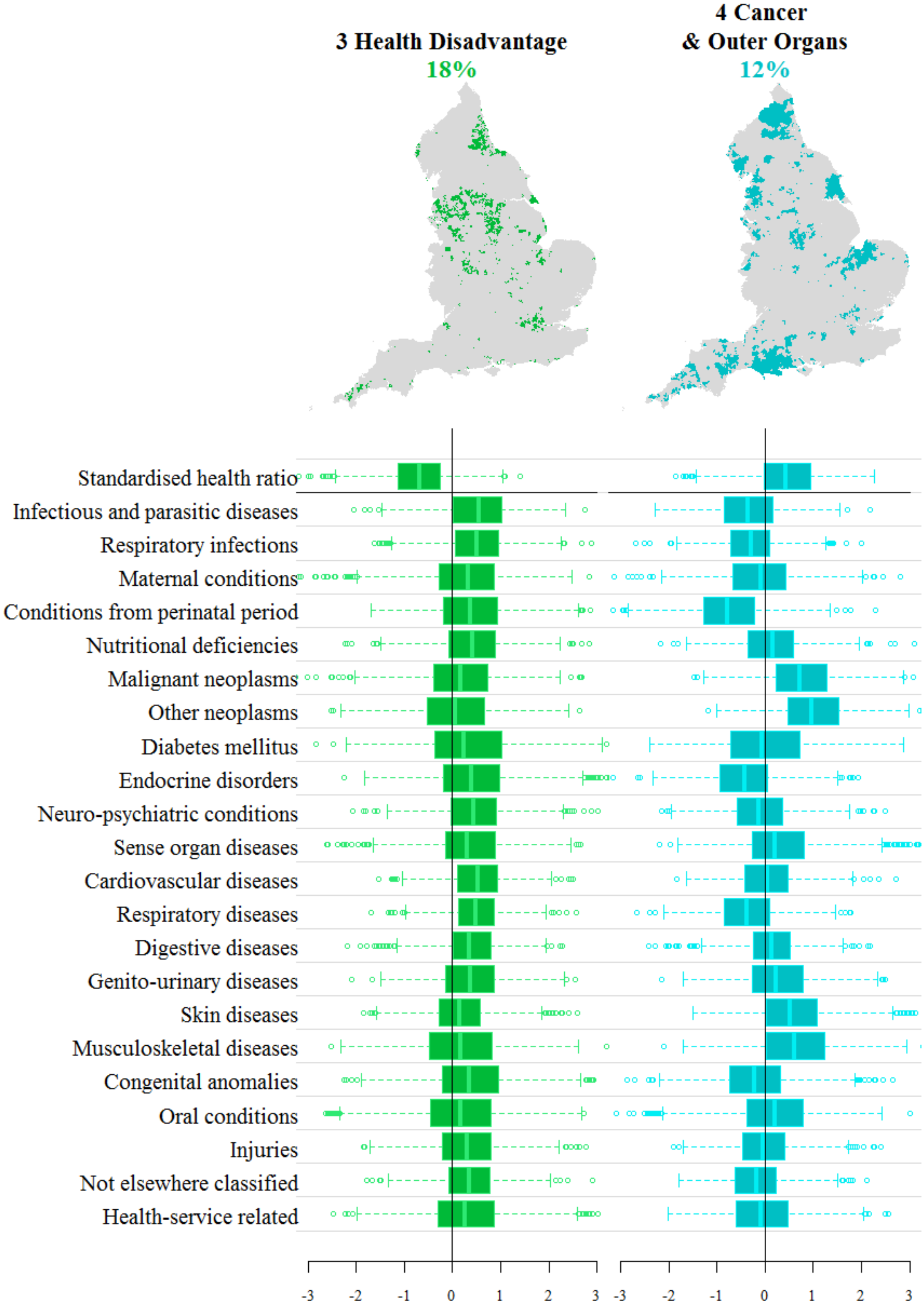



Figure 5.4 (continued)

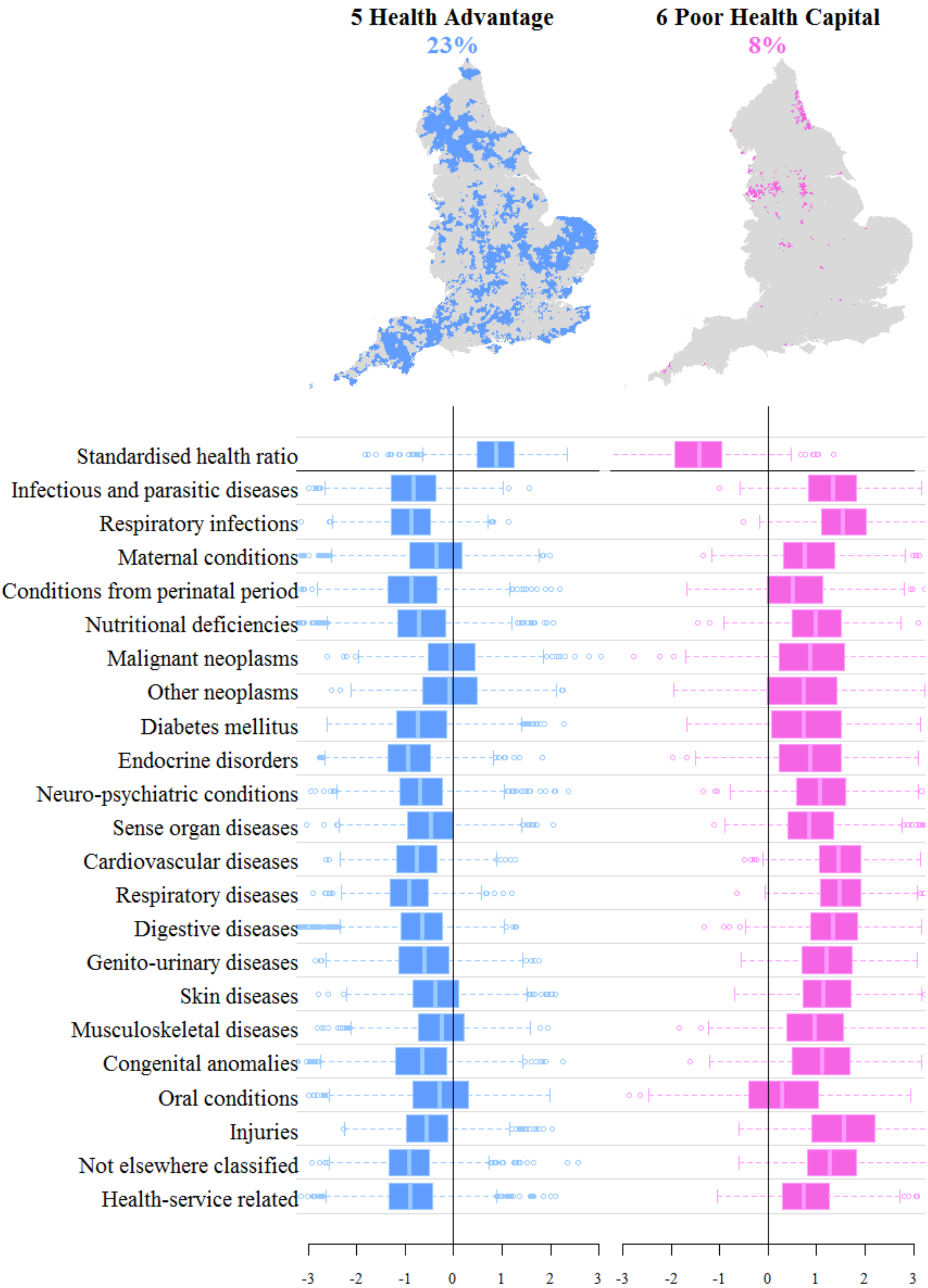



Figure 5.4 (continued)

larger agglomerations in the west Midlands, Yorkshire and the Humber as well as the north west. Some of these areas are close to urban centres such as Manchester, Leeds, Sheffield or Leicester. But in general, rural areas prevail in this group.

The third group comprises areas in which subjective health ratings are poorer and almost all health conditions lie above their England-wide mean values. Incidence rates are particularly high with regards to respiratory infections, other infectious and parasitic diseases as well as non-communicable respiratory, cardiovascular and digestive diseases. Neuro-psychiatric and endocrine as well as nutritional deficiencies also show higher incidences. These patterns may suggest a stronger behavioural or lifestyle influence in these areas' ecological health outcomes. In contrast to the first two groups, this group comprises mainly urban areas. Newcastle, Manchester, Leeds, Sheffield, Birmingham and London emerge as high risk areas as regards this group of areas.

The fourth group possess a more differentiated pattern, in which cancers and conditions related to outer organs – musculoskeletal and skin diseases – occur more often than in other areas. Oral conditions and nutritional deficiencies have marginally higher incidence rates, too, while self-rated health is better. The geography of this group is based on a number of different spatial clusters in the South West, notably Devon and Dorset, Norfolk in the east and Northumberland, Lancashire and East Riding in the north. Most of these areas are rural and close to the coast.

The fifth group exhibits tendencies of better self-rated health and lower incidence rates of most conditions. Cancer-related conditions show a diverse distribution; their incidences are not significantly different from the English average. Infections as well as cardiovascular and respiratory diseases occur less often in these areas; and so do uncommon and service-related conditions. The group comprises 23 per cent of all wards and is the largest in terms of the number of wards. Most of these wards are rural, and none of them are close to the northern cities, be it the agglomeration of Newcastle or the urban corridor between Liverpool and Leeds. The north west, the east and the south west of England show the largest contiguous spatial clusters of this group.

Finally, the sixth group indicates the poorest health on all counts. While self-rated health is low, all except one incidence rate is significantly above average. Risks associated with injuries, respiratory and cardiovascular diseases as well as infections are particularly high. The group is reminiscent of the first Principal Component and may be best described with poor health capital. The geographical pattern is very concentrated; it includes mainly central areas of Northern English cities, including, Newcastle, Gateshead, Liverpool, Manchester, Leeds, Sheffield and Birmingham. The combination of health outcomes suggests that in these areas multiple pathways are at work, including ones that are related to the operation and quality of health care services.

All in all, the groups and their associated geographies show a significant degree of regional concentration, suggesting distinct dynamics of population health in north and south England, coastal and more continental areas as well as urban and rural differences based on individual incidence rates. In classifying areas, an alternative approach would be to use the PCA component scores and classify areas based on the scores. But since the six components only account for little more than half of the variation, a typology based on component scores would only provide a partial picture.

Nevertheless, in conjunction with the above-described six groups, the patterns associated with clustering component scores are informative and aid the interpretation of the six groups. To briefly summarise, clustering component scores yields seven groups of areas with distinct tendencies in terms of health risks and the geography thereof. The first group emphasises lower risk or higher health capital with a large dominating cluster south of London in Kent and a sparse scattered pattern in the rest of England. The second cluster indicates low health capital and service and birth-related conditions and lower risk of infections and injuries. The group is overwhelmingly concentrated in Inner London with only very few additional areas elsewhere. The third group emphasises cancer and is similar to the fourth group above in terms of the spatial distribution in England. A fourth, high risk group emphasises the northern conurbations. The fifth group highlights a distinct pattern of risk with respect to nutritional deficiencies and infections in the north. The sixth group describes distinctly service and birth-related challenges in the southern half of rural England, largely excluding the north. The seventh group suggest high risks of infections and very low risks of nutritional deficiencies in four spatial clusters in the southern half of England with very few instances north of the Midlands.

In their reduced form, the area classification using component scores confirm, too, distinct dynamics in north and south England as well as rural and urban England. Yet, northern cities and Birmingham are different from London. It seems that regional patternings of health in England are defined by these larger geographical trends.

## 5.4 Urban health inequalities in England

The same sequence of analytical steps – summary of ecological and spatial health inequalities, PCA and area classification – is repeated for ten of England's metropolitan areas (the same as in chapter 4). Subsequently, a more detailed study at higher spatial granularity (LSOA) level is carried out for London.

### Health inequalities in English cities

The values for the deficit-related indicators are combined in one measure weighted by their absolute incidence rates. Comparing their absolute, relative and spatial health inequalities reveals differences in magnitude between the ten English metropolitan areas [Table 5.5]. In terms of self-rated health, Middlesbrough, Manchester, Kingston-upon-Hull and Liverpool reveal the strongest absolute differences of between 14,500 and 15,500

Table 5.5: Health inequalities in across ten English metropolitan areas.

| | asset | | | | deficit | | | |
|---|---|---|---|---|---|---|---|---|
| | **ARD** | **QR** | **I** | **p(I)** | **ARD** | **QR** | **I** | **p(I)** |
| Birmingham | 12,797 | 1.18 | .653 | .000 | 1,267 | 2.17 | .507 | .000 |
| Kingston-upon-Hull | 14,951 | 1.21 | .448 | .000 | 1,982 | 2.24 | .494 | .000 |
| Leeds | 11,799 | 1.16 | .441 | .000 | 1,610 | 2.06 | .372 | .000 |
| Liverpool | 14,756 | 1.21 | .502 | .000 | 2,151 | 2.28 | .457 | .000 |
| London | 12,070 | 1.15 | .642 | .000 | 1,058 | 2.30 | .623 | .000 |
| Manchester | 15,284 | 1.22 | .522 | .000 | 1,573 | 2.18 | .416 | .000 |
| Middlesbrough | 15,362 | 1.23 | .511 | .000 | 1,802 | 2.17 | .403 | .000 |
| Newcastle-upon-Tyne | 12,606 | 1.18 | .305 | .000 | 2,419 | 2.45 | .208 | .058 |
| Sheffield | 14,352 | 1.21 | .511 | .000 | 1,399 | 2.10 | .480 | .000 |
| York | 9,785 | 1.12 | .407 | .000 | 880 | 1.82 | .268 | .027 |

**columns**: ARD = absolute rate difference per 100,000; QR = quantile ratio; I = Moran's I of ward standardised rate ratios; p(i) = p-value of Moran's I

cases per 100,000 people. This translates into a relative ratio (QR) of a 1.2-fold difference of the level of self-rated health between the 95th healthiest and the fifth healthiest percentile of wards. The metropolitan areas with the lowest level of health inequalities are York with under 10,000 cases, followed by London and Leeds. They QR shows a difference of 1.16 or less. Although this is not very different from the more unequal cities, the magnitude is important given the high number of observations. The level of health inequalities does not translate into spatial health disparities in the same way. The highest degrees of spatial clustering can be found in Birmingham (.653) and London (.642). Newcastle and York exhibit the lowest level of spatial health disparities with values of .305 and .407 respectively.

As for the deficit measures, the most unequal cities are to be by far Newcastle and Liverpool in terms of absolute measures. In terms of relative health inequalities, Newcastle shows a 2.45-fold difference between the top and bottom quantiles, followed by London and Liverpool with approximately 2.3. The level of spatial disparities is in general lower than for self-rated health. This time, London reveals strongest disparities and Newcastle again the lowest. The remaining cities lie in between but remain distant from the values of the two extremes.

This comparison demonstrates that urban health inequalities exist in all of England's metropolitan areas. These patterns suggest the existence of a global trend that affects the way people with differential vulnerability sort into neighbourhoods, thus generating geographical patterns of health inequalities. Yet, the magnitude and spatial manifestation of urban health inequalities differs between England's metropolitan areas, suggesting locally specific pathways of health advantage and disadvantage.

The cluster analysis of urban areas produces five groups with different health profiles and discernible geographical patterns. The first may be interpreted as strong health capital and shows similar tendencies to the one found for all of England. The SHR is significantly above the average while all other conditions are significantly below. The cluster

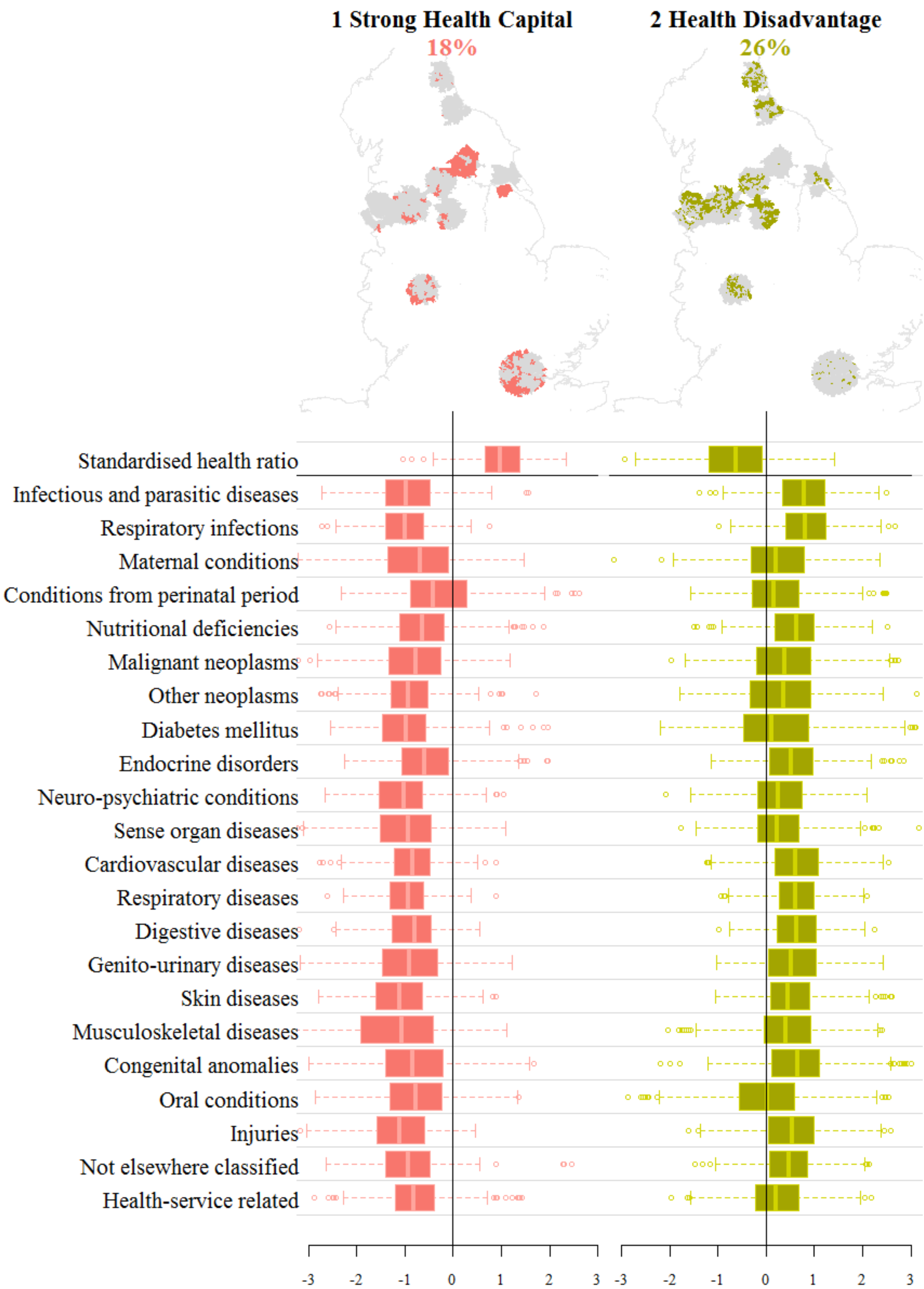



Figure 5.5: Geography and profiles of urban health area types across ten metropolitan regions.

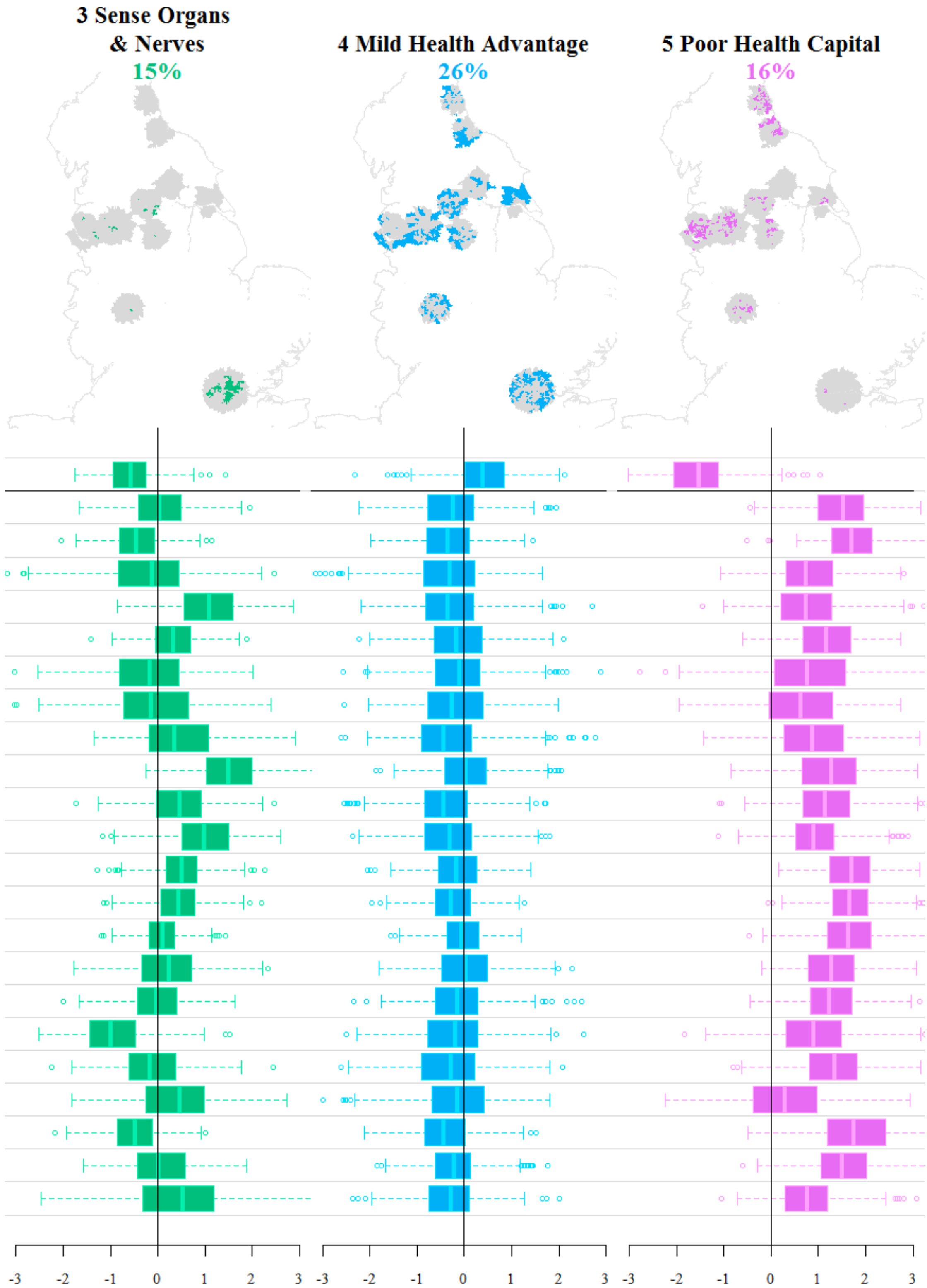



Figure 5.5 (continued)

comprises mainly suburban areas. It nearly covers all outer areas in the metropolitan region of York and the parts of Kingston's region that are located south of the river Humber. A high number of suburban wards in south west and north west London as well as Birmingham also belong to this group. The cluster hardly occurs in Newcastle, Middlesbrough or Liverpool.

The second group represents a degree of health disadvantage, in particular with respect to infectious diseases. It is widespread in all metropolitan regions except in York and London. It compares to the cluster of health disadvantage in England, but the more pronounced incidence or infections appears to be a specifically northern urban variant. The distinctiveness of the third group results from high incidence of three conditions: conditions arising during the perinatal period, sense organ diseases and endocrine disorders, with low incidence of musculoskeletal diseases and injuries. This differentiated pattern has a large concentration in inner London with a few instances in Liverpool, Manchester and Leeds and no instance in the northern and western most metropolitan regions.

A fourth group expresses mild health advantage alongside average tendencies with only the standardised health ratio and injuries being significantly different from the average. The cluster is widespread in mainly suburban locations that comprise particularly large parts in Middlesbrough, Kingston-upon-Hull, Leeds and the outer boroughs in London. Areas of poor health capital are nearly identical to the England-wide equivalent in terms of their profile and geography. This seems to be a distinctively urban cluster that can be found predominantly in the northern conurbations of Liverpool, Manchester and Sheffield as well as in the very north. There is no instance in York and only a very small number of instances in London. This area type appears to comprise poorer areas in regions that were home to heavy industries in the past.

Unlike population structure, many health challenges are at a first glance rarely specific to particular cities; they appear in multiple cities at a time. Only London has its own specific expression of health disadvantage in form of diseases connected to the nervous system. Across cities, the intra-urban distribution of clusters show similarities: it is possible to distinguish suburban from central-urban area types. Across England, inner-city wards tend to express health disadvantage while certain suburban areas indicate advantage. Yet, there are specific variants. Health disadvantage is virtually absent in York in comparison. All areas in York belong to the group of strong health capital or mild health advantage. Liverpool, Middlesbrough and to a degree Newcastle-upon-Tyne do not have areas of strong health capital. In these metropolitan regions, we may observe an interaction effect. While the patterns across all appear to reflect a generic societal process that generates inner-city urban disadvantage and suburban advantage, locally specific conditions moderate the outcomes of this process such that advantage translates in some cities only in mild health advantage rather than strong health capital. These conditions also appear to translate into an attenuation or inflation of urban health inequalities. Disadvantage can also take on the specific forms in interaction with local characteristics, as is the case in London.

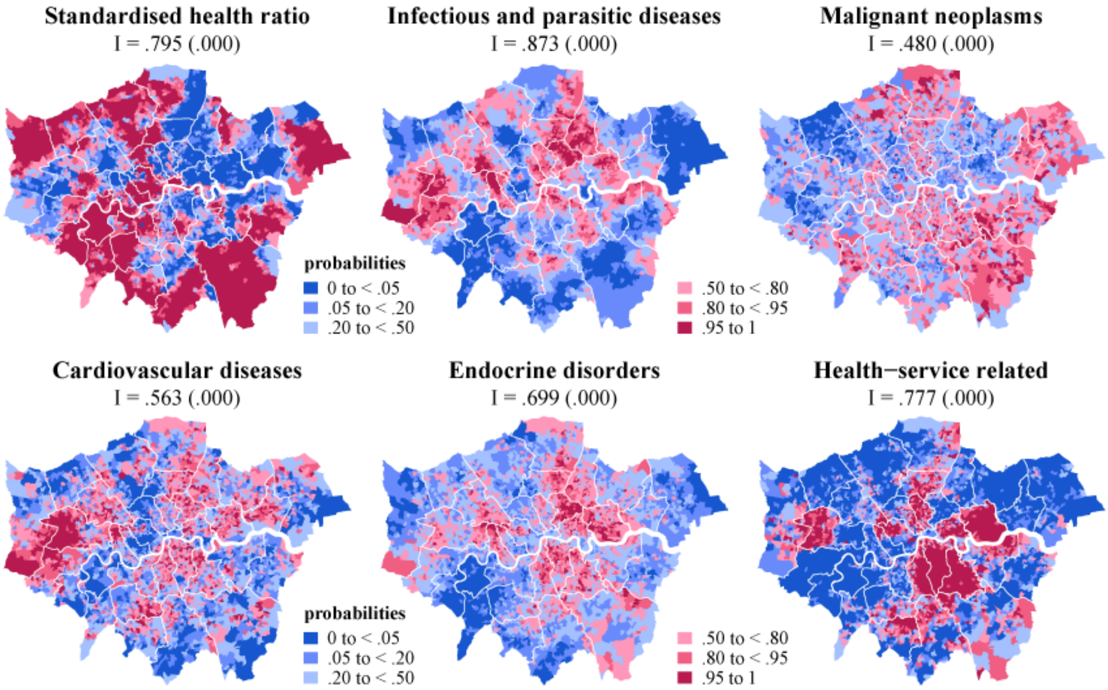



Figure 5.6: The probability of smoothed incidence rates for self-rated health and selected conditions being above average in London.

### Urban health inequalities: the case of London

The comparison of ten metropolitan regions in England provides a relevant context for a more detailed investigation of urban health disparities in Greater London. For this purpose, the spatial granularity is increased to the level of Lower Level Super Output Areas (LSOAs), the second most granular Census geography in England and Wales. There are 4,838 LSOAs in London; most of them have a size of between 1,000 and 1,500 people and thus represent a scale that may correspond more to the cognitive scale of neighbourhoods than census wards, which in London often have more than 10,000 people. Just under 1.6 million records of the HES 2008/09 pertain to persons that live in London and are used to calculate LSOA level age-and sex standardised incidence rates along with the Census-derived Standardised Health Ratio.

Geographical disparities are significant after smoothing and exhibit a strong degree of spatial clustering [Figure 5.6]. Smoothed rates for self-rated health broadly divide London into west and east and a northern and southern suburban fringe. Here, rates of good self-rated health are higher. With a value of .795, the degree of spatial clustering is very high. Infectious and parasitic diseases exhibit three large clusters (.873), one across Hillingdon, Hounslow and Ealing, one in Hammersmith and Fulham and one in north-east London, in the boroughs of Hackney, Haringey and Islington. Malignant neoplasms are more dispersed across Greater London (.480); risks are higher in northeast, east and southeast London. Cardiovascular diseases cluster in Hillingdon and Ealing as well as

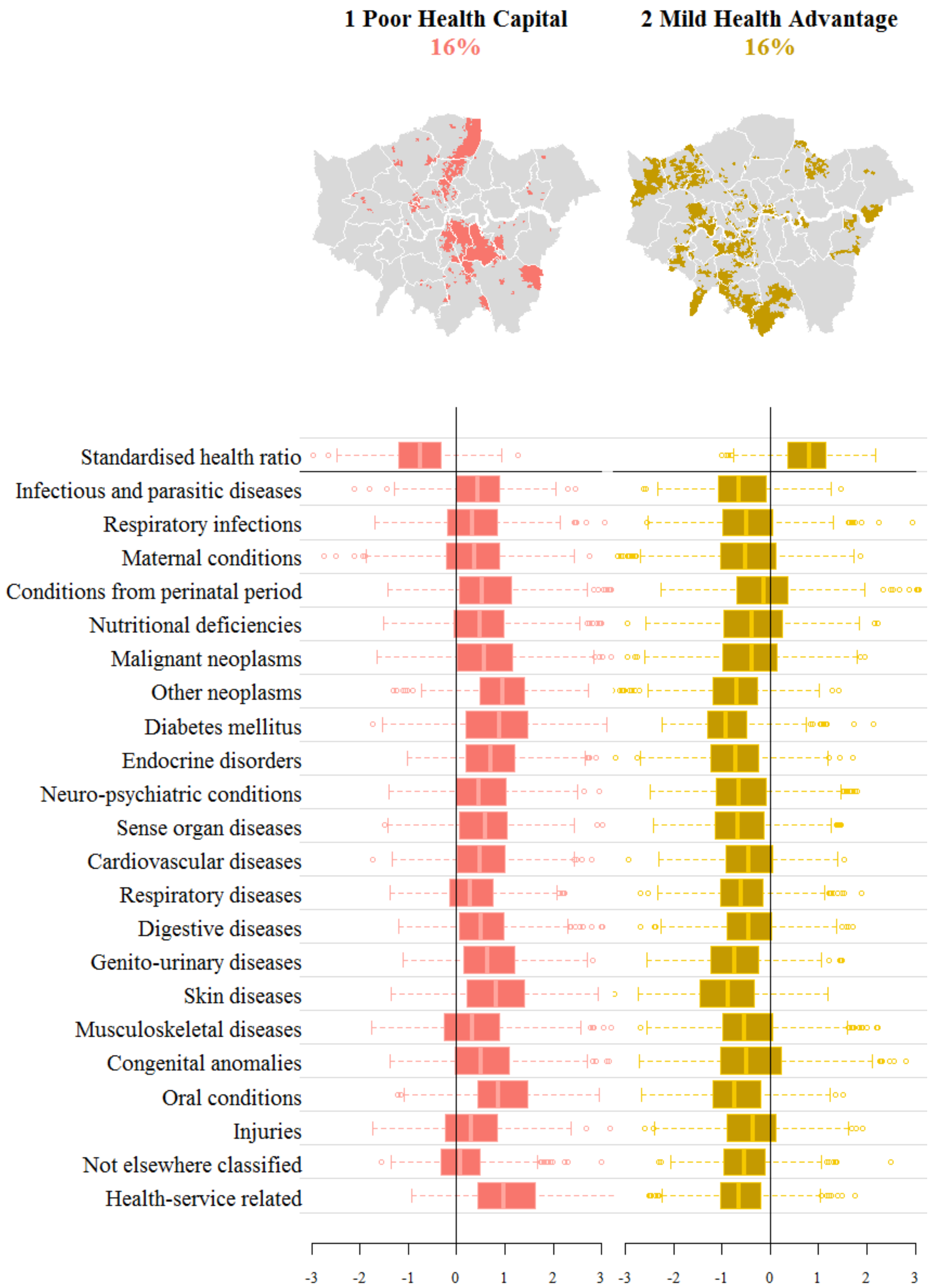



Figure 5.7: Geography and profiles of health area types in London.

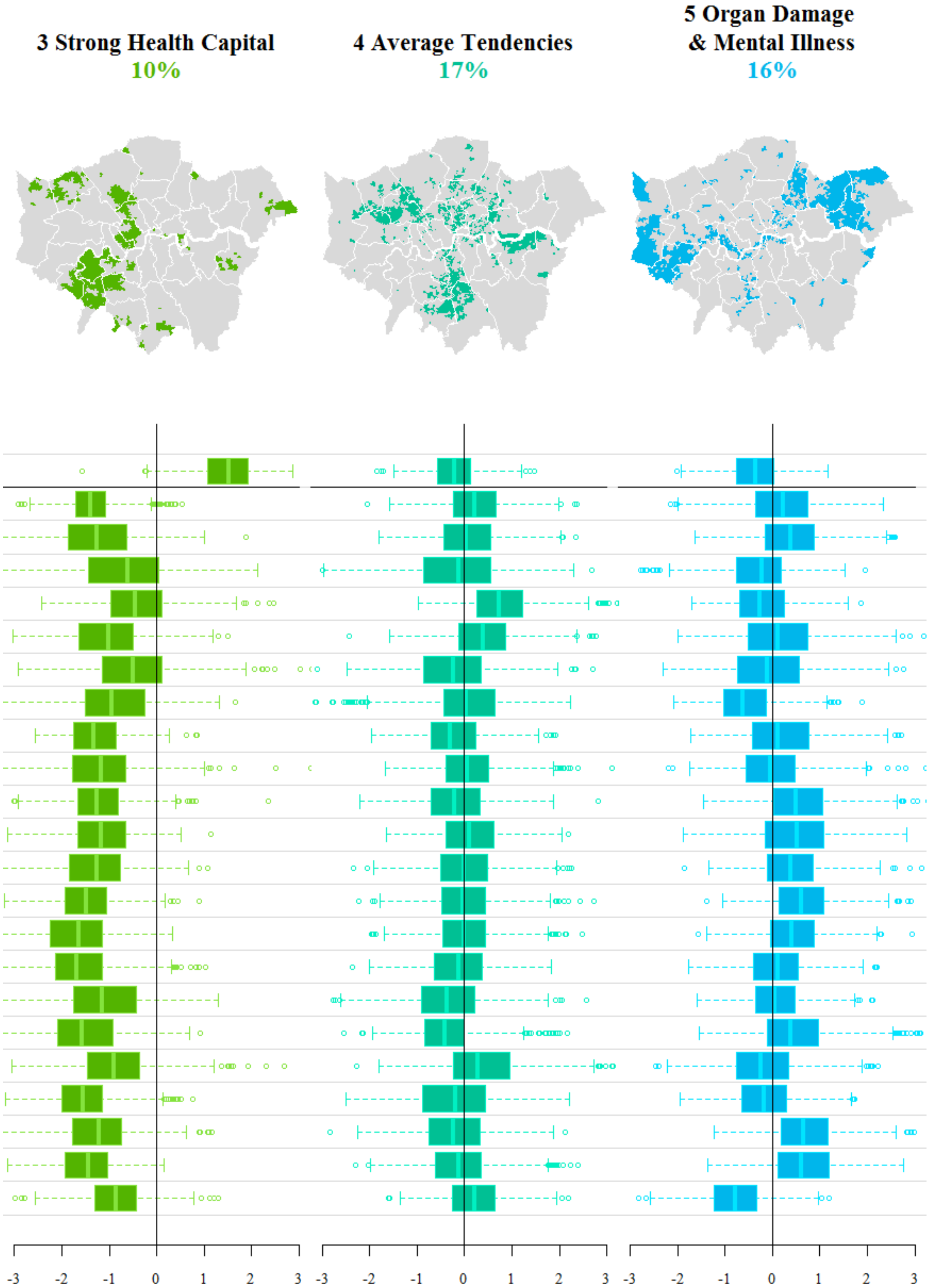



Figure 5.7 (continued)

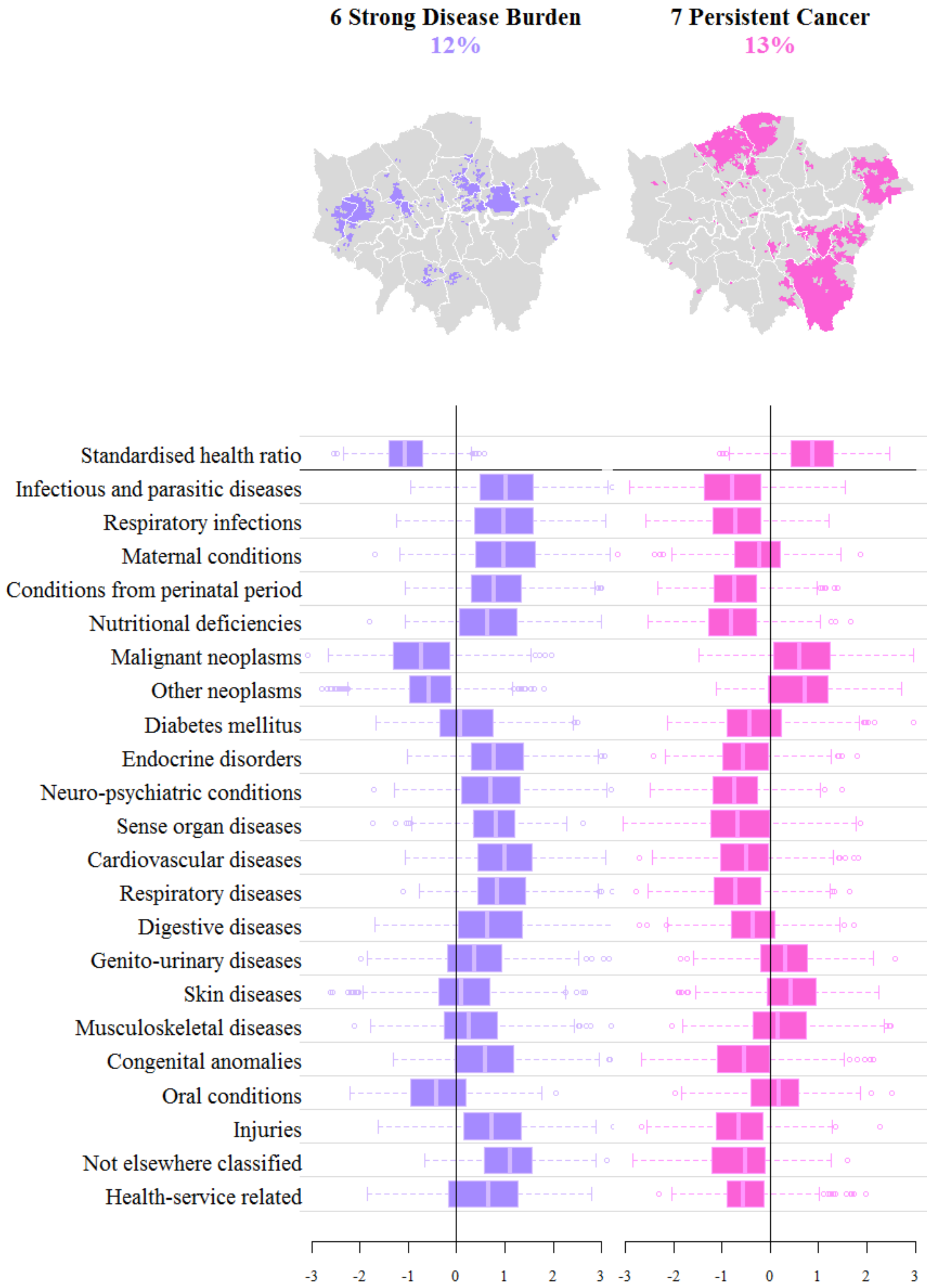



Figure 5.7 (continued)

Table 5.6: PCA solution for London LSOAs.

| condition | PC 1 | 2 | 3 | 4 | 5 | u'ness |
|---|---|---|---|---|---|---|
| **asset** | | | | | | |
| Self-rated health (SHR) | -.539 | -.105 | -.366 | -.460 | -.348 | .231 |
| **deficit** | | | | | | |
| Infectious and parasitic dis's | .455 | .058 | .271 | .147 | .645 | .278 |
| Respiratory infections | .622 | .028 | .232 | .113 | .301 | .456 |
| Maternal conditions | .260 | .171 | .264 | .081 | .108 | .815 |
| Cond's from perinatal period | .126 | -.097 | .649 | .138 | .177 | .503 |
| Nutritional deficiencies | .312 | -.016 | .365 | .211 | .313 | .627 |
| Malignant neoplasms | .033 | .473 | -.003 | -.027 | -.193 | .737 |
| Other neoplasms | -.115 | .774 | .115 | .030 | .185 | .339 |
| Diabetes mellitus | .254 | .330 | .189 | .614 | .024 | .414 |
| Endocrine disorders | .295 | .188 | .360 | .258 | .423 | .502 |
| Neuro-psychiatric conditions | .577 | .104 | .186 | .171 | .271 | .519 |
| Sense organ diseases | .328 | .032 | .204 | .709 | .196 | .309 |
| Cardiovascular diseases | .557 | .139 | .237 | .313 | .172 | .486 |
| Respiratory diseases | .665 | .047 | .198 | .323 | .182 | .379 |
| Digestive diseases | .646 | .349 | .211 | .113 | .112 | .391 |
| Genito-urinary diseases | .425 | .593 | .137 | .175 | .050 | .416 |
| Skin diseases | .205 | .640 | .032 | .116 | .304 | .442 |
| Musculoskeletal diseases | .580 | .417 | -.043 | .098 | -.100 | .468 |
| Congenital anomalies | .224 | .127 | .481 | .120 | .064 | .683 |
| Oral conditions | .230 | .541 | .209 | .162 | .014 | .584 |
| Injuries | .723 | .109 | .109 | .058 | .067 | .446 |
| Not elsewhere classified | .668 | .030 | .170 | .318 | .086 | .416 |
| Health-service related | .085 | .295 | .800 | .080 | .016 | .259 |
| SS loadings | 4.481 | 2.521 | 2.258 | 1.713 | 1.328 | |
| Proportion Var | .195 | .110 | .098 | .074 | .058 | |
| Cumulative Var | .195 | .304 | .403 | .477 | .535 | |

**interpretation**: poor health capital (1); cancer & outer organs (2); birth & service-related (3); special conditions (4); infections (5); u'ness = variate uniqueness

in London's eastern boroughs (.563). Endocrine disorders are mainly concentrated in east London as well as in more central northern and southern boroughs (.699); they are less common in the suburbs. Service-related conditions are significantly concentrated (.777) in five boroughs: Newham, Tower Hamlets, Lewisham, Southwark and Lambeth with an additional cluster in western Ealing.

Other conditions with distinctive spatial patterns are maternal conditions, with higher likelihood of increased incidence in the boroughs of Newham, Greenwich and Bexley as well as South London and Hillingdon, other neoplasms, which cluster the strongest and exhibit two large regional concentrations in the northwest and the south east, each capturing nearly four entire boroughs. The patterns are nearly identical for men and women viewed separately (not applicable to maternal conditions) except for genito-urinary diseases, whose probability of incidence appears to be higher in the suburbs for men and in central locations for women.

### An area classification of health in London

Exploration of ecological patterns of the health indicators through PCA highlight specificities for London [Table 5.6]. Self-rated health correlates inversely with most components in London, indicating a stronger role of health-relevant assets in health inequalities in London. The most important component, which may be interpreted as poor health capital, exhibits higher incidences of a range of diseases yet not all diseases. Sense organ diseases and endocrine disorders, the two conditions that formed a separate cluster in the urban comparison, depart from the general pattern of health disadvantage and show high loadings with other components. Sense organ diseases lead a highly specific fourth component indicating special health challenges notably risk of diabetes. The highest loading for endocrine diseases can be found on a fifth component of infections and a third component of birth and service-related conditions. In England, these two conditions appeared as part of a general component of health disadvantage; informally, this confirms the presence of specific expressions of health disadvantage in London. Together, the five components for London account for 54 per cent of the variance.

Clustering self-rated health and standardised SMbRRs produces seven clusters for London [Figure 5.7]. The first cluster shows general health disadvantage with higher risk of all conditions and lower self-rated health, which may be interpreted as poor health capital. Service-related conditions, injuries, oral conditions and other neoplasms are particularly high in this group. The cluster is strongly spatially concentrated; it stretches from the boroughs west to Lea Valley – Enfield and Haringey – southward towards Islington and further south across Lambeth, Southwark and Lewisham.

The second cluster reflects mildly healthy tendencies with higher self-rated health and lower SMbRRs. In geographical terms, it is more dispersed and can be found more often in suburban west and southwest London, including Harrow, Hillingdon, Ealing, Kingston, Sutton and Croydon. The third cluster represents a stronger version of the previous one; it is the healthiest cluster with low SMbRRs, in particular for infectious and parasitic diseases and a range of chronic conditions. The ratio of good self-rated health is high and permits the interpretation of representing strong health capital. Areas in west and south west London, from Westminster to Richmond belong to this cluster. The fourth cluster shows average levels of risk with diverse tendencies and comprises intermediate areas that are adjacent to the City to the north, notably Camden and Islington, to the south covering a region between Lambeth and Croydon and further outside, in Brent and in Greenwich.

The remaining three clusters depart from the more uniform, generic patterns: slightly elevated risk of some non-communicable diseases related to organ impairment, injuries, mental conditions and conditions not elsewhere classified with lower self-rated health and service-related risk (cluster five); strong burden of disease excluding cancer (cluster six); and high incidence of cancer despite indication of good health capital (cluster seven). The more specialised cluster five is located mainly at the western and eastern extremes of London, encompassing south Hillingdon and Hounslow in the west and

Barking and Dagenham, Redbridge and Havering. The cluster covers Hounslow and Barking and Dagenham almost in their entirety. Cluster six, the strong disease burden excluding cancer, prevails in the borough of Newham and in parts of Tower Hamlets as well as in a cross-boundary region between east Hillingdon and west Ealing. Specific challenges of cancers and outer organs (cluster seven) appear in three major geographical concentrations: the first and largest in the south, extending across Bromley, Greenwich and Bexley, the second in Brent and Enfield and the third in Havering.

The different geographies of the clusters show various patterns. They appear to sometimes follow borough boundaries – some clusters seem to be map on the characteristics of borough populations. At the same time, most clusters encompass two or three boroughs or parts thereof, suggesting cross-borough challenges with their own, emerging spatial extent. The profiles and geography of clusters one and six confirm again a specific expression of disadvantage in London with higher incidence of disorders related to the nervous system and – unlike in other cities – lower incidence of cancer. Conditions related to cancer appear to follow a distinct geography in London, straddling diverse areas in deprived neighbourhoods and affluent suburbs.

## 5.5 Synthesis: health environments as emergent geographies of vulnerability

### The merit of this classification

The England-wide classification shows both similarities and important differences with the comparable studies conducted by Shelton et al [2006] and Green et al [2014]. Like the former study, this classification identifies geographically contiguous areas across England with a special role of London, urban England and coastal areas. Yet, due to the higher spatial resolution of this work, intra-regional and intra-urban differences could be discovered that refine the results of Shelton et al. [2006]. Unlike in theirs, the role of individual conditions is less prominent in this study, which highlights more generic social determinants of disadvantage.

With respect to area health profiles, this work resembles more the study by Green et al [2014], although they find very little discriminatory power of mortality causes. Yet, this may be a result from their use of finer mortality data, rather than more coarsely grouped morbidity categories and also of their statistical design because they weight conditions in the clustering by their incidence rates. Hence, more common conditions have a larger impact on the cluster solutions in their study, whereas here, in the absence of weight, a special role of a number of diseases, including cancer, outer organs, disorders related to the nervous system and infections can be identified.

This study differs from Green et al in the identification of geographically contiguous patterns. While both classifications identify an urban-rural divide, Green et al do not identify a north-south divide, which appears very prominently among the results above.

Again, in this respect the findings are more consistent with Shelton et al [2006], who identify the south east and former mining areas (north) as different from each other.

The geographical pattern of health profiles is also shaped by the spatial-structural model applied to estimate health indicators. Spatial autocorrelation is used as a source of information to increase the robustness of local estimates. Given an average death number of 275 cases per MSOA and 63 categories in Green et al's study, applying a Bayesian framework might have been advisable, since individual area class memberships may be prone to temporal fluctuation due to sparse data. The authors acknowledge the uncertainty associated with their classification and recommend Bayesian methods as a way of addressing this. It can be expected that a Bayesian approach will have wider implications for the development of geodemographic classifications, since little is known about the temporal stability of geodemographic classifications. The Bayesian method produces so-called credible intervals, which are based on the residuals that are measured in terms of the spatially structured component and the unstructured component. As more years become available, the credible intervals will reduce and the robustness can be improved further, in theory also within longitudinal models of local population health. In theory, simulations drawing values for each area and each clustering variable from the posterior marginals (the posterior distribution of area rates) may be used to assess the robustness of the classification and quantify the uncertainty thereof.

The geodemographic classifications for England, England's metropolitan areas and London thus incorporate explicit spatial context. The clustering of smoothed estimates causes the convergence of neighbouring areas and thus similar cluster memberships. As a consequence, regional agglomerations of clusters tend to extend their boundaries to encompass similar areas in their vicinity. This may be one way of grounding geodemographic classification more strongly in spatial context, as Singleton and Longley [2009] advocate. In addition, uncertainty can be incorporated as another dimension of information; this will be further explored in a later chapter.

Finally, the comparative framework applied on ten metropolitan regions in England further contextualisesd health profiles and revealed regionally and locally specific expressions of health advantage and disadvantage. Some general trends of disadvantage in inner-urban areas and advantage in suburban areas could be disentangled from specific guises by viewing the health profiles and their explicit spatial distributions. Within encompassing comparative frameworks, a common, 'global' process is assumed to produce instantiations across different contexts that interact with local conditions [Robinson 2011]. This framework is used in an exploratory manner here; it reveals the presence of a process that generates differently structured environments with respect to health assets and deficits, placing advantage in suburbs and disadvantage in inner cities. In consequence, south western suburbs in London appear more similar to York or south Kingston-upon-Hull than to north eastern London suburbs, which resemble more inner York, inner Kingston or south Middlesbrough. Some local conditions appear to attenuate or intensify the divide between inner cities and suburbs.

The comparative framework also highlights specific expressions of disadvantage in London, which would not have been identified without comparison. Viewing specificities in an encompassing framework may compel us to rethink the meaning of area classifications. It may be said that the interaction between global processes and local conditions generate distinct environments in which health patterns emerge in specific ways. These *health environments* represent connected urban outcomes that are neither wholly determined by the global process nor by local conditions; they emerge out of the interaction of both and find their manifestations in singular local phenomena. Robinson refers to those as "singularities" [2015, 10]: undetermined outcomes as manifestations of processes that cannot be deduced from pre-existing general categories but only discovered and interpreted through unprejudiced, empirical method. This understanding applied to health environments suggests that each type contains a specific set of political, biological, social and physical vectors that modify a social distribution of assets and risks in particular ways and thus produces specific forms and expressions of vulnerability. But, following the notion of singularity, classified units (neighbourhoods) that belong to the same class must possess unique characteristics generating not identical but similar, possibly related outcomes. This complex notion of health environments is crucial to the design of effective policy responses to improve population health.

### The limitations and possible extensions of the classification

In substantive terms, the clusters need to be interpreted with caution. First, the deficit-related indicators rest on hospital admission data. Hospital admission can only provide a partial picture of health as they fail to capture incidences that are not treated in inpatient hospital care. These include any less severe health events treated in primary care as well as outpatient care and accident and emergencies, which cover a large number of injuries. In addition, there is evidence that some groups are more likely to seek health care at different stages of diseases progression than others [Scambler 2008]. Hence, hospital admissions may not be an accurate indicator of need, although this limitation may be less applicable to inpatient data.

Nevertheless, it can be assumed that different types of health issues are correlated and that poor health capital can manifest in health outcomes that are treated in other settings of health care. A general challenge in measuring health is that measurement typically occurs when a condition crosses a certain level of severity to be detected, especially in a system that does not emphasise health prevention. Ways of incorporating health consequences that are not yet symptomatically manifest or clinically detected represents a relevant aspect of everyday population health. This touches on a narrow range of data pertaining to the asset view of health, more specifically subjective health and well-being.

Second, an ecological classification alone does not reveal any causal pathways. Through geodemographic classification it is possible to view common characteristics of similar cases or understand the structure of health of one area by viewing the properties of the class it is assigned to. This is useful in terms of exploration, or, as Byrne would have it, "system-level description" [Byrne 1998]. But any further leading causal analysis would have to rely on the validity of causal linkage at the aggregate level. Without systematic contextualisation of geodemographic classes and their geography, the social meaning of geodemographic classes remain vague and the implications for policy elusive.

Third, the disease grouping itself according to WHO typology may be contested as it groups together diverse diseases with potentially very different consequences and impacts on people's lives. Tuberculosis has different causes and consequences than HIV/AIDS and yet the two belong both to the same second order disease category of infectious and parasitic diseases. Although further refinement of diagnosis using HES data would have been possible, it would have magnified the occurrence of small numbers and thereby inflated statistical uncertainty in exchange for a reduction in substantive uncertainty. This trade-off has to be made when analysing health records and it can only be optimised by pooling records over several years. For finer categorisations of diseases, there are other typologies that may be used, notably the Clinical Classification Software (known as CCS) published by the US Agency for Health Research and Quality with 285 mutually exclusive disease categories [Elixhauser et al. 2010]. Alternatively, the most common diseases identified in the latest Global Burden of Disease Report could be used [Murray et al. 2012].

Fourth, it can be argued that not all conditions are equally important in affecting population health. Incidence rates for each of the 22 conditions can range between 93 (diabetes mellitus) and 2,300 (digestive diseases) cases per 100,000 people. Green et al [2014] therefore decide to weight each mortality cause by its incidence in the population so as to account for the structure of mortality in the region of interest. Indeed, a weighted classification carried out here as part of sensitivity testing reveals that more common diseases such as digestive diseases, maternal conditions, musculoskeletal diseases and in fact unclassified disease have a larger impact on the clustering. The resulting clusters are defined by the variation of these few conditions while the remaining conditions are pushed towards the population average. Despite this behaviour, the resulting clusters and their tendencies are similar to the unweighted variant. Cross-tabulation reveals that the majority of clusters (between 50 and 70 per cent) remain in the same clusters in both alternative classifications. The remaining cases differ in their subtleties, some may have higher incidences of cancer-related conditions, others may have higher incidence if diabetes mellitus. Yet, it is those subtleties that distinguish area-specific health challenges. Rare diseases can have a profound impact on local service commissioning since they may require specialised health care provision. Absolute incidence of conditions may hence not be the decisive criterion in warranting attention to local health needs.

Fifth and finally, the area classification includes conditions that only apply to sub-groups of the population, specifically maternal conditions (women) and conditions arising during the perinatal period (newly born). The substantive argument in favour of including these conditions would be that the social environment may still affect the incidence of those, that is social pathways in health may still be active, and hence those conditions may be as informative about an area's health as another condition that can be found throughout the population. These and particularly rare diseases have been removed from the clustering in the sensitivity tests; the results are nearly identical with 70 to almost 100 per cent of cases classified similarly.

A potential extension of this research that may increase the substantive certainty of the classification is a joint investigation of the geographies of morbidity and mortality as both ONS mortality data linked to HES records become available. Using mortality data would allow to develop alternative health metrics, such as Disability Adjusted Life Years and related indicators, although at a small area level, neighbourhood turnover would have to be accounted for. Combining mortality and morbidity data in a single classification would be the first study of its kind and may generate potential to support local strategic efforts to assess health needs and develop health and well-being policies.

In conclusion, the preceding study identifies different health environments, which reflect different forms and degrees of vulnerability of populations, households and individuals. Yet, while the health environments highlight geographically varying challenges, they do not reveal the precise pathways that may be at work across and within them. These pathways need to be identified through further contextualisation, notably by ascertaining who are the inhabitants of the health environments.

# 6 Vulnerable social milieus: the individual level

Who, then, lives in these health environments? The individual level is by definition central to health as the concept derives from the state of physical and mental well-being of humans. While health environments emphasise structural and contextual conditions, health is not wholly a deterministic outcome of circumstances but also modified by the action of individuals. Some form of agency at the individual level is therefore fundamental to the understanding of vulnerability. Hence, heuristics are required to characterise the inhabitants of health environments in terms of their properties and practices.

## 6.1 Subjective orientations and health practices

Health scientists have studied the role of individual agency by measuring instance and intensity of relevant individual routines, known as health-related behaviours [Locker 2008]. While the term can signify both lifestyle risk factors [Blaxter 1990] and sometimes propensity to seek care [Scambler 2008], the World Health Organisation has recently ascertained that behavioural risk factors have come to be the single most threat to human health as non-communicable diseases ascend to the leading cause of death worldwide [WHO 2014]. They recommend at various places in their 2014 report to design policy interventions that focus on influencing behaviour with respect to smoking, drinking, unhealthy eating and sedentariness. The fundamental assumption is that these habits are guided by free will and are hence modifiable.

Health-related behaviours encompass all those routinised actions that have a direct impact on health. Stringhini and colleagues [2010], for example, investigate the impact of four behavioural risks – drinking, smoking, healthy eating, exercising – on mortality in the widely cited Whitehall II study, which focusses on health among civil servants employed in London‘s government district. Measured on an ordinal scale of intensity, the health-relevant habits are treated as random, explanatory variables in a regression model. Statistically speaking, the four behaviours are conceived of as independently varying with an impact over and above social and demographic circumstances. The authors conclude in their study that it is this behavioural pathway that accounts for much of the link between socio-economic position and mortality. Other researchers have studied the impact of health-relevant behaviours on a variety of outcomes, including mortality [Ford et al. 2012; Khaw et al. 2008; Loef & Walach 2012], mental disorders [Cerimele & Katon 2013; Vancampfort et al. 2013; Vermeulen-Smit et al. 2015], physical non-communicable diseases [Long et al. 2015; Shankar et al. 2006; Wu et al. 2015] or self-rated health [Blaxter 1990; Eriksen et al. 2013; Kasmel et al. 2004]. In most studies, it is found that behavioural risks are significantly associated with health outcomes; indeed, often they account for the majority of the variation.

In health geography, behavioural risk is typically treated as response rather than control or exposure variable. Akin to neighbourhood effects research, studies investigate associations between density, street layout, segregation or unemployment and alcohol consumption, physical activity or smoking rates [e.g. Carpiano 2007; Ewing et al. 2013; Jongeneel-Grimen et al. 2014; Kearns & Mason 2015; Moon et al. 2012; Pearce et al. 2009; Turrell et al. 2013]. The study by Carpiano draws on Bourdieu‘s work to assess in a multilevel regression model the impact of neighbourhood social capital on individual smoking and drinking habits. He controls for individual demographics. Health behaviours are presented as outcomes of neighbourhood and individual conditions; this representation contrasts with the aforementioned epidemiological studies, which emphasise the voluntary nature of health behaviours. Hence, a structure-agency divide becomes apparent, within which one is privileged over the other depending on the discipline and outcome of interest.

The current approaches to studying health-relevant behaviours are strongly analytical and variable-focussed. Although multiple behavioural risks are found to be associated [Blaxter 1990; DeRuiter et al. 2014; Loef & Walach 2012], most quantitative studies do not acknowledge the social situation in which behaviours occur either consciously or unconsciously. Drawing on Bourdieu’s theory, behaviours should be conceived of as part of practices structured by and constituent of habitus, mergeing agency and structure into routinised actions. These practices are instantiations of social milieus; they do not occur in a neutral context. Health-relevant behaviours thus become health-relevant, socially informed practices that are deeply rooted in an individual’s conscious and unconscious modes of living as parts of everyday life routines that follow their own context-specific, pragmatic logics [cf. Williams 1995, 582]. These pragmatic logics also extend to subjective notions of health, well-being, a healthy lifestyle, a healthy body.

For example, working class members have been found to emphasise physical functioning and absence of disease in their concept of health, whereas middle class members stresses an intrinsic value of fitness as a component of long-term well-being [Blaxter 2003, 71; Bourdieu 1984, 177]. The behavioural implications of these subjective orientation may differ: the former individuals prefer food that is filling rather than healthy and perhaps exercise to compensate smoking or drinking habits, while the latter eat healthy and smoke or drink only occasionally, exercise, too, but with a different motivation.

Some scholars hold that failure to account for these embodied logics has generated ineffective policy interventions [Baum & Fisher 2014]. Health promotion campaigns based on ’universal’ rationality have to date disappointed, if not achieved the unintended effect of exacerbating health inequalities as the narrative of campaigners matched the logics of economically and culturally resourced. As a consequence, individuals who were already better-off quitted smoking or volunteered for health screening programmes and not those who were in more urgent need [Baum & Fisher 2014; Horrocks & Johnson 2014, 215]. This finding also raises an important question for geodemographics: if the technique is to inform policy effectively, subjective dimensions and structural properties need to be balanced.

Bourdieu's work has inspired a generation of studies designed to identify and characterise individual lifestyle milieus. In the private domain, customer segmentation is a well established technique applied to this end, often complementing commercial geodemographics [Webber 2004]. Indeed, the explicit purpose of geodemographic classifications is to infer lifestyle milieus, and commercial products provide a highly detailed characterisation of values, subjective orientations and activity patterns [Burrows & Gane 2006; Parker et al. 2007; Savage & Burrows 2007].

In social science, milieu studies have been employed to study social class under a stronger aspect of identity-forming elements of daily practice. Bourdieu's own study of "Distinction" draws on survey and interview data to characterise different milieus in France based on moral values, taste and leisure activities [Bourdieu 1984]. More recently, in the UK, Bennett et al [2009] have applied a similar approach to investigate and characterise lifestyle milieus based on cultural capital. In a similar vein, Savage et al [2013] propose an alternative class model for Britain. There are other applications of Bourdieu's concepts in more specific domains, such as education [see Serre & Wagner 2015 for a review]; in social epidemiology and health geography they are still rare and are often carried out in qualitative designs.

Some of the measured aspects in those studies – sports, exercising and diet – directly link to health, but, essentially, they serve as revealed expressions of embodied, attitudinal and behavioural dispositions that render certain directly health-relevant practices more or less likely. An individual that is socially integrated, perhaps attached to the local neighbourhood, and engaged in the community is likely to engage in different practices, encounter different resources, and exercise different levels of claims and control than another one that is more isolated. Associated behavioural and psycho-social pathways are therefore differentially activated or deactivated within their social and behavioural context. Since a milieu study approach admits a place for subjective orientations in wider representations of populations, there may be potential in combining it with neighbourhood classifications and thus enhance the value of geodemographics as a socio-spatial hermeneutic and strategic input into policy.

## 6.2 Data and methods to specify vulnerable milieus

### Components of subjective vulnerability

In view of the rarity of milieu studies for health research, there is no agreed method to identify empirical dimensions of behavioural milieus. Here, social health pathways are chosen as a reference. Brunner and Marmot [2006] distinguish three types of social pathways: material, behavioural and psycho-social. Material pathways predominantly refer to structural conditions – economic capital, quality of physical environment, access to health care; the remaining two refer to individual behaviour and psychological conditions. Brunner and Marmot refer to the idea of anatomy and argue that there

Table 6.1: Dimensions of vulnerability and their position to health.

| pos | dimension | examples | pathway | studies |
|---|---|---|---|---|
| 1 | health-related activities | smoking habits, diet, physical activity | behavioural | Blaxter 1990; Stringhini et al. 2010 |
| 2 | perceived social security | financial prospects, job prospects, life satisfaction | psycho-social | Wilkinson & Pickett 2010 |
| 3 | perceived social position | control over life, job task | psycho-social | Marmot et al. 1991; Stansfeld et al. 1997 |
| 4 | social support | confiding emotional support, local support | psycho-social | Carpiano 2006, 2007; Marmot et al. 1991 |
| 5 | social participation | visits of friends, relatives | behavioural | Marmot et al. 1991 |
| 6 | civic orientation | political competence, interest, perceived benefits | psycho-social | Frohlich & Abel 2014 |
| 7 | civic participation | community engagement, voting, volunteering | behavioural | Frohlich & Abel 2014 |
| 8 | cultural participation | leisure, visit of events, museums | behavioural | Bennett et al. 2009; Bourdieu 1984 |
| 9 | communication | language, ICT, media use | behavioural | Bourdieu 1984 |

are more "proximal" and "distal" determinants of health. This typology is useful to arrange subjective orientations on a continuum of positions or degrees of direct impact on health [Table 6.1].

Within a loose concept of lifestyles, the health-related activities are most proximal to health, since smoking and drinking directly affect organs and thus may become manifest in health outcomes. A broad category of social stress encompasses perceived prospects in material and psychological terms. Fear of immediate loss, perceived risk of insecurity and life satisfaction can have a direct impact on well-being and mental health, although it may not lead directly to organ impairment as health-related activities do [Marmot et al. 1991]. Thus, social stress is positioned one step away from those activities.

A sense of autonomy results from reflexive perceptions of one's own position in society. Inequality can be experienced in a number of ways, for example through lack of control over domestic or professional affairs, such as decision making in the household or job tasks [Wilkinson & Pickett 2010]. Sense of autonomy may directly influence well-being or encourage behaviours that do not submit to external forces over one's own personal and bodily needs. The impact on health is less direct than those of the previous two dimensions, yet it is an important element of psycho-social pathways. There is evidence that social support – or social capital in Bourdieu's terms – discourage unhealthy practices and have protective effects on health for some. Social support could come from the closest person in the form of confiding emotional support, from wider networks such as neighbours, friends or relatives or may be derived from local social cohesion. Neighbourhood attachment can be an expression of neighbourhood social capital with implications for healthy behaviour [Carpiano 2006, 2007]. Social participation is closely related: it describes actually chosen routines of an individual rather than perceived characteristics of the social environment. Visiting friends and relatives as part of leisure activities may indicate an active approach to maintaining social networks, an expression of embodied social position and a sense of social belonging.

Civic orientation operates at a more general level of political integration. The perceived ability to influence higher level decision making in one's own interest may again be an important expression of embodied social standing. Alienation, isolation and perceived powerlessness may not only induce social stress but also pose the risk that specific needs of certain social groups are not being cared for at a systemic level in the long run. This aspect is closely connected to actual civic participation and therefore constitutes an element of behavioural pathways. The two aspects are not independent in practice, but in theory they represent a psychological dimension of social integration and control vis-a-vis a behavioural dimension of active participation and the ability to air one's needs. Frohlich and Abel [2014] point at the institutional environment of neighbourhoods that allow people to make certain choices and take charge of their own needs given the various forms of capital available to them. They demonstrate the importance of civic participation in addressing systematic inequities in the resources for health and well-being [ibid., 208].

One step further away is cultural participation as an expression of individual orientations and taste. Cultural consumption is a more distal determinant that includes visits to museums, libraries or concerts and contributes to short-term pleasure, pursuit of personal interests as well as formation of cultural identity and belonging [Bennett et al. 2009; Bourdieu 1984]. Finally, communication can be a distal determinant of health and well-being. Language barriers (such as difficulties to speak the local language) may impede the expression of one's needs or attract discriminative treatment by others. Bourdieu has shown that news consumption varies across classes and is differently motivated [Bourdieu 1984, 440], for example, by the goal of entertainment or obtaining information. Knowledge about autonomous and healthy living is a by-product of news consumption in addition to general orientations in everyday life. Social media and Internet use have become central to modern life, and the digital dimension of being socially integrated is increasing. In addition, ICT becomes a potential channel for health-promoting marketing interventions, and while being distal, form part of the an individual's practical life context.

### Data and methods: The UK Understanding Society social survey

The ESRC-funded Understanding Society survey is a longitudinal lifestyle survey that has been conducted annually in the United Kingdom since 2009. The second wave of the survey collected data about more than 50,000 individuals in over 30,000 households.

Table 6.2: Relevant variables in the ESRC Understanding Society study.

| pos | composite | variable | description | module |
|---|---|---|---|---|
| 1 | nutrition | Usdairy | Usual type of dairy consumption | nutrition |
| 1 | nutrition | Usbread | Type of bread eats most frequently | nutrition |
| 1 | nutrition | Wkfruit | Days each week eat fruit | nutrition |
| 1 | nutrition | Wkvege | Days each week eat vegetables | nutrition |
| 1 | smoke | Ncigs | Usual no. of cigarettes smoked per day | smoking |
| 1 | smoke | Smcigs | Ever smoked cigarettes regularly | smoking |
| 1 | smoke | Smncigs | Number of cigarettes smoked in past | smoking |
| 1 | walk | Wlk30min | Number of days walked at least 30 minutes | physical activity |
| 1 | sports | Sportsfreq | Moderate intensity sports frequency | leisure, culture and sport |
| 1 | sports | Sports3freq | Mild intensity sports frequency | leisure, culture and sport |
| 4 | advice | Scopngbhc | Advice obtainable locally | neighbourhood (self-compl) |
| 4 | belong | Scopngbha | Belong to neighbourhood | neighbourhood (self-compl) |
| 4 | borrow | Scopngbhd | Can borrow things from neighbours | neighbourhood (self-compl) |
| 4 | dark | Crdark | Feel safe walking alone at night | local neighbourhood |
| 4 | friends | Scopngbhb | Local friends mean a lot | neighbourhood (self-compl) |
| 4 | improve | Scopngbhe | Willing to improve neighbourhood | neighbourhood (self-compl) |
| 4 | stay | Scopngbhf | Plan to stay in neighbourhood | neighbourhood (self-compl) |
| 4 | talk | Scopngbhh | Talk regularly to neighbours | neighbourhood (self-compl) |
| 4 | close | Closenum | How many close friends | social network |
| 4 | family | Simfam | Proportion of friends who are also family members | social network |
| 4 | local | Simarea | Proportion of friends living in local area | social network |
| 4 | network | Simage | Proportion of friends with similar age | social network |
| 4 | network | Simrace | Proportion of friends of same race | social network |
| 4 | network | Simeduc | Proportion of friends with similar level of education | social network |
| 4 | network | Simjob | Proportion of friends who have a job | social network |
| 5 | socnet | Netcht | Hours spent interacting with friends through social websites | social network |
| 6 | civic | Civicduty | Sense of civic duty | political engagement |
| 6 | polcomp | Poleff1 | Qualified to participate in politics | political self-efficacy |
| 6 | polcomp | Poleff2 | Better informed about politics | political self-efficacy |
| 6 | polcost | Polcost | Cost of political engagement | political engagement |
| 6 | polcyn | Poleff3 | Public officials don't care | political self-efficacy |
| 6 | polcyn | Poleff4 | Don't have a say in what government does | political self-efficacy |
| 6 | polinf | Perpolinf | Perceived political influence | political engagement |
| 6 | polit | Vote6 | Level of interest in politics | politics |
| 6 | voteben | Perbfts | Personal benefit in voting | political engagement |
| 6 | voteint | Voteintent | Voting intention | political engagement |
| 6 | votenorm | Votenorm | Voting as a social norm | political engagement |
| 7 | org | Orgm | Which organisations member of | groups and organisations |
| 7 | org | Orga | Active in organisations | groups and organisations |
| 7 | org | Orgmt | Member of organisations NSC | groups and organisations |

Table 6.2: (continued)

| pos | composite | variable | description | module |
|---|---|---|---|---|
| 7 | org | Orgat | Active in organisations NSC | groups and organisations |
| 7 | volun | Volfreq | Frequency of volunteering | voluntary work |
| 8 | arts1 | Arts1freq | Arts activities frequency | leisure, culture and sport |
| 8 | arts2 | Arts2freq | Arts events frequency | leisure, culture and sport |
| 8 | hist | Herfreq | Historical sites frequency | leisure, culture and sport |
| 8 | lib | Libfreq | Library frequency | leisure, culture and sport |
| 8 | musm | Musfreq | Museum frequency | leisure, culture and sport |
| 9 | news | Newsource | Sources of News | news and media use |
| 9 | tv | Tvhours | Hours of TV per weekday | news and media use |

The survey consists of core modules that are asked at each wave and additional modules that are asked less frequently [Knies 2014]. The majority is conducted through Computer-aided Telephone Interviews (CATI) and a subset of questions is asked by self-completed questionnaires. In 2013, a nurse visited a subset of respondents to take physical health measures for the development of biomarkers, which includes genetic sequencing. The survey provides detailed characteristics on respondents' social and economic circumstances, including household and geographical context.

The second and third waves of the survey provide a number of modules that are relevant as components of the subjective dimension of vulnerability [Table 6.2]. The survey holds information on self-reported activites: the frequency of smoking, some characteristics of the diet, frequency of walking and sports. The way this data is operationalised follows the approach of Stringhini et al [2010]. As for the remaining components, the survey provides a range of variables that may not cover the components completely but nevertheless provide some useful information to describe life context. Variables that pertained to the same aspects were combined to scales. The one dimension that is visibly underrepresented is social participation. One can merely take the extent of interaction with friends through social media; the frequency of visits or spontaneous face-to-face interactions are not recorded in the survey.

Some dimensions could not be included due to a high number of missing data. Questions about control over job tasks covers the dimension "perceived social position" and therefore only relates to respondents that are working – approximately half of the sample. On perceived social security, questions on general satisfaction are asked through separate method of self-completed questionnaire, for which the response rate is lower. Respondents which have missing observations on any of the relevant variables are excluded from further statistical processing, resulting in an analytical sample of 41,639 [Table 6.3].

A final list of 56 variables are combined to 35 composites for further analysis. First, Principal Components Analysis (PCA) is run on the scales. Strongly loading variables (with communalities of more than .3) tend to discriminate observations within a sample. It is also possible to use communalities of variables as inclusion criteria to form new, consistent scales [cf. Gorsuch 1983, 29]. Those scales are sometimes used as input

Table 6.3: The survey sample before and after application of exclusion criteria.

| | **wave 2** | **wave 3** | **common** |
|---|---|---|---|
| households | 30,508 | 27,782 | 26,137 |
| respondents | 54,597 | 49,739 | 44,178 |
| ... no missing responses | – | – | 41,639 |
| ... no missing geography | – | – | 41,559 |

variables in sample segmentation. In this application, PCA is used as a way to shortlist variables for the subsequent milieu segmentation.

The PCA generates five components with an eigenvalue of one or more [Table 6.4]. The first component shows high loadings on neighbourhood characteristics, indicating perceptions of high local support and social cohesion. High-loading variables include statements in relation to the possibility to seek advice, borrow or talk in the neighbourhood and the degree of personal local attachment. The second component indicates high levels of cultural participation. Visiting arts exhibitions and events, historical sites and museums are highly correlated. Two health-related activities correlate with the frequency of exercising. The component appears to measure a general orientation towards activity. A third component reflects positive civic orientation. Interest and knowledge in politics is combined with a strong belief in voting and a sense of civic duty. A fourth component indicates digital communication combined with a sense of safety in the local neighbourhood and higher levels of exercising. A fifth component reflects again civic orientation wherein political competence is coupled with political interest. The sixth component has an eigenvalue of below one, but nevertheless is useful in highlighting civic participation through volunteering and organisational membership, which does not seem to be directly correlated with civic orientation or cultural participation.

The PCA confirms that some of the identified empirical dimensions emerge from the survey‘s variables. They do not, however, contribute to much of the explained variance. A hybrid approach of variable selection has therefore been taken: those variables with communalities below .3 are candidates for exclusion because they can be expected to discriminate poorly. They are, however, retained, if they contribute to the selection of variables by plausibly capturing one of the empirical dimensions of vulnerability [see Table 6.1]. Informal sensitivity analysis based on alternative cluster solutions are carried out to validate the results. Overall the cluster solutions remain stable, but their interpretability improves after application of these exclusion criteria.

Subsequently, cluster analysis is applied to segment the analytical sample, who are present in both the second and third waves of the survey and do not have any missing observations on any of the variables. The chosen clustering consists of two steps: Ward's hierarchical clustering to identify cluster centres, which are then used as input for *k means* clustering [see technical specifications on page 86]. The resulting clusters are characterised with respect to respondent demographics, socio-economic and other circumstances including geographical distribution, housing context, alcohol consumption and health and well-being. Since, in the survey, alcohol consumption is only collected

Table 6.4: The principal components (PC) for all composite variables (n=41,639).

| **pos** | **composite** | **PC 1** | **2** | **3** | **4** | **5** | **6** |
|---|---|---|---|---|---|---|---|
| 1 | nutrition | .079 | .342 | .138 | .064 | .055 | .128 |
| 1 | smoke | -.018 | -.136 | -.094 | -.084 | .018 | -.082 |
| 1 | sports | -.002 | .352 | -.024 | .356 | .067 | .136 |
| 1 | walk | .016 | .096 | -.020 | .052 | .005 | .021 |
| 4 | advice | .774 | .015 | .018 | -.094 | -.011 | .009 |
| 4 | belong | .749 | .002 | .073 | -.066 | .007 | .038 |
| 4 | borrow | .639 | .047 | .003 | .156 | .025 | -.009 |
| 4 | close | .138 | .194 | .024 | .090 | .089 | .179 |
| 4 | dark | .103 | .063 | -.008 | .450 | .105 | .031 |
| 4 | family | .046 | -.080 | .024 | -.201 | -.038 | -.033 |
| 4 | friends | .798 | .017 | .055 | -.137 | .011 | .046 |
| 4 | improve | .485 | .081 | .097 | .156 | .051 | .071 |
| 4 | local | .192 | -.094 | -.019 | -.136 | -.045 | .030 |
| 4 | network | -.013 | .039 | -.018 | -.025 | .038 | .107 |
| 4 | stay | .530 | -.026 | .086 | -.068 | -.044 | .052 |
| 4 | talk | .687 | .022 | .059 | -.032 | -.017 | .055 |
| 5 | socnet | -.117 | .044 | -.092 | .341 | -.016 | -.034 |
| 6 | civic | .080 | .183 | .717 | .013 | .082 | .127 |
| 6 | polcomp | .011 | .148 | .273 | .175 | .617 | .073 |
| 6 | polcost | .024 | -.106 | -.016 | -.061 | -.297 | -.040 |
| 6 | polcyn | -.038 | -.134 | -.158 | -.139 | -.163 | -.059 |
| 6 | polinf | .038 | -.023 | .408 | -.050 | .134 | -.045 |
| 6 | polit | .011 | .215 | .405 | .096 | .661 | .100 |
| 6 | voteben | .069 | .050 | .617 | -.055 | .113 | .041 |
| 6 | voteint | .069 | .166 | .597 | -.015 | .230 | .064 |
| 6 | votenorm | .145 | -.040 | .191 | -.022 | -.076 | -.034 |
| 7 | org | .128 | .267 | .111 | .128 | .137 | .657 |
| 7 | volun | .067 | .204 | .035 | .061 | .039 | .418 |
| 8 | arts1 | .005 | .469 | .069 | .048 | .071 | .083 |
| 8 | arts2 | -.049 | .547 | .005 | .298 | .080 | .077 |
| 8 | hist | .024 | .652 | .064 | .180 | .124 | .096 |
| 8 | lib | .025 | .332 | .047 | .022 | .024 | .066 |
| 9 | musm | .015 | .639 | .069 | .148 | .123 | .084 |
| 9 | net | -.093 | .210 | -.038 | .608 | .082 | .037 |
| 9 | news | -.001 | .244 | .106 | .212 | .248 | .186 |
| 9 | tv | .025 | -.173 | .020 | -.394 | -.036 | -.056 |
| - | SS loadings | 3.355 | 2.213 | 1.830 | 1.418 | 1.199 | .828 |
| - | Proportion Variance | .093 | .061 | .051 | .039 | .033 | .023 |
| - | Cumulative Variance | .093 | .155 | .206 | .245 | .278 | .301 |

**interpretation:** social support (1); cultural participation (2); civic orientation (3); digital communication (4); political orientation (5); civic participation (6)

for a sub-sample, it is investigated as a contextual variable instead of being included as a cluster variable. The socio-demographic and contextual investigation is based on $\chi$-squared tests for categories and oneway ANOVAs coupled with Tukey post-hoc tests for comparing group means.

## 6.3 The social health milieus

After a series of tests, a ten cluster solution appears to best fit the data. Each cluster is investigated with respect to, first, their attitudinal and activity profiles [Figure 6.1] and, subsequently, their socio-demographic attributes, economic and other contextual characteristics [Table 6.5]. The evidence emerging from this data is used to suggest tentative profiles of each group, which serve to infer health pathways that may be activated and decide cluster labels.

### Cluster 1: Enduring Isolation

Individuals that belong to the *Enduring Isolation* milieu show unhealthy activities with higher than average smoking rates, lower scores on healthy nutrition and lower levels of physical activity in terms of both walking and exercising. Members of this cluster do not indicate any particular trend with respect to local support in their neighbourhood and they report the lowest level of close friendships compared to all clusters. They are not engaged in civic activities, do not deem political engagement beneficial, do not intend to vote and do not volunteer. Their cultural participation is low. They do not use the Internet or follow the news often; instead, their TV consumption is higher.

More than three fourths of this group are 45 or older, 35 per cent are 65 or older. Two in three have no educational qualification, one in five have completed GCSE level. They tend to receive lower earnings – half of the *Enduring Isolation* milieu receive 850 pounds or less per month and often have a larger welfare component. One in three is in full or part-time employment; the jobs tend to be elementary occupations. Two in five live in social housing. The milieu has the highest rate of separations with one in six being separated or divorced. One in four provide unpaid care for a person either living inside or outside the household.

Overall, the *Enduring Isolation* milieu shows an introvert, less physically and socially active lifestyle with signs of social exclusion but also limits to time budgets and leisure opportunities imposed by low incomes and care obligations. Activity patterns indicate unhealthy practices and the experienced exclusion likely enforces psycho-social pathways, which added to their material conditions contribute to a quick depletion of biological capital.

### Cluster 2: Unconcerned Starters

*Unconcerned Starters* are less physically active than average and score lower on the nutrition scale. They report lower levels of local support and neighbourhood cohesion and consistent with that lower local attachment. Their civic orientation is introvert, they neither express political interest nor inclination to vote. They do not volunteer or belong to any organisation. Whereas they sometimes engage arts activities and events, they do not visit historical sites or museums. Their media use is focussed on the Internet with an average TV and lower news consumption.

More than half are between 16 and 34, only six per cent are 65 or older. The most common level of attainment is GCSE level with 43 per cent, followed by no formal qualification with 19 per cent. Lower incomes dominate with more than 60 per cent earning 1,150 pounds or less per month. 56 per cent are in employment and 15 per cent are unemployed. One in ten are full-time students. Nearly half work in lower status occupations. 30 per cent live in social housing, 35 per cent pay mortgage on their home. More than half of the milieu members are single and another 32 per cent are married or in a civil partnership – the lowest share among all clusters.

*Unconcerned Starters* are younger individuals that show little engagement with their local social environment, the civic or political sphere. They tend to pursue activities that are personal (arts activities) rather than communal (museums). They are not concerned about healthy living. Altogether, it is likely that behavioural and psycho-social pathways of vulnerability are activated in this group. Perceived disadvantage and perhaps stigmatisation in social interactions may contribute to a quicker decline of health and well-being as they progress through the life course.

#### Cluster 3: Retiring Generation

The *Retiring Generation* milieu shows diverse scores on the nutrition scale and has low smoking rates. They are physically inactive with low frequency of walking and exercising. They are locally attached and attest a supportive social environment in their neighbourhoods albeit not overly positive. They show an average level of civic orientation – with a strong intention to vote – and tend to be politically informed. They do not volunteer or are members of any organisations. Their cultural participation is very low, too, and in the area of communications TV consumption dominates, whereas news and Internet use play no part.

Nearly half are 75 years or older, three quarters are at retirement age. Two thirds do not have an educational qualification; this generation could not benefit from the educational expansion, which probably began after they reached adulthood. 65 per cent are women, more than half are married and 30 per cent are widowed. Their personal incomes are low with 68 per cent receiving less than 1,150 pounds per month. 80 per cent are retired and still 13 per cent are in employment. Their social status is derived from lower-status occupations, such as routine work. The majority (60 per cent) are outright owners of their property and one in four live in social housing.

The inactive lifestyles of the *Retiring Generation* milieu becomes comprehensible when viewed alongside their socio-demographics. Activity and behavioural patterns reflect the overlay between social position, partly historically determined, and age. The infor-

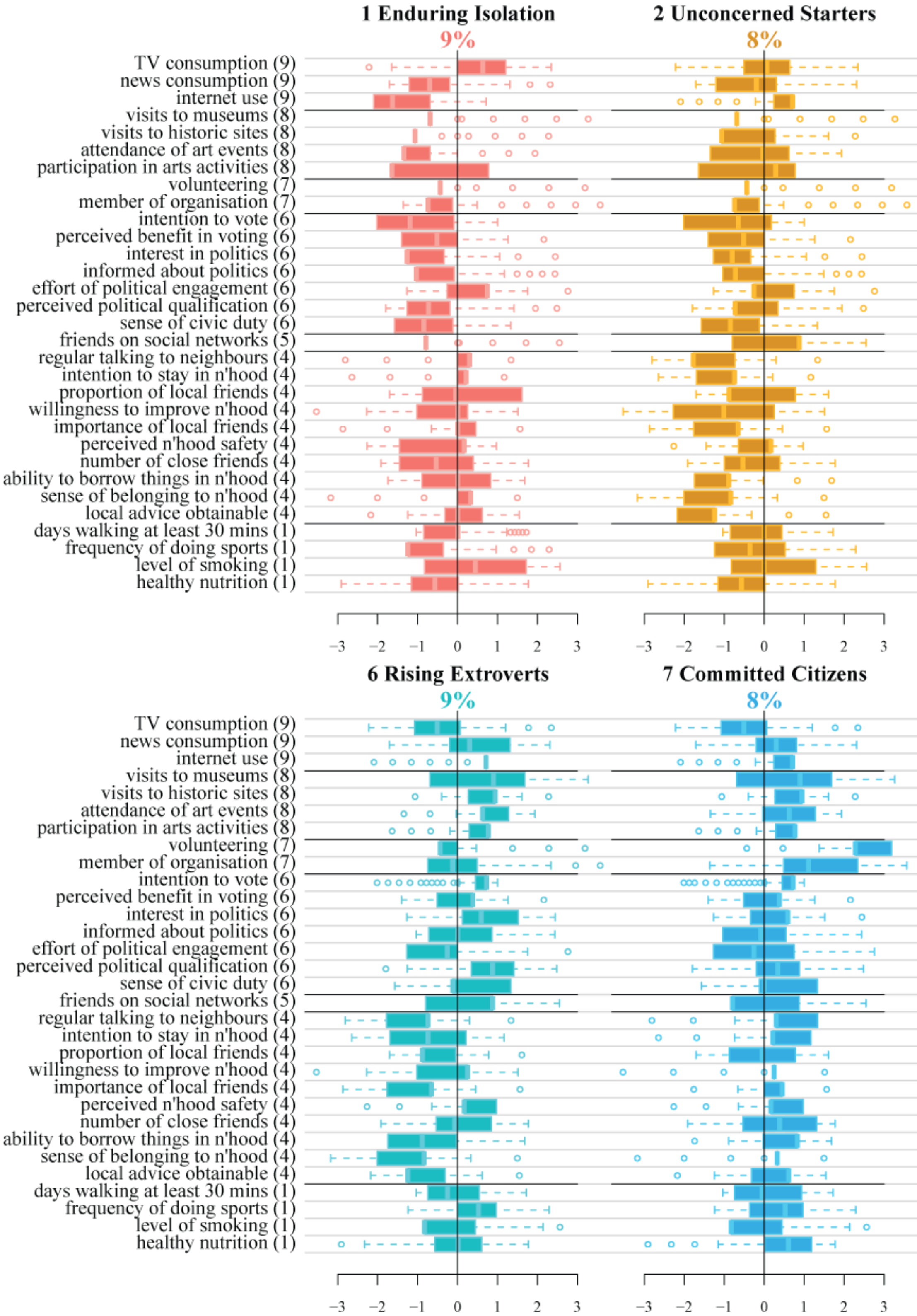



Figure 6.1: Attitiudinal and behavioural profiles of milieus.

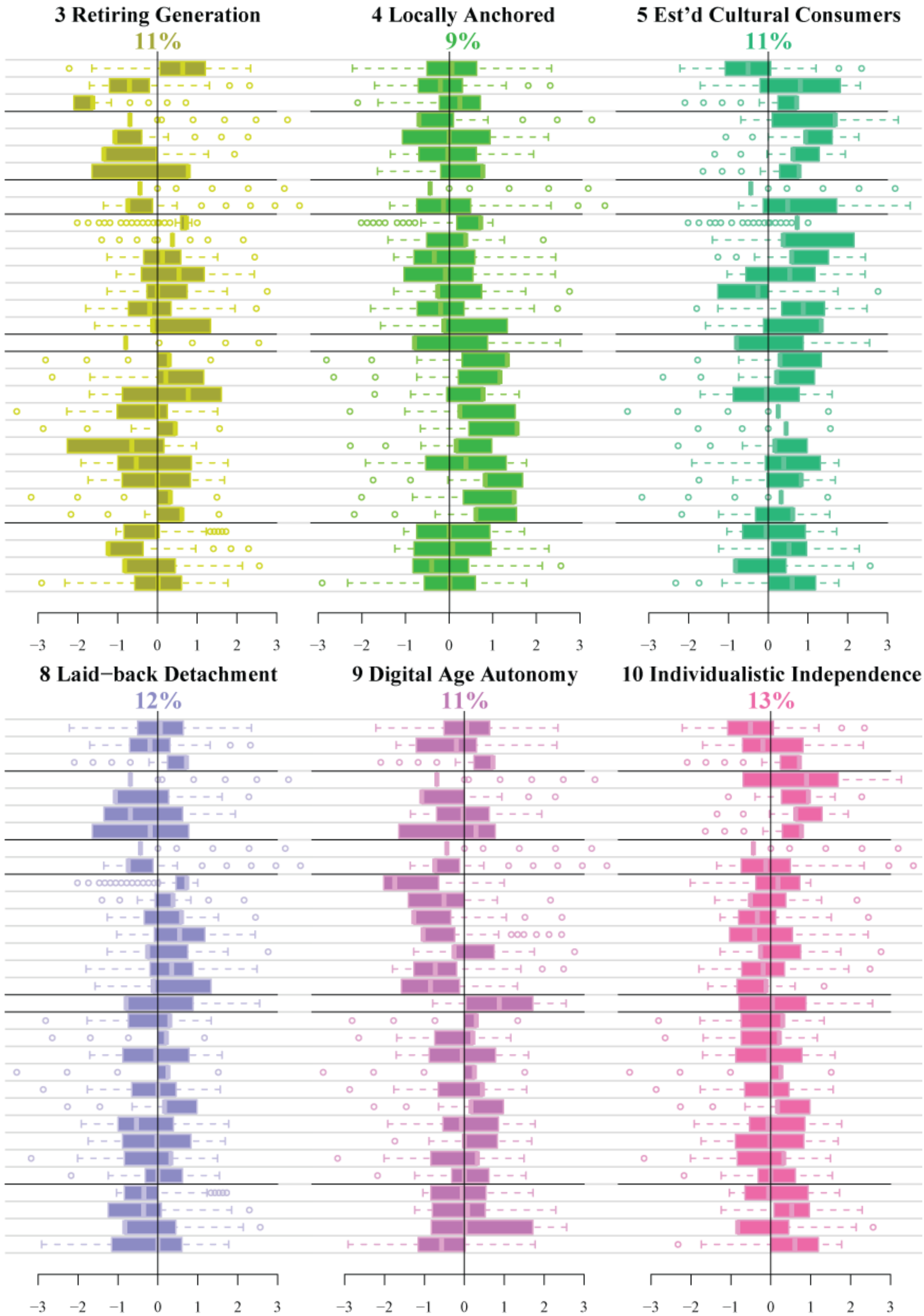



Figure 6.1 (continued)

Table 6.5: The milieus' demographic, economic, household and social characteristics.

| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **p** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | **age** | | | | | .000 |
| 16-24 | 3.4 | 29.2 | 0.4 | 7.0 | 4.2 | 22.6 | 5.1 | 7.6 | 27.4 | 12.1 | |
| 25-34 | 7.2 | 22.3 | 0.9 | 11.0 | 11.6 | 24.5 | 6.9 | 16.1 | 27.6 | 18.9 | |
| 35-44 | 11.5 | 17.8 | 2.6 | 18.8 | 22.2 | 18.5 | 15.1 | 20.9 | 22.5 | 24.2 | |
| 45-54 | 21.1 | 16.7 | 6.2 | 21.0 | 21.4 | 18.5 | 18.5 | 22.1 | 14.7 | 20.9 | |
| 55-64 | 21.6 | 7.9 | 16.1 | 19.4 | 21.9 | 10.6 | 21.7 | 18.7 | 5.8 | 15.0 | |
| 65-74 | 17.6 | 3.6 | 26.9 | 16.1 | 13.1 | 4.1 | 21.6 | 10.0 | 1.7 | 7.1 | |
| 75 or more | 17.5 | 2.5 | 46.9 | 6.9 | 5.7 | 1.3 | 11.0 | 4.7 | 0.4 | 1.9 | |
| | | | | | | **ethnicity** | | | | | .000 |
| white British | 89.9 | 87.4 | 89.3 | 91.2 | 90.1 | 86.2 | 91.7 | 81.0 | 89.9 | 89.1 | |
| white other | 4.5 | 4.2 | 4.0 | 3.3 | 5.8 | 6.1 | 4.3 | 4.4 | 3.8 | 4.9 | |
| Asian | 3.5 | 4.1 | 4.6 | 4.0 | 2.1 | 3.8 | 1.9 | 9.1 | 3.6 | 4.1 | |
| African, Caribbean | 1.7 | 3.4 | 1.5 | 0.9 | 1.4 | 2.7 | 1.7 | 4.2 | 2.1 | 1.2 | |
| other | 0.5 | 0.9 | 0.6 | 0.6 | 0.6 | 1.1 | 0.4 | 1.3 | 0.5 | 0.7 | |
| | | | | | | **household characteristics: children** | | | | | .000 |
| 1 child | 18.6 | 18.6 | 10.8 | 17.4 | 16.3 | 15.1 | 12.0 | 20.5 | 20.5 | 17.7 | |
| 2 or more | 16.6 | 22.6 | 5.7 | 26.0 | 22.9 | 16.1 | 22.3 | 26.8 | 32.1 | 26.6 | |
| | | | | | | **household obligations: unpaid care** | | | | | .000 |
| in household | 15.5 | 8.1 | 17.0 | 8.3 | 5.1 | 4.3 | 7.5 | 9.2 | 6.2 | 4.9 | |
| outside | 9.2 | 7.9 | 8.8 | 14.6 | 16.2 | 9.2 | 20.0 | 11.0 | 9.6 | 13.2 | |
| | | | | | | **marital status** | | | | | .000 |
| single | 22.5 | 52.5 | 8.4 | 20.3 | 22.2 | 49.5 | 16.4 | 24.9 | 52.0 | 32.2 | |
| married/CP | 46.0 | 32.1 | 51.8 | 62.5 | 63.2 | 38.8 | 67.3 | 61.8 | 35.6 | 55.6 | |
| separated | 16.8 | 12.4 | 10.3 | 10.8 | 9.7 | 10.4 | 8.5 | 10.2 | 11.2 | 10.4 | |
| widowed | 14.7 | 3.0 | 29.5 | 6.4 | 4.8 | 1.4 | 7.8 | 3.0 | 1.2 | 1.9 | |
| | | | | | | **sex** | | | | | .000 |
| female | 56.9 | 57.9 | 64.6 | 57.6 | 50.1 | 45.0 | 58.0 | 41.2 | 57.5 | 56.9 | |
| | | | | | | **tenure** | | | | | .000 |
| owned | 31.1 | 14.6 | 60.2 | 40.6 | 39.0 | 20.9 | 51.8 | 29.8 | 11.8 | 28.7 | |
| mortgage | 17.6 | 35.6 | 9.5 | 40.5 | 48.1 | 46.5 | 35.3 | 44.4 | 43.3 | 52.2 | |
| social rented | 40.6 | 29.6 | 24.6 | 11.9 | 3.7 | 9.5 | 5.7 | 14.5 | 26.9 | 6.6 | |
| private rented | 9.7 | 19.5 | 4.9 | 5.9 | 8.5 | 22.2 | 6.0 | 10.2 | 17.2 | 11.6 | |

mation suggests that members of this milieu literally retire from a long life of work and hard-earned achievements of sufficient material security. It seems that their health is likely an outcome of past behavioural necessities and to a lesser degree psycho-social factors.

### Cluster 4: Locally Anchored

Average health 'behaviours' with lower levels of smoking prevail in the milieu of *Locally Anchored* citizens. Their ratings of the social environment express strong local attach-

Table 6.5 (continued)

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | p |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | **economic activity** | | | | | | .000 |
| employed | 31.2 | 56.1 | 12.9 | 57.9 | 66.4 | 70.0 | 47.3 | 63.9 | 62.1 | 71.8 | |
| unemployed | 21.7 | 15.8 | 6.1 | 6.0 | 2.6 | 4.8 | 3.9 | 9.3 | 13.2 | 4.3 | |
| retired | 37.7 | 7.6 | 76.6 | 26.0 | 23.1 | 6.6 | 38.8 | 16.8 | 3.1 | 11.6 | |
| student | 0.4 | 11.2 | 0.1 | 3.0 | 3.5 | 14.7 | 3.3 | 3.8 | 10.1 | 6.8 | |
| other | 8.9 | 9.3 | 4.3 | 7.0 | 4.3 | 3.9 | 6.6 | 6.2 | 11.5 | 5.4 | |
| | | | | | **education** | | | | | | .000 |
| postgraduate | 0.7 | 4.6 | 1.0 | 5.2 | 22.3 | 19.9 | 15.3 | 8.4 | 2.7 | 10.9 | |
| professional | 4.4 | 17.5 | 8.1 | 20.7 | 41.7 | 37.0 | 40.8 | 21.5 | 14.3 | 31.8 | |
| A levels | 2.8 | 11.6 | 2.3 | 8.1 | 9.2 | 15.9 | 9.9 | 9.7 | 9.3 | 11.8 | |
| GCSE | 20.5 | 42.6 | 11.5 | 33.1 | 16.4 | 21.3 | 20.2 | 34.3 | 50.8 | 32.4 | |
| lower or other | 71.6 | 23.7 | 77.1 | 32.8 | 10.4 | 6.0 | 13.9 | 26.1 | 23.0 | 13.1 | |
| | | | | | **monthly gross income (pounds)** | | | | | | .000 |
| 0 -550 | 17.7 | 14.9 | 11.7 | 11.2 | 4.8 | 7.7 | 7.3 | 9.6 | 17.3 | 6.7 | |
| >550 – 850 | 30.7 | 27.8 | 30.7 | 20.8 | 10.7 | 12.9 | 16.5 | 22.4 | 27.4 | 16.9 | |
| >850 – 1,150 | 25.7 | 19.6 | 26.1 | 21.3 | 14.6 | 15.6 | 18.5 | 20.4 | 20.5 | 17.8 | |
| > 1,150 – 1,450 | 12.4 | 14.0 | 15.1 | 16.0 | 15.5 | 13.8 | 16.9 | 15.7 | 13.7 | 16.4 | |
| > 1,450 – 1,950 | 9.6 | 12.5 | 10.5 | 15.8 | 19.4 | 17.2 | 17.9 | 15.8 | 12.2 | 16.5 | |
| more than 1,950 | 4.0 | 11.2 | 5.9 | 14.9 | 34.9 | 32.9 | 23.0 | 16.1 | 9.0 | 25.6 | |
| average income | 991 | 1,166 | 1,094 | 1,326 | 1,949 | 1,791 | 1,581 | 1,333 | 1,078 | 1,601 | .000 |
| statistically same as cluster | 3 | 9 | 1,9 | 8 | - | - | 10 | 4 | 2,3 | 7 | |
| | | | | | **occupational status (NSSEC)** | | | | | | .000 |
| senior professional | 12.9 | 20.4 | 19.6 | 28.7 | 54.4 | 48.3 | 46.6 | 32.8 | 18.6 | 41.5 | |
| intermediate | 8.0 | 11.8 | 9.4 | 13.0 | 10.9 | 11.9 | 11.6 | 11.5 | 10.7 | 13.9 | |
| small employer | 10.4 | 6.7 | 11.4 | 11.1 | 8.9 | 4.9 | 9.2 | 9.8 | 7.9 | 8.5 | |
| lower skilled | 9.5 | 7.7 | 8.6 | 9.2 | 4.5 | 4.0 | 5.2 | 7.8 | 9.3 | 5.9 | |
| routine | 55.8 | 39.6 | 46.8 | 32.5 | 15.5 | 15.5 | 22.0 | 32.4 | 41.6 | 22.0 | |
| unemployed | 3.1 | 12.4 | 3.6 | 4.3 | 4.7 | 14.9 | 4.8 | 4.9 | 10.6 | 7.5 | |

**milieus in columns**: 1 Enduring Isolation, 2 Unconcerned Starters, 3 Retiring Generation, 4 Locally Anchored, 5 Established Cultural Consumers, 6 Rising Extroverts, 7 Committed Citizens, 8 Laid-back Detachment, 9 Digital Age Autonomy, 10 Individualistic Independence.

ment and engagement in a rooted community. In particular, they value local friends and a desire to stay and remain in their neighbourhoods. They have diverse civic orientations with a tendency towards disinterest in wider political affairs, although they intend to vote. While they participate in arts activities, museums are not popular in this milieu. Their media use is diverse; they frequently use the Internet, though less often for social interaction.

There is a slight majority of women in this cluster with 58 per cent. 60 per cent are between 35 and 64. Seven in ten live with their partners. Nearly half of this group have one or more children and one in five provide unpaid care for another person in or outside the household. The distribution of qualifications is split into thirds: one third has lower or no qualifications, one third has GCSE and another has A levels or above. Their incomes are at a medium level: 60 per cent earn between 550 and 1,450 pound per month. 60 per cent are in employment and one in four is retired. Eight in ten either own their property or pay a mortgage.

*Locally Anchored* citizens exhibit a strong local orientation in their lifestyle and are engaged in local communities. This may be a voluntary action or an outcome of their care obligations. As they are approaching retirement, they exhibit a more active lifestyle, in which personal health may not be the most important priority. Behavioural pathways are very relevant to their health, while psycho-social aspects tend to work in their favour: their local integration and strong engagement contributes to a high level of subjective well-being.

### Cluster 5: Established Cultural Consumers

The milieu of *Established Cultural Consumers* show higher levels of exercising and a healthy diet. Their levels of smoking are very low, with the majority indicating that they do not smoke at all. They value characteristics of the local social environment as positive and show a mild degree of local orientation. Local friends are important but the majority of their friends are not local, pointing towards action spaces that transgress the local neighbourhood. Their civic orientation stresses involvement and being informed. They intend to vote at the next elections, although they do not necessarily support the idea that voting is associated with personal satisfaction. Their cultural participation is high on all counts: arts activities, events, historical sites and museums. The Internet is a central component of their media use and so is their following the news, while their TV consumption is below average.

Two in three are between 35 and 64 and more than 70 per cent live with their partner. Their levels of educational attainment are high: 64 per cent have a qualification from a university, one in five have a postgraduate degree. Their mean income is the highest compared to all clusters, one in three earn 1,950 pounds per month or above. Nearly 90 per cent live in their own property either as outright owners or mortgage payers. Two thirds of *Established Cultural Consumers* are employed, one fourth are retired. 70 per cent have jobs with senior functions.

*Established Cultural Consumers* form an affluent milieu that have managed to consolidate their privileged position through the life course. High qualification and a successful career have enabled them to lead a materially comfortable and active lifestyle marked by high levels of cultural consumption. They are able to build assets to maintain good health. Psycho-social pathways operate upwards: through personal success and superior positions, the milieu provides grounds for positive outlooks of autonomy and control over one's own life.

### Cluster 6: Rising Extroverts

*Rising Extroverts* rarely smoke and show diverse scores on the nutrition scale. They walk less often compared to other clusters, which they compensate by high levels of exercising in their free time. They show little positive rating of their local environments; overall, they are not locally orientated. Local friends are not important to them and they report that their friends are generally not local. They are politically interested and view themselves as competent in political matters. They express the intention to vote at the next election. They do not volunteer, although some are members of organisations. They participate in cultural events and activities. Their Internet use is high and their TV consumption is low; they follow the news regularly.

This is a younger milieu with two thirds below the age of 45. Half of the milieu are single and another half live with their partner. Nearly three quarters of *Rising Extroverts* have A-levels qualification or higher, and 15 per cent are still full-time students. 70 per cent are in full or part-time employment. They earn higher salaries; the average income of this group is at nearly 1,800 pounds. Half of this milieu pay mortgage, one in five own their homes and 22 per cent live in private rental flats – this is the highest proportion of private renters compared to all other clusters.

This milieu expresses outward orientations that go beyond the local; indeed, the local environment is of minor importance. Their attitudes express an extrovert and confident manner of living with high level of activities to pursue personal interests. Sports is important in this milieu. Disinterest in the local and striving for success may contribute to valuation of strength and fitness in their idea of health. Psycho-social aspects play in their favour, their positive prospects generate potential for material and psychological well-being.

### Cluster 7: Committed Citizens

*Committed Citizens* tend to eat healthily, they do not smoke and many of them exercise regularly. They show a moderate orientation towards their local environments; they indicate that they can draw on local social support. Their civic orientation is more diverse; they view themselves as competent in political matters but they are not necessarily informed about them. What stands out about this milieu is their level of civic participation: volunteering is an important component in their lifestyles, and they often belong to multiple organisations. Their cultural participation is medium to higher than the average. The Internet plays an important part in their communication, and TV does not.

The age profile of this milieu is dominated by people between 45 and 74, who make up a proportion of more than 60 per cent. 58 per cent are women. 56 per cent have a degree from a university, but only 15 per cent have a postgraduate degree. Their income levels are medium to high: 40 per cent earn more than 1,450 pounds a month. 40 per cent are retired. 87 per cent of *Committed Citizens* live in their own properties; 60 per cent of them are mortgage-free. One in five provide unpaid care outside their household, and another 7.5 per cent cares for a person living inside the household. This level of providing care is the highest among all milieus. 70 per cent live with their partner, 22 per cent are with children. The vast majority (87 per cent) report 'good' health or better; yet, nearly 40 per cent have at least one disability.

*Committed Citizens* show high level of voluntary engagement; they do not only volunteer in organisations but are also providing the highest level of care for another person who typically lives outside the household. Their cultural participation is higher but not as high as other clusters, probably because a large part of the time budget is spent on volunteering and caring. They lead an active lifestyle with many commitments that are not for personal benefit in the first place. Behavioural pathways may play in their favour, while psycho-social pathways are not likely to be significant in this group.

### Cluster 8: Laid-back Detachment

The milieu of *Laid-back Detachment* show low levels of physical activity and low propensity to smoking. They score diversely on the nutrition scale. They express an average valuation of the local social networks, have fewer close friends and seem to be almost indifferent towards their immediate social environment. They indicate moderate levels of interest in civic matters and they regard voting as beneficial. Their cultural participation is lower, although there is some variation with respect to artistic activities. The Internet is important in this milieu's communication practices; their TV consumption is average.

43 per cent of this milieu are between 35 and 54. Another sixth are in the adjacent decennial age-bands each. With a proportion of 59 per cent, men slightly outbalance women. They have by far the highest share of people with Asian, African and Caribbean descent compared to all other clusters, amounting to a total of 13 per cent. The level of education is lower in this milieu: one in three have GCSE level and one in five have no formal qualification. This translates into their incomes: more than half gain 1,150 pounds or less per month. They are positioned at the middle range of the occupational scale: the majority has administrative, skilled trades, personal and customer service jobs. Three in four live in their own homes with the majority of them being on mortgage payments. Most often, members of this milieu live with their partners and nearly half have children.

The *Laid-back Detachment* milieu lead unhealthier lives with low levels of physical activity and low cultural participation. Media use, notably Internet, may constitute a major component of their leisure time. Their low orientation towards the local environment may indicate that sources of social support may not be readily available to them. Yet, they live in families with children more often than others, which also implies that, due to child care, time budgets are constrained to pursue other activities. Psycho-social pathways may be at work through relative detachment from the wider social environment. Limited time budgets and perhaps a domestic orientation may translate into low levels of walking and exercising.

#### Cluster 9: Digital Age Autonomy

The *Digital Age Autonomy* milieu scores low on the nutrition scale, indicating an unhealthy diet, and show diverse trends with respect to smoking and physical activity. They reveal a mildly positive rating of local social networks and moderate levels of caring about their neighbourhoods. They are not interested in politics and do not respond to the idea of voting as civic duty or act of wider social benefit. They do not volunteer or join organisations. Their more limited level of participation is focussed on arts activities, while museums are rarely frequented. Internet and social media form a central component in their communicative profile and probably drive their socialising.

More than half are between 16 and 34 – only two per cent are 65 or older. The milieu have the highest proportion of people with GCSE qualification (50 per cent), and another 22 per cent have lower or no qualifications. Their incomes are lower – two thirds earn up to 1,150 pounds per month. 62 per cent are in employment, 13 per cent are unemployed and another 12 per cent – the highest share compared to all other milieus – pursue other economic activity, which includes being a home maker. More than 60 per cent work in jobs associated with lower occupational status or are unemployed. Tenures distribute diversely on mortgage payment with 45 per cent, social housing with 27 per cent and the private rental sector with 17 per cent. While more than half are single, many of this milieu live with children either as unmarried couple or single parent. The proportion of single parents is highest in comparison: 27 per cent live either alone or in shared households. Another third live as a couple with children. In total, more than half are with children, one third with two or more – both constitute again the highest proportion in comparison.

Members of the *Digital Age Autonomy* use online technology, while other forms of social participation are absent. Their subjective orientations suggest a desire towards hedonism satisfied through ICT within a context of material constraints. In some cases, digital technology may constitute a distraction or substitute to unaffordable activities. At the same time, digital channels have come to be their principal connection to the social world, through which they emit and receive information to form their view of reality. In theory, ICT may be a promising channel for psychological interventions. Health is likely to be influenced through behavioural pathways that are specifically linked to communication and media use. Psycho-social pathways are also at work but may exert positive rather than negative effects because of perceived relative autonomy through digital technology.

### Cluster 10: Individualistic Independence

Members of the *Individualistic Independence* milieu lead healthier lives: nutrition is better, propensity to smoking is low, and physical activity is higher. They express less interest and attachment to their local social environments. Their civic orientation tends to de-emphasise collective values and moderate interest in political affairs. Voting intentions are diverse in this milieu. Volunteering is uncommon, while some indicate organisational membership probably to pursue personal pastimes. In terms of their cultural participation, arts activities and events as well as visits to historical sites feature highly and, to some extent, museums, too. Internet use is important in their communication, while the importance of social media differs within this milieu.

Nearly two thirds are between 25 and 54, and the gender distribution is mildly skewed towards women. Their education ranges from GCSE to university degrees, while the former and special professional qualifications are most common. Their incomes are medium to higher – one in three earns between 1,150 and 1,950 pounds per month and one in four more than that. With more than 70 per cent being in employment, they are among the most economically active milieus. Their occupational profile tends to be skewed towards higher-status occupations with two in three having administrative jobs or posts with more responsibility. More than half make mortgage payments – the highest share in comparison. The vast majority (70 per cent) live with their partner and 45 per cent have children.

Members of the *Individualistic Independence* milieu reveal an orientation towards personal freedom and autonomy rather than collective values and engagement in political affairs. They enjoy relative material success and have been able to progress into securer arrangements as regards property and family circumstances. Behavioural pathways are activated in positive ways – health and fitness is valued as something that expands life chances and personal prosperity. Similarly, psycho-social pathways act beneficially through experienced, successful choices and increased sense of being in control over their own lives.

### Alcohol consumption

Alcohol consumption varies significantly between the clusters. The milieus that consume least are *Retiring Generation* and *Enduring Isolation* with each more than half reporting that they drink alcohol on two days per month or less. Similarly, half of *Unconcerned Starters* drink monthly or less and a further 29 per cent drink a few times a week. They have the lowest share of respondents who drink on a daily basis. The reverse applies to *Established Cultural Consumers*; one in five report drinking on at least five days per week and another 23 per cent drink on three or four days per week. Altogether, with three in four *Established Cultural Consumers* drinking at least weekly, the milieu has the highest alcohol consumption compared to other milieus. The level of alcohol consumption is

Table 6.6: The milieus' alcohol consumption.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | p |
|---|---|---|---|---|---|---|---|---|---|---|---|
| at least 5 days/week | 11.6 | 7.0 | 13.2 | 15.5 | 21.1 | 13.6 | 18.5 | 14.7 | 8.4 | 13.9 | .000 |
| 3-4 days/week | 8.1 | 9.6 | 9.2 | 14.7 | 22.6 | 14.4 | 18.7 | 13.2 | 11.8 | 17.6 | |
| 1-2 days/week | 23.7 | 28.8 | 20.3 | 29.6 | 31.2 | 34.5 | 26.6 | 28.2 | 28.5 | 32.3 | |
| 1-2 days/month | 11.9 | 18.6 | 11.6 | 15.4 | 11.8 | 19.0 | 14.2 | 15.0 | 19.6 | 17.8 | |
| less often | 44.7 | 35.9 | 45.7 | 24.8 | 13.2 | 18.5 | 22.0 | 28.9 | 31.7 | 18.4 | |

**milieus in columns**: 1 Enduring Isolation, 2 Unconcerned Starters, 3 Retiring Generation, 4 Locally Anchored, 5 Established Cultural Consumers, 6 Rising Extroverts, 7 Committed Citizens, 8 Laid-back Detachment, 9 Digital Age Autonomy, 10 Individualistic Independence.

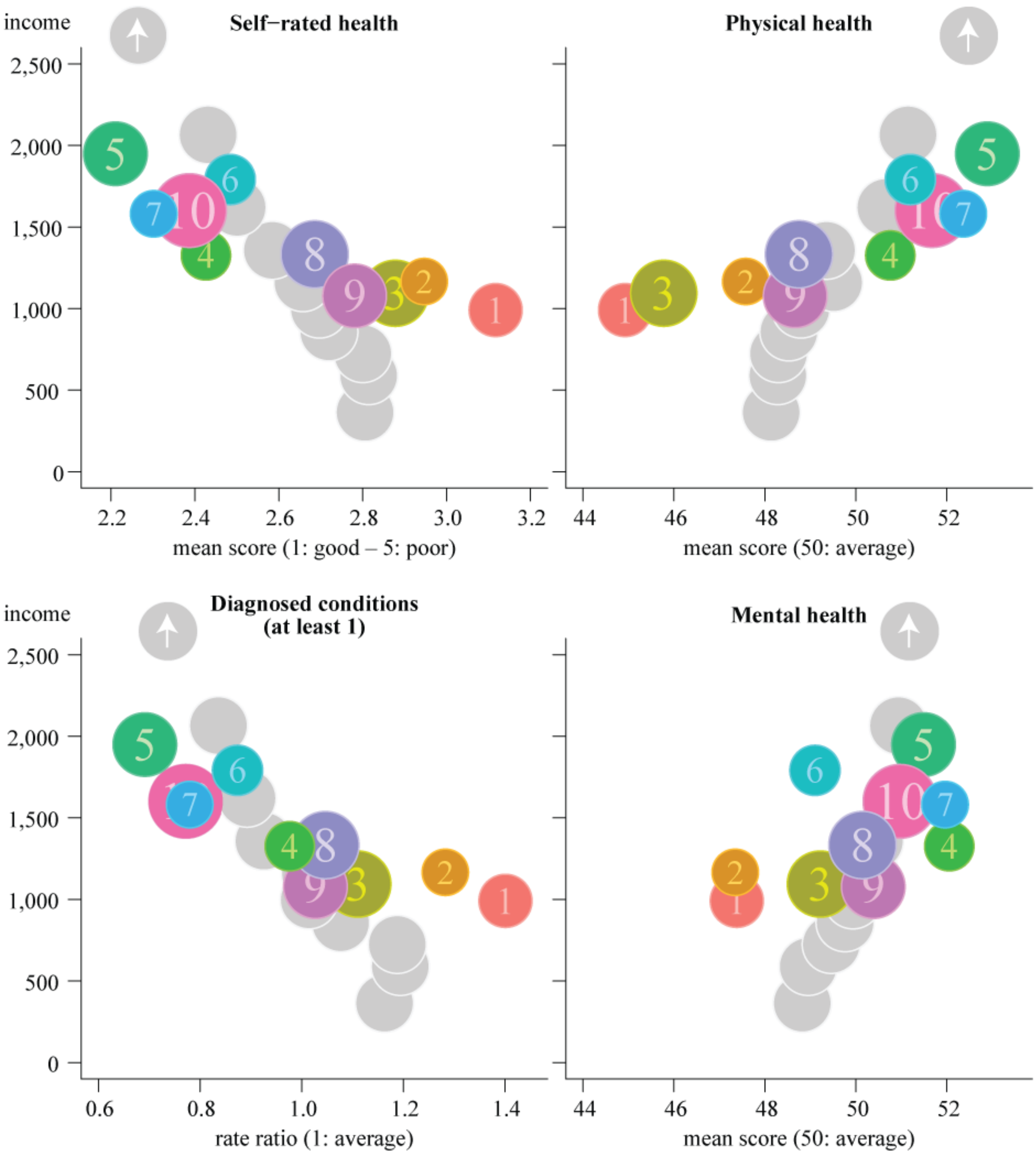
also higher among *Committed Citizens* and *Individualistic Independences*, whereas it is lower in the *Digital Age Autonomy* milieu. As a general rule, it seems more affluent groups consume alcohol more regularly than poorer groups, and differences in alcohol also correlate with the milieus' activity patterns. Yet, this result might be confounded by reporting bias, to which the groups may be unequally liable.

### Health milieus and the social gradient

The argument that subjective orientations shape health in addition to structural factors implies that health outcomes differ significantly between health milieus. Understanding Society provides variables of subjective and objective (though reported) health and well-being. The survey asks about self-rated health, hospital utilisation and any diagnosed conditions. Self-rated health ranges on a scale from one to five, where one is the best. The survey also includes a short health questionnaire (known as SF-12) with twelve questions on physical and mental functioning. The questions are, for example, whether current health limits certain activities, whether pain interfered with work, how a respondent has felt during the past four weeks. Respondents receive a score based on each question, which can be summed to a total standardised scale of physical or mental health with mean 50 and standard deviation ten [Ware et al. 2002]. Finally, a more objective health measure, the risk ratios of non-communicable conditions (NCC) can be obtained from a question of whether any of 16 selected conditions (asthma, diabetes, heart attack, stroke, etc) have been diagnosed in a respondent

Respondents are grouped into income deciles and alternatively into the ten milieus. Four age and sex-standardised health measures were selected to compare differential health outcomes of the ten milieus against what would be expected on the basis of the social gradient [Figure 6.2].

Self-rated health follows a nearly linear gradient: individuals with more income rate their health consistently better. The pattern observed with the milieu grouping exhibits additional differentiation between income groups. With a few exceptions, all pair-wise




**milieus**: 1 Enduring Isolation, 2 Unconcerned Starters, 3 Retiring Generation, 4 Locally Anchored, 5 Established Cultural Consumers, 6 Rising Extroverts, 7 Committed Citizens, 8 Laid-back Detachment, 9 Digital Age Autonomy, 10 Individualistic Independence.

Figure 6.2: Selected health indicators plotted against income deciles (grey) and milieus (coloured); the dots being scaled to group size.

average incomes between milieus are statistically significant, based on ANOVA Tukey post-hoc tests [cf also Table 6.5]. The *Locally Anchored* and *Committed Citizen* milieus show both better self-rated health than would be expected on the basis of the social gradient. *Rising Extroverts*, on the other hand, exhibit – relatively speaking – poorer self-rated health. There is also more differentiation in lower income milieus: based on similar incomes, *Enduring Isolation*, *Retiring Generation* and *Digital Age Autonomy* are far apart with the former revealing the strongest health disadvantage. The milieus also show more variation in health scores: the *Enduring Isolation* milieu reveals more inferior health (score 3.1) than the lowest income deciles (around 2.8), whose average incomes are between 350 and 750 pounds. The reverse is true for *Established Cultural Consumers*; on average, they report their health to be better than even the highest income decile with its average income of 3,700 pounds.

The physical health component reveals again a social gradient with only minor interruptions. The age and sex-standardised mean scores for the milieus, however, show a higher degree of variation. Similar to self-rated health, some milieus are better or worse-off than expected. The *Locally Anchored* and *Committed Citizens* as well as *Digital Age Autonomy* score higher on the physical health components than their similar income counterparts. As for mental health, the patterns are more diffuse. Here, *Locally Anchored* and *Committed Citizens* fare better in terms of mental health than even the highest income decile. Conversely, the *Enduring Isolation* and *Unconcerned Starters* fare worse than lowest income deciles. *Rising Extroverts* reveal again disadvantage, which seems to be more pronounced for mental health than for physical health. The *Digital Age Autonomy* milieu is on par with *Laid-back Detachment* despite lower income of the former.

Finally, in terms of diagnosed conditions, the milieus show more variation than income deciles, too. While the three lowest income deciles (earners of less than 789 pounds per month) are 1.2 times more likely to experience NCC than the average, the members of the *Enduring Isolation* milieu are 1.4 times more likely. *Unconcerned Starters* nearly have a 1.3-fold risk of NCC. The NCC risk ratio of more affluent milieus is below the average, but again *Rising Extroverts* are at a relative disadvantage given their income.

## 6.4 Regional variants of the social health milieus

The preceding study of health milieus are suggestive of differentially activated pathways within different health milieus in the UK-wide sample. In view of the evidence from the research on regional specificity [chapters 4 and 5], however, it may be expected that regional variants of the milieus may exist and produce specific expressions of vulnerability. National and regional classifications have both their respective merits. While the potentially free movement of citizens within a country might cause convergence of regional milieus so that a nation-wide classification is appropriate, the geographical distribution of surnames suggests that most people do not move over sufficiently long distances, and hence regions are likely to retain some distinctive characteristics.

The geography of isonymy is used to define regions by which the survey sample can be partitioned. Akin to the regionalisations presented previously, 2011 UK census wards are classified by surname compositions using 2011 Electoral Roll data, the wave of

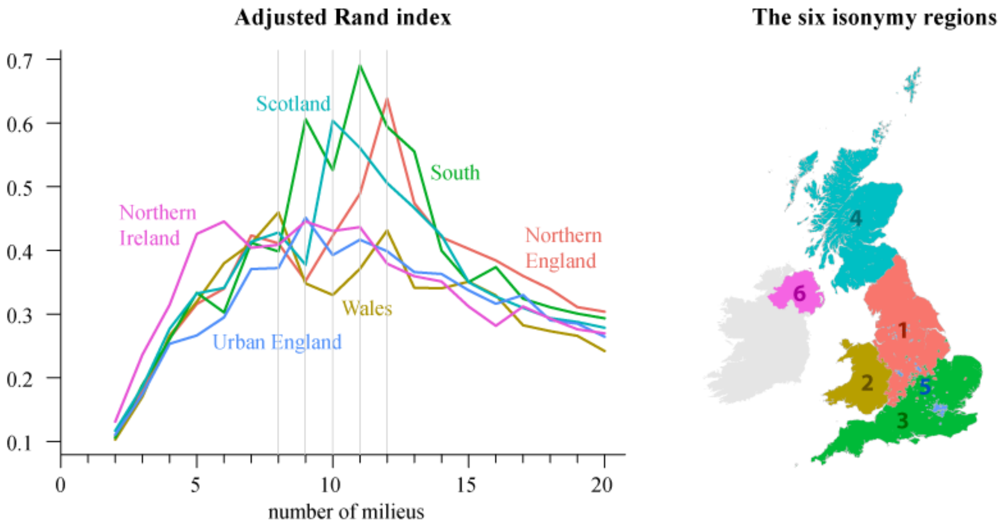



Figure 6.3: Agreement of different number of regional milieus with the national classification: Adjusted Rand index for different regional cluster solutions (left) and the isonymy groups for 2011 (right).

data which corresponds in time to waves two and three from Understanding Society. A number of internal clustering criteria for cluster solutions with $k$ groups varying between two and 40 suggests that six isonymy clusters produce the best solution [Figure 6.3 (right)].

The six emerging regions comprise a Northern Irish, Scottish, Welsh, Northern English and Southern English one. A sixth cluster is composed of non-contiguous English inner-city wards. A higher share of migrant names distinguishes English inner-city wards from their surroundings, suggesting that inner-city London, Birmingham, Leeds and Leicester are more similar in cultural, linguistic and ethnic terms. The centre and some suburban areas of Manchester are also part of this cluster, but not Liverpool, Newcastle, Glasgow, Southampton or Cardiff. On a side note, the corresponding interpretation of this cluster would be that English urban populations are culturally more similar across these cities (e.g. in inner Birmingham and inner London neighbourhoods) than the respective neighbouring population in their surrounding regions (central England, southern England). May this be evidence for the existence of a distinct English metropolitan identity [McDermott 2015; Savage et al. 2013; Webber 2007]? Linking observations in Understanding Society to surname geographies may provide insight into this question.

The strategy to select regional milieus is as follows. Each survey record is assigned to one of the six isonymy regions based on the geographical information provided in the survey. The sample is split accordingly into six regional subsamples. Each subsample is segmented into between two and 20 groups using the regionally centred (z score standardisation) milieu variables. The result is 19 milieu solutions for each regional subsample. In a next step, the overlap between regional milieu solutions and the ten-milieu national classification is determined by calculating common cases of each

Table 6.7: Statistical properties of regional milieus in relation to the national classification.

| region | sub-sample size (total=41,559) | best matching number of milieus | value of Adjusted Rand index |
|---|---|---|---|
| 1 Northern England | 12,248 | 12 | .638 |
| 2 Wales | 3,561 | 8 | .460 |
| 3 Southern England | 12,188 | 11 | .690 |
| 4 Scotland | 3,841 | 10 | .604 |
| 5 Urban England | 6,561 | 9 | .451 |
| 6 Northern Ireland | 3,160 | 9 | .445 |

class. The Adjusted Rand index [see chapter 4.2] is used to assess the correspondence between two partitions, in this case the regional milieus and the national classification [Figure 6.3 (left)]. In this way, for each isonymy region, the most agreeing milieu solution (defined by the number of milieus) with the national classification can be identified.

The regions in which milieu solutions tend to agree most with the nation-wide ten milieus are the larger central and southern English subsamples – they peak at $k = 12$ and $k = 11$ respectively with a value above .6. Scotland also peaks at the same level at $k = 10$. The remaining regions, Wales and Northern Ireland, as well as urban England show lower levels of agreement; their peak values lie below a value of .45 throughout the solutions.

Since the national classification is defined as the optimal reference, all regional subsamples have been clustered with the value of $k$ classes where the index peaks. Simple summary statistics as well as cluster distances reveal which individual regional milieus are closest to national milieus [Table 6.7]. It is then possible to focus on the national milieus and explore the additional regional variation.

In a next step, after having identified the most agreeing regional cluster solutions vis-à-vis the national classification, it needs to be determined, which regional milieus are more or less equivalent to which national milieu. This is done in two ways: by cross-tabulation to determine for each regional milieu the proportion of survey respondents that are classified in each national milieu [Figure 6.4 (left)] and by measuring the cluster centre distances – in statistical space – between each regional milieu and each national milieu [Figure 6.4 (right)]. Regional clusters are re-ordered to best match the order of national classes (one to ten) by their similarity in terms of shared respondents.

In the region of Northern England, the milieu solution that agrees most with the national milieus is one with twelve groups. More than 90 per cent of those respondents that belong to the first Northern England milieu also belong to *Enduring Isolation* (1) of the national classification [Figure 6.4(a)]. Approximately six per cent of respondents belong to the national *Digital Age Autonomy* (9).

Most of the Northern England milieus share at least 70 per cent of respondents with

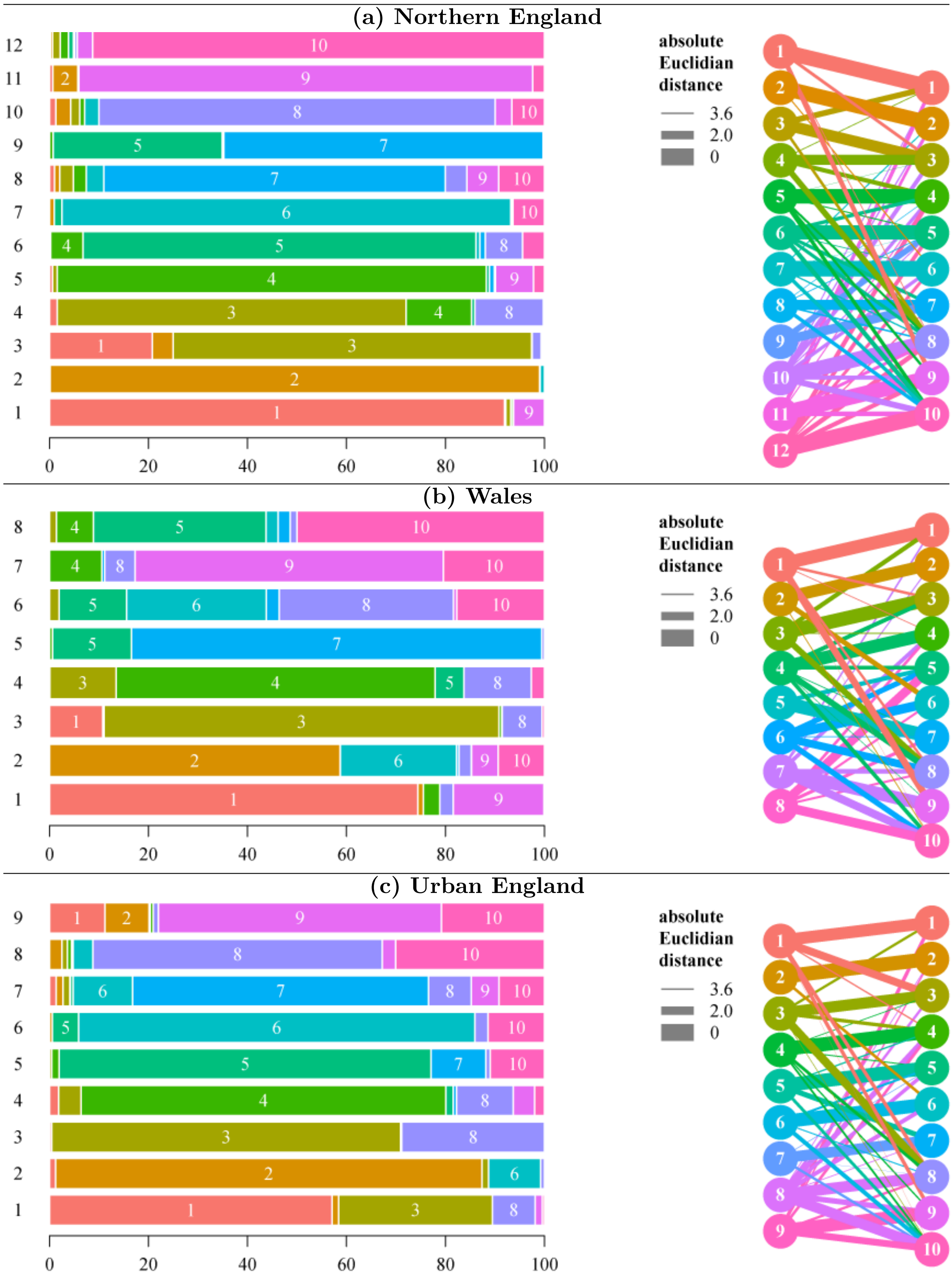



Figure 6.4: Similarity measures between selected regional milieus and the national reference classification: percentage of cross-classified cases (left) and absolute Euclidean distance from cluster centres (right).

any one milieu of the national classification. The only exception to this pattern are the eighth and ninth Northern England milieus, which indicate a split of *Committed Citizens* (7) into two types in Northern England. A closer look at these two milieus reveals that, while both groups stand out by their high degree of volunteering, in the ninth milieu this occurs with more cultural participation and a higher degree of local attachment [Figure 6.5]. The health-related activities of the latter are also significantly healthier than the other less active eighth milieu. A look at their socio-demographics reveals that the age and sex profiles of these two clusters are nearly identical. Their education and income levels, however, differ substantially: the more active ninth milieu occupies a higher position in this respect than the less active counterpart.

The similarities between milieus is also reflected in the cluster centre distances. The eights and ninth Northern England milieus both have strong links with the national *Committed Citizens* (7) milieu, although the ninth regional milieu also exhibits an additional link to *Established Cultural Consumers* (5).

The Welsh classification, peaking at $k = 8$, [Figure 6.4(b)] shows a lower level of agreement with the national clusters. While most of the national clusters reappear in the Welsh sample, *Rising Extroverts* (6) and *Laid-back Detachment* (8) do not appear in the same form in Wales. *Established Cultural Consumers* (5) and *Individualistic Independence* (10) are combined in the eighth Welsh cluster. The second Welsh cluster is most similar to the national *Unconcerned Starters* (2) and accommodates some of the *Rising Extroverts*. The cultural and economic capital of these milieus rise therefore in comparison to the national equivalent, while their age profile remains skewed towards younger ages. The sixth Welsh milieu combines cases from various national clusters with a high degree of civic orientation. In Wales, affirmative civic orientation tends to be a stronger driver across behavioural and socio-demographic contexts than in the remainder of the UK. The eighth Welsh milieu combines the most culturally participating respondents with healthy activities and diverse civic orientations. Civic and cultural orientations distinguish milieus in Wales more than in the United Kingdom as a whole.

For Urban England [Figure 6.4(c)], a nine cluster solutions compares best to the national classification. The *Enduring Isolation* (1) and *Retiring Generation* (3) milieus are combined in the first clusters and the third combines the *Retiring Generation* and *Laid-back Detachment* (8). This combination spreads the distinct age profiles of the former two milieus to the point that the label *Retiring Generation* may no longer be adequate for Urban England. The English urban *Enduring Isolation* milieu now shows complete absence in cultural activities with little variation. This applies to exercising, too. The third cluster shows higher scores on the nutrition scale, physical inactivity but is significantly more informed with respect to politics than urban English sample as a whole. This suggest the existence of a maturer (in age terms) and politically conscious working class in English cities.

There are other socio-demographic patterns that are specific to urban England [Table 6.8]. The income distribution between poorer and more affluent milieus indicates higher


Figure 6.5: Selected regional milieus indicating regional specificity.

inequalities. While in the national classification, *Established Cultural Consumers* receive an average income (including benefits) of 1,950 pounds and the *Enduring Isolation* 990 pounds, these figures are respectively 970 and 2,238 in Urban England. The *Rising Extroverts*' income, too, increase by 300 pounds per month and exceed 2,000. Higher living cost faced by the Urban England sub-sample may further exacerbate these inequalities. In addition, ethnicity is now a more significant component of each milieu's socio-demographic profile. The most affluent fifth and sixth clusters have with 70 per cent the highest share of white British respondents, this share drops to 48 per cent in the *Retiring Generation* equivalent and 45 per cent in the *Laid-back Detachment* milieu. One in three of the latter and in the *Locally Anchored* milieu are of Asian descent.

Comparing selected health summary measures of the urban English milieus against their income levels reveals further specific variations of the social gradient in health [Figure 6.6]. As for self-rated health, the urban-English milieus indicate similar levels of health scores – even the affluent milieus with their higher nominal incomes. This pattern is repeated for the physical health component scores. On mental health component scores, the range increases and in contrast to the national milieus, *Unconcerned Starters* fare significantly worse than *Enduring Isolation*. The score drops below 45 indicating stronger inequalities in mental health in urban England than nation-wide. The

Table 6.8: Selected characteristics of milieus in urban England

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | p |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | **age** | | | | | .000 |
| 16-24 | 3.4 | 25.1 | 2.3 | 17.9 | 3.8 | 15.5 | 12.4 | 19.9 | 23.4 | |
| 25-34 | 9.9 | 24.9 | 6.7 | 14.6 | 15.4 | 31.7 | 14.7 | 26.4 | 29.7 | |
| 35-44 | 13.5 | 19.3 | 14.2 | 19.5 | 24.0 | 21.6 | 19.8 | 23.7 | 18.2 | |
| 45-54 | 15.7 | 12.9 | 15.5 | 18.4 | 24.5 | 17.1 | 17.5 | 14.6 | 14.5 | |
| 55-64 | 17.9 | 11.6 | 18.7 | 13.3 | 17.8 | 10.6 | 17.0 | 9.2 | 8.4 | |
| 65-74 | 15.7 | 3.9 | 19.6 | 10.0 | 9.5 | 2.4 | 12.1 | 4.0 | 3.8 | |
| 75 or more | 23.8 | 2.4 | 23.0 | 6.2 | 5.0 | 1.1 | 6.6 | 2.1 | 2.0 | |
| | | | | | **ethnicity** | | | | | .000 |
| white British | 56.1 | 62.2 | 48.2 | 53.2 | 73.4 | 70.6 | 64.4 | 44.6 | 64.3 | |
| white other | 10.1 | 9.5 | 6.1 | 4.1 | 11.6 | 10.7 | 7.3 | 8.2 | 7.6 | |
| Asian | 24.5 | 15.5 | 31.3 | 32.2 | 8.4 | 8.8 | 14.0 | 32.1 | 15.1 | |
| African, Caribbean | 7.3 | 10.5 | 11.9 | 8.1 | 5.2 | 7.3 | 12.0 | 11.8 | 10.6 | |
| | | | | | **monthly disposable income** | | | | | .000 |
| average | 972 | 1,116 | 1,048 | 1,142 | 2,238 | 2,079 | 1,519 | 1,328 | 1,111 | |
| stat. same as clusters | 3,4 | 3,4,8,9 | 1,2,4 | 1,2,3,9 | 6 | 5 | 8 | 2,7,9 | 2,4,8 | |

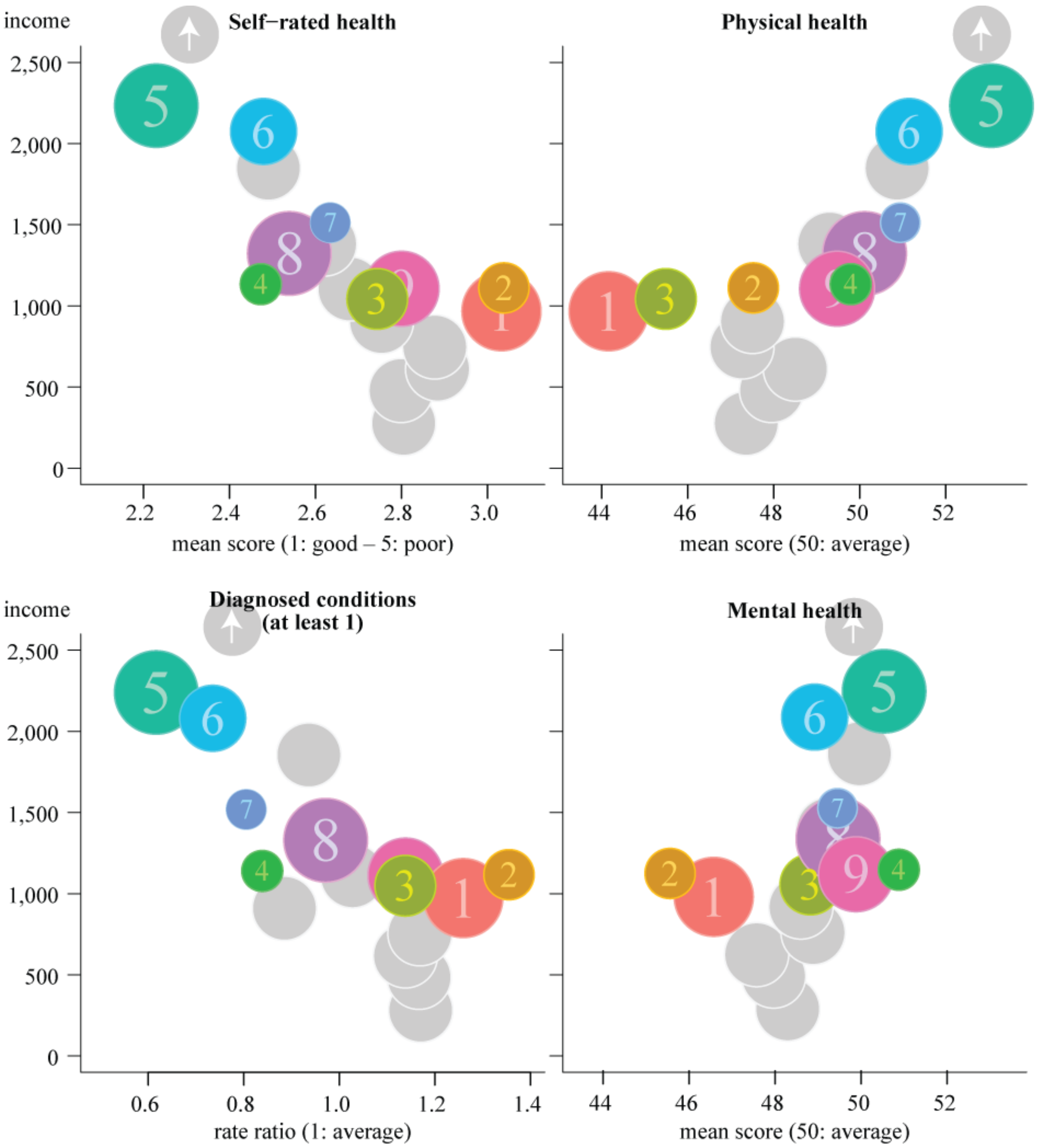
**milieus in columns**: 1 Enduring Isolation, 2 Unconcerned Starters, 3 Retiring Generation/Laid-back Detachment, 4 Locally Anchored, 5 Established Cultural Consumers, 6 Rising Extroverts, 7 Committed Citizens, 8 Laid-back Detachment/Individualistic Independence, 9 Digital Age Autonomy.

higher disadvantage of *Unconcerned Starters* reappears in non-communicable disease risk. The two affluent milieus *Established Cultural Consumers* and *Rising Extroverts* indicate lower risks than their nation-wide counterparts. This again indicates increased inequalities with specific expressions of vulnerability in urban England.

## 6.5 Synthesis: subjective orientations, health and geodemographics

### The merit of the social health milieu classification

The social milieu approach in studying health practices confirms the existence of the social gradient in health yet unmasks significant variations with strong implications for policy. Within a given income category more active and socially engaged milieus tend to experience a health and well-being advantage. Health-related activities are associated with milieu-specific health outcomes albeit to different degrees depending on the wider lifestyle context. Milieus that engage in social networks – the reception and provision of social support – are also associated with better health. This is similar for those who are actively participating in the civic domain. Certain milieus, notably *Locally Anchored* and *Committed Citizens* even fared better than the highest income deciles on some health measures. Conversely, the most socially withdrawn milieus tended to show worse health scores than income deciles with even lower levels of income. This points towards milieu-specific activations of psycho-social and behavioural pathways, which




**milieus**: 1 Enduring Isolation, 2 Unconcerned Starters, 3 Retiring Generation/Laid-back Detachment, 4 Locally Anchored, 5 Established Cultural Consumers, 6 Rising Extroverts, 7 Committed Citizens, 8 Laid-back Detachment/Individualistic Independence, 9 Digital Age Autonomy.

Figure 6.6: Selected health indicators plotted against income deciles (grey) and urban England milieus (coloured); the dots being scaled to group size.

remain concealed by looking at health differentials across income deciles or material factors only – the perspective of conventional epidemiology and health geography.

The milieu approach suggests that the same detrimental or beneficial health-related activity can occur in different milieu contexts and therefore for different reasons. Drawing on some of the literature [Nettleton & Green 2014; Veenstra & Burnett 2014; Williams 1995] and viewing the correlations between activities and orientations in the milieus, some speculations about the underlying drivers concerning these practices are possible. *Rising Extroverts* are likely to exercise for fitness as end in itself and *Established Cultural Consumers* as part of leisure activities [Blaxter 2003]. If it occurs among *Unconcerned Starters*, however, it is likely to support functional motives (fitness as a means) and may be encouraged in this way. Vice versa, the absence of physical activity in the *Enduring Isolation* milieu may be attributed to a psycho-social barrier to participation, whereas in the *Laid-back Detachment* milieu it may be driven by a behavioural orientation towards convenience. The conceptualisation of health practice – as opposed to randomly varying health behaviours in presumed neutral social space – renders implausible interventions that draw on one-size-fits-all narratives targetting the physically inactive or the average materially 'deprived' in general.

The milieus also draw attention to multi-directional causality. Does lack of social participation increase health disadvantage or does ill-health reduce participation? It seems that causal modes operate differently in different milieus, and some speculations based on limited additional data shall be offered. Individuals of the *Enduring Isolation* milieu are likely to be born into deprived circumstances, which expose them to limited life chances and opportunities for participation. Parental education and occupation among these individuals confirm this hypothesis: here, deprivation and lack of participation may become an intergenerational determinant of health, which itself limits life chances even more [Table 6.9]. The reverse may be true among individuals of the *Retiring Generation* milieu. Hard labour in more routine and elementary occupations have caused depletion of health capital, which limit social participation and mobility. The parental education and occupation of *Rising Extroverts* suggests that they were born into privileged circumstances. Yet, they indicate health disadvantage – in particular with respect to mental health – and it is possible that they experience limits to social and economic participation as they move through the life course. Additional data from Understanding Society, also by linkage to administrative or consumer data, may advance the investigation of the milieus, their specific forces and probable trajectories.

Although socio-demographics are not part of the clustering, clear patterns emerge. On the one hand, this finding makes the case for conventional geodemographics: demographic and socio-economic characteristics of residents, aggregated to zonal statistics, are associated with different lifestyle contexts, because socio-demographics are themselves determinants of lifestyles. On the other hand, the numerous attitudinal and behavioural nuances and associated pathways will be missed, especially as they emerge in regionally focussed clusters. With conventional geodemographics alone, it will not be possible to determine which behavioural tendencies may be most representative of a given area. This difficulty becomes more acute as the social gradient – though easy to identify in census statistics – is evidently not sufficient to represent local pathways and inform interventions that target people through neighbourhoods.

Table 6.9: The milieus' parental background.

| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **p** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | **parental education** | | | | | | .000 |
| higher degree | 1.5 | 11.8 | 1.8 | 7.8 | 20.8 | 26.2 | 17.0 | 10.7 | 9.4 | 16.8 | |
| specialised | 14.6 | 29.0 | 16.2 | 26.9 | 33.4 | 33.3 | 33.2 | 26.9 | 30.2 | 32.4 | |
| school | 16.1 | 28.6 | 12.9 | 23.3 | 22.7 | 25.2 | 21.4 | 25.5 | 33.3 | 25.6 | |
| lower, none | 67.7 | 30.6 | 69.2 | 42.0 | 23.1 | 15.3 | 28.4 | 36.9 | 27.1 | 25.2 | |
| | | | | | **parental occupation** | | | | | | .000 |
| senior | 8.3 | 14.7 | 8.7 | 15.3 | 21.9 | 23.9 | 19.5 | 17.4 | 16.2 | 19.0 | |
| professional | 2.2 | 8.6 | 2.7 | 7.4 | 18.8 | 18.9 | 17.6 | 8.8 | 6.0 | 13.2 | |
| associate, admin | 10.8 | 22.6 | 12.6 | 18.5 | 23.7 | 26.9 | 22.1 | 19.7 | 19.4 | 24.5 | |
| skilled | 27.2 | 19.3 | 30.6 | 24.6 | 17.5 | 13.8 | 20.8 | 22.5 | 21.1 | 20.1 | |
| services, operatives | 31.3 | 24.3 | 29.0 | 24.7 | 13.0 | 12.6 | 14.6 | 21.9 | 27.3 | 17.3 | |
| elementary | 20.2 | 10.5 | 16.4 | 9.5 | 5.1 | 3.8 | 5.4 | 9.8 | 9.9 | 6.1 | |

**milieus in columns**: 1 Enduring Isolation, 2 Unconcerned Starters, 3 Retiring Generation, 4 Locally Anchored, 5 Established Cultural Consumers, 6 Rising Extroverts, 7 Committed Citizens, 8 Laid-back Detachment, 9 Digital Age Autonomy, 10 Individualistic Independence.

The use of surnames as a method for regionalisation may be productive in identifying regionally specific variants of the milieus. Since surnames correlate with culture, they may offer an appropriate basis for regionalisation to describe localised subjective milieu contexts and their linked practices. A potential, non-contiguous region that may represent English metropolitan cultures could thus be discovered, which would not have been discovered by using administrative regional boundaries, for example.

The regional patterns that reflect specific variations of health, social patterns and associated inequalities refine the national classification and promise, in theory, to inform regionally adapted policy interventions. The national reference classification still proved useful as a benchmark against which regional specificities can be scaled within a nation-wide approach to public health and health care, notably within the National Health Service. The different levels of the regional Rand indices for Wales, Northern Ireland and Urban England suggests that, in some regions, milieus are sufficiently different to justify locally devolved approaches to public health. Considering the evidence from milieus, a nation-wide approach applies best to Scotland and also England excluding urban areas, while specificities seem stronger in Wales and Northern Ireland. It could be worthwhile in future applications of this work to compare the emerging social milieus from administrative boundaries and surname geographies to determine where health care devolution may be most beneficial and where devolution merely results in "post-code lottery".

### The limitations of the classification

Some limitations of the presented approach relate to the data and various parameters that need to be decided. First, Understanding Society only provides a limited number of items on health-related activities; a wider range that cover alcohol consumption of the entire sample, types of exercising, more information on nutrition and direct attitudes towards healthy living and health care would add more subjective, health-relevant information to the milieus. In addition, data on the frequency of visiting friends or relatives, use of cultural institutions such as churches or community centres as well as general cultural values would reveal more about individual rootedness in social networks and inclination towards autonomy, independence and control. This information would enhance the interpretation of the variables and milieus and attenuate the inevitable partiality of the researcher that stems from his own social situatedness and cultural background. Additional data may challenge current interpretations; hence, the qualitative profiles offered here are necessarily tentative.

The additional information may also contribute to the debate on how successful current government initiatives that emphasise individual responsibility and self-monitoring are likely to be. In terms of measuring health, the survey relies on self-assignment of health outcomes, causing a risk of reporting bias. Understanding Society has recently released biomarkers taken from physical measurements by a nurse. This may offer another layer of information that can be used to describe the milieus in the future (although this information is only available for a sub-sample of 10,000 respondents [Benzeval et al. 2014]).

Second, the problem of choosing the correct number of clusters is itself a debate in the statistical clustering literature. Here, a simple between-group per within-group-variance comparison has been applied, but this approach still contains a subjective element in the choice of cluster solutions, which may rest on very subtle differences between cluster solutions. As for the regional milieus, the most agreeing cluster solutions with reference to the national classification is selected in order to preserve its adequacy in representing all regions. This may be optimal in terms of referencing but not necessarily in terms of representing the data structure of the regional sub-samples. Although the approach is still robust in terms of the complete presence or absence of one milieu in a region (the Rand index is not sensitive to absence or presence of a milieu), one may experiment with using regional cluster solutions that optimally respond to internal cluster criteria and then apply the distance measures in the same fashion. While this would probably improve the regional representations, the value of the national reference classification as a summary schema would fade.

Third, in addition to a classification of individuals, a household-level classification may be informative in terms of health milieus. A household-level classification would have certain limitations, however, because the activity patterns and orientations are not recorded for all household members, and heuristics to infer household level patterns would first need to be developed. They may hence have to be based on demographic and economic characteristics, but these are already partly considered in the comparison of the milieus' contextual circumstances. Moreover, since health is an individual-level outcome, households could only be considered as an additional ecological level rather than a substitution of the individual level. It seems unlikely that a separate household-level classification will enhance sufficiently the understanding of health milieus to justify its creation.

### The health milieus and social class

One of the assumptions underlying the study of health milieus is that relevant social practices are indicative of various proximal and distal social determinants of health. Health milieus are classified by the frequency or intensity of relevant practices, which each receive equal weight, resulting in various health-relevant profiles. Different traces of capitals are found retrospectively associated with these profiles, which, viewed jointly, form the notion of milieus. This approach differs from another UK study that employ Bourdieu's approach: the study by Savage et al [2013], who identify seven milieus (or classes) based on different forms of capitals rather than social practices. Incidentally, the health milieus show some similarities to their classes. Their Elites, Traditional Working Class, Precariat and Emerging Service Workers easily correspond to *Established Cultural Consumers*, *Retiring Generation*, *Enduring Isolation* and *Digital Age Autonomy* respectively, judging from their capitals and socio-deomographics. Parts of *Individualistic Independence* and *Committed Citizen* may constitute their class of Technical Middle Class, and other parts together with *Rising Extroverts* can easily be envisaged as their Established Middle Class. *Unconcerned Starters* might be found in the Precariat and the Emerging Service Workers, which seems consistent with the age profile of the latter. *Locally Anchored*, however, do not directly appear in Savage et al's seven class model.

At least informally, then, the health milieus seem consistent with more recent work on class in British society. Nevertheless, it should be noted that Savage et al's model is more generic and based on capitals as input information to Latent Class Analysis, in which the authors accord a more substantialist reading to capitals, whereby individuals possess more or less cultural or social capital according to additive composite measures. Yet, composites of capital indicators tend to conceal the relational significance of, for example, cultural capitals, which as such "cannot be reified because its uses and forms vary according to the issues that are relevant to the world in which it is put to use." [Serre & Wagner 2015, 447].

To illustrate this point, social capital inherent in the local environment appears to be central to status and health of the *Locally Anchored*, whereas it is unimportant for *Rising Extroverts*, who probably possess social capital with a wider spatial range (for example, friends in other cities or countries), or *Laid-back Detachment*, who are more focussed on the domestic environment. The same argument may be applied to economic capital or occupation-based status, whose significance can only be determined in conjunction with other forms of capitals. This relational role of capitals may explain the finding here that the milieus show larger variation in health measures than alternative groupings by income. Therefore, a definition of milieus by disaggregate practices appears to lend itself more easily to the study of the significance of milieu-specific capitals in defining social position and shaping health.

This reasoning has some implications for neighbourhood effects studies, too, which tend to reify certain forms of capital and disregard their relational character [see, for example, Carpiano 2006, 2007]. The consequence are questionable calls for quantifiable "dosage-response" [Galster 2012] approaches to policy based on population-wide effects averaging over *Laid-back Detachmant*, *Locally Anchored* and other milieus with very different predispositions outside the remit of control variables. Dosage-response policies can probably only be effectively designed after a precise determination of who lives in a given neighbourhood and what pathways link neighbourhood characteristics to the health of heterogeneous groups of residents.

With respect to geodemographics, the milieu approach provides a substantive enrichment as it adds a subjective dimension of health along a continuum of social determinants in addition to structural conditions shaping the social gradient in health. Yet, unlike geodemographics, the social milieu approach lacks spatial precision. Methods to combine the coverage of geodemographics with the variegated subjective information and regional specifications are likely to add significant value in comprehensively representing the vulnerability of populations.

# 7 Simulated geographies: the spatial context of health milieus

The milieu approach is aspatial and the sparseness of survey data does not allow direct inference of the milieus' spatial distributions. In order to turn the milieus into a building block of geodemographics, it is necessary to estimate, at the spatial granularity of geodemographic systems, where individuals of different social health milieus live.

## 7.1 Techniques of spatial statistical matching

Making the milieu approach accessible to geodemographics requires that the social milieus be translated into an explicit geographic distribution with full spatial coverage. The translation of sparse survey data into a spatial distribution can be framed as a problem of statistical matching. Statistical matching is a term for a set of technique that combines information held in different datasets based on common information found in them [Raessler 2002, 2]. This approach is technically similar to record linkage, but the assumptions of the two are fundamentally opposed. Whereas record linkage assumes that records of two different datasets pertain to the same individuals (for example, when linking survey respondents to their hospital episodes), statistical matching presupposes that there are no individuals common to the two datasets. It is then necessary to create a new dataset out of the two with hypothetical individuals based on common or equivalent information that can be found in both datasets.

Ascribing geographies to social milieus can be framed as a problem of missing data to be addressed through a framework of statistical matching [ibid., 7]. For example, let us suppose that a survey asks about smoking habits and some respondents refuse to answer the question. In this case, missing answers may be imputed using other information in the survey. The sample is split into respondent and non-respondent cases, and the information available about the non-respondent cases is used to predict their missing answers drawing on the answers of respondent cases and their characteristics. If a survey records the occupation of a respondent in addition to smoking habits, it is possible to infer answers to the smoking question by using the occupation as observed correlated data, if the two are associated. Unobserved data can hence be imputed, if it is associated with some observed data.

If, however, missingness depends on the nature of the unobserved data itself, for example, when smoking is highly stigmatised, this leads to a problem of response bias, which can be partially addressed by imputation or case weights depending on the nature of the data and sampling design. Statistical matching assumes that the situation resembles a scenario of *conditional independence* [ibid.], whereby the missing data (missing responses to smoking) is associated with observed data (occupation) but independent of unobserved data (smoking habits). If data of interest is missing simply because it has not been collected or asked for, the assumption of conditional independence is satisfied.

Geodemographics in the UK heavily relies on the UK Census in constructing a neighbourhood classification that is informative about local lifestyles. The UK Census covers almost the entire UK population and provides highly granular spatial data. Yet, the census questionnaire does not collect data that could be used for a comprehensive characterisation of health milieus. In the present case, the health milieus and their constituent variables are missing and need to be derived from an alternative dataset – the Understanding Society survey. Since health milieus are found to be associated with respondents' demographics, social and economic context, and some of those demographics are also asked in the Census, inferring geographical distributions of unobserved data based on conditional independence is a special case of a missing data problem.

Techniques of spatial microsimulation address the special case of geographic inference of unobserved data. Rather than merging two micro-datasets, spatial microsimulation handles missing data within an ecological research framework, in which the data is doubly constrained. First, the spatial dataset (for example, Census) does not collect the data of interest and, second, the spatial information is only made available in form of aggregate statistics. Spatial microsimulation approaches this problem by generating a synthetic micro-dataset of a zonal population which can be used to either describe small area populations in generalised terms (static) or to simulate the impact of events such as policy interventions (dynamic).

Harland et al [2012] distinguish three kinds of spatial microsimulation: deterministic weighting, conditional-probabilistic attribution and simulated annealing. Deterministic weighting assigns weights to individual survey records according to their representativeness of a given zone based on matched variables in the survey and census zonal statistics. The unobserved data can then be summarised by means of weighted statistics. Conditional-probabilistic attribution begins with an empty micro-dataset covering the population (each resident has an empty record) of each zone. The matched variables are then assigned in turn based on their relative frequency in each zone and their conditional probability resulting from previous variable value assignments as each variable is successively estimated. The technique operates with Bayesian inference where ultimately the probabilistically assigned variables act as prior distributions for the posterior estimates of the unobserved data. Simulated annealing takes iteratively random draws from the survey sample for each zone and calculates the difference between the thus created synthetic dataset and zonal statistics based on common variables. It then iteratively replaces individuals from the synthetic dataset with others from the original survey sample until the difference between synthetic records and zonal statistics is minimised. Each technique has strengths and weaknesses and the choice depends on the research objective, the size of the study area, the format of the data and computational resources available.

Harland et al [ibid.] do not discuss model-based deterministic approaches, which estimate unmeasured small-area variables through predictive models from a representative survey, whereby model coefficients of co-variates are applied to equivalent co-variates found in a population-wide dataset, such as the Census [Twigg et al. 2000]. These model-based approaches are purpose-oriented: instead of creating synthetic populations based on the widest possible range of information available, they simulate area variables of interest using the most correlated covariates found in both survey and Census. Twigg and Moon [2002] apply a variation of a deterministic technique to estimate smoking and drinking prevalence in selected UK wards. They attest potential of their technique to rank wards according to the degree of smoking and drinking prevalences but remain sceptical of the possibility to generate precise estimates of prevalence rates. The strenghts of their method lies, however, in its ability to account for multi-level structure of social phenomena in the predictive models and its computational efficiency.

As computers and statistical algorithms improve their handling of complex modelling problems, spatial microsimulation models have become increasingly popular [Harland et al. 2012]. As a representative example of emerging studies in health geography, Clark et al [2014] apply the simulated annealing technique to combine 2001 UK Census statistics with English Longitudinal Survey of England to predict future morbidities among elderly citizens in UK's local authorities for the following ten years. They argue that spatial microsimulation help planners to assess future health needs and provide for health care more efficiently. Riva and Smith [2012] employ the deterministic weighting method to infer small-area rates of psychological distress and alcohol consumption. They conclude that the method generates plausible spatial patterns of these prevalences at LSOA level. An equivalent study on smoking in New Zealand ascertains success in estimating small area smoking prevalences based on deterministic weighting [Smith et al. 2011]. There are further studies of similar kind with similar encouraging findings [e.g. Hermes & Poulsen 2013; Morrissey et al. 2010, 2013; Smith et al. 2011].

In spite of the positive experience with spatial microsimulation in health research, there are limitations to these studies. First, many applications still focus on health or health-related activities as independent random variables in isolation. No efforts have been made to socially contextualise these phenomena, although the generation of spatially referenced microdata would permit it. As argued earlier, a case-based rather than variable-based focus is necessary to align the models more realistically to the embodied drivers of behaviours.

Second, the assumption that matched variables and unobserved variables are associated to the same degree everywhere in the study region is not always realistic. Education may be associated in a national survey with smoking and hence it may present itself as a suitable variable. Yet it may be the primary driver in one region but a secondary one in another. The recent literature discusses the implications of this form of place effects. Smith et al [2009] recommend to estimate a series of local models instead of one single model. Different orders of constraint variables could be defined for each regional sub-sample, type of neighbourhoods or rural versus urban areas. Their own application suggests that this may reflect local specificities better. Another way of bringing place into spatial microsimulation are introducing aggregation constraints on unobserved data taken from other sources. Morrissey et al [2013] align their spatial microsimulation model of hospital admissions in Ireland's small areas such that the sum of hospital admissions match county level admission totals. In so doing, they adjust their models for unobserved regional effects.

A third challenge arises when the unobserved data is associated directly with characteristics of location. Physical activity may not just be associated with milieu context but also with quality of the local environment. Here, spatial microsimulation would produce biased estimates that need to be appraised and adjusted. Birkin and Clarke [2011] experiment with microsimulation calibrations for different geodemographic classes in Leeds. They find that variables that depend on neighbourhood context (in their case, car ownership) are better predicted when validated against Census small area statistics. All these points need to be considered, too, in applying spatial microsimulation to the health milieus.

## 7.2 Matching health milieus and small area statistics

In order to overcome the first two challenges and at least offer conclusions with respect to the third one, the potential of spatial microsimulation in localising milieus will be combined with the idea of surname geographies as cultural indicators to inform the definition of regional models. Applied to milieus, the focus lies on a probabilistic spatial distribution of milieus rather than a representative synthetic microdataset. Deterministic weighting seems most efficient to achieve this objective, since it abstains from creating a microdataset and allows to generate weighted small area statistics by means of zone-wise respondent weights.

Deterministic weighting is implemented through an algorithm known as Iterative Proportional Fitting (IPF) [Lovelace & Ballas 2013]. This algorithm determines the degree to which a survey individual is representative of a population based on multiple common variables. Individual records are iteratively weighted until a final weight is found that best fits the distribution of variables in the population. In this iterative process, IPF produces results that fit the first variables used in this iterative process better than the ones included later. Therefore, using deterministic weighting requires, first, the identification of common and suitable variables, where suitable means that the variables are likely to be associated with the unobserved data (in this case, the milieus).

Second, the identified matched variables need to be ordered according to their strength of association. This can be achieved in several ways. The ten social milieus can be treated as a multinomial outcome and some form of multinomial regression or discriminant analysis can be run using milieu membership as response and common variables of the Understanding Society survey and the 2011 UK Census as independent variables. The regression coefficient can then be transformed into a ranking of variables, indicating the order of variables in the IPF.

A second approach for multinomial data is to use descriptive statistics to identify suitable matched variables and their relative importance. $\chi^2$-based tests (e.g. Cramer's V) or similarity indices comparing partitions (e.g. Adjusted Rand) draw on contingency tables of the categorical variables and the milieu memberships produce strengths of associations that can be compared across variables, if they are normalised for sample size and degrees of freedoms. An additional requirement in this case, however, is to check inter-correlation between matched variables. For example, education and social status are correlated, and therefore the relative importance of one variable decreases once the other is introduced in the IPF iterations. Highly correlated variables can be identified again by contingency tables and support the inclusion or exclusion of variables in the IPF process.

The major difference between the two approaches is that the former accounts for intercorrelations between variables, while the latter only shows bi-variable relationships between any given variable and milieu membership. Yet estimating the overall discriminant effect by using the first approach can be complex, if the independent variables are categorical themselves. The first approach generates an association parameter for each category (for example, age category) rather than of the variable itself. Variable intercorrelation (collinearity) may not be an important statistical issue in microsimulation models, although strong redundancies may affect the fit of variables that are more weakly associated with the unobserved data. Consequently, the second, descriptive approach to selecting matched variables has been chosen.

The different similarity indices, which all range from -1 to +1, reveal high consistency with regards to covariates for the national classification [Figure 7.1]. Age and education have the highest index values followed by job status; these three variables seem to 'explain' milieu membership best. The next three variables are tenure, household type and marital status. The $\chi^2$-based Cramer's V finds that disability has higher explanatory potential than other variables. Ethnicity, sex, household size and aid (unpaid care) discriminate less well. Social status as measured by NSSEC is less associated with milieu membership than other variables. Choosing the Normalised Mutual Information as ordering criterion, the order of matched variables to be chosen begins with education, age and job status, proceeds with a number of variables related to social context, health, economic circumstances and social obligations and ends with sex.

Some of the variables are intercorrelated: job status with its categories 'employment', 'pensioners' or 'students' is associated with age [Table 7.1]. Age also correlates with household types and to a lesser degree with tenure; both variables appear to reflect an individual's life stage. Education – the best discriminator – does not show a strong association with any other variable. Job status correlates with social status, but the latter is not associated strongly with age. Between-variable correlations are therefore limited and except for household size and household type, there is no strong redundancy that may have an impact on the spatial microsimulation model.

The same comparisons of indices have been run for each of the regional milieus. Taking again the NMI as criterion, it can be observed that education is the superior discriminator in almost all regions except in Wales and Southern England, where age is the

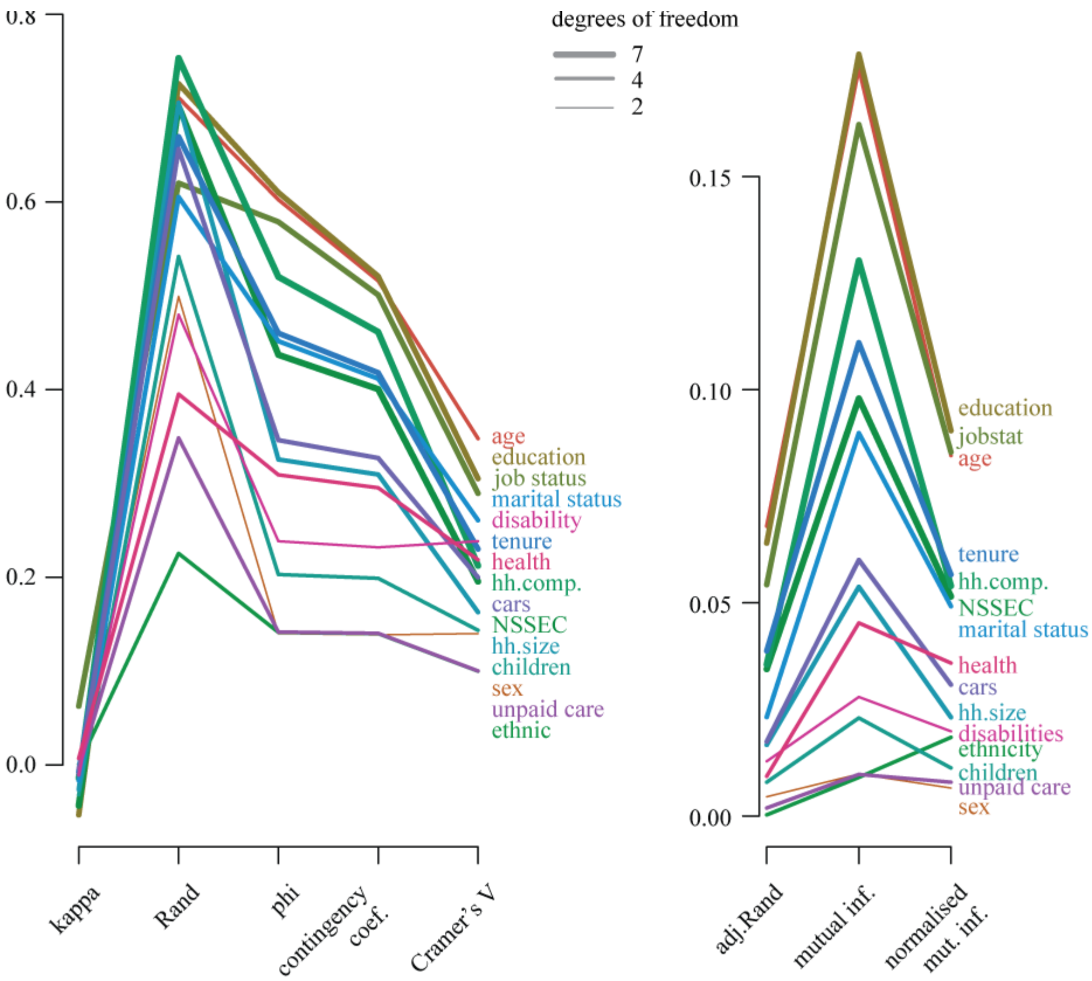



Figure 7.1: Selected indices measuring the categorical association of potential match variables and milieus.

best discriminator [not illustrated]. In Wales, job status is a more strongly associated covariate of milieu membership than education. In urban England, education and job status discriminate more strongly than age, which is on par with social status. This is followed by marital status in urban England, thus marital status is a stronger predictor than household type, whereas the reverse applies in all other regions. The next most important predictor of milieu membership is ethnicity, which in urban England is even more important than tenure. In all other regions, ethnicity occupies the lowest rank.

The index comparison produces regionally specific ranks of variables with regards to their potential discriminatory power in milieu membership. The regional differences suggest that just using the order of matched variables for the national classification would not lead to optimal models within each regional context. Consequently, in line with what Smith et al [2009] propose as ways to bring place into spatial microsimulation, local models should be calibrated by using regionally specific orders of variables. Each

Table 7.1: Matched variables, their intercorrelations and Normalised Mutual Information with milieu category.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | NMI | rk |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 age | | | | | | | | | | | | | | | .083 | 2 |
| 2 aidhrs | .011 | | | | | | | | | | | | | | .007 | 14 |
| 3 cars | .025 | .004 | | | | | | | | | | | | | .034 | 9 |
| 4 children | .137 | .000 | .016 | | | | | | | | | | | | .012 | 13 |
| 5 disab | .051 | .005 | .014 | .011 | | | | | | | | | | | .021 | 11 |
| 6 edu | .058 | .004 | .030 | .008 | .021 | | | | | | | | | | .093 | 1 |
| 7 ethnic | .022 | .002 | .020 | .013 | .010 | .004 | | | | | | | | | .014 | 12 |
| 8 hhsize | .108 | .005 | .071 | .387 | .026 | .017 | .036 | | | | | | | | .025 | 10 |
| 9 hhtype | .298 | .008 | .088 | .320 | .031 | .033 | .034 | .537 | | | | | | | .053 | 5 |
| 10 jobstat | .292 | .012 | .048 | .100 | .055 | .066 | .026 | .089 | .186 | | | | | | .081 | 3 |
| 11 marital | .186 | .007 | .061 | .045 | .022 | .029 | .014 | .132 | .214 | .120 | | | | | .050 | 6 |
| 12 NSSEC | .066 | .003 | .021 | .018 | .006 | .069 | .007 | .010 | .021 | .252 | .042 | | | | .036 | 7 |
| 13 sex | .000 | .003 | .002 | .006 | .000 | .003 | .001 | .000 | .014 | .022 | .010 | .009 | | | .006 | 15 |
| 14 sfhealth | .026 | .004 | .027 | .004 | .170 | .033 | .000 | .013 | .017 | .059 | .014 | .009 | .000 | | .035 | 8 |
| 15 tenure | .122 | .007 | .099 | .048 | .024 | .039 | .024 | .050 | .108 | .126 | .056 | .024 | .001 | .025 | .058 | 4 |

**columns**: numbers refer to variables in rows; NMI = Normalised Mutual Information; rk = rank of variable based on NMI

small area will be assigned to one of the six isonymy regions [see chapter 6.4] and subsequently the microsimulation with the specific, regional order of variables will be run. In a second step, it can be decided which milieu classification is to be predicted: the national or the regional ones.

## 7.3 Spatial distributions of urban health milieus

The matched variables [see Table 7.1] are used to estimate the prevalence of health milieus in Lower Super Output Areas (LSOAs) in London. LSOAs appear to provide a suitable geographic resolution because they are sufficiently small to approximate the scale of neighbourhoods (most LSOAs in London have between 1,000 and 1,500 people), and the Census provides sufficiently detailed categorisations of variables at this level. Age bands, household compositions, tenure and others are much more broadly defined for the highest level of resolution, Output Areas.

Based on the isonymy regions, the Greater London Authority jurisdiction splits into two regions of local surname compositions [Figure 7.2]. The one with the urban English composition stretches in a crescent from the west near Heathrow through Brent, Barnet, Camden, Islington to the east and continues towards south London. The region that belongs to the southern isonymy group appears in wedges reaching from the south western borough of Richmond towards the centre, covering large parts of Hammersmith and Fulham, Kensington and Chelsea and Westminster. In the east, Havering, Bexley and Bromley emerge as a large contiguous area with southern surname composition. It can be observed again that, although geography is not part of the clustering of surname

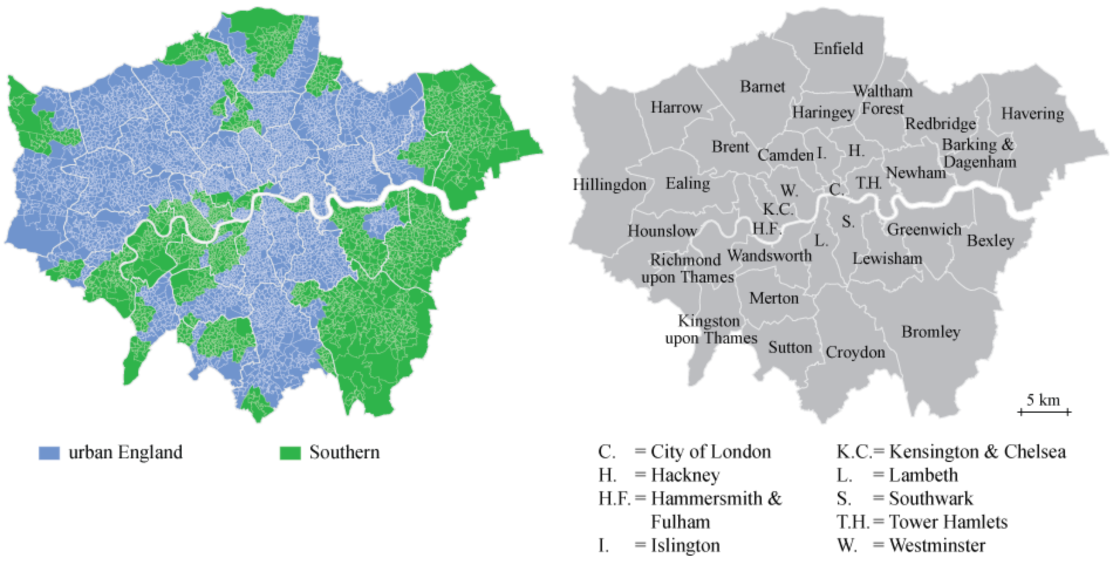



Figure 7.2: London LSOAs assigned to one of six national surname clusters (left) and London borough jurisdictions (right).

regions, a clear regionalisation emerges whereby London is split into two types of nearly contiguous regions.

#### Results of the locally calibrated simulation

Three microsimulation models are calibrated: a global model, a combined local model and a split local model. The global model takes the entire sample of the United Kingdom, and microsimulation is performed using the order of variables that best differentiates the milieus for the entire United Kingdom to derive the IPF weights for London's LSOAs. The second, combined local model takes the sub-sample of the survey with only those respondents that fall into the isonymy regions of Urban England and Southern England [see Figure 6.3]. Variables are ordered according to their differentiating power in the sub-sample. In the third, split local approach, two microsimulation models are calibrated for each of the two isonymy regions: one using the survey sub-sample that is resident in the Southern England isonymy region and another resident in the Urban England region. The sub-sample-specific ordering of matched variables is used to derive the IPF weights. In all cases, the generated weights are used to calculate weighted frequencies of each LSOA, which directly translate into estimated proportions of the ten milieus in each LSOAs. Validation of the three models reveals which approach performs best in terms of reflecting local characteristics. The validation results are presented in the subsequent section.

As will be shown below, the combined local approach performs best, and hence the results of this model is presented. Mapping the modelled geographical distribution of each milieu within London reveals that each milieu exhibits a distinct and spatially clustered geographical distributions [Figure 7.3]. The Global Moran's I, which measures the degree of spatial clustering on a scale of -1 (extreme dispersal) and +1 (extreme concentration), shows very high values ranging between .5 and .8. The most clustered milieu in London is the *Laid-back Detachment* milieu with a Moran's I of .83. The milieu mirrors the crescent geography of the urban English surname composition in a non-contiguous way with large centres in Ealing, Barnet, Newham and north Croydon. These nodes emerge in fainter form in the distribution of *Unconcerned Starters*, although Brent does not appear here. A similar observation can be made for the *Enduring Isolation* milieu. The three milieus appear to be co-located with some subtle differences, however. Since the *Laid-back Detachment* milieu have the highest share of Asians and the surname geography mirrors ethnicity in the census, its geographical projection reproduces London's ethnic geography.

The milieu of *Individual Independence* represents the reverse. Here, concentrations can be found where the map of surname compositions indicates high prevalence of Southern English names and their correlates. Consequently, Richmond and Hammersmith and Fulham in west London show concentrations of this milieu as well as the outer Eastern and Southern suburbs. The *Committed Citizens* closely follow this pattern though with less density in the eastern suburbs. The two milieus – *Committed Citizens* and *Individual Independence* – do not share many similarities with respect to their socio-demographics except their economic context, notably income. Income is not available in the census, but the consistency between similar income groups suggests that other variables such as car ownership or social stats (NSSEC) are effective substitutes.

*Established Cultural Consumers* and *Rising Extroverts* also exhibit similar geographic distributions. The model suggests that they reside predominantly in west and central London with a southern belt running across Lambeth and Southwark towards Greenwich. *Established Cultural Consumers* are also represented in some suburban locations, notably Bromley. *Rising Extroverts* are more concentrated in central locations, in particular in boroughs close to the City.

The *Digital Age Autonomy* milieu show high prevalence in more peripheral locations with a larger concentration straddling the river in the east. This concentration stretches ten kilometres north and south into the boroughs of Havering and Bexley. Another concentration can be found in the borough of Hillingdon in the west, in Kingston and Sutton in the south as well as Enfield to the north. The geography reflects their young age and low educational profile with mainly English ethnicity. These peripheral locations of London provide almost a geographic metaphor for their position in society and imply area of special attention for supportive policy. A milieu that follows best that of *Digital Age Autonomy* is the *Locally Anchored* milieu, which, in age terms, is broadly generation ahead, suggesting itself as the parents' milieu of their younger co-locators. Yet, a closer look reveals differences with concentrations in Bromley and a nearly complete absence in central locations. In addition, the attitudinal profiles of the two groups in question are not very consistent, although this does not rule out some relationship.

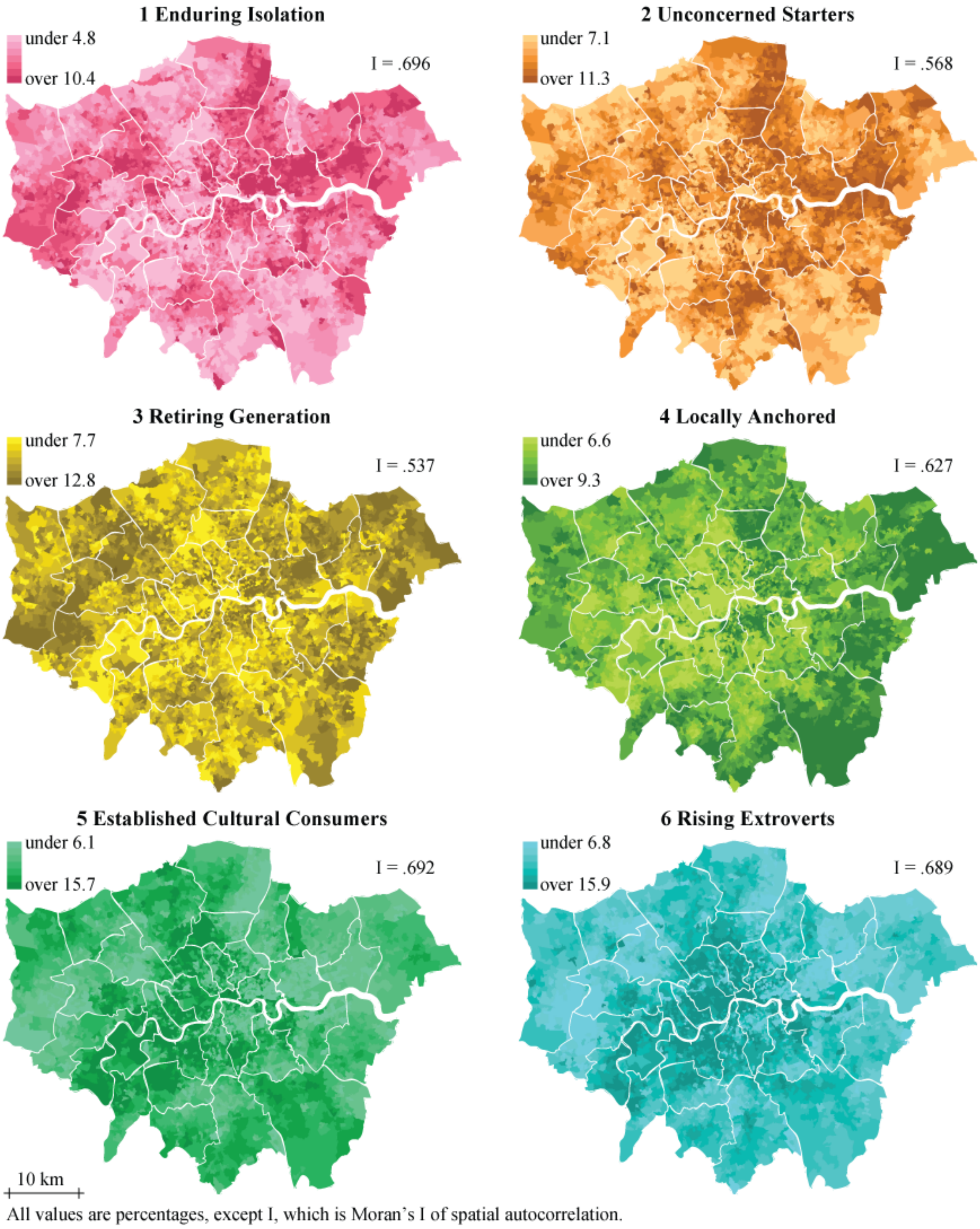



Figure 7.3: Inferred spatial distribution of health milieus in London.

Finally, the most dispersed milieu is *Retiring Generation*. Their age structure is the single most discriminant and their low education coupled with high ownership rates produces a very different geography encompassing east and west, poorer and more affluent parts of London. This seems plausible, since their high ownership rates is presumably

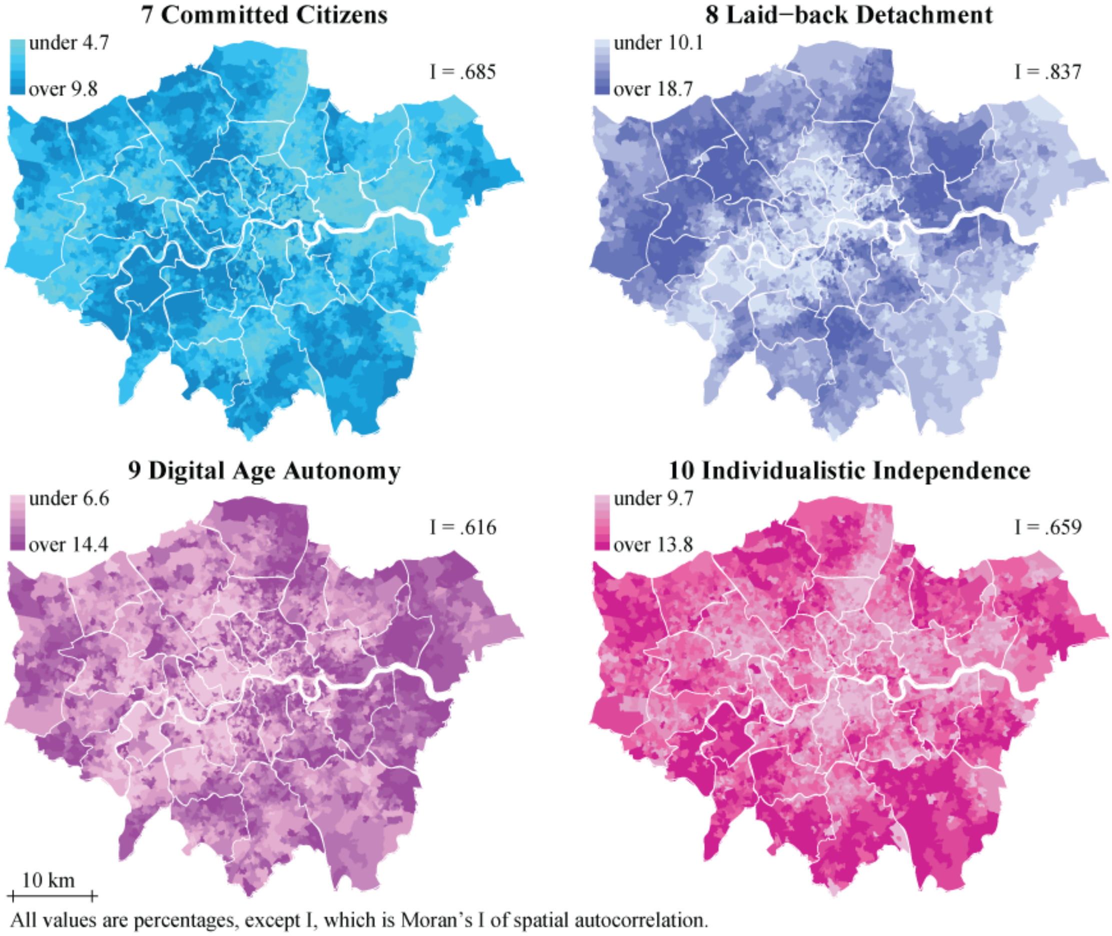



Figure 7.3 (continued)

due to earlier acquisition of homes when this was more affordable even for economically less established groups. The spatial distribution exhibits the remnants of the phenomenon with high concentrations in central neighbourhoods within the boroughs of Camden, Islington, Hackney, Tower Hamlets and Southwark.

While the spatial patterns are distinctive, the variation of relative frequencies are sometimes limited. The interseptile range varies between five and nine per cent. This may not seem substantial, but given the relative share of the milieus in the entire population – one tenth on average – this variation often translates into a factor of two. In other words, if the location quotient of the milieus were used, values ranging between half and twice the average milieu prevalence would be common. The location quotient could be interpreted as an internally normalised score of milieu likelihood within each LSOA. This enables place-relational statements, such as: "The *Digital Age Autonomy* milieu is twice as likely to live in Barking than in Whitechapel". It is impossible to make the same assertions with conventional geodemographics. Thus, rather than taking relative

Table 7.2: Average variable classification errors of the three models in per cent.

| # | age | car | child | disab | jobstat | edu | ethnic | hhcomp | marital | NSSEC | tenure |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 11.7 | 0.0 | 0.0 | 1.1 | 9.5 | 0.0 | 0.0 | 9.6 | 0.0 | 0.0 | 6.6 |
| 2 | 11.7 | 0.0 | 0.0 | 1.1 | 0.0 | 0.0 | 0.0 | 9.6 | 6.6 | 0.0 | 6.8 |
| 3 | 16.4 | 13.4 | 9.0 | 8.0 | 12.5 | 8.9 | 14.8 | 23.7 | 10.9 | 13.3 | 21.5 |

**models**: 1 global, 2 combined local, 3 split local.

frequencies as direct indicators of milieu prevalence, this study may be more suitable to rank-order localities according to locational propensities of different milieus.

## 7.4 Triangulation of health milieu geographies

Due to multiple weighting of individual cases according to a high number of variables, weights never perfectly mirror the zonal characteristics [Harland et al. 2012; Lovelace & Ballas 2013]. Consequently, in matching survey samples with geographical areas, errors occur, and it is necessary to validate microsimulation models by estimating and characterising the error and to triangulate the results. Several strategies to do so are adopted here. First, all spatial microsimulation models – global, combined local, split local – are compared to assess potential variations of outcomes and appraise alternative model results. Second, as a way to estimate substantive uncertainty, cluster distances are mapped across London. Third, the survey itself is taken as a source of validation by estimating the spatial distribution of the milieus directly from the geo-located sample. Fourth, the results are compared to an external data source for consistency and plausibility; in this case, the 2011 London Output Area Classification (LOAC) is used.

### Classification errors compared

Errors can be calculated as the classification error, that is the proportion of people that are misclassified (e.g. GCSE qualification instead of A levels) on each variable for each area. Comparing the error from the three models reveals that the combined local model is marginally superior to the global model [Table 7.2]. The classification errors range from zero to twelve per cent, with age showing the highest error. Household composition, marital status and tenure also show higher errors of between six and ten per cent. The global model performs similarly with the difference that there is an error in estimating job status in 9.5 per cent of the population and no misclassification with respect to marital status. It seems that households differ in London in the combinations of these characteristics compared to the rest of the UK and the Southern and Urban England. The split local model performs worst with high classification errors on all variables. This is in part surprising as, for example, in case of the Urban England sub-sample, there is more ethnic diversity compared to the total sample. Hence, their ethnic profile should match the profile of the areas classified as Urban England better

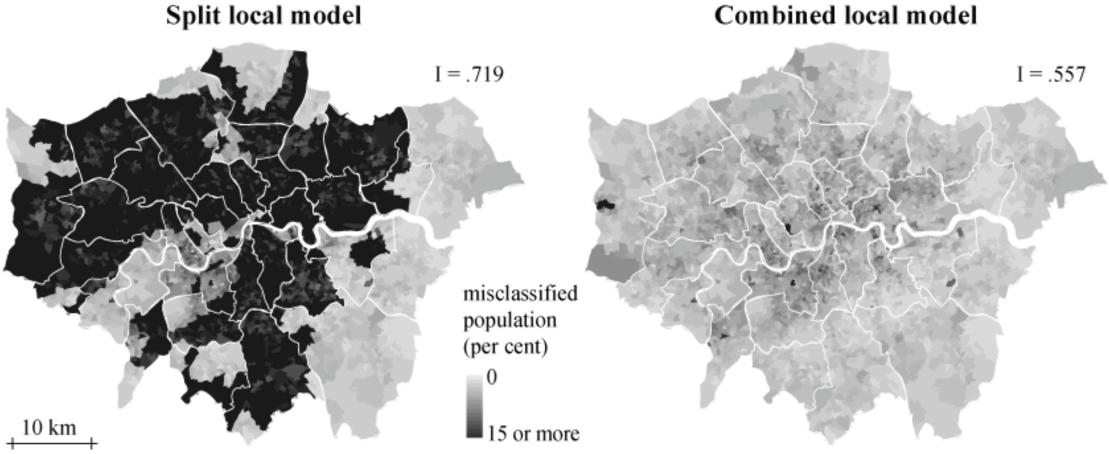



Figure 7.4: Average classification error – measured in per cent – in the split local (left) and the combined local (right) calibrated models.

than the total sample. On the other hand, the sample design of Understanding Society does not establish representativeness for the exact surname regions, especially not for a highly peculiar geography as the one observed for Urban England.

The distributions of average classification errors, that is the average error across matched variables for each LSOA, differ between the locally and the globally calibrated model [Figure 7.4]. With regards to the split local model, the maps clearly reveal the source of the classification error: high error values predominantly fall within the geography of wards that have been classified as Urban England due to their surname composition. Errors exceeding 20 per cent are common in this subregion, whereas the majority of LSOAs that fall within the southern isonymy region show values of up to two per cent. The results suggests that it is the Urban England sub-sample that does not match well the characteristics of the population of Urban England.

The combined local model, on the other hand, generates a different conclusion. High errors, which are at a level of five per cent or above, are concentrated in inner London. These areas are more populated than other areas. The global model exhibits a very similar pattern to the combined local model [not illustrated]. Yet, the split local model appears to fare better in those central areas in London that belong to the Southern England isonymy group. This implies that the southern sub-sample of the survey better reflects the population of those parts in London than either the total sample or the Urban England sample.

### Substantive uncertainty in inferring milieu geographies

The antecedent results do not imply, however, that the same error observed for socio-demographic match between simulated and actual population also applies to the milieus.

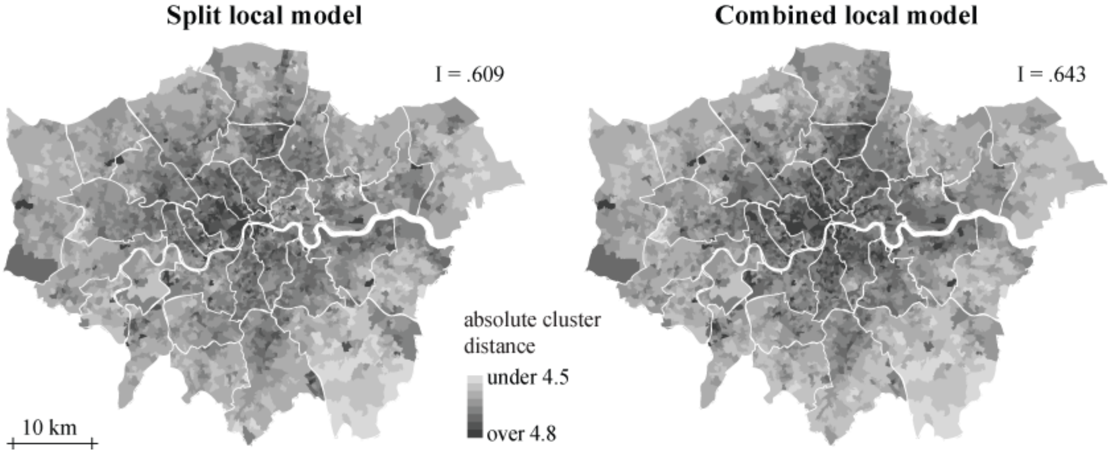



Figure 7.5: Average absolute Euclidean cluster distances of simulated individuals in the split local (left) and the combined local (right) models.

The milieus may still be locally more specific and better captured by the Urban England sub-sample than by the total sample. Comparing socio-demographics is only as good as they are associated with milieu membership independent of place. The milieus themselves bear some inherent uncertainty that arises from the clustering method. Clusters are groupings of observations in statistical space according to their proximity. Cluster centres are the points of gravity for each group, at the location of the mean values for each variable in each cluster. The distance (here: absolute Euclidean distance in statistical space) between each observation (survey respondent) to the centre of its assigned clusters is hence a direct measurement of uncertainty of cluster membership. In the context of spatial microsimulation, in which cases are re-weighted according to their local representativeness, it may be possible that more certain or more uncertain cases receive more or less weight. Some areas may hence be predominantly composed of cases with high uncertainty of cluster membership, others with lower.

While the overall variation of cluster distances is very low in all three models, there is a clear spatial pattern [Figure 7.5]. Most of the uncertainty can be located in central London, with a large concentration in Westminster and adjacent boroughs. Ascribing milieus to areas here is associated with more uncertainty. In the global and combined local models, higher levels of uncertainty extend eastward towards Barking and Dagenham. Another concentration of uncertainty can be found west of the Lea Valley in the boroughs of Haringey and Enfield. The spatial pattern observed in the split local model differs here: cluster distances are lower in east London and in the north. Here, the models select cases that are closer to their milieu centre. It can be concluded that although the split local model fits less the socio-demographic characteristics of the actual population, this does not necessarily translate into weaker representations of local milieus. Yet, determining the accuracy of estimated milieu prevalence requires triangulation of the resulting estimates.

### Internal triangulation of health milieu geographies

A way of triangulating the local milieu prevalences is to use the geographical information available in the survey itself. The survey records the coordinates of the exact residential location of each respondent, which is available under secure access arrangements. The residential LSOA is available under special licence conditions and have been used here to explore the spatial distribution of respondents classified by their health milieu. This distribution can be compared to the modelled distribution of milieus that results from spatial microsimulation.

The spatial distribution of survey respondents has been estimated using kernel density estimation (KDE) of respondents' LSOA centroids. For visualisation purposes, the coordinates vary randomly within a buffer of 200 metres [Figure 7.6]. A bandwidth of approximately 2,700 metres was chosen because it covers 99 per cent of nearest neighbour distances between sampling points. The distributions of sample respondents that belong to a given milieu differ. Respondents that pertain to the *Enduring Isolation* milieu spatially cluster in east London with two weak sub-clusters in Sutton and Brent. All spatial microsimulation models detected these concentrations, too. Respondents assigned to the milieu of *Unconcerned starters* are concentrated further east in the borough of Newham with sub-clusters in Haringey, Tower Hamlets and Southwark. The geography emphasises similar centres and exhibits sparseness in London's west. The KDE of *Retiring Generation*, too, shows a distribution that shares similarities with the simulated one, in particular the sub-centres on Southwest London and in Brent and Ealing. Consistent with the simulation, the KDE of the *Locally Anchored* milieu emphasises more suburban locations, albeit to a less extreme extent. As with the simulated geographies, the KDEs of *Established Cultural Consumers* and *Rising Extroverts* changes significantly and extends further west along the river towards Richmond. The latter also emphasises central location but extends further east than would be expected. The *Laid-back Detachment* milieu distributes across the east, Brent and south London with a centre in Tower Hamlets, again, very consistent with the simulated geography.

All in all, this type of internal triangulation suggests that location results from the milieus and their associated socio-demographic characteristics. Health milieu geographies predominantly result from sorting processes rather than local causalities, such as neighbourhood effects. This finding lends credibility to the spatial microsimulation technique in this research context. At the same time, the milieu-wise KDEs are necessarily based on sparse observations, so that conclusive assessment of actual and simulated geographies are impossible. Another caveat in relation to this point is that the KDEs are not adjusted for the population density of LSOAs. But since LSOAs do not vary much in size, the extent of this bias should be limited.

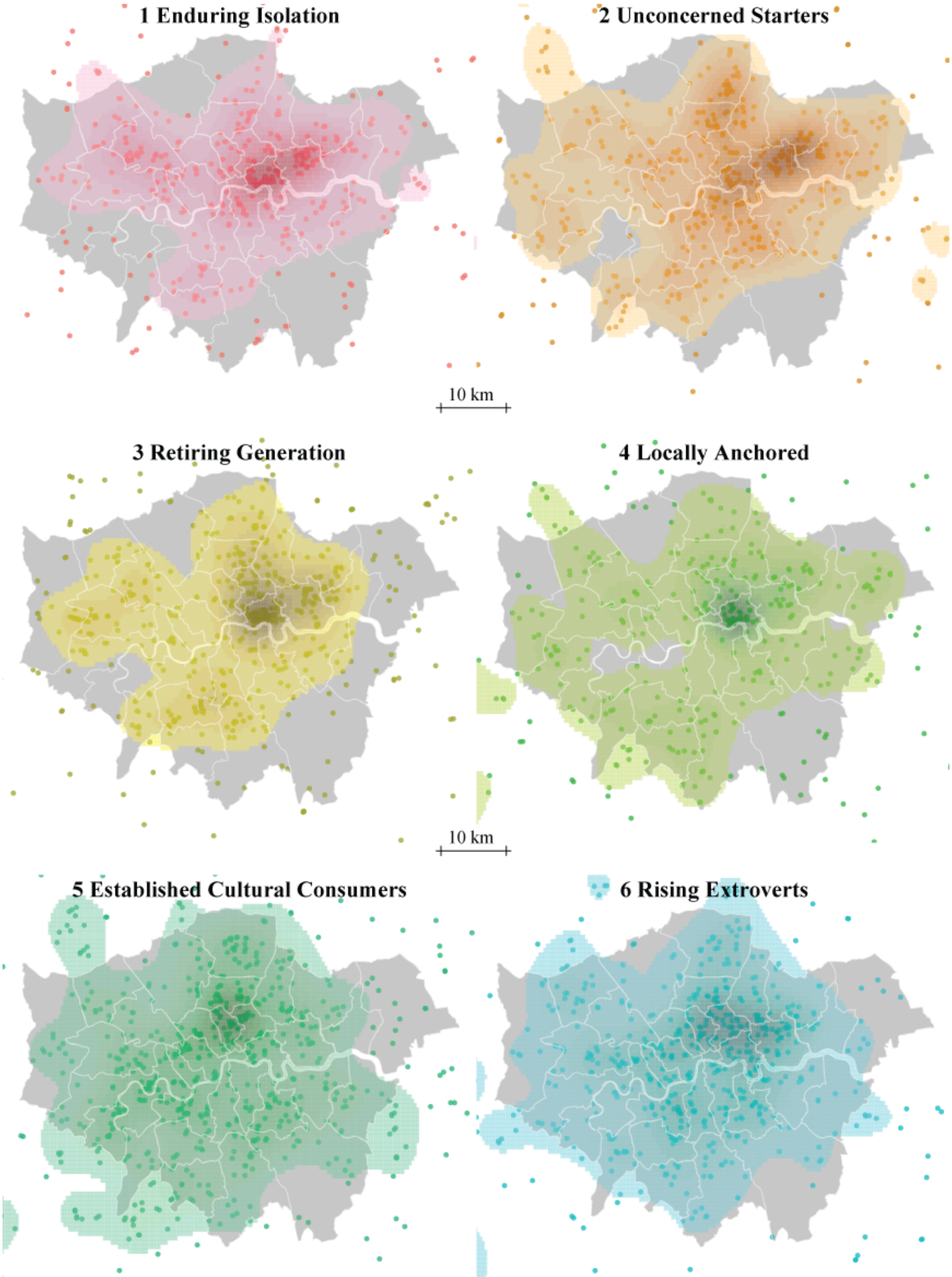



Figure 7.6: Kernel density estimates of the milieus based on direct assignment of survey respondents.

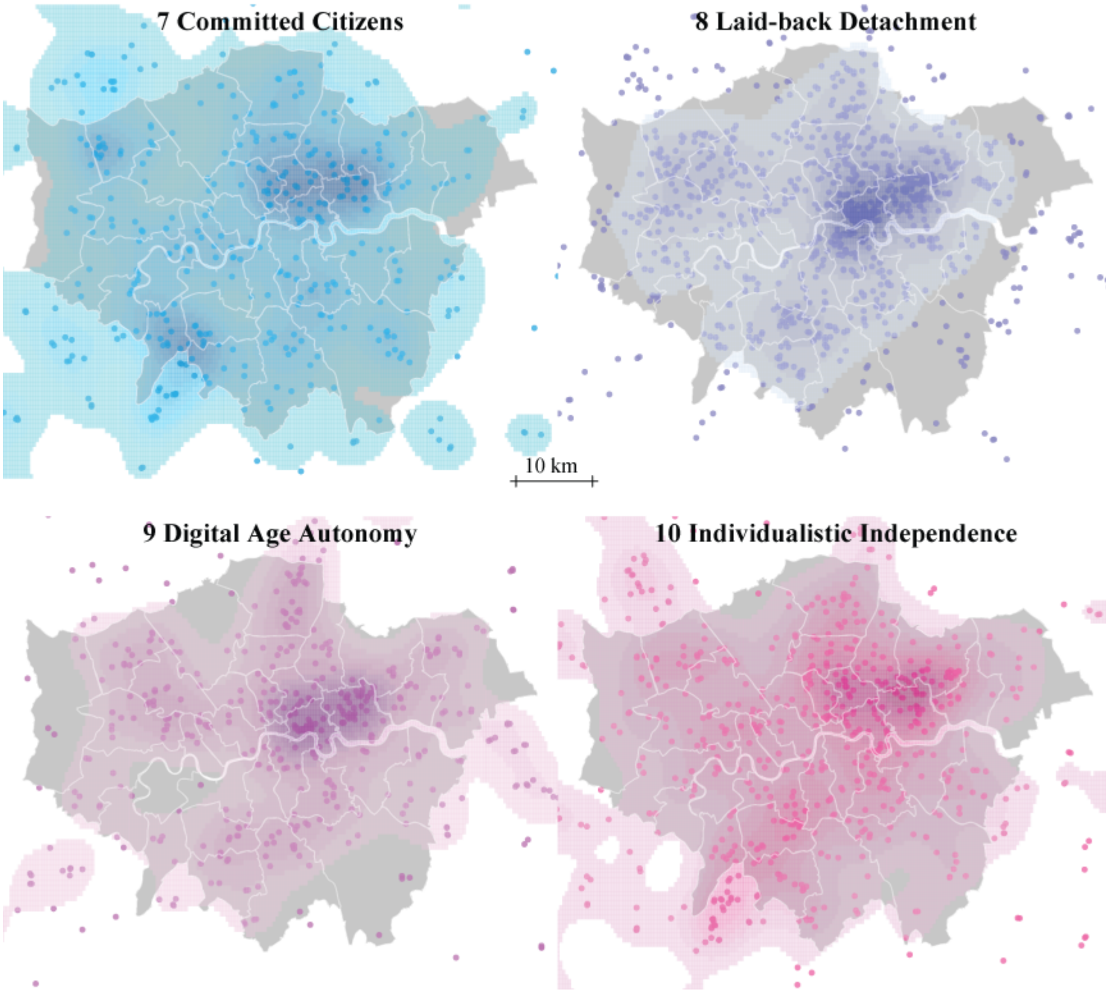



Figure 7.6 (continued)

### External triangulation with the London Output Area Classification

The results can also be triangulated using an external data source. The London Output Area Classification (LOAC) is a neighbourhood classification based on census statistics. It thus draws on the same data source as the microsimulation models. LOAC classifies London Output Areas – a geography below LSOAs – into eight groups and provides short profiles for each [Figure 7.7].

Since the health milieus are simulated to the LSOA level, the LOAC is aggregated by assigning to each LSOA the most common LOAC group – measured by number of Output Areas – within each LSOA. The LSOA-level milieu prevalences derived from the combined local model are converted into the location quotient (LQ), and the average LQ is estimated for each LOAC group at LSOA level. Oneway ANOVAs indicate that

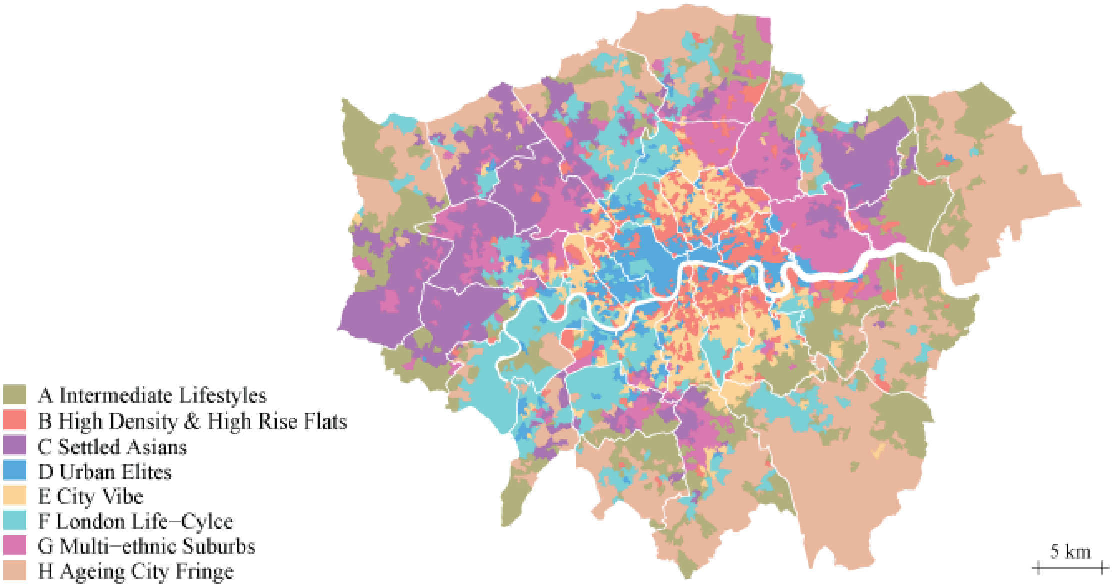



Figure 7.7: The London Output Area Classification aggregated to LSOA level.

associations between milieu-wise LQs and LOAC group are statistically significant [Table 7.3].

The milieu prevalences associated with each LOAC group correspond plausibly. The most prominent consonances can be observed with Settled Asians and the *Laid-back Detachment* milieu, where the share of Asians is higher. The Urban Elites, on the other hand, correspond mostly to the geography of *Established Cultural Consumers*, *Rising Extroverts* and *Committed Citizens*. These milieus are also more prevalent in areas classified as Urban Vibe and London Life-cylce, two types of areas which are described as more affluent and economically active in LOAC. Multi-ethnic suburbs tend to be composed of *Enduring Isolation* and *Laid-back Detachment* milieus in more deprived

Table 7.3: Average location quotients of milieus in each LOAC category.

| | **Milieus** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **LOAC category** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| A. Intermediate Lifestyles | 1.14 | 1.12 | 0.99 | 1.06 | 0.78 | 0.77 | 0.82 | 0.95 | 1.45 | 0.99 |
| B. High Density & High Rise | 1.30 | 1.21 | 1.07 | 1.09 | 0.65 | 0.87 | 0.72 | 1.01 | 1.28 | 0.88 |
| C. Settled Asians | 1.12 | 0.92 | 1.24 | 0.99 | 0.79 | 0.64 | 0.92 | 1.39 | 0.86 | 1.01 |
| D. Urban Elites | 0.76 | 0.90 | 0.83 | 0.77 | 1.51 | 1.76 | 1.20 | 0.69 | 0.60 | 0.95 |
| E. City Vibe | 0.92 | 0.99 | 0.89 | 0.95 | 1.21 | 1.30 | 1.11 | 0.83 | 0.88 | 0.94 |
| F. London Life Cycle | 0.62 | 0.79 | 0.84 | 0.86 | 1.49 | 1.29 | 1.40 | 0.82 | 0.71 | 1.16 |
| G. Multi-Ethnic Suburbs | 1.21 | 1.16 | 1.03 | 1.05 | 0.69 | 0.77 | 0.74 | 1.24 | 1.16 | 0.92 |
| H. Ageing City Fringe | 0.75 | 0.83 | 1.00 | 1.17 | 1.14 | 0.85 | 1.24 | 0.91 | 0.97 | 1.18 |

**milieus in columns**: 1 Enduring Isolation, 2 Unconcerned Starters, 3 Retiring Generation, 4 Locally Anchored, 5 Established Cultural Consumers, 6 Rising Extroverts, 7 Committed Citizens, 8 Laid-back Detachment, 9 Digital Age Autonomy, 10 Individualistic Independence.

contexts. In the Ageing City Fringe, the indeed maturer and more established milieus of *Committed Citizens* and *Locally Anchored* reside along with the demographically more diverse *Individual Independence* milieu. Hints of inconsistency can be found with respect to areas described as Intermediate Lifestyles: they coincide with neighbourhoods in which the younger *Digital Age Autonomy* milieu is more prevalent. The highest yet weaker location quotient in High Density and High-rise Flats is the *Retiring Generation*, followed again by *Digital Age Autonomy* and more materially deprived milieus.

The external triangulation shows a high level of correspondence but also some indications of inconsistency. While LOAC as a neighbourhood classification may suit the purpose of triangulation, it should be noted that both the health milieu geographies and LOAC draw on the same data source and are therefore not independent. Nevertheless, if LOAC does not fully establish a validation of health milieu geographies, it may conversely lend credibility to LOAC by using milieus as guides for interpretation.

## 7.5 Synthesis: milieu geographies and geodemographics

### Advantages of the models over conventional geodemographics

The local microsimulation models of the ten milieus offer an alternative, contextual picture of inequalities in London beyond mapping of social class or deprivation, highlighting emergent geographies of vulnerability, associated activity patterns, orientations and their differential expressions. Various forms of triangulation confirm that these geographies are plausible in London. Accounting for place effects as they may be manifest in culturally delineated regions improves the estimated geographies. Furthermore, a closer look at the geography of error associated with the split local model reveals that it is the urban English sub-sample that maps poorly to the local population, while the southern one is at least as good as the global sample if not better.

As far as geodemographics as a hermeneutic is concerned, the milieu approach in conjunction with spatial microsimulation offers some advantages over neighbourhood classifications that are based on aggregate statistics only. First, individuals rather than areas are classified; this permits a more thorough theoretical and interpretative engagement with the topic of lifestyles, their contexts and behavioural tendencies. Individual sample segmentation can be directly framed within social theory. So, too, can neighbourhood classifications be linked to theory, but they require some kind of ecological reasoning, which is often difficult to verify empirically as experienced in the neighbourhood effects literature and often assumes axiomatic or metaphorical character ("Birds of a feather flock together").

Second, since the milieu approach is directly informative about individuals, the nevertheless ecological nature of synthetic population distributions produces directly 'testable' geographies. The kernel density estimates of survey respondents permits comparison of geographies based on like-with-like. A similar large scale sample pertaining to individuals can be used to verify geographies based on what is known about the individual observation in both source datasets. This kind of testing is more complicated with neighbourhood classifications because a direct match between respondent characteristics and area statistics transcends ecological levels. Therefore, while milieu geographies are direct spatial hypotheses or inferences derived from aggregation, geodemographic classifications are ecological descriptions based on aggregate statistics. Which of the two is more pertinent depends on the research problem; in the context of health, where the interest is in people, the former seems to be more appropriate.

Third, the milieu approach estimates area compositions while geodemographics classifies their aggregate manifestations. The result of the former is a continuous geography measuring degrees, while the result of the latter is a discrete geography with sharp boundaries. In the milieu approach, it is thus possible to disentangle different sources of uncertainty: the sample segmentation and subsequent lifestyle characterisation as a source of substantive uncertainty and the process of aggregation resulting in ecological uncertainty. In conventional geodemographics, both types of uncertainty collapse into the latter. Hence, in the milieu approach it is possible to explore the distribution of substantive uncertainty by mapping individual cluster distances and separately ecological uncertainty by mapping classification errors. The results for London demonstrate that the spatial patterns can be different. Further evidence of substantive uncertainty can be obtained from estimations of the explanatory power of socio-demographic variables in individual milieu membership. This can be achieved in logit models for the sample, and if geo-located, it can be extended to a spatial, geographically weighted variant.

#### Uncertainty inherent in the model

The triangulation would benefit from further comparisons with external data sources. Triangulation and validation is a common difficulty for microsimulations because the absence of adequate data establishes the need to simulate in the first place. Candidate datasets for triangulation are the English Indices of Multiple Deprivation and commercial neighbourhood classifications (such as MOSAIC or ACORN), but since they draw on the census, too, they do not ensure statistical independence. Indeed, classifications that depend highly on the census may themselves be in need of validation as regards their qualitative inference of lifestyles from area characteristics. The inconsistency between health milieu geographies and LOAC, for example, may be due not to errors in the milieus but to some form of ecological fallacy committed in this way. Triangulation often works both ways and benefits from the breadth of information used.

Another opportunity for triangulation may consist in social media or consumer data. Their large sample sizes offer opportunities for contents mining, and if that can be geo-located, a direct comparison with what we would expect based on milieu profiles and actual behaviours may yield further insight. Such a dataset may itself constitute a layer for geodemographic representations in addition to milieu geographies.

Triangulation can address some of the shortcomings of conventional geodemographics, albeit as a post-hoc measure and not as an internal component of the classification. Survey samples or other individual datasets can be explored with respect to their geodemographic context. This exploration can then be used to confirm or modify generalised pen profiles of neighbourhood categories. Multi-level models, as often applied in neighbourhood effects research, may be suitable for this purpose. The aim of these models would not be, however, to explain, but to assess the adequacy of the qualitative contents of geodemographic categories. Depending on the kind of triangulation data available, a milieu approach may still be a preferable alternative.

In both the milieu approach and geodemographics, it is possible to incorporate place effects. The simplest way in geodemographics is to copy the approach proposed here and develop local classifications that are nested in, for example, regions of surname compositions and link regional classifications to a national reference classification. The theoretical assumption would be that neighbourhoods as outcomes of urban policy are subject to the same national strategic policy regime, which, however, translates differentially through regional regimes, political and social practices. This framing of geodemographics may invite experiments to include additional data related to jobs, physical conditions, accessibility in order to better contextualise and characterise place specific outcomes. One may also consider to use the milieu approach as a layer in a wider geodemographic system that incorporates direct geo-spatial and other thematic information, such as health care.

### Possible extensions: modelling urban dynamics in health

Finally, the milieu approach offers the possibility to assess impacts of policy through a dynamic variant of spatial microsimulation. Bhattacherjee and colleagues [2015], for example, employ dynamic microsiumlation in Scottish regions to simulate spatially differential impact on health after a hypothetical educational policy intervention. They find that such a policy would produce beneficial outcomes in some but not all of the regions. Although the authors do not consider individual context – health-related activities, health and education are treated again as independent variables – the study showcases an approach with direct policy relevance.

In the context of the health milieus, one could test, for instance, the impact of influencing smoking habits in the *Enduring Isolation* milieu on smoking prevalence in the general population or in selected localities. It is also possible to estimate resulting geographies of health conditions or health care utilisation, since they are directly included in the survey. Latent choice models for the sample could generate hypotheses as to how health-relevant choices are being made in different milieus. The expected population-wide and spatial effect of altering those choices could then be estimated. In some cases, this could resolve existing conundrums as to why the same health campaigns are successful in some places but not in others [Baum & Fisher 2014].

But hypothetical spatial patterns emerging from static microsimulation could also be used in a different way. By asking the questions, where are the good schools in London, and, how are they associated with the health milieu geographies, it may be possible to gain insight into the drivers of residential mobility of more economically secure milieus with children (*Locally Anchored*, *Individualistic Independence*). Here, space provides contextual information as a basis for a new hypothesis for drivers of residential choice that may shape vulnerability in the long-run through unequal access to cultural capital and life chances.

Another hyothetical question, more directly linked to health, could be whether the areas where *Laid-back Detachment* live are walkable. Together with evidence of their subjective orientations, it may then be estimated how likely these individuals are to walk after some kind of public health intervention. Policy-makers could thus frame interventions within different elements of evidence that pertain to causal pathways, discrete spatial context as well as expected effectiveness given milieu-specific preferences, needs and constraints. Methodologically, a combination of dynamic microsimulation – also with its potential to weight potentially unstructured 'Big' Data – and recently advanced forms of latent choice models [Hess et al. 2013] could deliver significant research and policy benefits, if explored further.

# 8 Towards an advanced geodemographic framework of vulnerability

Health milieus and health environments each reflect two different levels of vulnerability: the individual in context of the social environment and areas based on aggregate observations. Within an advanced geodemographic framework, the information at the two levels can be viewed jointly in order to contextualise health inequalities and inform urban policy responses.

## 8.1 Health milieus and health environments: a compositional view

Drawing on the inferred spatial distribution of health milieus, the individual perspective is combined with that of health environments through ecological linkage at the level of Lower Layer Super Output Areas (LSOAs). To briefly recollect, seven health environments have been identified earlier in the thesis:

1. Poor Health Capital. Likelihood of good self-rated health is low, risk of all conditions is increased.

2. Mild Health Advantage. Self-rated health better than average, risk of all conditions are slightly lower than average.

3. Strong Health Capital. Self-rated health significantly better, while risk of all other conditions is lowest

4. Average Tendencies. All indicators except conditions arising from the peri-natal period are around the average in their distributions.

5. Organ Damage & Mental Illness. Self-rated health is lower, some other conditions are less common, whereas conditions related to injuries, specific organ damage and psychology are more common.

6. Strong Disease Burden. Risk of poor self-rated health and all conditions is strongly increased, except for cancer-related conditions, which are less common.

7. Persistent Cancer. Opposing tendencies between self-rated health and risk of cancer-related conditions as well as conditions concerning some outer organs. All other conditions are less common.

Each LSOA is associated to a health environment and each milieu's likely local prevalence is taken, using the previously developed estimates [see chapter 7]. The distribution

Table 8.1: Associations between health milieus (rows) and health environments (columns).

| | Poor HC. | Mild HA. | Strong HC. | Avg. T. | Org.D.-MI. | Str.DB. | Pers.C. | test |
|---|---|---|---|---|---|---|---|---|
| | (a) | (b) | (c) | (d) | (e) | (f) | (g) | |
| **milieus' residential structure (per cent)** | | | | | | | | |
| End.I. (1) | 18.0 | 12.9 | 6.1 | 17.8 | 18.2 | 16.6 | 10.3 | .000 |
| Unc.S. (2) | 18.5 | 14.5 | 8.0 | 17.7 | 16.5 | 13.3 | 11.5 | .000 |
| Ret.G. (3) | 15.2 | 15.6 | 8.4 | 17.0 | 17.0 | 14.2 | 12.6 | .000 |
| Loc.A. (4) | 16.8 | 15.4 | 8.6 | 16.9 | 16.2 | 12.5 | 13.6 | .000 |
| Est.CC. (5) | 13.1 | 20.1 | 14.7 | 17.0 | 13.4 | 7.2 | 14.5 | .000 |
| Ris.E. (6) | 15.1 | 18.8 | 13.6 | 18.1 | 14.0 | 8.2 | 12.2 | .000 |
| Com.C (7) | 13.2 | 19.4 | 13.8 | 16.4 | 14.0 | 8.3 | 14.8 | .000 |
| Lai.D. (8) | 16.2 | 15.1 | 8.2 | 17.8 | 16.3 | 15.2 | 11.1 | .000 |
| Dig.AA. (9) | 20.1 | 13.4 | 6.5 | 16.0 | 17.7 | 13.7 | 12.5 | .000 |
| Ind.I. (10) | 14.3 | 17.3 | 11.0 | 16.3 | 15.3 | 11.3 | 14.5 | .000 |
| **milieus' location quotient of milieus** | | | | | | | | |
| End.I. (1) | 1.13 | 0.79 | 0.62 | 1.04 | 1.15 | 1.37 | 0.81 | .000 |
| Unc.S. (2) | 1.16 | 0.88 | 0.80 | 1.04 | 1.05 | 1.10 | 0.90 | .000 |
| Ret.G. (3) | 0.95 | 0.96 | 0.84 | 0.99 | 1.08 | 1.17 | 0.99 | .000 |
| Loc.A. (4) | 1.05 | 0.95 | 0.87 | 0.99 | 1.02 | 1.04 | 1.08 | .000 |
| Est.CC. (5) | 0.80 | 1.24 | 1.49 | 0.99 | 0.84 | 0.59 | 1.14 | .000 |
| Ris.E. (6) | 0.92 | 1.13 | 1.35 | 1.04 | 0.87 | 0.68 | 0.95 | .000 |
| Com.C (7) | 0.82 | 1.20 | 1.40 | 0.96 | 0.88 | 0.69 | 1.16 | .000 |
| Lai.D. (8) | 1.02 | 0.93 | 0.83 | 1.04 | 1.03 | 1.25 | 0.88 | .000 |
| Dig.AA. (9) | 1.28 | 0.83 | 0.67 | 0.95 | 1.14 | 1.16 | 0.99 | .000 |
| Ind.I. (10) | 0.89 | 1.07 | 1.11 | 0.96 | 0.97 | 0.94 | 1.14 | .000 |

**columns**: (a) Poor Health Capital, (b) Mild Health Advantage, (c) Strong Health Capital, (d) Average Tendencies, (e) Organ Damage & Mental Illness, (f) Strong Disease Burden, (g) Persistent Cancer; **rows**: (1) Enduring Isolation, (2) Unconcerned Starters, (3) Retiring Generation, (4) Locally Anchored, (5) Established Cultural Consumers, (6) Rising Extroverts, (7) Committed Citizens, (8) Laid-back Detachment, (9) Digital Age Autonomy, (10) Individualistic Independence.

of health environments among the residences of each health milieu is determined in order to contextualise health milieus and health environments and characterise, for each milieu, the most likely residential structure in terms of health environments. In addition, each health environment is investigated with respect to their composition, or how likely each milieu is to occur in each of the environments. $\chi$-squared and ANOVA-based tests are applied to determine the statistical significance of results [Table 8.1]; they reveal that there is a significant association between health milieu prevalences and health environments, although the two measures draw on different datasets and are technically independent.

The most common health clusters among the residences of the population attributed to the *Enduring Isolation* milieu are the Organ Damage & Mental Illness, Poor Health Capital, Average Tendencies and the Strong Disease burden environments. 70 per cent of this milieu live in these four health environments. The milieu are 1.37 times more likely to live in the Strong Disease Burden environment than other milieus. They are least likely to live in the Strong Health Capital environment.

The most common residential environment of *Unconcerned Starters* are the Poor Health Capital, Average Tendencies and Organ Damage & Mental Illness environments with more than half of the milieu living in those. They are 1.16 times more likely to live in the Poor Health Capital environment and least likely to reside in the Strong Health Capital counterpart.

Similar environments are common within the *Retiring Generation* milieu with nearly half of the milieu residing in environments Poor Health Capital, Average Tendencies and Organ Damage & Mental Illness. In addition, 16 per cent reside in the Mild Health Advantage environment despite their relative social disadvantage. This reflects their being located in more affluent areas perhaps enabled through earlier acquisition of property. On the whole, they are most likely to live in the Strong Disease Burden environment with an average location quotient of 1.17. As the other two previous milieus, they are least likely to live in the Strong Health Capital environment.

The majority of the *Locally Anchored* milieu live in the same environments as *Retiring Generation*. Two in three of this group live in environments Average Tendencies, Poor Health Capitals, Organ Damage & Mental Illness or Mild Health Advantage. Yet they are eight per cent more likely to live in the cancer high prevalence environment, while they are least likely to reside in the Strong Health Capital environment, too.

One in five *Established Cultural Consumers* live in the Mild Health Advantage environment and a further 17 per cent in the Average Tendencies environment. Yet, they are nearly 1.5 times more likely to live in the Strong Health Capital environment than other milieus. They are least likely (.60) to reside in the Strong Disease Burden environment, which is consistent with their observed individual health. Furthermore, they are only .80 times as likely to live in the Poor Health Capital and Organ Damage & Mental Illness environments.

Similar tendencies as with *Established Cultural Consumers* can be observed for both *Rising Extroverts* and *Committed Citizens*. 15 per cent of the *Rising Extroverts* live in the Poor Health Capital environment, which may be due to their distribution in recently gentrified areas in London with the health data dating back from 2008/2009. 15 per cent of the *Committed Citizens* live in the Persistent Cancer environment. Both milieus are most likely than other milieus to live in the Strong Health Capital environment, however, and least likely to reside in the Strong Disease Burden environment.

Half of the *Laid-back Detachment* milieu live in the Average Tendencies, Organ Damage & Mental Illness and Poor Health Capital environments, therefore in presumably less healthy areas. They are 1.25 times more likely to reside in the Strong Disease Burden environment and hence exhibit a similar residential pattern than more disadvantaged milieus.

20 per cent of the *Digital Age Autonomy* live in the Poor Health Capital environment and a third is distributed between Organ Damage & Mental Illness and the Average Tendencies environments. They are 1.28 times more likely to live in the Poor Health Capital environment, followed by Strong Disease Burden and Organ Damage & Mental Illness.

Members of the *Individualistic Independence* milieu distribute more evenly across residential environments. The share of all residential environments vary between eleven and 17 per cent with Mild Health Advantage being the most common environment. They are 1.14 more likely to live in the Persistent Cancer environment, owing to their concentration in the southeastern outer suburbs. They are also eleven per cent more likely to live in the Strong Health Capital environment, which is consistent with their more affluent socio-demographic context.

The distribution of health milieus across different health environments corresponds to what could be informally expected from the milieus based on the information collected about them in previous chapters. Disadvantaged milieus tend to be associated with in environments that indicate health disadvantage and, vice versa, more affluent milieus live in areas indicating better health. The most dividing health environments are Strong Health Capital and the Strong Disease Burden and to a degree Poor Health Capital. The Average Tendencies environment is likely to host all milieus equally, which also corresponds to its interpretation. The different likelihoods of milieus residing in those environments may confirm the presence of social pathways linking social background to health at the level of small areas.

The health environments could therefore be validated to some extent by testing the plausibility of their social composition, and the inferred milieu geographies could be validated by examining the plausibility of health contexts. Both sets of information confirm and thereby contextualise each other, which is a necessary step to design effective policy interventions.

## 8.2 The spatial structure of neighbourhood vulnerability

The previous descriptive analysis of milieus and health environments shows some associations between milieus and neighbourhood health, suggesting that the milieus may possess some predictive power with regards to area health challenges. In epidemiology, these association are typically estimated by means of statistical models, in which the significance of covariates indicates the presence of particular pathways. This type of investigation may potentially be useful in identifying starting points for policy; yet, their privileging of variables has invited the criticism detailed at the beginning of this thesis. Here, a hybrid approach that combines case-based variables with predictive techniques is applied. Local milieu prevalence are used as covariates. Although they are treated as covariates in statistical terms, in substantive terms, they are case-focussed and may describe local populations more fully than a range of separately specified variables, such as

Table 8.2: Ecological correlations between conditions and milieus.

| | | End.I. | Unc.S. | Ret.G. | Loc.A. | Est.C. | Ris.E. | Com.C. | Lai.D. | Dig.A. | Ind.I. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (a) | (b) | (c) | (d) | (e) | (f) | (g) | (h) | (i) | (j) |
| **asset** | | | | | | | | | | | |
| S.H.R | (1) | -.834 | -.699 | -.320 | -.275 | .280 | .490 | .782 | -.402 | -.562 | .533 |
| **deficit** | | | | | | | | | | | |
| Inf.P.D. | (2) | .484 | .373 | .149 | .060 | -.135 | -.248 | -.447 | .334 | .192 | -.416 |
| Resp.I. | (3) | .463 | .345 | .124 | .104 | -.139 | -.205 | -.407 | .143 | .270 | -.284 |
| Mat.C. | (4) | .323 | .265 | .189 | .301 | -.152 | -.530 | -.418 | .514 | .401 | -.029 |
| C.perin. | (5) | .239 | .276 | -.146 | -.100 | -.085 | -.017 | -.317 | .181 | .052 | -.287 |
| Nutr.D. | (6) | .435 | .303 | .211 | .062 | -.130 | -.254 | -.400 | .345 | .151 | -.330 |
| M.neop. | (7) | -.031 | .142 | -.143 | .118 | .027 | .119 | .021 | -.288 | .235 | -.025 |
| O.neop. | (8) | .043 | .184 | -.068 | .124 | -.016 | -.053 | -.059 | .032 | .236 | -.074 |
| Diab.m. | (9) | .418 | .395 | .126 | .213 | -.135 | -.293 | -.420 | .249 | .369 | -.331 |
| End.D. | (10) | .444 | .416 | .136 | .151 | -.127 | -.183 | -.431 | .235 | .242 | -.413 |
| N.Psych. | (11) | .488 | .441 | .110 | .111 | -.121 | -.154 | -.450 | .082 | .355 | -.345 |
| S.O.D. | (12) | .552 | .360 | .263 | .188 | -.143 | -.378 | -.481 | .372 | .246 | -.362 |
| C.V.D. | (13) | .538 | .356 | .283 | .170 | -.136 | -.389 | -.465 | .324 | .315 | -.325 |
| Resp.D. | (14) | .623 | .454 | .251 | .183 | -.161 | -.328 | -.528 | .172 | .401 | -.386 |
| Dig.D. | (15) | .512 | .414 | .235 | .242 | -.119 | -.383 | -.456 | .229 | .467 | -.265 |
| Genit.D. | (16) | .431 | .385 | .235 | .349 | -.134 | -.476 | -.423 | .344 | .501 | -.135 |
| Sk.D. | (17) | .236 | .244 | .059 | .101 | -.051 | -.130 | -.193 | .045 | .262 | -.190 |
| Musc.D. | (18) | .403 | .300 | .213 | .339 | -.094 | -.403 | -.331 | .134 | .513 | -.074 |
| Cong.D. | (19) | .285 | .265 | .027 | .113 | -.082 | -.155 | -.310 | .179 | .225 | -.220 |
| Oral | (20) | .217 | .326 | .030 | .273 | -.040 | -.125 | -.198 | .016 | .387 | -.227 |
| Injur. | (21) | .381 | .307 | .101 | .086 | -.085 | -.163 | -.316 | .041 | .325 | -.249 |
| N.E.C. | (22) | .584 | .410 | .302 | .157 | -.151 | -.430 | -.530 | .358 | .356 | -.275 |
| H.Serv. | (23) | .274 | .345 | -.031 | .069 | -.093 | -.073 | -.339 | .122 | .203 | -.288 |

**columns**: (a) Enduring Isolation, (b) Unconcerned Starters, (c) Retiring Generation, (d) Locally Anchored, (e) Established Cultural Consumers, (f) Rising Extroverts, (g) Committed Citizens, (h) Laidback Detachment, (i) Digital Age Autonomy, (j) Individualistic Independence.
**rows**: (1) self-rated health (SHR); (2) infectious and parasitic diseases; (3) respiratory infections; (4) maternal conditions; (5) conditions from perinatal period; (6) nutritional deficiencies; (7) malignant neoplasms; (8) other neoplasms; (9) diabetes mellitus; (10) endocrine disorders; (11) neuro-psychiatric conditions; (12) sense organ diseases; (13) cardiovascular diseases; (14) respiratory diseases; (15) digestive diseases; (16) genito-urinary diseases; (17) skin diseases; (18) musculoskeletal diseases; (19) congenital anomalies; (20) oral conditions; (21) injuries; (22) not elsewhere classified; (23) health-service related.

age, sex or socio-economic status. Local milieu prevalences are then put into a predictive statistical model.

Ecological correlations between local milieu prevalence and individual diseases are in parts consistent with earlier findings, in parts they reveal some divergent trends that reflect the nature of the data [Table 8.2]. The milieu that shows the strongest correlation with area-level incidence rates is *Enduring Isolation*. The top three diseases in terms of correlations are respiratory diseases, unclassified conditions and sense-organ diseases. The correlation with self-rated health is negative, with a strong coefficient of .834. The reverse applies to the *Committed Citizens* milieu; they show strong negative correlation with all conditions except malignant and other neoplasms. *Established Cultural Consumers* also show negative correlations, but to a considerably lower level than *Committed Citizens*. This may be due to their location in areas that can be adjacent to areas with more disadvantaged population groups. *Unconcerned Starters* and *Digital Age Autonomy* tend to follow the pattern of *Enduring Isolation*, whereas *Rising Extroverts* and *Individualistic Independence* are more similar to *Committed Citizens*.

### A spatial-structural model of health milieus and health environments

The ecological correlation, as they are unadjusted for spatial autocorrelation, reflect the different geographical distributions of the milieus. The similarity between some milieus indicates a strong spatial covariance among milieus. In predictive models, this covariance can cause large standard errors and thus imprecise estimates of associations. The multicollinearity between milieus is tested using the Variance Inflation Factor (VIF), which measures the extent of covariation between variables in predictive models. Based on a step-wise inclusion of each milieu in turn, the VIF is calculated and at each iteration the milieu showing a value of three or more (high inflation) is excluded, resulting in six uncorrelated milieus: *Unconcerned Starters*, *Retiring Generation*, *Locally Anchored*, *Laid-back Detachment*, *Digital Age Autonomy* and *Individualistic Indpendence*. The predictive spatial-structural model [Besag et al. 1991; see chapter 5.2 and equation 5.2] is expanded by the covariates representing the prevalence of those milieus that are uncorrelated.

The strength of the spatial-structural model is that it accounts for spatial autocorrelation by modelling the impact of neighbouring areas on the outcomes of a given area. As has been found earlier, all conditions and self-rated health are marked by spatial structure. The expanded models with the milieu covariates show that the milieus account for additional variation for almost all conditions [Table 8.3]. In all models, the Deviance Information Criterion (DIC) reduces significantly. The smallest reduction occurs for malignant and other neoplasms, the largest for self-rated health. In the latter case, an additional 87 per cent of variance is accounted for compared to the same model without covariates.

Self-rated health is positively associated with the presence of the *Locally Anchored* milieu and negatively related to all other milieus except *Individualistic Independence*. Despite low regression coefficients, the explanatory power is strong. This is because there are many observations on self-rated health (the entire population) and hence credible intervals are small. Using the statistical metaphor of "dosage-response", the coefficient suggests that if the share of *Locally Anchored* doubled in an area, there would be an increase of good self-rated health by 3.1 per cent (risk ratio of $exp(.031)$ of a change of LQ of one). This is very little, yet given the number of observations, a high level of confidence can be attributed to the existence of this effect.

Table 8.3: Explanatory models results of regression health conditions against milieus. Underlined coefficients are statistically significant at $p \leq .05$

| | | **Unc.S.** | **Ret.G.** | **Loc.A.** | **Lai.D.** | **Dig.A.** | **Ind.I.** | **DIC** | **ΔDIC** | **ΔpD** | **Var** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (a) | (b) | (c) | (d) | (e) | (f) | (g) | (h) | (i) | (j) |
| **asset** | | | | | | | | | | | |
| S.H.R | (1) | -.068 | -.071 | .031 | -.023 | -.077 | .004 | 47,705 | -3,216 | -1,612 | .867 |
| **deficit** | | | | | | | | | | | |
| Inf.P.D. | (2) | .193 | .115 | -.123 | .348 | .281 | -.082 | 21,659 | -177 | -218 | .408 |
| Resp.I. | (3) | .301 | .205 | .071 | -.036 | .302 | .124 | 25,937 | -215 | -203 | .350 |
| Mat.C. | (4) | -.080 | .216 | .022 | .516 | .405 | .334 | 34,914 | -317 | -344 | .409 |
| C.perin. | (5) | .143 | -1.901 | -1.491 | 1.934 | .323 | -1.648 | 20,422 | -113 | -168 | .250 |
| Nutr.D. | (6) | .428 | .259 | -.064 | .335 | .145 | -.007 | 18,219 | -105 | -107 | .385 |
| M.neop. | (7) | .099 | .026 | .003 | -.225 | .145 | .037 | 27,020 | -53 | -62 | .240 |
| O.neop. | (8) | .071 | .068 | -.006 | .048 | .169 | .101 | 23,408 | -51 | -42 | .077 |
| Diab.m. | (9) | .581 | .370 | -.335 | .223 | .395 | -.258 | 14,782 | -183 | -159 | .362 |
| End.D. | (10) | .268 | .192 | .029 | .193 | .278 | .013 | 21,728 | -177 | -168 | .374 |
| N.Psych. | (11) | .390 | .295 | -.171 | -.085 | .386 | .120 | 26,998 | -346 | -295 | .415 |
| S.O.D. | (12) | .178 | .175 | -.046 | .309 | .144 | -.197 | 27,120 | -242 | -236 | .423 |
| C.V.D. | (13) | .220 | .219 | -.066 | .091 | .192 | -.030 | 30,372 | -220 | -243 | .434 |
| Resp.D. | (14) | .382 | .404 | -.101 | -.027 | .388 | .021 | 25,774 | -421 | -421 | .545 |
| Dig.D. | (15) | .188 | .239 | -.038 | .021 | .268 | .097 | 32,666 | -399 | -417 | .427 |
| Genit.D. | (16) | .189 | .218 | .060 | .090 | .245 | .115 | 30,362 | -336 | -327 | .398 |
| Sk.D. | (17) | .116 | .202 | -.030 | .028 | .279 | .212 | 24,286 | -159 | -156 | .219 |
| Musc.D. | (18) | .109 | .270 | .047 | -.100 | .343 | .147 | 29,699 | -419 | -433 | .421 |
| Cong.D. | (19) | -.064 | -.218 | -.140 | .454 | .301 | -.231 | 17,552 | -87 | -84 | .253 |
| Oral | (20) | .195 | .363 | .142 | .092 | .440 | .259 | 25,054 | -448 | -391 | .367 |
| Injur. | (21) | .109 | .160 | -.007 | -.098 | .311 | .025 | 30,263 | -325 | -318 | .365 |
| N.E.C. | (22) | .252 | .259 | -.123 | .125 | .286 | .103 | 33,553 | -528 | -527 | .474 |
| H.Serv. | (23) | .069 | -.171 | -.133 | .445 | .276 | -.019 | 30,842 | -195 | -222 | .180 |

**columns**: (a) Unconcerned Starters, (b) Retiring Generation, (c) Locally Anchored, (d) Laid-back Detachment, (e) Digital Age Autonomy, (f) Individualistic Independence, (g) Deviance Information Criterion, (h) effective number of parameters, (i)-(j) difference to model without covariates, (k) additional variance accounted for;
**rows**: (1) self-rated health (SHR); (2) infectious and parasitic diseases; (3) respiratory infections; (4) maternal conditions; (5) conditions from perinatal period; (6) nutritional deficiencies; (7) malignant neoplasms; (8) other neoplasms; (9) diabetes mellitus; (10) endocrine disorders; (11) neuro-psychiatric conditions; (12) sense organ diseases; (13) cardiovascular diseases; (14) respiratory diseases; (15) digestive diseases; (16) genito-urinary diseases; (17) skin diseases; (18) musculoskeletal diseases; (19) congenital anomalies; (20) oral conditions; (21) injuries; (22) not elsewhere classified; (23) health-service related.

Malignant neoplasms co-occur with the *Unconcerned Starters* and *Digital Age Autonomy* milieus, *Laid-back Detachment* is associated with a reduction of risk by more than 20 per cent (risk ratio of $exp(-.225)$), if the share of the latter milieu doubled in an area. These results may seem unexpected: younger milieus are associated with higher incidence of cancer. Yet this may be explained by the potential co-location of related milieus or a temporal effect, if the two milieus remain sedentary. Onset of cancer may be more common in these milieus at a later life stage, or related higher generation milieus are more prone to developing cancer.

Maternal conditions, on the other hand, are positively associated with the presence of *Laid-back Detachment* – doubling the share leads to an increase in risk by 68 per cent – and negatively associated with *Unconcerned Starters*. The risk for neuro-psychiatric conditions is lower in areas with more residents of the *Locally Anchored* and *Laid-back Detachment* milieus and higher for all other milieus. Musculoskeletal diseases and injuries are lower in areas with more residents of the *Laid-back Detachment* group, higher for *Unconcerned Starters* and *Retiring Generation*.

Viewed all at once, the *Digital Age Autonomy* milieu show the strongest associations with the conditions, indicating that area health is worse where this milieu is located. Similar though weaker tendencies can be observed for *Unconcerned Starters* and *Retiring Generation*. The presence of *Locally Anchored* and *Individualistic Indpendence* tends to be associated with health advantage, although there are exceptions, notably cancer with its distinct geography.

Of course, doubling the share of one social milieu or other implies in reality that either existing milieus become like other milieus or certain milieus are attracted while others are displaced. Thus, these associations are not directly informative for policy interventions that seek to reduce population vulnerability. The models merely highlight ecological associations between health indicators and the inferred presence of health milieus representing a combination of a range of neighbourhood characteristics (derived from the Census). Here, not the individual associations are of interest but their explicit geographies.

### Local specificity of milieu-health associations: health spaces

The associations can be viewed as weak evidence that local population health is an outcome of generic categories of social similarity drawn together in local lifestyle milieus. As with all statistical models, true causality cannot be ascertained. Yet, the residuals reveal where other unmeasured factors outside the range of Census variables may account for health disparities. In spatial-structural models, the variance that is not accounted for can be divided into a spatially structured and a global, unstructured component. The former determines variance that can be attributed to local, spatial factors; the latter to other factors that operate universally without spatial pattern. Each type of residual can be summarised as a residual rate ratio with the same statistical properties as area-level incidence rate ratios (SHR or SMbRRs). Consequently, for each LSOA and each disease, the residual rate ratios resulting from spatial structure on the one hand and other unobserved factors on the other hand can be estimated.

For instance, plotting the two components of residuals for respiratory diseases reveal spatial pattern for the residual rate ratios captured in the spatial-structural component [Figure 8.1]. In some parts of London, notably the south east, there are higher incidence rates of respiratory diseases than expected based on local lifestyles (in addition to age and sex). Along with other concentrations of under and overestimation, the model

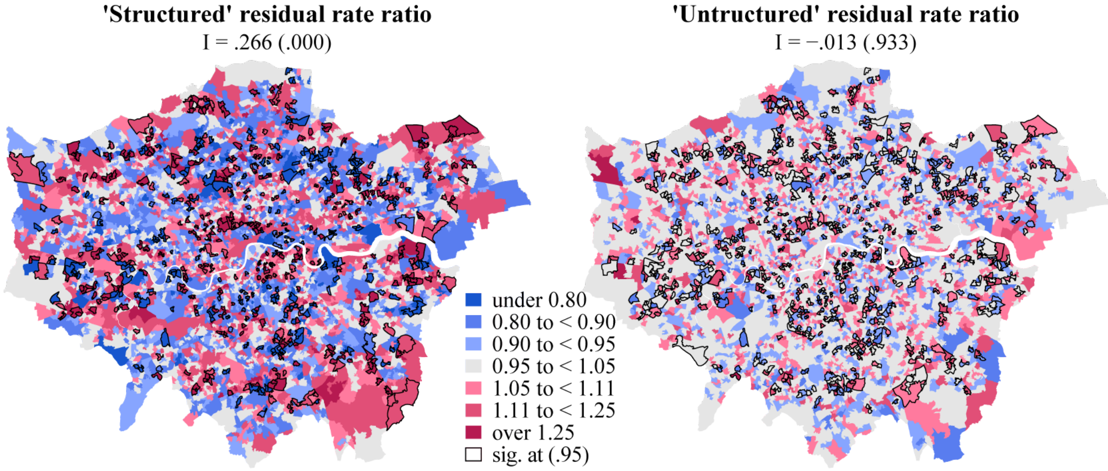



Figure 8.1: Residual incidence rate ratios of respiratory diseases for the spatially structured component (left) and the unstructured component (right).

exhibits significant spatial patterning (Moran's I of .266). In contrast, the residual rate ratios captured in the unstructured component seem to be randomly, spatially distributed (Moran's I is close to zero) throughout London with most areas showing little difference from one (no residual). The probability that a residual rate ratio is significantly different from one can be calculated based on the posterior marginals of the random effects.

The implications of high, absolute deviations from one are that the level of local vulnerability cannot be attributed to local lifestyle but is likely to be related to unmeasured local factors, which may include other characteristics of the local population, specific, local environmental conditions or poor access to health care or other relevant services. In those area, one is to conclude that an area needs special attention and that further investigation of local conditions may be required in order to discover the causes for specific local health challenges. The spatially structured residuals indicate that unobserved factors are spatially variant, whereas the unstructured residuals indicate a global (London-wide) effect.

Areas can be clustered by similar patterns of spatial-structural effects. The standardised spatial-structural residual ratios are used as input in the two-stage clustering procedure. The result is a classification of areas by different types of spatial effects on incidence rate ratios that cannot be attributed to local lifestyles but potentially other spatially concentrated factors. Five types of health spaces emerge; their ordering reflects the degree to which they deviate from expectation (no residual) as a whole [Figure 8.2]. The general interpretation of these clusters is the more the cluster variables deviate from zero – signifying no spatial effect on an uncentered z score – the more specific is the cluster and the more it requires cautious interpretation and focussed investigation.

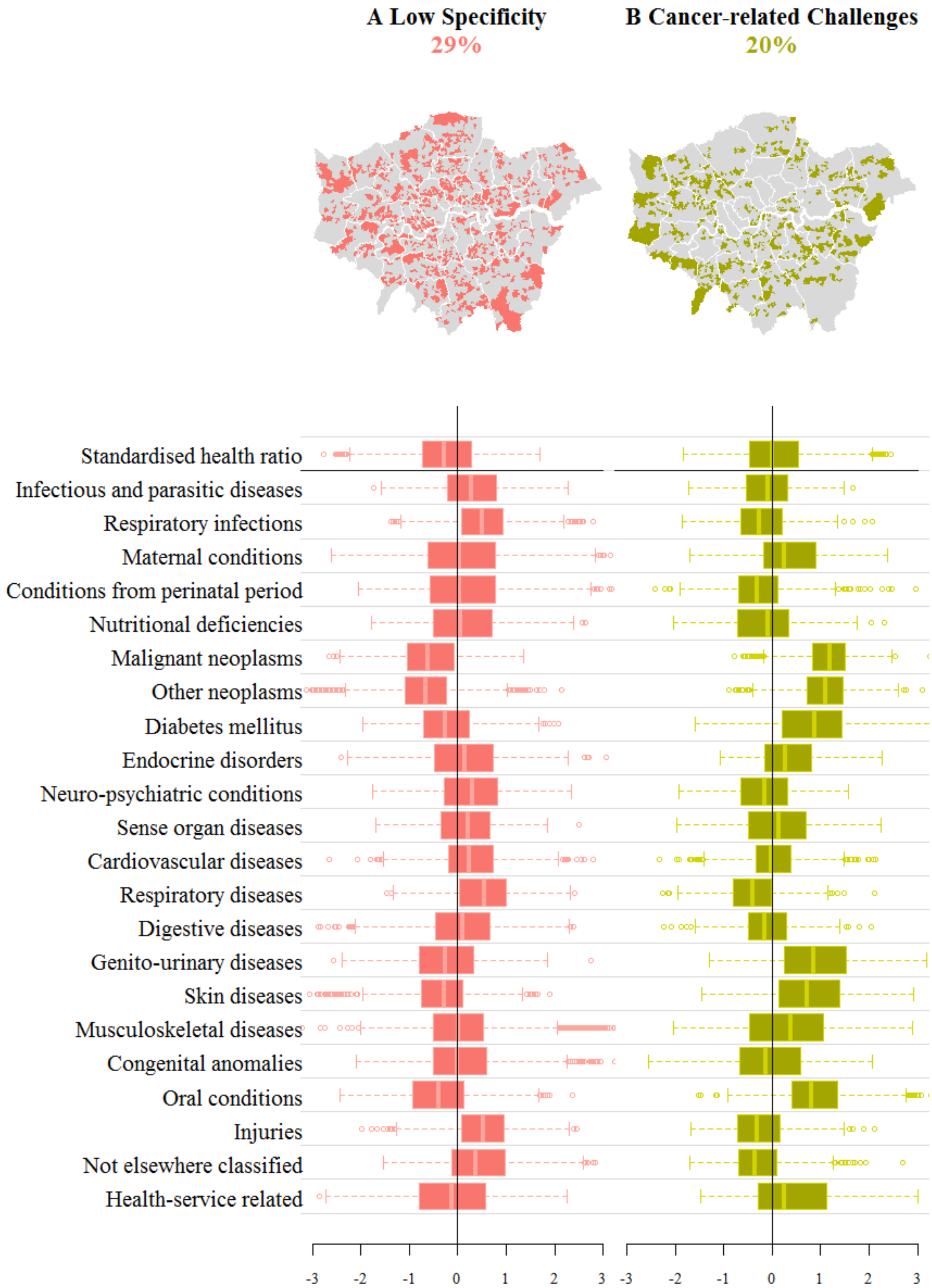



Figure 8.2: Geography and profiles of health spaces in London.

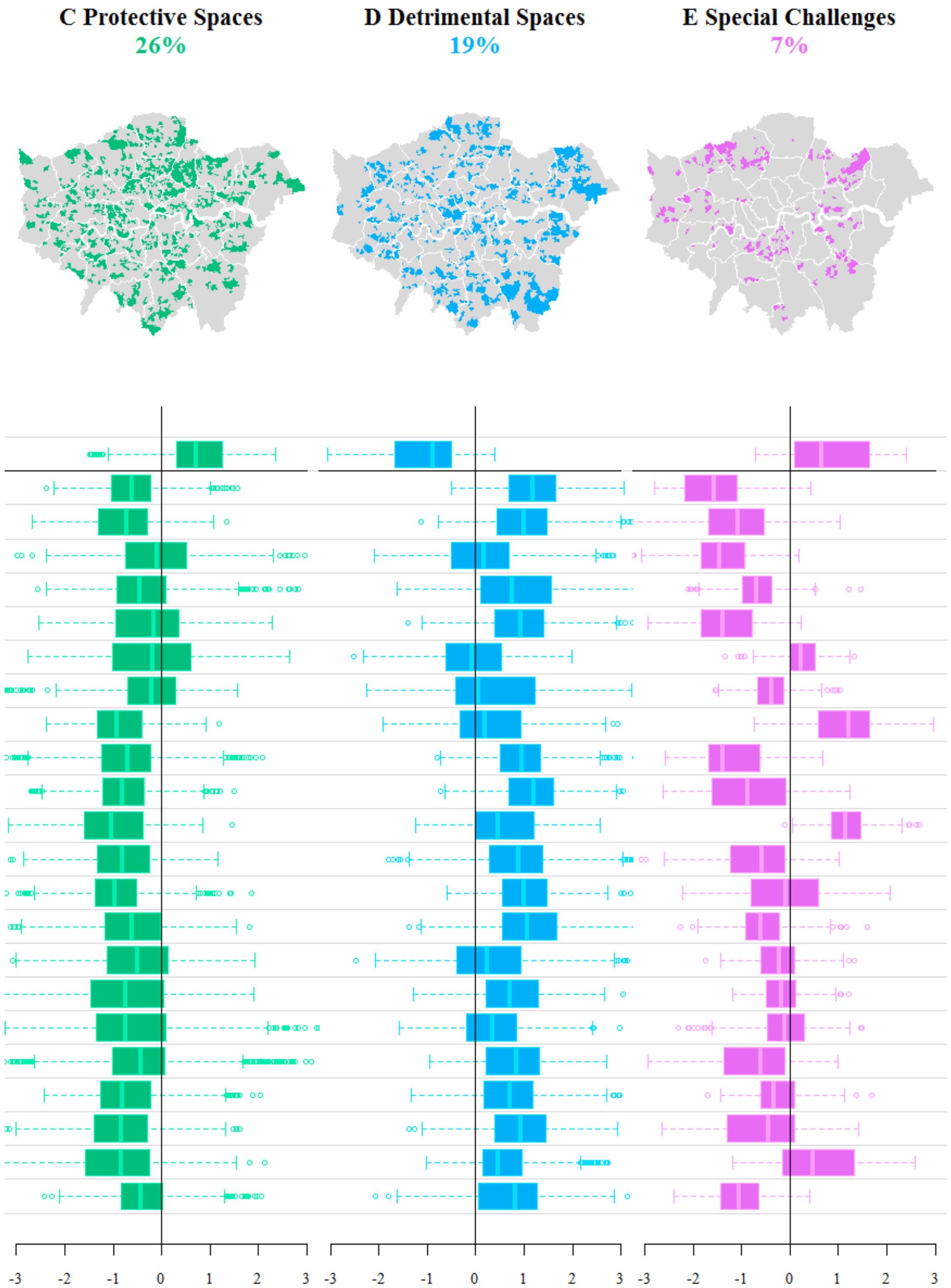



Figure 8.2 (continued)

The first cluster indicates locally specific increased risk for injuries, unclassified conditions, respiratory diseases and respiratory infections. At the same time, it shows locally specific protection from cancers. 30 per cent of areas belong to this cluster and are distributed across Greater London. The second cluster indicates increased risk of a number of conditions including diabetes, malignant and other neoplasms. 20 per cent of areas fall into this category; most of them are located in outer London and in south of the river Thames in inner London. The third cluster suggests protective spatial effects on nearly all condition. The cluster encompasses 26 per cent of all London areas, which are widely distributed with two larger concentrations in north and west London. The fourth cluster shows the opposite pattern of the third; in this group, local factors account for an increase in health risk across diseases, with particular effects on infections, neuro-psychiatric and endocrine conditions. The respective areas constitute 19 per cent of all and can be found in more affluent, central areas and suburbs; they are less common in an intermediate ring around Inner London. The fifth cluster shows the most distinctive patterns of locally specific health factors. Infections and neuro-psychiatric and endocrine conditions are less common than would be expected, while the risk of diabetes, sense-organ diseases and malignant neoplasms is increased. Only seven per cent of areas fall into this category; they can be found in Lambeth, Wandsworth and some selected suburban locations.

In summary, health space represent groups with similar traces of local place effects, which cannot be attributed to neighbourhood characteristics as collected by the Census and matched to the health milieus. The health spaces highlight those areas where other, locally specific factors are at work over and above variations at the level of individuals.

## 8.3 An advanced classification of urban vulnerability

While London is composed of distinct health environments, the preceding analysis suggests that they appear in locally specific guises. These guises are defined here by the degree to which health environments can be associated with the local prevalence of health milieus as inferred from Census neighbourhood characteristics. The combination of this information generates a picture of geographically varying health challenges, summarised in health environments, with locally specific levels of uncertainty. $\chi$-squared tests between health environment and health spaces confirm that they are statistically independent, with some weak exceptions observed for the Strong Disease Burden environment and the Special Challenges health space [not illustrated]. As a general rule, they represent different phenomena.

The combined geography displays the spatial distribution of health environments and health spaces, which may be interpreted as local variants of health environments [Figure 8.3]. Local variants are shaded by their order or degree of specificity. The first health environment, Poor Health Capital, exhibits weak specificity throughout London, particularly in the southern tip of the northern concentration and the exclaves in the southern suburbs. The southern concentration shows a network of protective health spaces, which

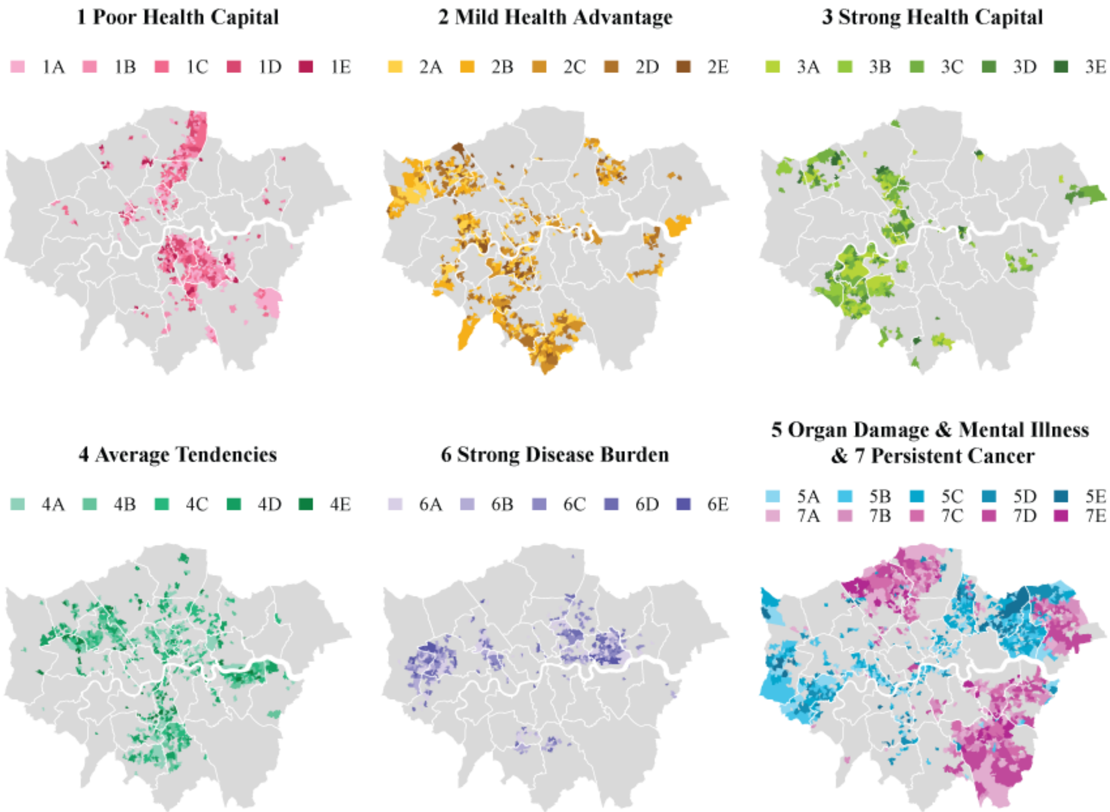



Figure 8.3: Health environments and their local variants.

are interspersed with their detrimental counterparts. In south London, protective and detrimental variants appear to co-exist in a more granular pattern. Strong specialised needs exist in single and more isolated locations within the geography of the cluster, in Lewisham, Bromley and Brent, as well as throughout Lambeth.

In the Mild Health Advantage environment, weak specificity can be found in the western concentrations. Specific cancer-related challenges are more common in this environment than in others and remain co-located to areas with weaker specificity. Protective spaces prevail in the southern and more central parts of the environment, notably in the City of London and the boroughs of Croydon and Sutton. These are co-located with detrimental health spaces in the south. Spaces with special health challenges occur in central, western and northwestern parts of the environment: Wandsworth, Hammersmith and Fulham and Harrow.

Within the Strong Health Capital environment, local variation is weaker in southwestern areas, Richmond and Merton. Cancer-related challenges and protective variants of the environment often cluster together, notably in Richmond and Kingston as well as in the more isolated parts in eastern and southeastern London. Detrimental health spaces and special local challenges occur in the more central parts of the environment, such as in Westminster and Canary Wharf, as well as in the suburbs, in Havering and Harrow.

Most areas in the Average Tendencies environment exhibit lower degrees of local specificity. Yet, some subtle patterns can be identified. Almost all parts of the environment that fall into the borough of Waltham Forest belong to the protective local variant and so do many of the central-northern areas. The number of local variants with more specific challenges increases as one proceeds further away from the centre in all directions. As one approaches Croydon, however, the trend reverses again.

A complex geography of health spaces can be observed in the environment of Organ Damage & Mental Illness. While the sparse central parts of the environment tend to exhibit lower specificity, with a higher frequency of the cancer-related variant, the outer parts reveal concentrations of specialised needs. These concentrations appear to have a centre from which specificity decreases. The eastern part of the environment that extends across Redbridge and Havering shows a large concentration of spaces with special, unexplained challenges (diabetes, sense-organ diseases and malignant neoplasms) and detrimental effects. The environment comprises a similar yet smaller cluster in Hillingdon, and, in Hounslow, it possess a concentration of detrimental health spaces.

Two thirds of areas in the Strong Disease Burden environment show either the lowest local specificity or belong to the protective variant. Hence, while the environment shows a stronger burden of disease in general, it has a high number of areas in which the burden is lower than expected given the information of local lifestyle milieus. The variant with special challenges exist in areas close to Hillingdon in the west and in the borough of Barking and Dagenham in the east. In those parts of London, the strong disease burden seems further increased due to local conditions.

In the Persistent Cancer environment, special local variants are more common than spaces which lie within the expectations given the characteristics of local lifestyle milieus. Still, many parts of the environment show weak local specificity, since it is the most common variant in London. But all three large, geographically separate subparts of the environment display larger concentrations of detrimental health spaces, sometimes located next to protective health spaces. Highly special challenges primarily exist in the boroughs of Barnet and Greenwich, which suggests that persistent cancer may often result from unobserved characteristics.

## 8.4 A speculative urban policy programme for health

The joint view of health milieus and health environments provides the ground for the design of policy interventions [Tables 8.4 to 8.6]. The following suggestions for strategic policy responses are based on the evidence that has emerged so far; they are speculative, meant to be illustrative and neither exhaustive nor definitive. The policy programme takes each of the ten health milieus as target group and it is structured as follows.

First, the most salient characteristics of each milieu's health profiles are summarised and subsequently the social pathways (material, behavioural, pyschosocial) are briefly characterised. In a next step, the most frequent health environments of each milieu are stated. Together, the health profile, pathways and health environments serve to identify the general policy priority for each milieu, which is summarised as a brief key word and more detailed policy goals.

Next, interventions are deduced from the strategic policy priority. A distinction is made between generic interventions and area-baed interventions. The former are suggestions for policy responses that address health challenges independent from where the milieus are located in London, focussing on the systemic, social "root causes" of health inequalities [Colburn & Jepson 2012; Scambler 2012]. Area-based interventions comprise suggested responses that are targeted to the areas where members of the milieus live, accounting for interactions of the milieu with the different health environments. Both types of policy interventions are listed in the order of their importance given the evidence of the pathways; yet, they should be regarded as examples rather than an exhaustive list.

Finally, a brief qualitative assessment is made of the potential impact the interventions would have on the organisation of health care. This is broken into short-term impact and long-term impact. In addition, an assessment of the current milieu's impact without policy implementation is made in order to provide a sense of priority for action across the milieus.

The approach to policy development is deductive: starting from the health profiles and context of milieus, strategic priorities are defined and detailed objectives formulated. In so doing, Bourdieu's theory of social practice and research on the sociology of health and illness are drawn on for inspiration [Baum & Fisher 2014; Blaxter 1990, 2003; Nettleton & Green 2014; Veenstra & Burnett 2014; Wilkinson & Pickett 2010; Williams 1995].

The health profiles and the likely activated pathways within the *Enduring Isolation* milieu [Table 8.4] suggest that they are trapped in a vicious circle of health disadvantage through interaction of psycho-social and material pathways. Hence the priority of policy responses is to address all capitals – economic, cultural and social – in order to improve access to health-supporting assets. Action on both the systemic and local levels is required, which should include better access to income based on employment creation as well as stronger and fairer redistribution and access to education, training and skill development. The required actions are therefore not in the remit of public health, but in the remit of economic and fiscal policy. At the local level, efforts to connect the milieu to society need to be made primarily by establishing access to social capital. If this is successful, the group may be enabled to add more extra-domestic activities to their lifestyles, which may lead to a more active lifestyle. Actions to encourage the use of community services and engage in social interaction need to respond to their everyday life logic, which is compounded by their psycho-social situation, persistent experience of exclusion, quality of life and time budgets.

Table 8.4: An urban policy programme viewed by milieus (one to three).

| Enduring Isolation (1) | Unconcerned Starters (2) | Retiring Generation (3) |
|---|---|---|
| **Health profile**<br>* poor mental and physical health<br>* risk of disability<br>* lower level of well-being<br>* unhealthy practices | **Health profile**<br>* poor mental and physical health<br>* risk of disability<br>* lower level of well-being<br>* poorer nutrition and lower levels of physical activity | **Health profile**<br>* weaker physical health in addition to impact of age<br>* low physical activity<br>* better subjective health |
| **Pathways**<br>* psycho-social: experienced marginalisation and social exclusion<br>* material: barriers to health assets | **Pathways**<br>* behavioural: low levels of concern and knowledge about health<br>* psycho-social: experience of being less successful than others | **Pathways**<br>* material: outcome of physical labour<br>* behavioural: pragmatic attitudes from material situation |
| **Health environments**<br>* Strong Disease Burden (6)<br>* Organ Damage & Mental Illness (5)<br>* Poor Health Capital (1) | **Health environments**<br>* Poor Health Capital (1)<br>* Strong Disease Burden (6) | **Health environments**<br>* Strong Disease Burden (6)<br>* Organ Damage & Mental Illness (5) |
| **Health policy priority: connect**<br>* facilitate more active lifestyle<br>* reduce social exclusion<br>* support development of economic and cultural capital<br>* support access to social capital | **Health policy priority: guide & empower**<br>* support interest in health<br>* support interest in social and political affairs<br>* increase cultural capital<br>* support access to social capital | **Health policy priority: facilitate**<br>* maintain economic capital<br>* maintain health and improve where possible<br>* increase mild physical activity |

The overall disadvantaged health profile of *Unconcerned Starters* is likely to be a result of behavioural and psycho-social pathways, which are in turn reinforced by health disadvantage. Yet the milieu offer starting points for policy that may be used to improve their health outlook. They are younger and often still in education; they may be amenable to active, preventive health-enhancing strategies. Their cultural participation reveals creative tendencies which may be leveraged to support interest in health and also social and public affairs. The latter would support inclusion in society, which can support empowerment and actions to take more conscious control of their health at a later stage. Although their health care utilisation is already higher than would be expected based on their demographic structure and income, the focus should not be on improving the quality of health care. A focus on the social sphere, guidance in taking charge and encouragement of their reflective practices are likely to respond better to their situation. Community services and social work would be well placed to offer support here; less so the health care service, including general practitioners.

Table 8.4 (continued)

| Enduring Isolation (1) | Unconcerned Starters (2) | Retiring Generation (3) |
|---|---|---|
| **Generic interventions**<br>* fiscal and economic policy to improve access to income, transfer payments, housing<br>* implementation of London Living Wage<br>* support of personalised education and training<br>* improve child care and social work | **Generic interventions**<br>* support creative potential<br>* support reflection on healthy and unhealthy practices | **Generic interventions**<br>* maintain concessions to transport and cultural activities<br>* maintain access to specialised health and social care throughout London |
| **Area-based interventions**<br>* create local employment opportunities<br>* improve housing conditions<br>* build and improve community spaces and services (centres, affordable leisure) to facilitate more active lifestyle<br>* strengthen health care service according to characteristics of health environment (specialised services)<br>* improve environmental conditions where health space suggests this | **Area-based interventions**<br>* build and improve community spaces and services (centres, affordable leisure) to facilitate more active lifestyle<br>* carefully designed social marketing on health issues<br>* improve environmental conditions where health space suggests this<br>* access to personalised education and training | **Area-based interventions**<br>* "Lifetime Neighbourhood" policies<br>* improve walking conditions<br>* barrier free residential environment<br>* strengthen health care service and facilitate independent living according to characteristics of health environment (specialised services) |
| **potential impact on health services: high**<br>* short-term: high<br>* long-term: high<br>* current impact: high | **potential impact on health services: medium**<br>* short-term: low<br>* long-term: high<br>* current impact: medium | **potential impact on health services: low**<br>* short-term: low<br>* long-term: low<br>* current impact: high |

The health profile of *Retiring Generation* has to be viewed in the context of their life stage; it is a result of material and behavioural pathways of past economic activity which mainly took place in manual jobs before or during shifts in the occupational structure. Their health and well-being is at a stable level, but it may quickly be eroded if material conditions change to the worse. Efforts to allow this group to maintain their lifestyle, be as mobile as possible and participate in cultural and social activities are crucial factors to their well-being. Given their age and the characteristics of their health environments, access to health services is important. Area-based interventions may further enable this group to live independently and perhaps mildly increase their physical activity. The "Lifetime Neighbourhoods" programme proposed by the Department for Communities

Table 8.5: An urban policy programme viewed by milieus (four to six).

| **Locally Anchored (4)** | **Established Cultural Consumers (5)** | **Rising Extroverts (6)** |
|---|---|---|
| **Health profile**<br>* very good mental health and subjective well-being<br>* medium prevalence of physical inactivity | **Health profile**<br>* healthy and active lifestyle<br>* very good physical health<br>* good mental health<br>* high levels of well-being<br>* alcohol consumption | **Health profile**<br>* good physical health<br>* high level of physical activity (exercising)<br>* poorer mental health compared to their income group |
| **Pathways**<br>* psycho-social: strong local, social integration<br>* behavioural: pragmatic attitude and convenience in suburban context | **Pathways**<br>* material: high economic capital provides access to health assets<br>* psycho-social: experience of success relative to others<br>* behavioural: conscious choices | **Pathways**<br>* material: high economic capital provides access to health assets<br>* psycho-social: experience of success relative to others<br>* behavioural: intrinsic value fitness |
| **Health environments**<br>* Persistent Cancer (7)<br>* all other except Strong Health Capital (3) | **Health environments**<br>* Strong Health Capital (3)<br>* Mild Health Advantage (2)<br>* Persistent Cancer (7) | **Health environments**<br>* Strong Health Capital (3)<br>* Mild Health Advantage (2) |
| **Health policy priority: sustain**<br>* protect economic and social capitals<br>* encourage more physical activity and better nutrition | **Health policy priority: sustain**<br>* help sustain healthy lifestyles<br>* reduce alcohol consumption | **Health policy priority: sustain & encourage**<br>* maintain physical health<br>* improve mental health<br>* support interest in health beyond fitness |

and Local Government [Bevan & Croucher 2011] offers relevant policies here, which include investment in public space with an emphasis on barrier free access.

The *Locally Anchored* [Table 8.5] enjoy health advantage due to psycho-social pathways that rest on their strong local, social integration. This advantage is mainly conditional on their economic and social capital, which need to be protected. Their orientation reveals a more pragmatic attitude that puts local concerns first and may also favour convenience in some domains of life, leaving room for improving their diet and physical activity. Investments into the local physical environment as well as easier access to healthy food coupled with some social marketing may be an effective means to appeal to their more pragmatic orientation. Given that one of their common health environments is Persistent Cancer, health screening in those areas will be important, as many of this milieu are at a high risk age of cancer onset.

The health profile of *Established Cultural Consumers* expresses health advantage, which may be explained by their favourable material and psycho-social situation, coupled

Table 8.5 (continued)

| **Locally Anchored (4)** | **Established Cultural Consumers (5)** | **Rising Extroverts (6)** |
|---|---|---|
| **Generic interventions**<br>* ensure that they maintain financial conditions<br>* maintain access to affordable housing<br>———<br>**Area-based interventions**<br>* improve cycling and walking infrastructure<br>* work with local retailers to improve access to healthy food<br>* social marketing that appeals to orientation towards convenience<br>* increase health screening where they live in Persistent Cancer environments and spaces with special challenges<br>———<br>**potential impact on health services: low**<br>* short-term: low<br>* long-term: medium<br>* current impact: low | **Generic interventions**<br>* London-wide efforts to make non-alcoholic drinks more popular and accessible<br>———<br>**Area-based interventions**<br>* social marketing to inform about effects of alcohol consumption, appealing to utilitaristic attitude<br>* increase general practitioners' awareness on alcohol consumption<br>* specialised health care, screening and monitoring where they live in Persistent cancer environment<br>———<br>**potential impact on health services: low**<br>* short-term: low<br>* long-term: medium<br>* current impact: low | **Generic interventions**<br>* London-wide employer programmes to improve mental health<br>* access to healthy food at workplaces<br>———<br>**Area-based interventions**<br>* increased mental health screening by general practitioners<br>* work with local retailers to improve access to healthy food<br>* information at fitness centres about healthy nutrition and stress management<br>———<br>**potential impact on health services: low**<br>* short-term: low<br>* long-term: low<br>* current impact: low |

with conscious choices about healthy practices. Despite this, they show a high level of alcohol consumption, which may pose some risk to their health. Generic interventions to address this include city-wide programmes to offer non-alcoholic alternatives at alcohol outlets, including sites of cultural activities, such as theatres, museums, galleries and in gastronomy near workplaces. Area-based campaigns to remind them of healthier levels of alcohol consumption may be effective.

Although the social pathways work in the favour of *Rising Extroverts*, their health profile provides reasons for some concern. Despite healthier lifestyles, their physical and mental health outcomes are not as favourable as would be expected given their demographics and income. This may be connected to their ambitions and focus on their careers, which may compromise on other aspects in life. Some research suggests that employers can take a role in improving mental health in particular [for example, see Layard 2013], and a public health grounded London-wide approach may be effective in so doing. Some additional initiative may be expedient to highlight the importance of health independent of fitness. The milieu members are younger, and their interest in fitness may fade as they age. Here it is important to encourage them to continue keeping a focus on health rather than fitness. Stronger investment in mental health screening and consultancy in areas where they are likely to live or work (based on the types of jobs).

Table 8.6: An urban policy programme viewed by milieus (seven to ten).

| **Commited Citizens (7)** | **Laid-back Detachment (8)** | **Digital Age Autonomy (9)** | **Individualistic Independence (10)** |
|---|---|---|---|
| **Health profile**<br>* healthy lifestyle with mildly increased alcohol consumption<br>* good physical health<br>* good mental health<br>* high levels of well-being | **Health profile**<br>* low levels of physical activity<br>* medium to lower levels of subjective well-being | **Health profile**<br>* poorer nutrition<br>* medium to lower levels of subjective well-being | **Health profile**<br>* healthy lifestyle: good nutrition and physical activity<br>* better physical health<br>* good mental health<br>* high levels of well-being |
| **Pathways**<br>* behavioural: conscious choices | **Pathways**<br>* behavioural: low levels of concern about health<br>* psycho-social: some experience of exclusion | **Pathways**<br>* material: constrained access to health assets<br>* behavioural: focus on electronic media<br>* psycho-social: sense of autonomy | **Pathways**<br>* material: high economic capital provides access to health assets<br>* behavioural: conscious choices |
| **Health environments**<br>* Strong Health Capital (3)<br>* Mild Health Advantage (2)<br>* Persistent Cancer (7) | **Health environments**<br>* Strong Disease Burden (6) | **Health environments**<br>* Poor Health Capital (1)<br>* Strong Disease Burden (6)<br>* Organ Damage & Mental Illness (5) | **Health environments**<br>* Persistent Cancer (7)<br>* Strong Health Capital (3)<br>* Mild Health Advantage (2) |
| **Health policy priority: sustain**<br>* help sustain healthy lifestyles<br>* prevent increase of alcohol consumption | **Health policy priority: guide & enable**<br>* increase physical activity<br>* increase interest in health<br>* increase access to social capital | **Health policy priority: guide**<br>* increase economic and cultural capital<br>* increase interest in health<br>* improve nutrition | **Health policy priority: sustain**<br>* help sustain healthy lifestyles<br>* monitor risk of cancer |

*Committed Citizens* [Table 8.6] lead healthy lives; their health is better than would be expected based on their income. Their lifestyle results from conscious choices and hence utility-centred approaches may be effective in addressing an aspect of potentially minor concern: higher alcohol consumption. Their health environments suggests that cancer may be a threat, and, given their ages, health screening would be an effective way to identify early need for action.

The *Laid-back Detachment* milieu signal some health disadvantage due to an interaction of behavioural and psycho-social pathways. Their single most common health environment is Strong Disease Burden, and although this is not reflected in the direct health profile of the milieu, it is indicative of potential future disadvantage. This may be ag-

Table 8.6 (continued)

| **Commited Citizens (7)** | **Laid-back Detachment (8)** | **Digital Age Autonomy (9)** | **Individualistic Independence (10)** |
|---|---|---|---|
| **Generic interventions**<br>* London-wide efforts to make non-alcoholic drinks more popular and accessible | **Generic interventions**<br>* protect economic capitals | **Generic interventions**<br>* access to secure income either through employment or transfer payments<br>* health awareness initiatives using social media as channel, focussing on nutrition | **Generic interventions**<br>* enable choice, self-management and monitoring |
| **Area-based interventions**<br>* increase general practitioners' awareness of alcohol consumption<br>* specialised health care, screening and monitoring where they live in Persistent Cancer environment | **Area-based interventions**<br>* build and improve community spaces and services (centres, affordable leisure) to facilitate more active lifestyle<br>* strengthen health care service according to characteristics of health environment (specialised services)<br>* improve environmental conditions to facilitate and encourage physical activity | **Area-based interventions**<br>* access to personalised education and training<br>* nutrition advice by general practitioners | **Area-based interventions**<br>* specialised health care, screening and monitoring where they live in Persistent Cancer environment |
| **potential impact on health services: low**<br>* short-term: low<br>* long-term: medium<br>* current impact: low | **potential impact on health services: low**<br>* short-term: low<br>* long-term: medium<br>* current impact: medium | **potential impact on health services: medium**<br>* short-term: low<br>* long-term: high<br>* current impact: low | **potential impact on health services: low**<br>* short-term: low<br>* long-term: medium<br>* current impact: low |

gravated, if the group lose their thin economic base, and hence an important priority is to ensure that they can retain their economic capital such that material pathways are not triggered in the future. Connection to the community through facilitating access to relevant services may be a way to nurture a more conscious approach to health and ease the detachment to society. Awareness campaigns are not likely to be effective to achieve this. Improvements of local walking conditions may further contribute to easier incorporation of physical activity into their lives.

The *Digital Age Autonomy* also exhibit a disadvantaged health profile that may contribute to an early depletion of their biological capital. Their material circumstances limit access to health-enhancing assets; yet, unlike other disadvantaged milieus, there are elements of behavioural and psycho-social pathways that may work in their favour. While on the one hand their interest in digital media may disincline them to engage in other activities that potentially entail health and social benefits, they experience a high level of autonomy through actively engaging in social interactions online and mastering the technology. This may be a transferable skill to the area of health [Nettleton et al. 2005], but also, social media may offer a potential channel for soft intervention. Conventional area-based campaigns may be ineffective. Self-management and monitoring of health may also be encouraged by building on their experience of autonomy.

The health profile of *Individualistic Independence* reflects advantage based on material pathways and conscious choices as part of their lifestyles. Yet their age structure and location in London suggests some imminent or future risk of cancer. Corresponding health screening seems particularly important for parts of the milieu that resides in Persistent Cancer environment. Since their orientations suggests that they take conscious choices in most areas of life, they can also be expected to move to areas where they can satisfy their health-oriented needs. The inferred geography of this milieu suggests that these preferred locations are primarily suburban. Given their probable appreciation of independence, this group seems particularly apt to take control of their own health actively, opening up possibilities for self-management and monitoring.

## 8.5 Synthesis: the spatial character of health and vulnerability

The integrated spatial view of health in London offers some starting points for the future directions of health-relevant urban policy in London. Interpreting the phenomena in terms of the theory of social practice and the social determinants approach to health compels policy to take a multisectoral approach that transgresses boundaries of public health, ideally within a framework of integrated spatial planning and coordination. Based on this premise viewed in context of the preceding analysis, some general strategic responses for health environments and their health spaces may be formulated.

The Poor Health Capital environment requires some action that address common determinants across all types of conditions and subjective health. These are likely to be found within the remit of social root causes of health. The emphasis should not be on public health but on other policy areas, such as fiscal and economic policy, which in cities can be closely linked to policies of housing, land use and transport. The higher prevalence of vulnerable milieus suggests a need to address material and psycho-social pathways by helping them build stronger economic, cultural and social capitals (generic interventions), in particular in health spaces with less specificity. More local, environmental factors (area-based interventions), including the operation of health services, may be addressed in areas with higher specificity.

The Mild Health Advantage and Strong Health Capital environments represent areas of lower risk and therefore may not require strong interventions across social root causes. The focus should lie on addressing individual health concerns, and here public health and health care delivery may have a greater role. Alcohol consumption may be a challenge in these areas, but it may be dealt with within a public health approach. Some health

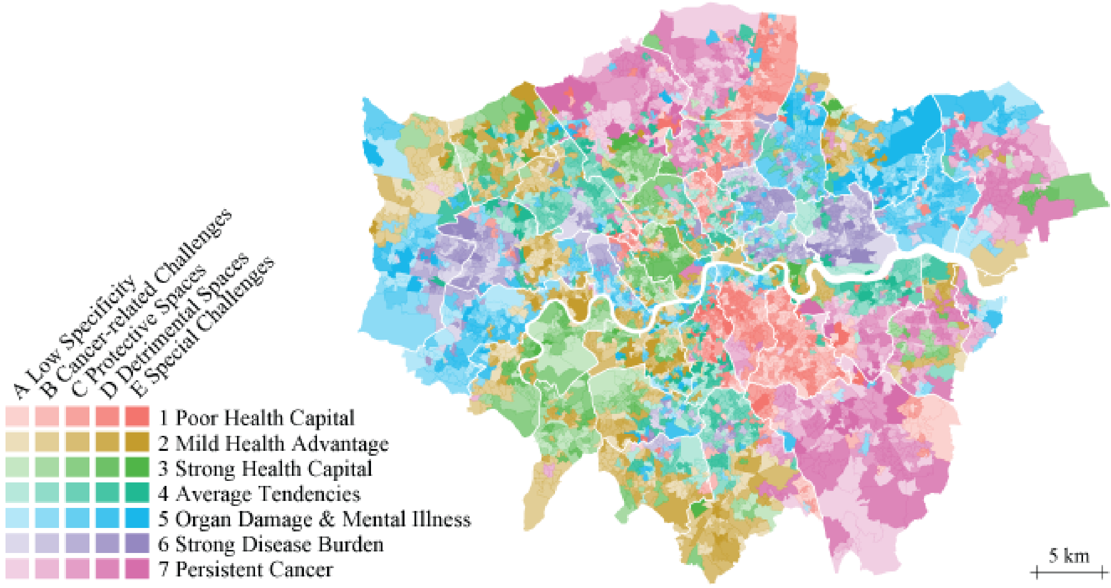



Figure 8.4: An integrated spatial view of health and vulnerability in London.

spaces – Detrimental Spaces and Special Challenges – may require closer attention in terms of the local operation of health care.

Similar conclusions can be drawn for the Average Tendencies environment. In terms of milieus, the population is the most heterogeneous; hence a view on health spaces may be a useful step to structure the agenda for this type of areas. Lessons may be learnt from Protective Health Spaces by taking a closer look at population and health care characteristics.

The environment of Organ Damage & Mental Illness requires a wider view that encompasses social policy and public health in equal measure. The level of risk justifies immediate response of the health service, in particular in health spaces of type D and E. In health spaces of type A and B, stronger action across social determinants is required. It may also be explored if neighbourhood safety could be a cause that affects the physical and mental well-being of the local population. This would be an action that applies to all health spaces, and again the Protective variant could be viewed as a learning example in the context of this health environment.

The Strong Disease Burden Environment requires a stronger focus on social determinants, as the risks of a great number of conditions including that of poor self-rated health are elevated. The prevalence of vulnerable milieus suggest similar actions to the one for Poor Health Capital, although the challenges are more differentiated. A hypothesis that may be tested by a focussed epidemiological investigation is that the risk of cancer is reduced because of the ethnic composition of these areas. The *Laid-back detachment* has a high share of Asians and resides more often in this health environment than in others. The effects of different nutrition may be observable here.

The Persistent Cancer requires a stronger response from the public health sector. Health screening and monitoring are important instruments to tackle the strong risk of cancer-related conditions. Given the high level of good self-rated health, the burden in this environment does not seem to be related to lack of access to health-enhancing assets; social root causes may be less relevant here.

In taking an explicitly multisectoral and spatial approach, the alternative urban policy programme proposed here differs from recent suggested programmes to tackle health and health inequalities in London, notably the London Health Commission's 2014 report [LHC 2014] and the mayor's Health Inequalities Strategy [GLA 2010]. This alternative programme provides some additional, strategic ideas.

First, some areas are more similar than others as regards their health challenges. The geography of health environments suggests that health challenges do not necessarily follow the well-known geography of deprivation in London but reveal some nuances with special challenges such as cancer, which appears to possess a suburban component affecting more affluent and ethnically less diverse areas.

Second, health environments are heterogeneous in terms of the causes of their health challenges. The distinction of five health spaces suggest different scales of pathways or processes that may be at work in shaping health. While it can be supposed that there is a common systemic process that generates different health environments, some health spaces produce specific phenomena in interaction with local conditions and hence require more specialised, area-based interventions in addition to generic interventions. In this abstract sense, health spaces of the same kind resemble each other more; they may require similar scales of the interventions. But what the interventions are in substantive terms should be defined by the nature of their health environment of which health milieus are part.

Third, the alternative spatial view of health also brings to bear that knowledge about local conditions and population varies geographically. Health spaces with strong local specificity need to be viewed with greater caution as regards appropriate interventions. The evidence arising from predictive models suggest that, in those spaces, vulnerability cannot be attributed very much to lifestyles. Yet the reverse is not true, that is that vulnerability found in low specificity spaces can necessarily be explained in terms of local lifestyles. Adding more data or more observations may well change the degree of uncertainty that the health spaces represent. Hence, causal interpretations with respect to lifestyles and health spaces remain necessarily tentative.

Fourth, the integrated spatial view also highlights the need for collaborations that transcend administrative geographies. Some health environments appear to be multi-borough phenomena, others connect subparts of boroughs and efforts to address the challenges arising in these health environments may require greater cross-boundary collaboration according to their emergent geography. Especially in health environments with more differentiated risks, Clinical Commissioning Groups or Local Health and Wellbeing Boards need to be sensitive to these geographies and allow the view that certain phenomena may be addressed in a more collaborative fashion. The London Plan can serve as a strategic basis to do so; currently the idea of cross-borough collaboration to address health is not prominent in London Plan, Health Inequalities Strategy or the London Health Commission's report.

In summary, the advanced geodemographic approach combines multiple perspectives on health in an integrated spatial view. It reflects the phenomenon of geographically varying health challenges, their contextualisation and associated uncertainty. It is designed to attend to the diversity of populations, more specifically their social milieus, and support multisectoral approaches that are founded on heterogeneity rather than homogeneity. It offers ground for hypothesis formulation regarding global and local pathways that may be at work to generate differential outcomes in both social and geographic space. And consequently, it may support the development of policy responses that recognise the diversity of lifestyles and their inevitable manifestations in diverse health outcomes. These outcomes cannot be addressed by attempting to prescribe, to converge or equalise the milieus – this is neither desirable nor possible – but to provide specific support to population groups in order to manage differential risks and mediate access to health-enhancing assets in the context of milieu-specific as well as spatial opportunities and constraints.

# 9 Conclusions: geodemographics, social science and urban policy

The advanced geodemographic framework developed here reflects the objective of the thesis to advance a multi-level description that aids interpretations of the social system with respect to vulnerability and its differential manifestations in health. It contributes to geodemographic research and policy development in a number of ways and offers opportunities to address current limitations therein. Yet the framework itself holds limitations that require debate and suggestions as to how they might be overcome.

## 9.1 Summary of the research

Building on the strength of geodemographics to integrate different data sources and describe systems ecologically in one single perspective, the advanced framework develops different layers of information in order to characterise populations in terms of their vulnerability at multiple levels. The classifications developed in this way presents vulnerability and resulting inequalities in health in the context of heterogeneously activated pathways as well as regional and local specificity and provides starting points for context-sensitive policy approaches.

I choose the target concept of vulnerability, as it cuts across different pathways, touches on a multitude of policy arenas and scales to any ecological level. Applied to health, vulnerability represents the chance of being healthy in potentially harmful events or circumstances. Drawing on different views on health, the causes of vulnerability are grouped into access to health-relevant assets and exposure to risk. These main components predispose individuals and populations towards a limited range of possible outcomes through embodiment of social relations, of which health is one expression.

In the thesis, vulnerability is first explored in terms of regional specificity with a focus on population structure. The strategy is to combine a sample of DNA data with a population-wide dataset that include correlating ancillary information. A regionalisation of rural Great Britain based on the joint information in both datasets is presented, suggesting geographies of broad regional specificities in terms of population structure. The emerging regions provide a generalised biological and cultural foundation that may reflect place specificities in the relationship between socio-economic position and health.

Based on these findings, an urban investigation is pursued that considers the degree to which UK cities are shaped by their regional populations and exhibit similarities and differences. A series of spatial statistical heuristics suggests that UK cities are primarily composed of regional populations that reflect a broad division of the four UK countries with England further divided into north and south. Yet as soon as the heuristics increase in granularity, the differentiation increases, too, whereby regional patterning recedes and sub-city parts show a distinctive urban population, which is yet common to many cities, and presumably reflect mixed neighbourhoods with a high proportion of international immigrants. This applies in particular to London and Birmingham, with the former revealing unique forms of population diversity.

The characterisation of regions and UK cities then leads to a step of investigating geographical disparities in health drawing on the asset and the deficit view and corresponding indicators derived from hospital admission data and the UK Census. A regionalisation of England based on age and sex-standardised health outcomes is explored, which reveals regions with distinct health challenges, broadly reflecting general social causes of ill-health across a range of health outcomes, general health advantage and a few more specialised challenges that result from particular disease burden and the absence or presence of cancer. In geographical terms, different dynamics can be discerned between north and south and urban and rural England. Here, the innovation lies in the application of spatial structural models that incorporate explicit geographical context into health estimates. This results in more contiguous regions of differing health challenges, which reveal a weak yet significant correspondence with a regionalisation of England based on surnames as surrogates of population structure and culture. Yet, since other regional characteristics such as unemployment or demographics are not included, any reasoning on causality remains highly speculative at this level of granularity in terms of both geography and health categories.

Akin to the urban investigation of population structure, the classification of areas by health outcomes is focussed on ten metropolitan regions. By virtue of this comparison, it becomes clear that many cities in the UK share common health challenges, although the extent to which health inequalities are manifest geographically differ with respect to assets and deficits. From the comparison between regional and urban health inequalities, one may conclude that a certain variation of health disparities is to be expected for the whole population, but in some cities these are attenuated or exacerbated, which should be attended to in policy. A detailed small-area classification for London is then proposed at a higher geographic resolution, and it is found that some of the challenges identified in England's cities appear in a specific guise of advantage and disadvantage in London. The notion of health environments is introduced in order to represent neighbourhood-level singularities of connected urban outcomes.

Subsequently, a perspective of agency is added to the spatial structural view on vulnerability. Focusing again on the entire UK, data taken from the Understanding Society survey reveal different groupings in society that can be characterised by their activity patterns, subjective orientations, attitudes and everyday life routines. The resulting groups, called health milieus, are found to differ significantly in their health profiles as well as in terms of socio-demographics and economics. With the regionalisation based on surnames as contextual geographies, I explore regional differences between milieus and find some specific patternings of the milieus, which may reflect distinct cultural tendencies among various parts in the UK and urban England. Given the potential complexity resulting from different regional milieus, the UK-wide classification is taken as a reference classification from which regional divergences are assessed through comparison.

The geographical distribution of health milieus is modelled probabilistically for London using a technique that belongs to the class of spatial microsimulation. The matching of Census data and the socio-demographic characteristics from the survey produce distinct geographies for each milieu. This leads to a contiguous geodemographic representation reflecting individual level rather than aggregate level phenomena in conceptual terms.

In a final step, I draw all these different layers into a geodemographic framework that links the information ecologically. London is again used as a test case, which reveals significant spatial correspondence between inferred milieu prevalence and different health environments. The ecological correspondence encourages the development of spatial-structural predictive models to regress health on milieu prevalences. Yet, the focus is not on the contribution of area covariates to health but on different types of residuals and associated levels of uncertainty. The resulting spatial structure of residuals generates different spaces where health outcomes cannot be attributed strongly to local lifestyles but presumably other spatially clustered unmeasured factors. Another geodemographic layer that incorporates both health challenges and their uncertainty in attributing potential pathways to their manifestation is suggested.

Based on the information that is part of this advanced geodemographic framework, I identify cornerstones of an urban policy programme to address population vulnerabilities. While generic recommendations can be made for different health environments in London, the milieus and their locations serve as a basis to formulate generic policy interventions and area-based interventions, which focus on social root causes at a systemic as well as a spatially explicit local level respectively. By drawing on the different layers of geodemographics that have been developed previously, the policy programme provides a multisectoral and context-sensitive alternative to monosectoral and aspatial policy approaches.

## 9.2 The hermeneutic and political value of the framework

The result of this research programme is an interpretive geodemographic framework that is conceptually focused on health as expression of vulnerability, draws on social and clinical datasets, comprises multiple ecological levels, integrates spatial context, incorporates systemic and specific pathways through comparison and, in parts, accounts for uncertainty. In so doing, the framework does not develop a single classification but provides multiple ones appropriate to each ecological level from broad-scale regionalisation (population structure and regional specificity), neighbourhoods (health environments and health spaces) to the individual level (health milieus), each of which contribute to the hermeneutic and political value of geodemographics.

## Geodemographics as social science hermeneutic

A main characteristic resulting from this multi-level approach is that a conceptual distinction between the individual level and the area level is maintained. Whereas many conventional geodemographic classification collapse the two by inferring lifestyles from aggregate characteristics, the approach taken here reverses the logic by using survey data to confer meaning to aggregate neighbourhood statistics, whereby the health milieus provide evidence for the lifestyle implications of area-level variables. The focus of the variable selection lies therefore in the definition of the social milieus within the sample, while area-level variables serve as evidence of the presence of a milieu given aggregate socio-demographic and economic characteristics. As a consequence, uncertainty associated with the milieu classification can be managed within a statistical framework in terms of cluster distances in statistical space as well as the zonal estimates of milieu prevalence themselves, the probability that a milieu resides in a given zone in view of zonal evidence. In other geodemographic systems, the uncertainty of attributing lifestyle to a neighbourhood class remains with the researcher making a substantive judgement; it is impossible to estimate.

The second layer of the framework corresponds to neighbourhoods and health environments, classifies areas and hence resembles conventional geodemographics. Here, the interest is explicitly that of grouping areas by aggregation of health records and self-rated health. Although health is still conceptualised in terms of people, a single health record or health rating only derives its meaning when viewed along with other records. Only then can it be known whether an incidence should be judged as exception or as expression of social disadvantage. Here, the ecological level is needed in order to make sense of individual observations. Furthermore, the spatial-structural estimation of health turns an ecological framework into an explicitly spatial framework with a stronger role of geographical specificity. The precise implication is that the resulting classification is spatially more contiguous and statistically robust than classifications that use unsmoothed rates.

The conceptual distinction between individual-level social milieus and area-level health environments is needed analytically to identify different types of causal pathways that operate at different ecological levels. The milieus reveal generic, systemic workings of society in shaping vulnerability and health. The health environments provide evidence for the geographical manifestation of those workings and, when viewed in combination with the geographical distribution of social milieus, they highlight where pathways are likely to be activated and where not. In this way, the framework accounts for the locality and contingency of causal pathways.

Reviewing health milieus in their regional context and classifying neighbourhoods across cities is one way of addressing the need to compare urban outcomes [Robinson 2014] and disentangle their plural global, local and specific, contextual causalities. The multiple classes the framework generates presents a taxonomy of possible urban outcomes, or in Byrne's terms "attractor spaces" within the limits of a system [Byrne 2002, 101].

Although there is a danger of reinforcing extant categorisations of social phenomena notably through variable selection and interpretation, classification can highlight new, unpredicted outcomes or "attractors" in the system, which, in principle, do not depend on a preconception on the part of the researcher. The classes found in the health milieus and health environments require description and interpretation; their associated geographies are wholly emergent, since explicit geographic information is not part of the input into the taxonomy. As a consequence, unlike explanatory models, they force the researcher to engage with them as emergent phenomena and topologies, which may potentially be at odds with established conceptions of urban outcomes and encourage alternative thoughts about potential policy responses.

### Geodemographics for policy

The conceptual design is also instrumental in structuring the policy programme to address population vulnerability. Joining the evidence from the milieus, health environments and health spaces enable the development of policy interventions that cross sectoral and administrative boundaries and emphasise prevention rather than treatment by focussing on the social determinants of health. It also provides a counterfactual to policy responses that assume homogeneity rather than heterogeneity of populations. The milieus demonstrate that individuals of the same social class can show different tendencies towards engaging with health, indicating different opportunities and constraints for change.

The detection of class-internal heterogeneity challenges to some extent the appropriateness of blanket policy interventions, such as smoking bans or "sin taxes" [LHC 2014], which not only promise limited effectiveness (though perhaps fiscal revenues) but are also predicated on a coercive notion of policy, whereby unhealthy lifestyles are to be re-directed into healthier avenues in a top-down approach. Even if this addresses the health of population groups on the surface, it treats the symptom – health – rather than the cause – vulnerability and its specific constituents. Similarly, large-scale marketing campaigns to persuade people to walk more (as recommended by the London Health Commission [ibid.]) are only likely to effect change among a small number of milieus, which may not necessarily be those in need. Other questions, such as whether areas are walkable, time budgets allow incorporation of walking in daily life, or whether health is the main concern of groups that are faced with more imminent social stress or economic risks are likely to dominate the receptiveness for social marketing campaigns.

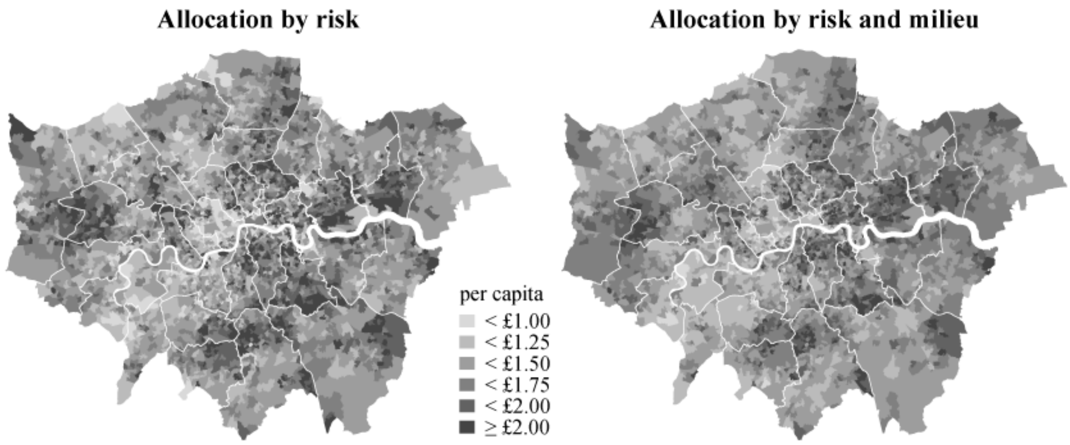
The evidence also suggests that the current policy emphasis on self-management and monitoring in health is appropriate only for a small subgroup of the population. So far, only the *Individualistic Independence* milieu suggests direct amenability to these measures. Equally affluent groups are likely to pursue other interests, whereas some other milieus do not possess the cultural capital to self-manage their health and probably would require inhumanely intense guidance to embody the logics of rational management. Some younger milieus such as *Digital Age Autonomy* may, however, be guided more easily, if such interventions can be connected to some self-managing experience they already possess, which in this case would be their autonomy within digital media. Although it is not possible to estimate directly the potential of adopting self-management in health care, the share of the population that would be able and willing to do this may be rather limited, in particular, if one considers the existence of multiple digital divides. Certainly, differentiated, context-sensitive approaches responding to milieu-specific orientations and logics are required to support the uptake of self-management in health.

A more critical stance should also be taken with respect to the often uttered argument that geodemographics can be used for resource allocation. Since small-area classifications do not quantify any order or intensity among classes, it is not obvious how geodemographic resource allocation can be specified. An exception may be the Index of Multiple Deprivation (IMD), which ranks rather than classifies areas. Yet, even in the IMD, the numerical relationship between areas is unclear, as the score reflects a ranking and therefore an ordinal scale, within which the distances of values are undefined. Moreover, the score of one area depends on the score of another, so in some sense a city's geography of the most deprived areas is sensitive to the level of deprivation in other cities in the UK. The London Plan defines areas that fall into the bottom quintile of the IMD as areas for regeneration without recognising that such a choice is sensitive to how well certain neighbourhoods in Liverpool or Middlesborough are doing [GLA 2011]. Here resource allocation becomes resource dedication based on an illusion of discrete area classes with sharp boundaries; this demonstrates that an insufficient debate about geodemographic classifications can misguide policy decisions with a potentially high level of impact.

In fact, the evidence from the milieus developed here confirms that, in terms of demographics, each milieu can occur in every LSOA in London to a certain minimum level of probability. Neighbourhood populations do not seem to fall into classes in the way clear-cut geodemographic categories suggest. This may not be an issue for consumer targetting, but it is important for area-based policy interventions with long-term goals. This finding emphasises again the need to triangulate neighbourhood classifications with other data, and with the emerging availability of high-resolution 'Big Data' that describe activity patterns in real-time, there may be increasing opportunities to do so in the future.

It can be said, then, that to date, there has not been a convincing demonstration as to how geodemographics can inform resource allocation. The here developed advanced geodemographic framework does not lend itself directly as a basis for resource allocation either, but since this appears to be a point often made about geodemographics, an attempt shall be made, using a fictional example to think through the discrete decisions and adjustments that would be necessary.

First, LSOAs could be ranked by the prevalence of vulnerable health milieus. This requires a decision as to which milieus should be considered vulnerable. Some strategies




| **borough** | **amount (£)** | **lifestyle (£)** | **special (£)** |
|---|---:|---:|---:|
| Barking and Dagenham | 248,813 | 202,521 | 46,292 |
| Barnet | 395,809 | 218,993 | 176,816 |
| Bexley | 276,688 | 189,027 | 87,661 |
| Brent | 388,352 | 311,281 | 77,071 |
| Bromley | 376,646 | 256,990 | 119,656 |
| Camden | 264,852 | 222,457 | 42,395 |
| City of London | 8,286 | 8,286 | 0 |
| Croydon | 458,573 | 354,149 | 104,424 |
| Ealing | 435,401 | 299,683 | 135,718 |
| Enfield | 375,710 | 263,602 | 112,108 |
| Greenwich | 309,083 | 223,607 | 85,476 |
| Hackney | 318,221 | 268,664 | 49,557 |
| Hammersmith and Fulham | 229,027 | 211,150 | 17,877 |
| Haringey | 313,966 | 257,083 | 56,884 |
| Harrow | 264,701 | 166,751 | 97,950 |
| Havering | 299,031 | 179,763 | 119,269 |
| Hillingdon | 334,608 | 264,937 | 69,671 |
| Hounslow | 303,693 | 209,832 | 93,861 |
| Islington | 276,790 | 226,187 | 50,603 |
| Kensington and Chelsea | 177,555 | 153,896 | 23,659 |
| Kingston upon Thames | 168,836 | 129,339 | 39,498 |
| Lambeth | 380,394 | 247,359 | 133,035 |
| Lewisham | 373,136 | 302,599 | 70,537 |
| Merton | 235,212 | 188,426 | 46,787 |
| Newham | 434,595 | 288,791 | 145,804 |
| Redbridge | 324,540 | 201,773 | 122,767 |
| Richmond upon Thames | 178,374 | 137,254 | 41,120 |
| Southwark | 379,063 | 246,332 | 132,731 |
| Sutton | 230,127 | 182,736 | 47,392 |
| Tower Hamlets | 337,377 | 277,618 | 59,759 |
| Waltham Forest | 316,257 | 274,915 | 41,342 |
| Wandsworth | 331,816 | 227,854 | 103,961 |
| Westminster | 254,466 | 176,079 | 78,386 |
| **total** | **10,000,000** | **7,369,932** | **2,630,068** |

Figure 9.1: A fictional example of resource allocation

could support the decision. Milieu-wise health outcomes derived from the survey could guide the decision, although a choice still needs to be made with respect to which indicator may be appropriate. Hospitalisations may lead to a different conclusion than using disability risk or self-rated health.

Second, the untransformed variables that define health environments could be used to calculate the aggregate risk across all indicators, weighted by the relative prevalence of the condition. Resource allocation could then be proportionate to the aggregate risk values. Here, one has to take into account that different diseases may require different resources for prevention or treatment – clinical equipment, excellence and expertise – independent of their relative prevalence. Disease weights could be modified accordingly. Finally, a score that combines the vulnerable milieu prevalence and the aggregate risk for each LSOA could be calculated. Unlike the IMD, this score would not be based on ranks; in statistical terms, the values each LSOA can attain are independent.

To provide a worked example, let us say the mayor of London had found some spare 10 million pounds and in a generous move decided to dedicate it to reducing population vulnerability. The estimated location quotient of the three most vulnerable milieus (*Enduring Isolation, Unconcerned Starters, Retiring Generation*) are put together and weighted by their population in each LSOA. The diseases weighted by their relative prevalence (and the inverse for self-rated health) are taken to calculate aggregate risk in each LSOA. Both milieu proportion and aggregate risk are then multiplied to allocate the weighted mean amount that is available for each LSOA ($£10m/4,835 = £2,068$) [Figure 9.1]. These amounts may be aggregated to borough level and further refined by using the proportion of the population that live in health space types D and E in each borough to allocate resources to lifestyle-focussed and special, locally focussed interventions. Mathematically speaking, this is an easy approach to resource allocation. Whether it is commendable in political terms cannot be decided by science; it needs to be debated by the public.

The spatially continuous representations of lifestyle milieus thus allows a more nuanced approach to resource decisions, which does not rely on discrete classes. This advantage is derived from the multi-level data integration of the framework. Similar integration approaches are probably common practice in commercial classifications [cf Webber 2004]. Yet, as any commercial product, commercial classifications are made to maximise short-term profit, their optimisation rests on some sort of cost-benefit analysis set to improve the operating margin of those who develop classifications and those who pay to use them. The long-term picture is of minor importance: so long classifications are up-to-date, they are useful. There is neither memory nor foresight in consumer targetting.

This is entirely different in policy. The most pressing social issues, such as health inequalities, appear to be firmly rooted in the social system and persistent. Health inequalities have received more attention in the United Kingdom than in other Euroean countries in the past, and nevertheless policy programmes have proven ineffective [Mackenbach 2011]. Complex, long-term problems require complex, long-term solutions, and the first step towards those solutions is measurement, description and diagnosis. Researchers and policy-makers alike need to know what social measurements are taken, how those measurements are processed, summarised and disseminated, in order to refine them for public goals and investigate their impact. Open debate and transparency of methods is particularly important in an information age, in which 'digital' cultural capital is mobilised unequally.

## 9.3 Future extensions of the framework

The interpretive geodemographic framework offers a number of starting points for potential future advancements that would overcome some limitations associated with this framework in particular and geodemographics in general. In addition, the framework may be embedded in wider comparative investigations that support the theoretical conceptualisation of vulnerability in social science and urban studies.

### Measurement and system dynamics

As for comprehensiveness of measurement, not all relevant aspects of vulnerability have been captured in this framework. The largest block of missing information relates to environmental conditions, including air pollution, noise, crime and what is sometimes referred to as therapeutic landscapes. General and spatial access to sufficient quality health care is another set of information that constitutes vulnerability. In terms of health, only hospital admissions are considered, not data arising from primary care, although some correlation may be supposed. Finally, consumer data can add health-relevant consumption patterns to the health profiles of milieus. Depending on the data that are available, either some statisical matching could be used to attribute consumption patterns to the milieus or to create a separate consumer classification that is viewed alongside the milieus as an additional geodemographic layer of information.

So far, the advanced framework only represents a snapshot of one point in time. Although uncertainty can be assessed in statistical terms, nothing is known about the temporal stability of the classification. This leaves a major shortcoming of current geodemographic approaches unaddressed. Some sense of temporal stability, however, may be derived from using the different layers of information that are part of the advanced geodemographic framework to validate the phenomena described in each against each other. For example, the health environments identified for London appear to correspond to the milieu geographies in substantively plausible ways. The spatial structural models confirmed the consistency of milieu geographies and the geography of various health indicators based on what one would expect given the health profile of each milieu. Additionally, the strong correspondence between self-rated health derived from the 2011 Census and 2008/2009 HES data (in absence of 2010/2011 HES data) suggests some stability of phenomena. This matching across different layers of information is indeed a form of triangulation that is inherent in the advanced geodemographic framework.

A simple approach to incorporate substantive uncertainty and temporal dynamics would be to measure neighbourhood (LSOAs) turnover using mid-year population estimates or other ancillary data. This would allow to quantify uncertainty in terms of neighbourhood change and attribute a measure of stability to each LSOA. Mid-year population estimates or other data sources could be used as a base population to which the milieus could be matched. Sensitivity tests in relation to the spatial microsimulation applied to London show that using only a few variables, notably age with coarse age bands, educational attainment and tenure, provides plausible though weaker geographic patterns of the population. These estimates may be based on a variety of official data sources (population mid-year estimates, incapacity benefits, council tax bands to name a few) as well as commercial or consumer data. The census years could then be used to validate and calibrate the between-years microsimulation models.

Since, in principle, the milieus lend themselves to dynamic simulation of policy interventions, the generic and spatial impact of policy could be assessed at an annual resolution. Recent advancements in latent choice modelling bear promise to advance geodemographics into this direction, enhancing the political value of the framework by the possibility of scenario development and ex-ante policy evaluation.

### Classification, complexity and the "urban now"

The comparative elements of the framework may also provide a basis for more detailed and explicit investigations of differentiated urban outcomes and nourish diverse conceptualisations of urban vulnerability. The framework offers suggestions as to what it is that may usefully be compared across cities. Robinson argues that alternative views on what constitutes a case to be studied (outcomes, processes, experiences) enables inquiry that draws on and expands knowledge of the widest possible range of urban experiences.

This framework provides multiple scales and starting points to re-think the objects of investigation in the study of urban vulnerability. For example, it may be decided to view individuals of certain milieus and research their manifestation in different cities. Neighbourhood ensembles of a particular kind may be revisited in different cities to explore their common causes and yet identify specific outcomes. Or cities with the most different regional populations may be selected as a starting point to study plural urban outcomes at the city-level. It can be investigated how dominant ideologies of neoliberal restructuring shape or modify vulnerabilities differentially in conjunction with local conditions and how they may thus be expressed at different ecological levels (such as in health milieus or health environments). But effects of more local urban policies or public health approaches may be compared across cities at different ecological levels in order to illuminate the active or inactive constituents of vulnerability. The framework may thus provide different grounds for comparison (genetic and generative as detailed in Robinson [2015]) in order to trace the manifestations of outcomes, processes and experiences in social and geographical space at different ecological levels.

The technique of classification, which has been central to the thesis, seems particularly suited to support various comparative tactics. We may classify across contexts, possibly by building on more culturally grounded geographies, or compare two different taxonomies to reason about their genesis, to view individual phenomena in their spatial and taxonomic context and explore their assemblage. Indeed, since classification places no pre-determined weight on global and local elements in their assemblage, the method may be part of a quantitative answer to empirically approach the reconceptualisation of urban outcomes as contextualised singularities. Urban outcomes are thus not by default assumed as being generated by a shared, global process; rather, they form part of a class of outcomes that share empirical similarities. Unlike linear explanatory models, which often merely specify what we more or less expect, classification demands a search for differentiated explanation and new concepts for emergent classes and their singular instances. It encourages interpretations that are in the spirit of studying the "urban now" [Robinson 2013b] through a practice of drawing together "elements from cities and places distant in both time and space, with leaps of explanation and connection reaching back in time as well as across other places to constitute the immanent interpretive space-times of globalizing urbanism." [ibid., 666].

One may envision, for example, an international comparison of London and Hong Kong (where broadly equivalent datasets to those used here currently become available). Using classification within a re-framing of urban vulnerabilities in terms of the urban now would orient the study towards social, systemic root causes and more profound investigation of how common global processes interact with the cities' diverse functionings to produce specific expressions of vulnerability within health milieus and health environments or at the city-level. We may interpret these expressions in the context of similar institutional roots formed during British colonialisation, flows of financial capital and labour as well as divergent political trajectories and urban policy regimes.

In such a project, which certainly requires a mix of different methods, we would benefit from the double role of classification: on the one hand, we learn about individual observations by their membership to a group of similar characteristics and the group's distinctiveness to other groups; on the other hand, we gain insight into the system of which the groups are part. Inasmuch as health milieus, health environments and health spaces represent distinct phenomena, so, too, are they all products of the social system as a whole. They are instantiations of the system's workings that distribute assets and risks in society and yet not wholly determined but shaped by local context, conditions and agency. If a temporal dimension is added, the double role of classification permits us to not only view how individual observations (people, places) change their position in the taxonomy but also how the taxonomy itself changes and how individual trajectories interact with these changes and generate emergent dynamics. It thus supports reasoning on cause not by reducing complexity but by accounting for it.

As novel, digital, routinely collected data of individuals' activities arise from diverse sources, the opportunities for social science to characterise and understand the urban now increase [cf Burrows & Savage 2014]. Linking this data into some shared, meaningful context in order to enable robust social measurement will be a central concern in this undertaking [Longley 2005b]; and the work of this thesis illustrates a potential step towards it. Yet as quantitative social science advances in analysing, producing and disseminating new information, greater interest needs to be taken in the social consequences of classifications, not just because of the possibly intensified double hermeneutic of classifications, but also because, as Bourdieu [1984, 483] alerts us, the classified is an outcome of the prejudices of the classifier, who belongs to a particular milieu in which classifying is learnt, specifically embodied and therefore never socially neutral. Open debate, verification and social impact assessment thus become empirical projects in their own right and should be critically pursued in both science and politics.

## References


Abbas, J, Ojo, a & Orange, S. 2009. "Geodemographics–a tool for health intelligence?" *Public health* 123.1, e35–9.

Albatineh, A. N., Niewiadomska-Bugaj, M. & Mihalko, D. 2006. "On Similarity Indices and Correction for Chance Agreement". *Journal of Classification* 23.2, pp. 301–313.

Amin, A. & Thrift, N. 2002. *Cities: reimagining the urban.* Cambridge: Polity.

Anderson, T. K. 2010. "Using geodemographics to measure and explain social and environment differences in road traffic accident risk". *Environment and Planning A* 42.9, pp. 2186–2200.

Asada, Y. 2005. "A framework for measuring health inequity." *Journal of Epidemiology and Community Health* 59.8, pp. 700–705.

Azar, D. & Rain, D. 2007. "Identifying population vulnerable to hydrological hazards in San Juan, Puerto Rico". *GeoJournal* 69.1-2, pp. 23–43.

Baggott, R. 2004. *Health and health care in Britain.* Basingstoke UK: Palgrave Macmillan.

Bankoff, G., Frerks, G. & Hilhorst, D. 2004. *Mapping vulnerability : disasters, development, and people.* London, Sterling VA: Earthscan.

Batey, P., Brown, P. & Pemberton, S. 2008. "Methods for the Spatial Targeting of Urban Policy in the UK: A Comparative Analysis". *Applied Spatial Analysis and Policy* 1.2, pp. 117–132.

Batty, M. 2013. "Big data, smart cities and city planning". *Dialogues in Human Geography* 3.3, pp. 274–279.

Baum, F. & Fisher, M. 2014. "Why behavioural health promotion endures despite its failure to reduce health inequities". *Sociology of Health & Illness* 36.2, pp. 213–25.

Beck, U. 1992. *Risk society: towards a new modernity.* London : Sage Publications.

Bennett, T. et al. 2009. *Culture, Class, Distinction.* London: Routledge.

Benzeval, M. et al. 2014. *Understanding Society The UK Household Longitudinal Study - Biomarker User Guide and Glossary.* Colchester UK.

Besag, J, York, J & Mollie, a. 1991. "Bayesian image-restoration, with 2 applications in spatial statistics". *Annals of the Institute of Statistical Mathematics* 43.1, pp. 1–20.

Best, N., Richardson, S. & Thomson, A. 2005. "A comparison of Bayesian spatial models for disease mapping". *Statistical Methods in Medical Research* 14.1, pp. 35–59.

Bevan, M. A. & Croucher, K. L. 2011. *Lifetime Neighbourhoods.* London: Department for Communities and Local Government.

Bhaskar, R. 2008 (1975). *A realist theory of science.* 3rd. London, New York: Verso.

Bhattacharjee, A., Maiti, T. & Petrie, D. 2015. "Regional Science and Urban Economics General equilibrium effects of spatial structure : Health outcomes and health behaviours in Scotland ". *Regional Science and Urban Economics* 49.2014, pp. 286–297.

Birkin, M. & Clarke, G. 2011. "The enhancement of spatial microsimulation models using geodemographics". *The Annals of Regional Science* 49.2, pp. 515–532.

Blas, E. & Kurup, A. S. 2010. *Equity, social determinants and public health programmes.* Geneva: World Health Organization.

Blas, E., Sommerfeld, J. & Kurup, A. S. 2011. *Social determinants approaches to public health: from concept to practice*. Ed. by E. Blas, J. Sommerfeld & A. S. Kurup. Geneva: World Health Organization.

Blaxter, M. 1990. *Health and lifestyles*. London: Routledge.

— 2003. "Biology, social class and inequalities in health: their synthesis in 'health capital'". In: *Debating biology. Sociological reflections on health, medicine and society*. Ed. by S. J. Williams, L. Birke & G. A. Bendelow. London, New York: Routledge. Chap. 5, pp. 69–83.

— 2010. *Health*. Cambridge: Polity.

Boattini, A. et al. 2012. "General Method to Unravel Ancient Population Structures through Surnames, Final Validation on Italian Data". *Human biology* 84.3, pp. 235–270.

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

— 1984. *Distinction: a Social Critique of the Judgment of Taste*. Cambridge: Harvard University Press.

— 1990. a. *In Other Words*. Stanford: Stanford University Press.

— 1990. b. *The Logic of Practice*. Cambridge: Polity Press.

— 1999. "Site effects". In: *The weight of the world. Social suffering in contemporary society*. Ed. by P. Bourdieu. Cambridge, Oxford: Polity Press, pp. 123–129.

— 2000. *Pascalian Meditations*. Stanford: Stanford University Press.

Bowker, G. C. & Star, S. L. 1999. *Sorting things out: classification and its consequences*. Cambridge MA; London: The MIT Press.

Breetzke, G. D. & Horn, A. C. 2009. "A geodemographic profiler for high offender propensity areas in the city of Tshwane, South Africa". *Environment and Planning A* 41.1, pp. 112–127.

Brown, P. & McCulloch, P. 2001. "Control sample design using a geodemographic discriminator: An application of Super Profiles". *Journal of Geographical Systems*.

Brunner, E. & Marmot, M. 2006. "Social organization, stress, and health". In: *Social determinants of health*. Ed. by M. Marmot & R. G. Wilkinson. Oxford: Oxford University Press.

Burgess, E. 1968. *The growth of the city: an introduction to a research project*.

Burrows, R. & Gane, N. 2006. "Geodemographics, Software and Class". *Sociology* 40.5, pp. 793–812.

Burrows, R. & Ellison, N. 2004. "Sorting Places Out? Towards a social politics of neighbourhood informatization". *Information, Communication & Society* 7.3, pp. 321–336.

Burrows, R. J. 2013. "The New Gilded Ghettos: The Geodemographics of the Super-Rich". *Discover Society* 3, pp. 1–5.

Burrows, R. J. & Savage, M. 2014. "After the crisis? Big Data and the methodological challenges of empirical sociology". *Big Data and Society* June, pp. 1–7.

Butler, T. & Robson, G. 2003. *London calling: the middle classes and the re-making of inner London*. Oxford: Berg.

Byrne, D. 1998. *Complexity theory and the social sciences: an introduction*. London: Routledge.

— 2002. *Interpreting quantitative data*. London, Thousand Oaks, Calif.: Sage.

CACI. 2013. *Acorn technical document*. Tech. rep. CACI Product Development Team.

Cao, Y., Hwang, S.-S. & Xi, J. 2012. "Project-induced displacement, secondary stressors, and health." *Social Science & Medicine* 74.7, pp. 1130–8.

Cardona, O. D. 2004. "The need for rethinking the concepts of vulnerability and risk from a holistic perspective: a necessary review and criticism for effective risk management". In: *Mapping vulnerability : disasters, development, and people*. Ed. by G. Bankoff, G. Frerks & D. Hilhorst. London, Sterling VA: Earthscan, pp. 10–24.

Carpiano, R. M. 2006. "Toward a neighborhood resource-based theory of social capital for health: can Bourdieu and sociology help?" *Social Science & Medicine* 62.1, pp. 165–75.

— 2007. "Neighborhood social capital and adult health: an empirical test of a Bourdieu-based model". *Health & Place* 13.3, pp. 639–55.

Cavalli-Sforza, L. L. 2001. *Genes, peoples, and languages*. London: Penguin.

Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. 1994. *The history and geography of human genes*. New Jersey: Princeton University Press.

Cerimele, J. M. & Katon, W. J. 2013. "Associations between health risk behaviors and symptoms of schizophrenia and bipolar disorder: A systematic review". *General Hospital Psychiatry* 35.1, pp. 16–22.

Chang, C. & Hsiao, H. 2007. "Establishing hazards-of-place model of vulnerability—a case of flood in the Shijhih City Taiwan". In: *2nd international conference on urban disaster reduction reduction, Taiwan*.

Chang, S. S. et al. 2011. "Geography of suicide in Taiwan: Spatial patterning and socioeconomic correlates". *Health and Place* 17.2, pp. 641–650.

Cheshire, J. A. & Longley, P. A. 2013. "Japanese surname regions". *Papers in Regional Science* online, pp. 1–17.

Cheshire, J. A., Longley, P. A. & Mateos, P. 2009. "Family names as indicators of Britain's changing regional geography". London.

Cheung, Y. T. D. et al. 2012. "Spatial analysis of suicide mortality in Australia: Investigation of metropolitan-rural-remote differentials of suicide risk across states/territories". *Social Science and Medicine* 75.8, pp. 1460–1468.

Clark, S. D., Birkin, M. & Heppenstall, A. 2014. "Sub regional estimates of morbidities in the English elderly population." *Health & Place* 27, pp. 176–85.

Coburn, D. 2012. "Health inequalities: a response to Scambler." *Sociology of Health & Illness* 34.1, 149–50; discussion 151–3.

Colburn, L. L. & Jepson, M. 2012. "Social Indicators of Gentrification Pressure in Fishing Communities: A Context for Social Impact Assessment". *Coastal Management* 40.3, pp. 289–300.

Collins, T. W., Grineski, S. E. & de Lourdes Romo Aguilar, M. 2009. "Vulnerability to environmental hazards in the Ciudad Juárez (Mexico)–El Paso (USA) metropolis: A model for spatial risk assessment in transnational context". *Applied Geography* 29.3, pp. 448–461.

Congdon, P. 2012. "Modelling small-area inequality in premature mortality using years of life lost rates". *Journal of Geographical Systems* 15.2, pp. 149–167.

Congdon, P. 2013. “Assessing the impact of socioeconomic variables on small area variations in suicide outcomes in England.” *International journal of environmental research and public health* 10.1, pp. 158–77.

CSDH. 2008. *Closing the gap in a generation: Health equity through action on the social determinants of health. Final report of the Commission on Social Determinants of Health.* Tech. rep. Geneva: World Health Organization.

Cummins, S. et al. 2007. “Understanding and representing 'place'in health research: a relational approach”. *Social Science & Medicine* 65.9, pp. 1825–38.

Cutter, S. L. 1996. “Vulnerablity to environmental hazards”. *Progress in Human Geography* 20.4, pp. 529–539.

Cutter, S. L. & Emrich, C. T. 2006. “Moral hazard, social catastrophe: The changing face of vulnerability along the hurricane coasts”. *The ANNALS of the American Academy of Political and Social Science* 604.1, pp. 102–112.

Cutter, S. L., Mitchell, J. T. & Scott, M. S. 2000. “Revealing the Vulnerability of People and Places: A Case Study of Georgetown County, South Carolina”. *Annals of the Association of American Geographers* 90.4, pp. 713–737.

Cutter, S. L., Boruff, B. J. & Shirley, W. L. 2003. “Social vulnerability to environmental hazards*”. *Social Science Quarterly* 84.2.

Cutter, S. L. et al. 2008. “A place-based model for understanding community resilience to natural disasters”. *Global Environmental Change* 18.4, pp. 598–606.

Dahlgren, G. & Whitehead, M. 1992. *Policies and strategies to promote equity in health.* Copenhagen: World Health Organization, Regional Office for Europe.

Darlu, P., Bloothooft, G. & Boattini, A. 2012. “The family name as socio-cultural feature and genetic metaphor: From concepts to methods”. *Human Biology* 84.2, pp. 169–214.

Dawkins, R. 2006(1989). *The selfish gene.* Oxford: Oxford University Press.

Debenham, J., Clarke, G. & Stillwell, J. 2003a. “Extending geodemographic classification: a new regional prototype”. *Environment and Planning A* 35.6, pp. 1025–1050.

Debenham, J., Stillwell, J. & Clarke, G. 2003b. “The estimation of self-containment and catchment size indicators for use in small area classification”. *International Journal of Population Geography* 9.3, pp. 253–271.

Dedman, D. et al. 2006. “Using Geodemographics to Illustrate Health Inequalities”.

Degioanni, A., Darlu, P. & Raffoux, C. 2003. “Analysis of the French National Registry of unrelated bone marrow donors, using surnames as a tool for improving geographical localisation of HLA haplotypes.” *European Journal of Human Genetics* 11.10, pp. 794–801.

DeRuiter, W. K. et al. 2014. “A Longitudinal Examination of the Interrelationship of Multiple Health Behaviors”. *American Journal of Preventive Medicine* 47.3, pp. 283–289.

Desrosières, A. 1998. *The politics of large numbers : a history of statistical reasoning.* Cambridge MA, London: Harvard University Press.

DeVerteuil, G. et al. 2007. “Mental health and the city: intra-urban mobility among individuals with schizophrenia.” *Health & Place* 13.2, pp. 310–23.

Dietz, R. D. 2002. “The estimation of neighborhood effects in the social sciences: An interdisciplinary approach”. *Social Science Research* 31, pp. 539–575.

Diez-Roux, A. V. 2004. “Estimating neighborhood health effects: the challenges of causal inference in a complex world”. *Social Science & Medicine* 58.10, pp. 1953–60.

Diez-Roux, A. V. 2007. “Neighborhoods and health: where are we and were do we go from here?” *Revue d’épidémiologie et de santé publique* 55.1, pp. 13–21.

Dipierri, J. et al. 2011. “A study of the population of Paraguay through isonymy.” *Annals of Human Genetics* 75.6, pp. 678–87.

Dowling, R. 2009. “Geographies of identity: landscapes of class”. *Progress in Human Geography* 33.6, pp. 833–839.

Elixhauser, A, Steiner, C & Palmer, L. 2010. *Clinical Classification Software (CCS).* Tech. rep. Rockville US.

Elliott, P. & Wakefield, J. 2001. “Bias and confounding in spatial epidemiology”. In: *Spatial Epidemiology.* Ed. by P. Elliott et al. Oxford: Oxford University Press.

Ellison, N. & Burrows, R. 2007. “New Spaces of (Dis)engagement? Social Politics, Urban Technologies and the Rezoning of the City”. *Housing Studies* 22.3, pp. 295–312.

Eriksen, L. et al. 2013. “The association between physical activity, cardiorespiratory fitness and self-rated health”. *Preventive Medicine* 57.6, pp. 900–902.

Evans, R. & Stoddart, G. 1990. “Producing health, consuming health”. *Social Science & Medicine* 31.12, pp. 1347–1363.

Everitt, B. 1974. *Cluster Analysis.* London: Heinemann Educational for the Social Science Research Council.

Ewing, R. et al. 2013. “Relationship between urban sprawl and physical activity, obesity, and morbidity - Update and refinement.” *Health & Place* 26C, pp. 118–126.

Experian. 2014. *Mosaic. The consumer classification solution for consistent cross-channel marketing.* Available online at http://publ.com/RUfDTGu.

Factor, R., Awerbuch, T. & Levins, R. 2013. “Social and land use composition determinants of health: variability in health indicators.” *Health & Place* 22, pp. 90–7.

Farr, M. & Evans, A. 2005. “Identifying ‘unknown diabetics’ using geodemographics and social marketing”. *Journal of Direct Data and Digital Marketing Practice* 7.1, pp. 47–58.

Ford, E. S. et al. 2012. “Healthy lifestyle behaviors and all-cause mortality among adults in the United States”. *Preventive Medicine* 55.1, pp. 23–27.

Fotheringham, A. S. 1997. “Trends in quantitative methods I: stressing the local”. *Progress in Human Geography* 21.1, pp. 88–96.

Frohlich, K. L. & Abel, T. 2014. “Environmental justice and health practices: understanding how health inequities arise at the local level”. *Sociology of Health & Illness* 36.2, pp. 199–212.

Galster, G. C. 2012. “The Mechanism(s) of Neighbourhood Effects: Theory, Evidence, and Policy Implications”. English. In: *Neighbourhood Effects Research: New Perspectives.* Ed. by M. Van Ham et al. Springer Netherlands, pp. 23–56.

Gee, G. C. & Payne-Sturges, D. C. 2004. “Environmental health disparities: a framework integrating psychosocial and environmental concepts.” *Environmental health perspectives* 112.17, pp. 1645–1653.

Giddens, A. 1982. *Profiles and critiques in social theory.* London : Macmillan.

— 1991. *Modernity and self-identity: self and society in the late modern age.* Cambridge: Polity Press.

GLA. 2010. *The London Health Inequalities Strategy.* April. London.
— 2011. *London Plan.* London: Greater London Authority, pp. 15–36.
Goldin, C. & Shim, M. 2004. “Making a Name: Women’s Surnames at Marriage and Beyond”. *Journal of Economic Perspectives* 18.2, pp. 143–160.
Gonzalez-Bailon, S. 2013. “Big data and the fabric of human geography”. *Dialogues in Human Geography* 3.3, pp. 292–296.
Goodchild, M. F. 1994. “GIS and geographic research”. In: *Ground Truth. The Social Implications of Geographic Information Systems.* Ed. by J. Pickles. The Guildford Press, pp. 31–50.
Gorsuch, R. L. 1983. *Factor analysis.* New Jersey: Lawrence Erlbaum Associates.
Goss, J. 1994. “Marketing the New Marketing: The Strategic Discourse of Geodemographic Information Systems”. In: *Ground Truth. The Social Implications of Geographic Information Systems.* Ed. by J. Pickles. The Guildford Press, pp. 130–170.
— 1995. “”We Know Who You Are and We Know Where You Live”: The Instrumental Rationality of Geodemographic Systems”. *Economic Geography* 71.2, pp. 171–198.
Graham, E. et al. 2009. “Mixing Housing Tenures: Is it Good for Social Well-being?” *Urban Studies* 46.1, pp. 139–165.
Graham, H. & Kelly, M. P. 2004. *Health inequalities: concepts, frameworks and policy.* London.
Graham, S. D. 2005. “Software-sorted geographies”. *Progress in Human Geography* 29.5, pp. 562–580.
Green, M. A., Vickers, D. & Dorling, D. 2014. “A neighbourhood level mortality classification of England and Wales, 2006-2009.” *Health & Place* 30C, pp. 196–204.
Grenfell, M. 2012. *Pierre Bourdieu: key concepts.* Durham: Acumen.
GRNUHE. 2010. *Improving urban health equity through action on the social and environmental determinants of health.* Tech. rep. July. London: Global Research Network for Urban Health Equity.
Guthman, J. & Mansfield, B. 2012. “The implications of environmental epigenetics: A new direction for geographic inquiry on health, space, and nature-society relations”. *Progress in Human Geography* 37.4, pp. 486–504.
Hahn, M. B., Riederer, A. M. & Foster, S. O. 2009. “The Livelihood Vulnerability Index: A pragmatic approach to assessing risks from climate variability and change—A case study in Mozambique”. *Global Environmental Change* 19.1, pp. 74–88.
Harland, K et al. 2012. “Creating realistic synthetic populations at varying spatial scales: a comparative critique of population synthesis techniques”. *Journal of Artifical Societies and Social Simulation* 15, pp. 1–24.
Harris, R., Sleight, P. & Webber, R. 2005. *Geodemographics, GIS, and neighbourhood targeting / Richard Harris, Peter Sleight and Richard Webber.* Chichester : Wiley.
Harris, R., Johnston, R. & Burgess, S. 2007. “Neighborhoods, Ethnicity and School Choice: Developing a Statistical Framework for Geodemographic Analysis”. *Population Research and Policy Review* 26.5-6, pp. 553–579.
Harvey, D. 1996. *Justice, nature, and the geography of difference.* Cambridge Mass, Oxford: Blackwell.
Hastie, T., Tibshirani, R. & Friedman, J. 2009. *The Elements of Statistical Learning Data Mining, Inference, and Prediction.* New York: Springer.

Hawe, P., Shiell, A. & Riley, T. 2009. "Theorising interventions as events in systems". *American journal of community psychology* 43.3-4, pp. 267–76.

Hedman, L. et al. 2013. "Cumulative exposure to disadvantage and the intergenerational transmission of neighbourhood effects". *Journal of Economic Geography* online.

Hellenthal, G. et al. 2015. "Fine scale genetic structure of the British population". *Nature* 519, pp. 309–314.

Hermes, K. & Poulsen, M. 2013. "The intraurban geography of generalised trust in Sydney". *Environment and Planning A* 45.2, pp. 276–294.

Herrera Paz, E. F. et al. 2014. "Surnames in Honduras: A study of the population of Honduras through isonymy." *Annals of human genetics* 78.3, pp. 165–77.

Hess, S., Shires, J. & Jopson, A. 2013. "Accommodating underlying pro-environmental attitudes in a rail travel context: Application of a latent variable latent class specification". *Transportation Research Part D: Transport and Environment* 25, pp. 42–48.

Hilhorst, D. & Bankoff, G. 2004. "Introduction: Mapping vulnerability". In: *Mapping vulnerability : disasters, development, and people.* Ed. by G. Bankoff, G. Frerks & D. Hilhorst. London, Sterling VA: Earthscan, pp. 1–9.

Hill, T. D., Ross, C. E. & Angel, R. J. 2005. "Neighborhood disorder, psychophysiological distress, and health." *Journal of Health and Social Behavior* 46.2, pp. 170–86.

Hiscock, R. et al. 2003. "Residents and Residence: Factors Predicting the Health Disadvantage of Social Renters Compared to Owner-Occupiers". *Journal of Social Issues* 59.3, pp. 527–546.

Hogan, D. J. & Marandola, E. 2007. "Vulnerability to Natural Hazards in Population-Environment Studies". In: November. Background paper to the Population-Environment Research Network (PERN) Cyberseminar1 on Population & Natural Hazards 5-19 November 2007, pp. 1–11.

Horrocks, C. & Johnson, S. 2014. "A socially situated approach to inform ways to improve health and wellbeing." *Sociology of Health & Illness* 36.2, pp. 175–86.

Hubert, L. & Arabie, P. 1985. "Comparing partitions". *Journal of Classification* 218, pp. 193–218.

Jackson, C. H., Richardson, S. & Best, N. G. 2008. "Studying place effects on health by synthesising individual and area-level outcomes." *Social Science & Medicine* 67.12, pp. 1995–2006.

Jenkins, R. 1992. *Pierre Bourdieu.* London: Routledge.

Jobling, M. A. 2001. "In the name of the father: surnames and genetics." *Trends in genetics : TIG* 17.6, pp. 353–7.

Johnson, O. 2012. "Toward a Theory of Place". In: *Research on Schools, Neighborhoods and Communities: Toward Civic Responsibility.* Ed. by W. Tate. Chap. 2, pp. 29–46.

Jongeneel-Grimen, B. et al. 2014. "The relationship between physical activity and the living environment: a multi-level analyses focussing on changes over time in environmental factors". *Health & Place* 26, pp. 149–160.

Jutte, D. P., Roos, L. L. & Brownell, M. D. 2011. "Administrative record linkage as a tool for public health research." *Annual Review of Public Health* 32, pp. 91–108.

Kasmel, A. et al. 2004. “Association between health behaviour and self-reported health in Estonia, Finland, Latvia and Lithuania”. *The European Journal of Public Health* 14.1, pp. 32–36.

Kearns, A. & Mason, P. 2015. “Regeneration, relocation and health behaviours in deprived communities”. *Health & Place* 32, pp. 43–58.

Kearns, R. & Moon, G. 2002. “From medical to health geography: novelty, place and theory after a decade of change”. *Progress in Human Geography* 5, pp. 605–625.

Khaw, K.-T. et al. 2008. “Combined Impact of Health Behaviours and Mortality in Men and Women: The EPIC-Norfolk Prospective Population Study”. *PLoS Med* 5.1, e12.

Kimura, Y. et al. 2011. “Geodemographics profiling of influenza A and B virus infections in community neighborhoods in Japan.” *BMC infectious diseases* 11.1, p. 36.

King, T. E. & Jobling, M. A. 2009. “What’s in a name? Y chromosomes, surnames and the genetic genealogy revolution.” *Trends in Genetics* 25.8, pp. 351–60.

Knies, G. 2014. *Understanding Society. The UK household longitudinal study Waves 1-4 – User Manual.* Tech. rep. Colchester UK: Institute for Social and Economic Research.

Koot, J., Mtung’e, R. & Miller, J. 2011. “Insecticide-treated nets in Tanzania mainland: challenges in reaching the most vulnerable, most exposed and poorest groups”. In: *Social determinants approaches to public health: from concept to practice.* Ed. by E. Blas, J. Sommerfeld & A. S. Kurup. Geneva: World Health Organization.

Krieger, N. 2005. “Embodiment: a conceptual glossary for epidemiology.” *Journal of Epidemiology and Community Health* 59.5, pp. 350–5.

Kumpulainen, S. 2006. “Vulnerability concepts in hazard and risk assessment: Natural and technological hazards and risks affecting the spatial development of European regions”. *Geological Survey of Finland* Special Papers.42, pp. 65–74.

Lash, S. 2002. *Critique of information.* London: Sage.

Lasker, G. 1977. “A coefficient of relationship by isonymy: a method for estimating the genetic relationship between populations”. *Human Biology* 49.3, pp. 489–493.

Lawder, R. et al. 2013. “Healthy Mixing? Investigating the Associations between Neighbourhood Housing Tenure Mix and Health Outcomes for Urban Residents”. *Urban Studies* 51.2, pp. 264–283.

Lawson, D. J. et al. 2012. “Inference of population structure using dense haplotype data.” *PLoS genetics* 8.1, e1002453.

Layard, R. 2013. *Mental health: the new frontier for labour economics.* CEP Discussion Paper. London: Centre for Economic Performance, London School of Economics and Political Science.

LHC. 2014. *Better health for London.* Tech. rep. London: London Health Commission.

Link, B. G. & Phelan, J. C. 1995. “Social Conditions As Fundamental Causes of Disease”. *Journal of health and social behavior* 35.May, pp. 80–94.

Locker, D. 2008. “Social determinants of health and disease”. In: *Sociology as applied to medicine.* Ed. by G. Scambler. 6th ed. London: Saunders Elsevier, pp. 18–37.

Loef, M. & Walach, H. 2012. “The combined effects of healthy lifestyle behaviors on all cause mortality: A systematic review and meta-analysis”. *Preventive Medicine* 55.3, pp. 163–170.

Long, G. et al. 2015. "Healthy behaviours and 10-year incidence of diabetes: A population cohort study". *Preventive Medicine* 71, pp. 121–127.

Longhurst, B. & Savage, M. 1996. "Social class, consumption and the influence of Bourdieu: some critical issues". In: *Consumption Matters: The Production and Experience of Consumption.* Ed. by S. Edgell, K. Hetherington & A. Warde. Oxford, Cambridge MA: Blackwell Publishers, pp. 274–301.

Longley, P. A. 2005a. "Geographical Information Systems: a renaissance of geodemographics for public service delivery". *Progress in Human Geography* 29.1, pp. 57–63.

Longley, P. A. 2005b. "Urban Studies". In: *Encyclopedia of Social Measurement.* Elsevier, pp. 921–926.

Longley, P. A. 2007. "Some challenges to geodemographic analysis and their wider implications for the practice of GIScience". *Computers, Environment and Urban Systems* 31.6, pp. 617–622.

Longley, P. A. & Singleton, A. D. 2009. "Classification through consultation: public views of the geography of the e-Society". *International Journal of Geographical Information Science* 23.6, pp. 737–763.

Longley, P. A., Webber, R. & Lloyd, D. 2007. "The quantitative analysis of family names: Historic migration and the present day neighborhood structure of Middlesbrough, United Kingdom". *Annals of the Association of American Geographers* 97.December 2005, pp. 31–48.

Longley, P. A., Webber, R. & Li, C. 2008. "The UK geography of the e-society: a national classification". *Environment and Planning A* 40.2, pp. 362–382.

Longley, P. A., Cheshire, J. A. & Mateos, P. 2011a. "Creating a regional geography of Britain through the spatial analysis of surnames". *Geoforum* 42.4, pp. 506–516.

Longley, P. A. et al. 2011b. *Geographic information systems & science.* 3rd. New Jersey: John Wiley & Sons.

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

Luers, A. L. 2005. "The surface of vulnerability: An analytical framework for examining environmental change". *Global Environmental Change* 15.3, pp. 214–223.

Lujala, P., Lein, H. & Rosvoldaune, R. 2014. "Quantifying vulnerability to flooding induced by climate change: The case of Verdal, Norway". *Norsk Geografisk Tidsskrift - Norwegian Journal of Geography* 68.1, pp. 34–49.

Maantay, J. & Maroko, A. 2009. "Mapping Urban Risk: Flood Hazards, Race, & Environmental Justice In New York"". *Applied Geography* 29.1, pp. 111–124.

Macintyre, S, Maciver, S & Sooman, A. 1993. "Area, class and health: should we be focusing on places or people?" *Journal of Social Policy* September 1990, pp. 213–234.

Macintyre, S., Ellaway, A. & Cummins, S. 2002. "Place effects on health: how can we conceptualise, operationalise and measure them?" *Social Science & Medicine* 55.1, pp. 125–139.

Mackenbach, J. P. 2011. “Can we reduce health inequalities? An analysis of the English strategy (1997-2010).” *Journal of Epidemiology and Community Health* 65.7, pp. 568–75.

Manley, D. & Van Ham, M. 2012. “Neighbourhood Effects, Housing Tenure and Individual Employment Outcomes”. English. In: *Neighbourhood Effects Research: New Perspectives*. Ed. by M. Van Ham et al. Springer Netherlands, pp. 147–173.

Marmot, M. G. et al. 1991. “Health inequalities among British civil servants: the Whitehall II study.” *Lancet* 337, pp. 1387–1393.

Marmot, M. 2005. “Social determinants of health inequalities”. *The Lancet* 365, pp. 1099–1104.

— 2010. *Fair Society, Healthy Lives*. Tech. rep. The Marmot Review.

Mateos, P. 2007. “A review of namebased ethnicity classification methods and their potential in population studies”. *Population, Space and Place* 263.May, pp. 243–263.

Maton, K. 2012. “Habitus”. In: *Pierre Bourdieu: key concepts*. Durham: Acumen, pp. 49–66.

McCulloch, P. et al. 2003. “The effects of an awareness-raising program for patients and primary care physicians on the early detection of gastro-oesophageal cancer”. *Surgery* 133.2, pp. 154 –161.

McDermott, J. 2015. “UK - An new order?” *Financial Times Magazine* 21/22 February, pp. 13–19.

McKenzie, R. 1924. “The ecological approach to the study of the human community”. *American Journal of Sociology* 30.3, pp. 287–301.

Moon, G. 1995. “(Re)placing research on health and health care”. *Health & Place* 1.1, pp. 1–4.

Moon, G. 1990. “Conceptions of space and community in British health policy.” *Social Science & Medicine* 30.1, pp. 165–171.

— 2009. “Health Geography”. *International Encyclopedia of Human Geography*, pp. 35–45.

Moon, G., Pearce, J. & Barnett, R. 2012. “Smoking, ethnic residential segregation, and ethnic diversity: a spatio-temporal analysis”. *Annals of the Association of American Geographers* 102.5, pp. 912–921.

Morrissey, K. et al. 2010. “Examining the factors associated with depression at the small area level in Ireland using spatial microsimulation techniques”. *Irish Geography* 43.1, pp. 1–22.

Morrissey, K. et al. 2013. “Using Simulated Data to Examine the Determinants of Acute Hospital Demand at the Small Area Level”. *Geographical Analysis* 45.1, pp. 49–76.

Muggli, E. E., McCloskey, D. & Halliday, J. L. 2006. “Health behaviour modelling for prenatal diagnosis in Australia: a geodemographic framework for health service utilisation and policy development.” *BMC health services research* 6.1, p. 109.

Murray, C. J. L. et al. 2012. “Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010”. *The Lancet* 380.9859, pp. 2197–2223.

Nettleton, S. & Green, J. 2014. “Thinking about changing mobility practices: how a social practice approach can help”. *Sociology of Health & Illness* 36.2, pp. 239–51.

Nettleton, S., Burrows, R. J. & O'Malley, L. 2005. "The mundane realities of the everyday lay use of the internet for health, and their consequences for media convergence". *Sociology of Health and Illness* 27.7, pp. 972–992.

NHS. 2011. *The joint strategic needs assessment.* London.

Nnoaham, K. E. et al. 2010. "Do geodemographic typologies explain variations in uptake in colorectal cancer screening? An assessment using routine screening data in the south of England." *Journal of public health (Oxford, England)* 32.4, pp. 572–81.

Norman, P. & Fraser, L. 2014. "Prevalence of life-limiting conditions in children and young people in England: Time trends by area type." *Health & Place* 26C, pp. 171–179.

Novotný, J. & Cheshire, J. A. 2012. "The surname space of the Czech Republic: examining population structure by network analysis of spatial co-occurrence of surnames". *PloS one* 7.10, e48568.

Oakes, J. M. 2004. "The (mis)estimation of neighborhood effects: causal inference for a practicable social epidemiology." *Social Science & Medicine* 58.10, pp. 1929–52.

Ojo, A., Vickers, D. & Ballas, D. 2013. "Creating a small scale area classification for understanding the economic, social and housing characteristics of small geographical areas in the Philippines". *Regional Science Policy & Practice* 5.1, pp. 1–24.

Olabarria, M. et al. 2014. "Daily mobility patterns of an urban population and their relationship to overweight and obesity". *Transport Policy* 32, pp. 165–171.

Oliver-Smith, A. 2004. "Theorizing vulnerability in a globalized world: a political ecological perspective". In: *Mapping vulnerability : disasters, development, and people.* Ed. by G. Bankoff, G. Frerks & D. Hilhorst. London, Sterling VA: Earthscan, pp. 10–24.

ONS. 2014. *ONS Area classifications.* Office for National Statistics, available online www.ons.gov.uk, April 2014.

Openshaw, S & Blake, M. 1995. "Geodemographic segmentation systems for screening health data." *Journal of Epidemiology and Community Health* 49.S2, S34–8.

Openshaw, S. 1983. "The modifiable areal unit problem". *Concepts and Techniques in Modern Geography* 38.

O'Reilly Radar Team. 2011. *Big Data now: current perspectives from O'Reilly Radar.* O'Reilly Media.

Parker, S., Uprichard, E. & Burrows, R. 2007. "Class Places and Place Classes: Geodemographics and the spatialization of class". *Information, Communication & Society* 10.6, pp. 902–921.

Parry, M. L. 2007. *Climate change 2007: impacts, adaptation and vulnerability.* Published for the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press.

Pearce, J et al. 2009. "The neighbourhood effects of geographical access to tobacco retailers on individual smoking behaviour." *Journal of Epidemiology and Community Health* 63.1, pp. 69–77.

Pearce, J., Barnett, R. & Moon, G. 2011. "Sociospatial inequalities in health-related behaviours: Pathways linking place and smoking". *Progress in Human Geography* 36.1, pp. 3–24.

Pearsall, H. 2012. "Moving out or moving in? Resilience to environmental gentrification in New York City". *Local Environment* 17.9, pp. 1013–1026.

Pendall, R., Theodos, B. & Franks, K. 2012. "Vulnerable people, precarious housing, and regional resilience: an exploratory analysis". *Housing Policy Debate* 22.2, pp. 271–296.

Perry, R. L. & Roberson, J. D. 2002. *Comparative analysis of nations: quantitative approaches.* Boulder, CO: Westview Press.

Petersen, J. et al. 2010. "Geodemographics as a tool for targeting neighbourhoods in public health campaigns". *Journal of Geographical Systems* 13.2, pp. 173–192.

Phelan, J. C. et al. 2004. ""Fundamental Causes" of Social Inequalities in Mortality: A Test of the Theory". *Journal of Health and Social Behavior* 45.3, pp. 265–285.

Pickles, J. 1994. "Representations in an Electronic Age: Geography, GIS and Democracy". In: *Ground Truth. The Social Implications of Geographic Information Systems.* Ed. by J. Pickles. The Guildford Press, pp. 1–30.

Propper, C. et al. 2005. "Local neighbourhood and mental health: evidence from the UK." *Social Science & Medicine* 61.10, pp. 2065–83.

R Core Team. 2014. *R: A language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing.

Raessler, S. 2002. *Statistical Matching. A frequentiest theory, practical applications, and alternative Bayesian approaches.* New York: Springer.

Rees, P. 1971. "Factorial ecology: an extended definition, survey, and critique of the field". *Economic Geography* 47, pp. 220–233.

Reibel, M. 2011. "Classification Approaches in Neighborhood Research: Introduction and Review". *Urban Geography* 32.3, pp. 305–316.

Riva, M. & Smith, D. M. 2012. "Generating small-area prevalence of psychological distress and alcohol consumption: validation of a spatial microsimulation method." *Social Psychiatry and Psychiatric Epidemiology* 47.5, pp. 745–55.

Robinson, J. 2006. *Ordinary cities: between modernity and development.* London: Routledge.

— 2011. "Cities in a World of Cities: The Comparative Gesture". *International Journal of Urban and Regional Research* 35.1, pp. 1–23.

— 2013. a. ""Arriving at" Urban Policies/the Urban: Traces of Elsewhere in Making City Futures". In: *Critical Mobilities.* Ed. by O. Söderström. London: Routledge, pp. 1–28.

— 2013. b. "The urban now: Theorising cities beyond the new". *European Journal of Cultural Studies* 16.6, pp. 659–677.

— 2014. "New geographies of theorising the urban: putting comparison to work for global urban studies". In: *Handbook for Cities of the Global South.* Ed. by S Parnell & S Oldfield. London: Routledge.

— 2015. "Thinking cities through elsewhere: Comparative tactics for a more global urban studies". *Progress in Human Geography*, pp. 1–27.

Rodríguez-Larralde, A. et al. 1998. "Isonymy and the genetic structure of Switzerland. II. Isolation by distance." *Annals of Human Biology* 25.6, pp. 533–40.

Rodríguez-Larralde, A., Morales, J. & Barrai, I. 2000. "Surname frequency and the isonymy structure of Venezuela." *American Journal of Human Biology* 12.3, pp. 352–362.

Rue, H. v., Martino, S. & Chopin, N. 2009. “Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71.2, pp. 319–392.

Rydin, Y. et al. 2012. “Shaping cities for health: complexity and the planning of urban environments in the 21st century.” *Lancet* 379.9831, pp. 2079–2108.

Savage, M. & Burrows, R. J. 2007. “The Coming Crisis of Empirical Sociology”. 41.5, pp. 885–899.

— 2009. “Some Further Reflections on the Coming Crisis of Empirical Sociology”. 43.4, pp. 762–772.

Savage, M., Bagnall, G. & Longhurst, B. 2005. *Globalization and belonging.* London: SAGE.

Savage, M. et al. 2013. “A New Model of Social Class? Findings from the BBC’s Great British Class Survey Experiment”. *Sociology* 47.2, pp. 219–250.

Scambler, G. 2007. “Social Structure and the Production, Reproduction and Durability of Health Inequalities”. *Social Theory & Health* 5.4, pp. 297–315.

— 2008. “Health and illness behaviour”. In: *Sociology as applied to medicine.* Ed. by G. Scambler. 6th ed. London: Saunders Elsevier, pp. 41–54.

— 2012. “Health inequalities”. *Sociology of Health & Illness* 34.1, pp. 130–46.

Serre, D. & Wagner, A.-c. 2015. “For a relational approach to cultural capital : a concept tested by changes in the French social space”. *The Sociological Review* 63, pp. 433–450.

Shankar, A., Klein, R. & Klein, B. E. K. 2006. “The Association among Smoking, Heavy Drinking, and Chronic Kidney Disease”. *American Journal of Epidemiology* 164.3, pp. 263–271.

Shelton, N. J., Birkin, M. H. & Dorling, D. 2006. “Where not to live: a geo-demographic classification of mortality for England and Wales, 1981-2000.” *Health & Place* 12.4, pp. 557–69.

Sheringham, J et al. 2009. “Monitoring inequalities in the National Chlamydia Screening Programme in England: added value of ACORN, a commercial geodemographic classification tool”. *Sexual Health* 6.1, pp. 57–62.

Simpson, E. H. 1949. “Measurement of diversity”. *Nature* 163, p. 688.

Singleton, A. D. & Longley, P. A. 2009. “Geodemographics, visualisation, and social networks in applied geography”. *Applied Geography* 29.3, pp. 289–298.

Small, M. & Feldman, J. 2012. “Ethnographic Evidence, Heterogeneity, and Neighbourhood Effects After Moving to Opportunity”. English. In: *Neighbourhood Effects Research: New Perspectives.* Ed. by M. Van Ham et al. Springer Netherlands, pp. 57–77.

Smith, D. M., Clarke, G. P. & Harland, K. 2009. “Improving the synthetic data generation process in spatial microsimulation models”. *Environment and Planning A* 41.5, pp. 1251–1268.

Smith, D. M., Pearce, J. R. & Harland, K. 2011. “Can a deterministic spatial microsimulation model provide reliable small-area estimates of health behaviours? An example of smoking prevalence in New Zealand”. *Health and Place* 17.2, pp. 618–624.

Smith, N. 2000. “What happened to class?” *Environment and Planning A* 32.6, pp. 1011–1032.

Solar, O. & Irwin, A. 2007. *A conceptual framework for action on the social determinants of health. Discussion paper for the Commission on Social Determinants of Health.* Geneva: World Health Organization.

Spiegelhalter, D. J. et al. 2002. “Bayesian measures of model complexity and fit”. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 64.4, pp. 583–616.

Spielman, S. E. & Thill, J.-C. 2008. “Social area analysis, data mining, and GIS”. *Computers, Environment and Urban Systems* 32.2, pp. 110–122.

Stansfeld, S. a. et al. 1997. “Work and psychiatric disorder in the Whitehall II Study.” *Journal of psychosomatic research* 43.1, pp. 73–81.

Starfield, B. 2001. “Basic concepts in population health and health care.” *Journal of Epidemiology and Community Health* 55.7, pp. 452–4.

Stephens, C et al. 1997. “Urban equity and urban health: using existing data to understand inequalities in health and environment in Accra, Ghana and Sao Paulo, Brazil.” *Environment and urbanization* 9.1, pp. 181–202.

Stringhini, S. et al. 2010. “Association of socioeconomic position with health behaviors and mortality.” *Jama* 303.12, pp. 1159–66.

Thrift, N. 2006. “Re-inventing invention: new tendencies in capitalist commodification”. *Economy and Society* 35.2, pp. 279–306.

Thrift, N. & French, S. 2002. “The automatic production of space”. *Transactions of the Institute of British Geographers* 27.3, pp. 309–335.

Turner, R. J. 2013. “Understanding Health Disparities The Relevance of the Stress Process Model”. *Society and Mental Health* 3.3, pp. 170–186.

Turrell, G. et al. 2013. “Can the built environment reduce health inequalities? A study of neighbourhood socioeconomic disadvantage and walking for transport.” *Health & Place* 19, pp. 89–98.

Twigg, L. & Moon, G. 2002. “Predicting small area health-related behaviour: a comparison of multilevel synthetic estimation and local survey data”. *Social Science & Medicine* 54.6, pp. 931–937.

Twigg, L., Moon, G. & Jones, K. 2000. “Predicting small-area health-related behaviour: a comparison of smoking and drinking indicators.” *Social Science & Medicine* 50.7-8, pp. 1109–20.

UKDH. 2007. *Guidance on Joint Strategic Needs Assessment.* Tech. rep. London: UK Department of Health.

Uprichard, E. 2009. “Introducing cluster analysis: what can it teach us about the case?” In: *The SAGE handbook of case-based methods.* Ed. by D. Byrne & C. C. Ragin. Los Angeles, London: SAGE, pp. 132–147.

Uprichard, E., Burrows, R. J. & Parker, S. 2009. “Geodemographic code and the production of space”. *Environment and Planning A* 41.12, pp. 2823–2835.

Valetas, M.-F. 2001. “The surname of married women in the European Union”. *Population & Sociétés* 367, pp. 1–4.

Van Ham, M. & Manley, D. 2012. “Neighbourhood effects research at a crossroads. Ten challenges for future research”. *Environment and Planning A* 44.12, pp. 2787–2793.

Van Ham, M. et al. 2012. “Neighbourhood Effects Research: New Perspectives”. English. In: *Neighbourhood Effects Research: New Perspectives.* Ed. by M. Van Ham et al. Springer Netherlands, pp. 1–21.

Vancampfort, D. et al. 2013. "Relationships between physical fitness, physical activity, smoking and metabolic and mental health parameters in people with schizophrenia". *Psychiatry Research* 207.1-2, pp. 25–32.

Veenstra, G. & Burnett, P. J. 2014. "A relational approach to health practices: towards transcending the agency-structure divide". *Sociology of Health & Illness* 36.2, pp. 187–98.

Vermeulen-Smit, E et al. 2015. "Clustering of health risk behaviours and the relationship with mental disorders". *Journal of Affective Disorders* 171, pp. 111–119.

Vinh, N. X. & Epps, J. 2009. "A Novel Approach for Automatic Number of Clusters Detection in Microarray Data Based on Consensus Clustering". *2009 Ninth IEEE International Conference on Bioinformatics and BioEngineering*, pp. 84–91.

Voas, D. & Williamson, P. 2001. "The diversity of diversity: a critique of geodemographic classification". *Area* 33.1, pp. 63–76.

Wacquant, L. 1999. "Urban marginality in the coming millennium". *Urban Studies* 36.10, pp. 1639–1647.

Wacquant, L. J. D. 2008. *Urban outcasts: a comparative sociology of advanced marginality.* Cambridge: Polity.

Ware, J. E. et al. 2002. *User's Manual for the SF-12v2 Health Survey with a supplement documenting SF-12 Health Survey.* Tech. rep. April. San Francisco: RI: QualityMetric Incorporated.

Webber, R. 2004. "Designing geodemographic classifications to meet contemporary business needs". *Interactive Marketing*, pp. 219–237.

— 2007. "The metropolitan habitus: its manifestations, locations, and consumption profiles". *Environment and Planning A* 39.1, pp. 182–207.

WHO. 1948. *Constitution of the World Health Organization.* London.

— 2004. *The global burden of disease 2004.* Tech. rep. Geneva: World Health Organization.

— 2011. *International Statistical Classification of Diseases and Related Health Problems. 10th Revision.* Tech. rep. Geneva: World Health Organization.

— 2014. *Global status report on noncommunicable diseases 2014.* Tech. rep. Geneva: World Health Organization.

WHO & UN-HABITAT. 2010. *Hidden cities: unmasking and overcoming health inequities in urban settings.* Tech. rep. Geneva: World Health Organization, UN-HABITAT.

Wilches-Chaux, G. 1989. *Desastres, Ecologismo, y Formación Profesional.* Popayán: Servicio Nacional de Aprendizaje (SENA).

Wilkinson, R. & Pickett, K. 2010. *The Spirit Level: Why Greater Equality Makes Societies Stronger.* London: Bloomsbury Publishing.

Williams, S. J. 1995. "Theorising class, health and lifestyles: can Bourdieu help us?" *Sociology of Health and Illness* 17.5, pp. 577–604.

Winney, B. et al. 2012. "People of the British Isles: preliminary analysis of genotypes and surnames in a UK-control population." *European Journal of Human Genetics* 20.2, pp. 203–10.

Wirth, L. 1938. "Urbanism as a Way of Life". *American Journal of Sociology* 44.1, pp. 1–24.

Wu, Y. et al. 2015. "Fruit and vegetable consumption and risk of type 2 diabetes mellitus: A dose-response meta-analysis of prospective cohort studies". *Nutrition, Metabolism and Cardiovascular Diseases* 25, pp. 140–147.

Zannas, a. S. & West, a. E. 2014. "Epigenetics and the regulation of stress vulnerability and resilience." *Neuroscience* 264, pp. 157–70.

# Appendix A: Data access and research ethics

This Appendix summarised details on data access applications and ethical approval for the datasets used in this thesis.

**Census, UK, 1881** data was obtained from the UK Data Service and used under the terms of Crown copyright held jointly with the Genealogical Society of Utah and the University of Essex. The usage is registered at the UK Data Service with reference 73249. Census boundary data were obtained from 'A vision of Britain through time' [http://www.visionofbritain.org.uk] and used under Creative Commons licensing. Prof Humphrey Southall and Dr Paula Aucott, University of Portsmouth, kindly provided a gazetteer for 1881 parishes in Scotland upon my request.

**Census, UK, 2001 and 2011** data including Census boundary data were obtained and used under the terms of the Open Government Licence from the following sources:

- Office for National Statistics, 2011 Census: Aggregate data (England and Wales) and Digitised Boundary Data (England and Wales)
- National Records of Scotland, 2011 Census: Aggregate data (Scotland) and Digitised Boundary Data (Scotland)
- Northern Ireland Statistics and Research Agency, 2011 Census: Aggregate data (Northern Ireland) and Digitised Boundary Data (Northern Ireland)

**Enhanced Electoral Roll 2007 and 2011** data were purchased from CACI Limited by UCL and its use is registered with the departmental Data Protection Officer.

**Hospital Episode Statistics (HES)** 1999/00-2012/13 were applied for from 2013 to 2015 with the outcome still pending. The intention was to obtain HES at a high spatial and temporal resolution with surname classification of patients. Ethical approval was obtained from the NHS Research Ethics Committee Bromley on 19 September 2013 (reference 13/LO/1355) and the Department for Health IG toolkit version 11 was passed on 30 August 2013. The Health Research Authority (Confidentiality Advisory Group) decided on 16 July 2015 that section 251 support was not required (reference 15/CAG/0159). Since then data access has been applied for at the NHS Health and Social Care Information Centre (HSCIC) with a decision still outstanding. The project was registered with UCL Data Protection Office (reference Z6364106/2013/04/37). Since the data did not arrive on time, HES data 2008/09 from a previous UCL project was used with the following licence and registration details: HSCIC NIC-33864-6226N/RU183 with ethical approval obtained from NHS Research Ethics Committee Bromley (reference 06/Q0705/2).

**People of the British Isles sample** was obtained as part of the project 'The People of the British Isles' co-investigated by Paul Longley. Ethical approval to conduct the research was obtained from Research Ethics Committee Leeds West (reference 05/Q1205/35) and the project is registered with UCL Data Protection Officer (reference Z6364106/2013/04/16).

**Understanding Society** microdata were obtained at two levels in addition to the standard end user licence. Special licence for inclusion of 2001 Lower Layer Super Output Areas (LSOA) of respondents' residences as well as the coding of Output Area Classification 2001. Secure Data Service was obtained by special user agreement and research accreditation on 24 September 2013. The usage reference at the UK Data Service is 68450 currently valid until 16 April 2016. The data use has been registered with the UCL Data Protection Officer with (reference Z6364106/2013/06/53).

The software for data management and processing used in the thesis are the open source products R [www.r-project.org] and PostgreSQL [www.postgresql.org].

# Appendix B: Glossary

**ANOVA** (used in chapters five and eight) stands for "Analysis of Variance" and is a statistical technique to compare population averages. For example, oneway ANOVA can be used to determine whether the average income between two or more groups is statistically significant based on their within-group variance of incomes compared to the between-group variance of incomes. A so-called F test is then employed to see whether the within/between variance ratio is significant. Tukey post-hoc tests are used to determine whether differences between any pair of groups is significant.

**Bayesian statistics** (used in chapters five and eight) has come to denote an approach to statistics that offers an alternative view of randomness and probability. In contrast to conventional, frequentist statistics, which regards randomness as inherent in nature (analogous to an ontological notion of chaos), Bayesian statistics views randomness and probability in epistemic terms, that is the degree of certainty we can ascribe to a phenomenon in view of imperfect measurement.

Frequentist statistics typically takes a Normal distribution (or related one from the Gaussian family) as the 'natural' distribution of repeated measurements. It postulates that if we were to measure height in a sample of a population and calculated the average height, the average would naturally vary each time we repeat the sampling. The degree of variation around the mean can be summarised as the standard error. Calculated means of new samples would considered to be 'unusual', if they fell further from the previous mean than the standard error. Probability reflects the likelihood to find a certain value of mean height given the true mean of the population. It reflects the proportion of times, we would expect a certain value of mean heights to be measured in repeated measurements (the frequentist notion).

Bayesian statistics, on the other hand, employs a notion of conditional probability. It requires us to specify what kind of distribution we would expect the data of heights to follow in the first place (called prior distribution). This distribution is then compared to the actual distribution in the data (the distribution of heights in the sample). The probability indicates the degree of certainty that the measurement results from the joint distribution (called posterior distribution) of the prior information and the distribution of the data. The choice and specification of the prior distribution can have a profound impact on the posterior distribution and thus the probabilities of measurements. If no prior information about a distribution of measurements is available, often a 'vague' prior is used, which broadly defines that each value in a given range of values is equally likely. This gives more weight to the distribution of the actual measurements in the posterior distribution.

In the thesis, Bayesian statistics is applied in the estimation of area-level rate ratios of various health indicators (self-rated health, incidence of infectious diseases etc). The data is assumed to be Poisson distributed and in the spatial-structural model, the observed rate of neighbouring areas is taken as additional evidence and a error term is assumed to reflect other unmeasured factors. Both the value of neighbouring areas and the error term are defined to be Normal distributed with a wide standard deviation to make them vague priors. These distributions along with the Poisson distribution and the observed data form the posterior distribution of the model. Based on this distribution, it is then possible to determine the probability that an area incidence rate ratio is above average given the joint evidence the prior beliefs and the data.

In a frequentist view, the probability (p-value) indicates the likelihood that the null hypothesis – an area rate ratio is above average – cannot be rejected, reflecting measured as the percentage of repeated experiments in which we would detect false positives, that is the area risk ratio is different from one although in reality it is not.

Desrosières [1998] suggests that the choice of the framework – Bayesian or frequentist – depends on the general scientific orientation of the researcher. Frequentist statistics originates from positivist approaches to interpreting quantitative information, conflating diversity to a notion of the 'average man' (or area or other subject of interest). In contrast, an epistemic, Bayesian framework corresponds to scientific realism in which the phenomena we observe are always incomplete and depend on the information and measurements available to us.

**Data sparseness** (mentioned in chapter five) refers to a phenomenon by which zero or a low number of observations occur very often. For example, in disease estimation, some areas may only have a low number of cases per year and therefore are prone to annual fluctuation and high level of statistical uncertainty. 'Bayesian smoothing' [see Bayesian statistics] is one way to address this issue and has been used in spatial-structural models of estimating incidence rate ratios in the thesis.

**Degrees of freedom** (widely mentioned in the thesis) in statistics is a short form of saying "degrees of freedom for error". The term relates to the number of values that can vary independently and thus constitute a source of error. In the thesis, the term is important in relation to $\chi$-squared tests when viewing the strength of association between two classifications. For example, we may classify people by age and tenure to identify whether there is any association between the two variables. If we had five age-bands and three types of tenure (rented, owner-occupied, mortgage), we may look at a two-way table of five rows (age-bands) by three columns (tenure). This makes 15 table cells. Given the total number of observations that fall into each tenures and each age band, the table cells are not independent of each other. If we had more observations in rented accommodation, we would necessarily have fewer observations in the other two tenure categories. Indeed, if we knew the number of observations in rented accommodation and on mortgage payment, we could simply calculate the number of owner-occupiers. Thus a three category variable has two degrees of freedom. A five category variable has four degrees of freedom. A two-way table of five by three cells has two times four equals eight degrees of freedom; we only need to know the population of eight table cells the infer the remaining seven. Degrees of freedom are important to determine the variation of values that we would expect if observations would be distributed 'randomly' between age-bands and tenure without any association. This forms the basis to determine whether two categorical variables are significantly associated.

**Dimension reduction** (widely used in the thesis) is a way to reduce vast information of a set of variables to fewer, underlying dimensions. There are different techniques to do this, which can be grouped by the information they process to achieve dimension reduction: one group processes pair-wise dissimilarities between observations, the other group processes covariance or correlations between variables.

**Multi-dimensional Scaling (MDS)** translates a matrix of pair-wise dissimilarities into coordinates. A simple example is, if we were to measure geographical distances between cities and we had a pair-wise matrix of distances between these cities, it is possible to translate the distances into coordinates and hence a position in two-dimensional space. The coordinates are new values that may indicated certain underlying dimensions. In the example of cities, it could be orthogonal geographical directions in two-dimensional space. The coordinates (or scores) can then be used in further analysis. In theory, as many dimensions are possible as there are observations. MDS is used in the thesis to estimate trend surfaces of isonymy and coancestry in geographic space.

**Principal Components Analysis (PCA)**, is similar to MDS, but instead of dissimilarities between observations, it processes covariances (or correlations) between variables. The technique measures intercorrelations between variables and thus identifies underlying components or dimensions. For example, we may measure for different cities their composition of employment by sector. We may find that across many cities banking is correlated with law firms, business services and financial advice and, independent from that, garment industry, commercial stores, sales and design. The first dimension indicates financial activity, the second dimension indicates fashion. In this way the information can be used to identify different functions across cities. The dimensions can be ranked by their importance which is derived from the proportion of which each dimension accounts for the variance of the entire data. The so-called eigenvalue is indicative of this proportion. It is also possible to compute dimension or component scores which indicate the degree to which each city reflects each of those components. City $A$ may indicate high affinity to financial activity, but little to fashion, city $B$ may indicate both and city $C$ only fashion. The score is the sum of standardised variable values weighted by variable loading (correlation) with each component. The variance of a variable that cannot be accounted for by any of the component is called uniqueness of a variable. In theory, there can be as many components as there are variables, but only those with an eigenvalue greater than one are typically considered as important, since a component with an eigenvalue below one indicates that it accounts for less variance than a single variable on average.

**Interdecile range, interquartile range and interseptile range** (widely used in the thesis) measure the distribution or variation of a sample of measurements, each emphasising different positions of data points. The interquartile range measures the value distance between the lower quartile and the upper quartile of a sample of measurements when it is ordered by its value. The interdecile and interseptile ranges do the same for the lower decile (tenth) and septile (seventh) respectively. We may, for example, order a sample of 100 heights taken from a population by the value of height itself. The interquartile range determines the range of values between the 75th and the 25th observation. This indicates a degree of the spread of heights and can also serve as a measure of inequality, e.g. when, instead of height, income or health is measured. Boxplots (used in chapters five, six and eight) visualise the interquartile range and, in addition, identify extreme outliers. The value where the second quartile touches the third is called the median, an estimate of the central summary value that is less sensitive to outliers than the mean.

**Interquantile ratio** (used in chapter five) determines the ratio between two chosen percentiles. It is equivalent to an interquantile range (see interdecile range), but instead of the range it calculates the ratio between the two percentiles. The interquantile ratio is in the thesis to estimate health inequalities by calculating the ratio between the 95th and 5th percentile of a health indicator.

**Inverse Distance Weighting (IDW)** see *spatial interpolation*.

**Kernel Density Estimation (KDE)** (used in chapters four and seven) is a statistical technique to estimate the density of observations in statistical (or geographical) space. Let us say we record the genotype for 2,000 individuals in the UK as well as the geo-coordinates of their residents. We receive a discrete sample of points distributed in space. We may now determine the density of these points any location in space. The simplest way is to count the number of points at each location which results in many zero values and a few ones, provided that none of the points shares the same location. If we begin to take into account the density values in close proximity, the density estimates become smoother. The distance and degree to which proximate estimates are considered in each estimate is controlled by the bandwidth: the higher it is, the more sensitive is each estimate to surrounding values that are further apart and the smoother becomes the distribution. The technique is useful in geography to generalise and highlight spatial distributions of discrete phenomena.

**Kriging** see *spatial interpolation*.

**Multi-Dimensional Scaling (MDS)** see *dimension reduction.*

**Principal Components Analysis (PCA)** see *dimension reduction.*

**Spatial autocorrelation** refers to the phenomenon by which observations in close proximity are more similar than observations at distant locations. Spatial autocorrelation is often treated as noise that needs to be controlled for so as to avoid overestimation of relationships. For example, the ecological association between crime rate and unemployment may be estimated in a regression model and summarised by a regression coefficient. Failing to acknowledge that unemployment and crime 'naturally' cluster spatially, that is neighbouring areas will have more similar unemployment and crime levels than more distant areas, would overestimate the strength of the association.

**Spatial interpolation** (used in chapter four) is a procedure that creates a spatial, continuous value surface from a set of spatially located data points. The objective is to infer values at any location from observed values at locations that are proximate. In the thesis, spatial interpolation is applied to infer continuous spatial surfaces of MDS scores [see dimension reduction] and regression residuals. Two measures have been applied for spatial interpolation:

**Inverse Distance Weighting (IDW)** is a method that at each position of a continuous spatial grid weights the observed values by their proximity to the position of estimation. The inverse of a squared Euclidean distance is often used as distance decay factor, but any power can be specified. IDW has been used to interpolate MDS scores of area isonymy based on a inverse linear weighting of distances to zonal centroids.

**Kriging** is a model-based spatial interpolation method, that is it supposes a statistical model of value decay by distance from each observation and applies it each position on a spatial grid. The statistical models are specified as variograms, which measures correlation between the absolute distance in values of each pair of observations relative to the geographic distance between those observations. In a fully autocorrelative model, the correlation between value distance and geographic distance is one, but in practice it is mostly observed that the relationship weakens as distance increases. In other words, there is typically a maximum limit of absolute distance between observation pairs. This is taken into account in a variogram; the function of the relationship between value distance and geographic distance is asymptotic to the maximum distance, which in the context of Kriging is called 'sill' or 'partial sill'. Another parameter is 'range', the maximum distance from which another observation is considered to be without influence. A final parameter is 'nugget', which defines the value fluctuation at very close locations that is considered to be zero. This parameter is useful, if some measurement error at close locations is to be expected. In contrast to IDW, Kriging is based on a statistical model and may be particularly useful where we deal with observations that are unevenly distributed in space.

**Tukey post hoc test** see *ANOVA*.

**Z score standardisation** (used widely in the thesis) refers to a procedure to translate variables that are measured on different measurement scales to a common scale such that they receive equal weight in any statistical follow-up procedure. For example, height, weight and blood pressure are measured on different scales with different units. If each of those were measured for a sample of 100 people and subsequently processed in a cluster analysis, the variables with the largest values (in this case blood pressure, assuming height is measured in metres and weight in kilograms) would have the largest impact on the clustering because it occupies a wider range in statistical space. Hence standardisation that adjusts for the different scales and also takes into account the variable-specific variation is necessary. Z score standardisation does just that in two steps: first, it subtracts each observed value of each variable by the variable's sample mean (centering) and, second, it divides it by the sample standard deviation (scaling).

For example, if a blood pressure of 105 is observed for case $A$, the sample mean is 130, and the sample standard deviation is 20, the resulting standardised value would be $(130 - 105)/20 = -1.25$, which indicates that case $A$'s blood pressure lies 1.25 times the standard deviation below the mean. If case $A$ weights 55 kg, the sample mean is 76 kg and the standard deviation is 14 kg, the standardised value for weight would be $(76 - 55)/14 = -1.5$. Hence, two very different scales are translated into comparable scales, which determine that case $A$ is a little bit more different in terms of weight than in terms of blood pressure than the sample in total.

**$\chi$-squared tests** see *degrees of freedom.*