Applications of Open Data in health care planning and assessing health inequalities

Jens Kandt[1,1a], Daniel J. Lewis[2], Paul A. Longley[1,1b]

[1]Department of Geography, University College London, Gower Street, London WC1E 6BT.
Tel. +44 (0)20 7679 0500, Fax +44 (0)20 7679 0565, URL www.geog.ucl.ac.uk
[1a]email: j.kandt.12@ucl.ac.uk
[1b]email: p.longley@ucl.ac.uk

[2] Department of Social and Environmental Health Research, London School of Hygiene and Tropical Medicine, 17-19 Tavistock Place, London, WC1H 9SH
email: daniel.lewis@lshtm.ac.uk

KEYWORDS: geodemographics, health care, health inequalities, open data, strategic planning

words: 1450

1. Introduction

Local authorities in the UK face the requirement to tackle health inequalities at a strategic level – and the recently introduced Joint Health and Well-being Strategies to be prepared by proposed local authority-level Health and Well-being Boards are expected to play a crucial part in this undertaking [cf. section 193, Health and Social Care Act 2012]. While the consultation of relevant data is at the heart of preparing these strategies, local authorities face the challenge of finding appropriate data sources to define summary measures that accurately describe the state of health of the local population, and highlight health inequalities at an appropriate spatial scale. But given the current political emphasis on austerity and increased efforts to cut spending, it can be argued that there is urgency to better assess local strategic needs, and so devise health care strategies and related policy interventions as to monitor and target these needs efficiently [cf. Lewis and Longley 2012, Smith et al. 2011].

The purpose of this paper is to provide an overview of currently available government Open Data, and discuss their potential extended application in the context of tackling health inequalities and health care planning. Particular attention will be paid to scale-based applications that can uncover trends in health at a range of geographical scales through the derivation of appropriate summary measures of health.

2. Open Data and planning for health and well-being

The recently enacted Health and Social Care Act 2012 makes clear that needs assessment, carried through Joint Strategic Needs Assessments (JSNAs), is a core element in devising Joint Health and Well-being Strategies. The overall objective of JSNAs is to improve health and well-being outcomes, which include the attenuation of persistent health inequalities [cf. NHS 2011]. Local authorities are expected to make use of a wide range of data in order to assess the local status quo with respect to health outcomes and inequalities, and to set out policy interventions that fulfil these objectives.

While, to date, little is understood as to what types of policy effectively address health

inequalities, the use of government Open Data may offer potential to better assess local strategic needs and inform strategic health care planning and related policy interventions. The NHS and the Department of Health offer some overall guidance on how to carry out JSNAs [DH 2012b, NHS 2011]; yet little is said about the potential role of Open Data on health, demographic and socio-economic as well as environmental characteristics.

Relevant Open Data can range from death registries collected by the Office of National Statistics (ONS), hospital episode statistics collected and maintained by the NHS, as well as data from surveys such as 'Understanding Society' or the Health Survey for England. While these are vast datasets collected for a variety of purposes, they can be of strategic interest to local authorities once they are viewed in conjunction with precise geographical information. The linkage of these datasets with socio-demographic and economic data of the population, notably geo-demographic classifications, can be of use in strategic health care planning - especially when this is understood as an application that can highlight different aspects of the current state of health and well-being such as demand for and utilisation of health care alongside direct strategic needs of the population.

In addition, extensive datasets on health and geodemographic characteristics may allow authorities to view a range of summary measures describing health and well-being in juxtaposition. Life expectancy has been a very popular indicator in the past - but its usefulness in policy has been limited, and a range of alternative indicators have been proposed [DH 2012a, LHC 2011, Marmot 2010, Starfield 2001, Landon 1996].

The power of linking Open Data through geographical referencing arises because users may view trends at multiple geographical scales, ranging from small area differences to regional and national health trends [cf. Manley et al 2005; Longley 2003]. A scale-based approach addresses the well-known MAUP issue and links to geodemographics may reduce the problem of ecological bias, since geodemographic classifications are by definition based on within-group homogeneity [cf. Petersen et al 2011; Abbas et al 2009; Openshaw and Blake 1995]. Hence a full-fledged overview of relevant Open Data and potentials for linkage may help to gauge future directions in strategic planning for health and well-being.

3. Strategic applications of Open Data in health

What, then, are the datasets that can be of strategic use in tackling health inequalities and promoting pro-active health care planning? This question will be addressed in the following with a view to linking datasets with explicit information on health and datasets containing contextual information such as geodemographic classifications.

3.1. Data on health

Open Data on health can be derived from two sources: records collected as part of administrative routine functions by various agencies and data from government-sponsored surveys. Table 1 shows a list of the most relevant datasets on health that are available for England, if not for the whole of the UK.

Table 1. Open Data on health

Dataset	Health aspect	Contents	Potential health measures	Lowest spatial level
Administrative data				
Registration of life events (General Register Office/ONS)	Mortality	Vital events (births and deaths) with demographic attributes	Life expectancy, age-standardised mortality ratios, pre-mature mortality, avoidable mortality	SOA (Super Output Area)
UK Census (ONS)	Self-reported health	Limiting long-term illness and demographic and socio-economic characteristics	DFLE (disability-free life expectancy)	OA (Output Area)
Hospital episode statistics (NHS)	Morbidity	Primary diagnosis, referring GP, patient characteristics	Prevalence and incidence rates, health expectancy (e.g. DALYS)	LSOA (Lower Super Output Area)
Admitted patient care (NHS)	Morbidity	Diagnosis, referring GP, patient characteristics	Prevalence and incidence rates, health expectancy (e.g. DALYS)	address
Prescriptions (NHS)	Morbidity, drug uptake	Drug prescription with classification code, prescribing GP	Prescriptions per person as indication of disease incidence	SOA (Super Output Area)
Benefits for diseases (Dept of Works)	Morbidity	Working-age people claiming incapacity benefit or severe disablement allowance with medical condition code	Prevalence and incidence rates	SOA (Super Output Area)
Special disease registers (e.g. Cancer Registry)	Morbidity	Various records for special diseases	Disease incidence depending on nature of dataset	Various
Survey data				
Understanding Society (ESRC/ISER)	Morbidity	Health outcomes, health-related behaviour, lifestyle, demographics, economic situation, attitudes	Estimates of disease prevalence and incidence, risk factors	Local authority
Health Survey for England (DH/NHS)	Morbidity	General health, health-related behaviour, lifestyle aspects; special topics	Estimates of disease prevalence and incidence, risk factors	Local authority
Specific surveys (DH/NHS)	Morbidity		Disease prevalence and – if longitudinal – incidence	Various

More information see websites: NHS Information Centre on health and social care, ONS Census 2011, Understanding Society

Administrative data are a useful resource to draw on, as they are comprehensive and, in principle, unbiased. The combination of unbiased data and consistent adherence to the disease coding recommended by the International Classification of Diseases (ICD) permits a wide range of applications, including more detailed analysis of population sub-groups or the incidence of specific diseases. All datasets include either information on the associated individual, including residential address, or reference the patient's GP surgery. This allows estimates and derivation of summary measures on a high spatial resolution. Users should be alert to the fact that some datasets allow for estimates of either disease incidence or prevalence. Both aspects are important in the context of strategic needs assessment – and

datasets will need to be selected accordingly.

Survey data, on the other hand, may be used to investigate generic trends across defined types of areas (e.g. ONS groups) or on a coarse geographical scale. The sample size and sampling design determines the extent to which trends can be viewed more locally. Large-scale surveys data may, however, offer insight into local trends when geodemographic classifications are appended. Record linkage can combine local specificity with longitudinal and subjective information on physical and mental health.

3.2. Data on context: geodemographic classifications

Geodemographic classifications can capture contextual factors pertaining to the social and physical environment people live in. Classification algorithms can draw on a range of datasets and differ according to their purpose and substantive focus (see table 2).

Table 2. Geodemographic classifications relevant to health

Dataset	Focus	Contents	Link to health	Lowest spatial level
Public				
Output Area Classification (ONS)	Generic	Demographics, household composition, housing, socio-economics, employment	Social determinants	OA (Output Area)
Index of Multiple Deprivation (DCLG)	Deprivation	Income, employment, health, education & training, barriers to housing & services, living environment, crime	Mortality, disease prevalence, access to health care, social determinants	LSOA (Lower Super Output Area)
Health Poverty Index (NHS)	Health poverty	Root causes, intervening factors, situation of health	Mortality, disease prevalence, access to health care, social determinants	Local Authority
Commercial				
ACORN (CACI)	Lifestyles, consumption	Housing, employment, family, lifestyle interests, media, biographical events	Social determinants, risk factors	Post codes
healthACORN (CACI/Kantar Healthcare)	Health, risk factors	Medical condition, lifestyle, eating behaviour	Health outcomes, risk factors	OA (Output Area)
Mosaic UK (Experian)	Lifestyles, consumption	Demographics, property characteristics, socio-economics, consumption, location, financial measures	Social determinants, health outcomes	Post codes, addresses

More information see DCLG [2011], Dibben et al. [2008], Experian [2009], Vickers and Rees [2007] as well as website: CACI

The Output Area Classification (OAC) is based on census data, and can be used in assessing general health needs of the population based on demographics. While this dataset directly links to social determinants of health in the context of small areas, its application in health may be limited, as it only captures a few aspects of deprivation. The Index of Multiple Deprivation (IMD) offers a more comprehensive picture by including access to services, education and prevalence of crime. The approach that is most closely aligned with a social determinants perspective is the Health Poverty Index (HPI) – a multi-level index that

includes broader regional trends such as economic development or resource allocation to health and social care alongside local factors. It draws on a diverse set of data and has to date only been produced for the level of local authorities. Nevertheless, it may offer a first step towards providing an 'upstream' picture of future health care needs – a direction that has been called for in geodemographics [cf. Singleton and Longley 2009].

Commercial providers have also developed classifications for marketing purposes. Although they do not count as Open Data, they can be regarded as resources that may provide useful contextual information with subjective domains such as lifestyles and attitudes and a presumably stronger reflection of behavioural risks.

4. Outlook

After a detailed review of the above-described datasets and their application in health care planning, this study will proceed with an assessment of the appropriateness and usefulness of various Open Datasets as well as the spatial resolution that can reasonably be obtained. The datasets pertaining to health on the one hand and geodemographic classifications on the other hand will be linked and analysed with a view to deriving useful summary measures in describing health and well-being needs on different spatial scales. It will be shown that certain approaches only provide a generic picture of health trends across population groups with certain socio-demographic, economic and contextual characteristics, while other approaches that utilise different datasets will uncover spatial trends in health at a geographically explicit and precise level.

The review will conclude with a discussion of what use Open Data and derived measures can be in strategic health care planning and interventions, what steps are needed to turn the data into information at a spatial scale that reasonably permits local authorities to fulfill their task to strategically provide for health care and address health inequalities, and how this information can be incorporated into the preparation of Joint Health and Well-being Strategies.

References

Abbas, J., Ojo, A., Orange, S. [2009]. Geodemographics – a tool for health intelligence?. *Public Health* 123, e35-e39.

DCLG [2011]. The English Indices of Deprivation 2010. London: Department for Communities and Local Government.

Dibben, C, Watson, J., Smith, T., Cox, M., Manley, D., Perry, I., Rolfe, L., Barnes, H., Wilkinson, K., Linn, J., Liu, L., Sims, A., and Hill, A. [2008]. *The Health Poverty Index.* Leeds: The NHS Information Centre.

DH [2012a]. *Healthy lives, healthy people: Improving outcomes and supporting transparency. A public health outcomes framework for England 2013-2016.* Department of Health policy document.

DH [2012b]. *Joint Strategic Needs Assessments and Joint Health and Wellbeing Strategies – draft guidance*. Department of Health consultation/discussion document.

Experian [2009]. Optimise the value of your customers and locations, now and in the future. Online at www.experian.co.uk/business-strategies/mosaic-uk.html, Nov 2012.

Landon, M. [1996]. Intra-urban health differentials in London – urban health indicators and policy implications. *Environment and Urbanization* 8(2), 119-128.

Lewis, D. and Longley, P.A. [2012]. Patterns of patient registration with primary healthcare in the UK National Health Service. *Annals of the Association of American Geographers* 102, 1–11.

LHC [2011]. *Fair London, Healthy Londoners?* London: London Health Commission.

Longley, P.A. [2003]. Geographical Information Systems: developments in socio-economic data infrastructures. *Progress in Human Geography* 27(1), 114-121.

Manley, D., Flowerdew, R., Steel, D. [2005]. Scales, levels, processes: Studying spatial patterns of British census variables. *Computers, Environment and Urban Systems* 30, 143–160.

Marmot, M. [2010]. *Fair Society, Health Lives. The Marmot Review.* London: The Marmot Review.

Marmot, M. [2005]. Social determinants of health inequalities. *The Lancet* 365, 1099-1104.

NHS [2011]. The joint strategic needs assessment. A vital tool to guide commissioning. *NHS Confederation Briefing* 221.

Olatunde O, White C and Smith MP [2010]. Life expectancy and disability-free life expectancy estimates for Middle Super Output Areas England, 1999-2003. *Health Statistics Quarterly* 47, 33-65.

Openshaw, S. and Blake, M. [1995]. Geodemographic segmentation systems for screening health data. *Journal of Epidemiology and Community Health* 46(S2), S34-S38.

Petersen, J., Gibin, M., Longeley, P.A., Mateos, P., Atkinson, P., Ashby, D. [2011]. Geodemographics as a tool for targeting neighbourhoods in public health campaigns. *Journal of Geographical Systems* 13, 173-192.

Singleton, A.D. and Longley, P.A. [2009]. Geodemographics, visualisation, and social networks in applied geography. *Applied Geography* 29, 289-298.

Smith, M.P., Olatunde, O. and White, C. [2011]. Disability-free life expectancy: comparison of sources and small area estimates in England, 2006–08. *Health Statistics Quaterly* 50, 1-39.

Starfield, B. [2001]. Basic concepts in population health and health care. *Journal of Epidemiology and Community Health* 55, 452-454.

Vickers, D. and Rees, P. [2007]. Creating the National Statistics 2001 Output Area Classification. *Journal of the Royal Statistical Society Series A* 170(2), 379-403.

Consulted websites

CACI, acorn user guide: www.caci.co.uk/acorn2009/CACI.htm, Nov 2012.

NHS Information Centre for health and social care: www.ic.nhs.uk, Nov 2012.

ONS Census 2011: www.ons.gov.uk/ons/guide-method/census/2011/index.html, Nov 2012.

Understanding Society: www.understandingsociety.org.uk, Nov 2012.

Biographies

Jens Kandt is a doctoral researcher at the Department of Geography, University College London, where his work focusses on health equity in cities. He is particularly interested in applying GIS, geo-demographics and other quantitative methods in the context of comparative urban research.

Daniel Lewis is a postdoctoral research fellow at London School of Hygiene and Tropical Medicine. Using GIS, Daniel's work on the environment and health contributes to a large longitudinal study of young people in East London which is evaluating the impact of the Olympic regeneration on them and their families.

Paul Longley's research interests concern the use of GIS and quantitative methods in urban analysis. He is a co-editor of Environment and Planning B and co-author of the book 'Geographic Information Systems and Science' and c. 150 other refereed publications.