

Socio-demographic data sources for monitoring locality health profiles and geographical planning of primary health care in the UK

Edgar Samarasundera^{1,2}, David Martin³, Sonia Saxena⁴ and Azeem Majeed⁵

¹Honorary Research Officer, Department of Primary Care and Social Medicine, Imperial College London, London, UK

²Research Associate in GIS, School of Geography, Geology and the Environment, Kingston University, Surrey, UK

³Professor of Geography, School of Geography, University of Southampton, Southampton, UK

⁴Clinical Senior Lecturer and Consultant in Primary Care, Department of Primary Care and Social Medicine, Imperial College London, London, UK

⁵Professor of Primary Care and Social Medicine, Department of Primary Care and Social Medicine, Imperial College London, London, UK

Aim: The aim of this article is to provide UK-based primary health care research and development workers with a review of the current range of published, aggregated socio-demographic indicators that can be combined with health and health care datasets, for the purposes of monitoring locality health profiles and planning primary health care. Non-UK readers should nevertheless find the review of some relevance to their own national contexts. **Background:** There is an increasing range of resources available for such purposes and many of these datasets are equally useful outside of geographic work. The 2001 census introduced important changes to what routine data are available, as will the 2011 census. These changes have been paralleled by developments in the availability of socio-demographic indicators and the increasing popularity of geographic information systems. Health data can now be combined with those from socio-demographic more efficiently to produce what are termed value-added datasets. **Methods:** We review recent and planned developments in key data sources currently available in the UK and examine they can be used to monitor inequalities in primary health care inequalities and their role in the integration of primary health care needs mapping and forecasting with the spatial planning of areas undergoing regeneration. **Conclusions:** Recent and planned developments in the availability of both socio-demographic datasets in tandem with parallel developments in spatial technologies have provided a flexible, potent geographical methodology for primary health care research and development. The current consultation process for the 2011 census provides those involved with primary health care research and development an opportunity to influence future developments.

Key words: census; deprivation; geodemographics; mapping; socio-economic

Received 22 March 2008; accepted 9 November 2009; first published online 14 May 2010

Introduction

Correspondence to: Mr Edgar Samarasundera, Department of Primary Care and Social Medicine, Imperial College London, Reynolds Building, St Dunstan's Road, London W6 8RP, UK. Email: e.samarasundera@imperial.ac.uk

Geographical analysis and planning of population-level health and health services has a long history, for example, monitoring equity in health

© Cambridge University Press 2010

outcomes and access to health services. Applications include, for example, the estimation of general practice (GP) catchment areas (eg, Jenkins and Campbell, 1996), population accessibility to primary health care facilities (Field, 2000; Luo and Wang, 2003; Jordan *et al.*, 2004a; Saxena *et al.*, 2007; Bottle *et al.*, 2008), monitoring socio-demographic variations in health screening uptake (Muggli *et al.*, 2000) and monitoring of primary health care inequalities (Sigfrid *et al.*, 2006; Strong *et al.*, 2006). Indeed much literature appearing in mainstream medical, epidemiological and health services journals are well within the scope of health geography (eg, Damiani *et al.*, 2005; Bowling *et al.*, 2006; Congdon, 2006). The last ten to fifteen years have seen substantial advances in data availability, methods and relevant technologies for these purposes, along with corresponding analytical boons and pitfalls.

Monitoring locality health profiles and planning population-level health care require routine, aggregated datasets, which allow the mapping and analysis of population variations in health status and use of health services. In the UK, there is a wide array of relevant published data sources, including Hospital Episode Statistics (HES), GP registers, census data, deprivation indicators and socio-economic classification schemes. With respect to the numerous sources of health and health care data that can be used for population studies and these have been described in depth elsewhere (Thiru *et al.*, 2003; Kalthenthaler *et al.*, 2004; Millett *et al.*, 2005; Gnani and Majeed, 2006). Hence, we focus on non-medical data sources that can be combined with such datasets to monitor locality health profiles and plan primary health care services effectively. Table 1 provides guidance on where to obtain data, products and services. In this review paper recent, current and future developments are discussed, with an emphasis on data issues, to provide not only an introduction for those new to the field but also a professional update for those already working in the domain.

Census data

The population census is the most commonly used source of socio-demographic data. In the UK, simultaneous decennial censuses are conducted

by the three national statistical agencies, the Office for National Statistics (ONS) in England and Wales, the General Register Office Scotland and the Northern Ireland Statistics and Research Agency (NISRA); similar censuses are carried out in a number of countries such as Australia, the Republic of Ireland and USA, though some countries including Finland and the Netherlands rely on administrative data in order to derive population statistics rather than carrying out a periodic census (Martin, 2006). The three UK censuses cover a broad range of demographic, social and economic topics, with most questions being consistent across the UK. Census results are made available as a variety of output products, which include individual responses aggregated across geographical areas. For the most recent census, that of 2001, the most geographically detailed products are the Key Statistics and Census Area Statistics produced for census units called Output Areas (OAs) each containing a mean of around 300 people; these OAs are aggregated by each respective census agency to form larger census units in a geographical hierarchy moving from the local to the national (Martin, 2002). Cross-tabulations of census results, known as Standard Tables, are available for the various tiers of government organisation from wards upwards. Since 2004 there has been data, based on the 2001 censuses, published on daily commuting patterns at the local authority level, including net flows by age band, enabling estimation of daytime population numbers at the local authority level. While net commuter patterns are not relevant to many health-related contexts they are relevant to some issues, for example, environmental exposures relating to respiratory diseases. The 2001 censuses also included two new questions that have relevance to health services planning, adding to the existing one regarding long-term illness: one on general health and another on the provision of unpaid care for people with limiting long-term illness and disabled people (Dixie and Dorling, 2002).

Although the ten years between censuses means that data are often out of date towards the end of the period, the highly detailed census counts are unmatched by any other data source, both in terms of small geographical areas and numerous cross-tabulated variables. It is this characteristic that causes census data to remain as

Table 1 A selection of internet sites providing data, products and services

Organisation	Internet address	Information available
Office for National Statistics	http://www.statistics.gov.uk/census2001	Census data
	http://www.statistics.gov.uk/StatBase	CAS; vital statistics; Health Authority Patient Register data; GPRD data; outpatient attendances; health services activity; deprivation indices
	http://www.statistics.gov.uk/methods_quality/ns_sec	NS-SEC
	http://www.statistics.gov.uk/geography/boundaries.asp	OA- and SOA-level boundary data for use in GIS
	http://www.neighbourhood.statistics.gov.uk	Online access to census key statistics for England and Wales, data from government administrative sources, area codes
Local Government Data Unit – Wales	http://www.statistics.gov.uk/geography/nrudp.asp	Urban and rural classification system for England and Wales
	http://www.statistics.gov.uk/about/methodology_by_theme/area_classification/default.asp	National Statistics Area Classification
General Register Office Scotland	http://www.areaclassification.org.uk/	OAC user support group
Scottish executive	http://www.lgdu-wales.gov.uk/eng/WimdProject.asp?id=2077	WIMD 2005 scores
	http://www.gro-scotland.gov.uk	Scottish census and SIMD 2006 scores
Northern Ireland Statistics and Research Agency	http://www.sns.gov.uk	Scottish neighbourhood statistics data
	http://www.scotland.gov.uk/Publications/2004/06/19498/38784	Urban and rural classifications for Scotland
Economic and Social Research Council Census Programme	http://www.nisra.gov.uk	Northern Ireland census statistics, neighbourhood statistics, NIMDM 2005 scores, urban and rural classifications
Department for Communities and Local Government	http://census.ac.uk	Online access to census area data; interaction data; boundaries; samples of anonymised records for academic use
Ordnance Survey	http://www.communities.gov.uk/communities/neighbourhoodrenewal/deprivation/deprivation2007	IMD 2007 scores
Market Research Society	http://www.ordnancesurvey.co.uk/oswebsite	GIS data, products and services
Geodemographics Knowledge base	http://www.geodemographics.org.uk	Links to commercial suppliers of census, geodemographic and postcode data, as well as to organisations dealing with data protection issues
	Example UK-based health-related websites offering geographic information	http://www.nwpho.org.uk/information
http://www.diabetes.org.uk/InfoBankUK/		
http://shape.dh.gov.uk/		
Dr Foster Intelligence	http://www.wmpho.org.uk/wmciu/GIS3.htm	Commercial value-added health data products and services
ESRI	http://www.drfooster.co.uk	GIS products and services
	http://www.esri.com	GIS products and services
MapInfo	http://www.mapinfo.com	GIS products and services

CAS = Census Area Statistics; GPRD = General Practitioner Research Database; NS-SEC = National Statistics Socio-economic Classification; OA = Output Area; SOA = Super Output Area; GIS = geographic information systems; OAC = Output Area Classification; WIMD = Welsh Index of Multiple Deprivation; SIMD = Scottish Index of Multiple Deprivation; NIMDM = Northern Ireland Multiple Deprivation Measure; IMD = Index of Multiple Deprivation; ESRI = Environmental Systems Research Institute.

an important foundation for government policy and resource allocation in preference to more recent but less comprehensive sources. The national census offices of the UK also produce

mid-year population estimates annually by modeling population change based on the most recent census as a base point and making use of the results of postcensus revisions, the mid-2007 estimates being

Primary Health Care Research & Development 2010; **11**: 287–300

the most recent, for example, the most recent small area estimates for Scotland are the Mid-2007 Population Estimates Scotland (General Register Office for Scotland, 2008).

The problem of census under-enumeration varies geographically as well as by age and sex. The census is considered to cover 98% of the total population of the UK but undercounting is marked amongst young men living in inner city areas; some estimates put under-enumeration among men aged between 25 and 29 years in the major metropolitan areas at around 20% (Office of Population Censuses and Surveys, 1994). The published 2001 census results represent the actual counts, combined with a sophisticated estimation methodology known as the 'One Number Census' (Steele *et al.*, 2002) intended to impute additional individuals and households most likely to be missing from each area. While coverage across the UK is considered to be good, particular enumeration problems were recognised in certain areas, for instance further corrections have to be applied to the counts for 15 local authorities in England post-2001 census, most notably for Westminster and Manchester. The census agencies of the UK are also making efforts to provide population estimates for small areas in the intercensal period, making use of the results of post-census revisions.

Preparations are now underway for the 2011 censuses (Martin, 2007) and a 2008 Census White Paper (Office for National Statistics, 2008) sets out the current proposals for England and Wales following an extended period of consultation on future question topics and holding field tests. England and Wales tested their pilot proposals in May 2007 covering 100 000 households in five local authority areas. Scotland piloted its proposals in April 2006 covering 50 000 households in five areas, and on 29th March 2009 conducted a full census rehearsal. Of particular significance to the health and healthcare research and development is the inclusion of a potential question on income, a variable that has been shown in a number of studies outside of the UK to be one of the strongest socio-economic indicators of health outcomes (Carstairs, 2000: 62–63; Gatrell, 2002: 126–127). However, in the 2011 pilot census income was deemed to have too great a negative impact on overall response rates in the 2006–07 test-bed populations' response rates in England,

and the question has been dropped in England and Wales. Intriguingly, the inclusion of the same question in the US census has not been a matter of public controversy and income data have been used in both epidemiological and health-care research for many years (Lynch *et al.*, 1998; Sanmartin *et al.*, 2003; Subramanian and Kawachi, 2006). It is also interesting to note that the inclusion of the same question in the 2006 test in Scotland did not result in a reduced response rate and an income question may still feature in the 2011 Scottish census (General Register Office for Scotland, 2007). All of the 2011 UK census proposals retain the 2001 questions on self-reported general health and limiting long-term illness, and also add new topics of relevance to aspects of locality health profiling. For example, the proposals include the questions on citizenship and national identity, as well as on second residences.

Socio-economic and behavioural indicators

Deprivation indices

There have been several area-based indices developed for measuring community deprivation using combinations of census variables, such as the Carstairs, Jarman and Townsend indices (Senior, 2002). The Carstairs index, for example, is a well-established gauge of material deprivation and uses four variables (overcrowding, the proportion of people in low social classes, lack of car ownership and unemployment rates). It should be noted that the Jarman Underprivileged Area index, the first deprivation index produced in the UK and incidentally originally created to predict GP workloads in London for primary health care research purposes rather than to measure material deprivation, is now no longer used due to criticisms regarding its limited applicability outside of inner metropolitan areas.

A newer generation of indices uses a combination of census and non-census data derived from government administrative information, of which the latest are the Indices of Deprivation (Neighbourhood Renewal Unit, 2004). Each part of the UK has its own variant tailored to its regional context (latest versions in brackets): England (Index of Multiple Deprivation, 2007), Scotland (Scottish Index of Multiple Deprivation, 2006),

Wales (Welsh Index of Multiple Deprivation, 2005) and Northern Ireland (Northern Ireland Multiple Deprivation Measure, 2005). Each of these indices is a composite of domains derived from constituent variables. For example, the English Index of Multiple Deprivation 2007 is composed of seven Domains (barriers to housing and services, crime, education, employment, environment, income and health) derived from 37 constituent variables, including a Health Deprivation and Disability Domain comprised of four community health indicators. The inclusion of non-census data offers an advantage over traditional census-based measures of deprivation, which become less accurate over time from the previous census. Examples of non-census datasets included in the current indices of deprivation are employment information (eg, number of persons on incapacity benefit) and educational information (eg, proportion of school-leavers aged sixteen) most of which are updated much more frequently than census data.

A key application of area deprivation data when combined with health and health care datasets is monitoring inequalities in primary health care delivery and quality. An illustrative case is Sigfrid *et al.* (2006) in which the relationships between Quality and Outcomes Framework diabetes exception reporting, Diabetes UK data and ID 2004 scores for Lower Super Output Areas (LSOAs) in the Brighton and Hove City Primary Care Trust (PCT) were statistically analysed. The findings suggested that although there was socio-economic equity in the achievement of diabetes management targets, exception reporting was higher in areas with a higher deprivation score. It should be noted that with the additional use of geodemographics data (discussed later in this review) in such analyses a geographically more precise identification of priority areas is possible.

The domains of the current generation of indices can be used individually or in combination, and thus give great flexibility. Research also suggests that this type flexible of index is more effective than the traditional indices at measuring deprivation in both rural and urban areas (Asthana *et al.*, 2002) particularly due to the inclusion of accessibility variables. Previous indices have been more effective at identifying deprivation variables most relevant to health outcomes in urban areas than rural areas (Cox, 1998; Barnett *et al.*, 2001)

and although the current range of indices incorporate variables more relevant to rural contexts nevertheless present indices are still more effective in the context of health care research and development for urban areas than rural areas (Jordan *et al.*, 2004b; Niggebrugge *et al.*, 2005). The current deprivation indices are calculated and made available for groups of OAs termed LSOAs. It should be noted that when used for health analyses different deprivation indices have shown a high degree of correlation with each other. When choosing measures of deprivation, both conceptual and practical considerations should be taken into account. For instance, the English Index of Multiple Deprivation 2007 includes a 'health and disability' domain, which means that analyses using the index may overestimate associations between deprivation and health (Adams and White, 2006), although the health domain can be excluded or individual domains used instead. There also are caveats to using deprivation indices in general. A notable obstacle has been the current lack of suitable population estimates for LSOAs (Morgan and Baker, 2006). As a result of small area analyses, particularly for rare health conditions, can be problematic in localities with transient populations such as inner city areas, resulting in a lack of robustness in denominator population estimates. To counteract these problems the ONS have now released an experimental series of small area population estimates, which are available at the LSOA level. The new dataset is based on a ratio-change methodology, which seeks to update census-based initial counts with change data drawn from multiple administrative sources, including the National Health Service (NHS) central register (Bates, 2006). However, while this methodological development is of great importance to small area studies it is essential to note that these inter-censal small area estimates are synthetic and are not actual census small area population counts.

Socio-economic classifications

Mapping and analysing the geographic distribution of socio-economic groups is an alternative and complementary approach to the use of deprivation indices. The traditional methods of classifying people into socio-economic groups were the Registrar General's Social Class (SC) based on

occupation and the less commonly used Socio-economic Groups (SEG). Limitations have been found in both measures by academics and policy makers due to changes over the decades in the composition of the workforce and the range of occupations available. This has resulted in the ONS producing the National Statistics Socio-economic Classification (NS-SEC) a more contemporary system to reflect present-day society (Rose and O'Reilly, 1998). One of the prime motivations for the structuring of the NS-SEC was to classify society into groups that better predict health outcomes than either the SC or the SEG did. In particular, small areas with a significant component of the NS-SEC class 8 also have high levels of healthcare utilisation. Table 2 compares the SC and NS-SEC approaches.

Geodemographic datasets

Whereas deprivation indices seek to place every small area on a single deprivation scale, there has also been extensive interest in area classifications, which allocate each small area to a neighbourhood type termed geodemographic classifications (Brown and Batey, 1994; Birkin and Clarke, 1998; Birkin *et al.*, 2002; Harris *et al.*, 2005). Geodemographics is in effect an area-based form of segmentation, to use social marketing terminology, which the NHS is showing increasing interest in (eg, Abbas *et al.*, 2009). While geodemographics has its roots in

academic geography it has been of particular interest to the marketing industry, which has made use of census and non-census data derived from credit-checking agencies, retail behaviour and consumer surveys, amongst other sources, to create 'lifestyle' segments. As a result the majority of current geodemographic classifications have been produced in the commercial sector; examples include ACORN, MOSAIC, PinPoint and Super Profiles. The use of these commercial datasets is growing in the public health sector (eg, Aveyard *et al.*, 2002; Birkin *et al.*, 2005; Powell *et al.*, 2007). Nevertheless, there are non-commercial classification systems, for example following the 2001 census ONS produced an entirely new, census-based geodemographic classification of OAs, which is freely available (Vickers and Rees, 2007). The ONS produced this classification for public sector research and planning, termed the National Statistics Output Area Classification (OAC), which has its own user support group (see Table 1). However, it should be noted that this dataset though available at a finer spatial scale than deprivation indices is still at a coarser scale than commercially available geodemographic classifications.

All of these neighbourhood classifications, both commercial and non-commercial, provide a complementary resource base for health needs assessment, giving indirect insight into geographic patterns in health-related behaviours and attitudes (eg, socio-economic gradients in diet and

Table 2 The SC and the ONS NS-SEC summarised

Classification scheme	Number of categories	Categories
SC	6	I – professional II – managerial and technical IIIN – skilled non-manual IIIM – skilled manual IV – partly unskilled V- unskilled
NS-SEC	9	1.1 – large employers and senior managers 1.2 – higher professional 2 – lower management and professional 3 – intermediate 4 – small employers and self-employed 5 – supervisory and technical 6 – semi-routine 7 – routine 8 – never worked and long-term unemployed

SC = social class based on occupation; ONS = Office for National Statistics; NS-SEC = National Statistics Socio-economic Classification.

smoking). Geodemographic data can be combined with health information for mapping disease risk, monitoring health profiles and access to services, for example Webber (2004), examined HES admissions by neighbourhood using MOSAIC. An advantage of commercial systems using non-census data is that their socio-demographic classifications can be estimated for areas even smaller than census OAs, hence theoretically capture diversity within small areas, and additionally are continually updated using a variety of both commercial and administrative data sources (Muggli *et al.*, 2000; Longley, 2005). An example of the application of geodemographic datasets to monitoring locality health profiles is Powell *et al.* (2007) in which a commercial geodemographic system was used to identify highly localised pockets of populations at high risk of developing diabetes mellitus within a PCT's boundaries. As Powell *et al.* (*ibid.*: 33–34) note, the identification of such pockets can aid in the social marketing of behavioural interventions.

However, there is much disagreement over the value of the use of geodemographic datasets in health contexts. Much of this relates to methodological issues with some geographers considering that because the indicators used in commercial geodemographic datasets were not developed for the purpose of analysing retail behaviour and not for identifying health-related behaviours their use is theoretically unsound (eg, Twigg *et al.*, 2000: 1110); other researchers considering that intra-group variation is greater than inter-group variation in published, commercially available geodemographic classifications (Voas and Williamson, 2001). There is also the fact that commercially available geodemographic classification systems sold to healthcare organisations have an underlying commercial motivation behind them and primary health care researchers and developers should always take this factor into account when evaluating the merits of purchasing such classifications. The 'jury is still out' on the exact contribution that geodemographics has to make to not only primary health care but also public health in general. Nevertheless, its potential merits further research and development as the OAC has shown. It is also important to note that with the current emphasis on lifestyles and healthy choices geodemographics is likely to play an increasing role in primary health

care research and development (Department of Health, 2004).

Area and rural–urban classifications

For those performing analyses requiring an understanding of the overall socio-economic character of areas the national statistics offices of the UK provide two forms of classification, the first summarising the overall economic and demographic character of areas and the second classifying areas on an rural to urban spectrum. The former is termed the National Statistics 2001 Area Classification and has been produced to cover all of the UK, the OAC discussed in the previous section forming its foundations (see Table 1). It is available at different levels of geographic aggregation including Health Areas in the form of primary care organisations. These Health Areas summarise area socio-demographic character and are structured into three geographic levels of classification: supergroup, group and subgroup. Example classes include manufacturing towns, coastal and countryside and industrial hinterlands. Unsurprisingly such classifications are of particular interest to human geographers and sociologists, but are also of relevance to the health sector analysis as Doran *et al.* (2006) have shown in their comparison of health outcomes across local authorities in England.

The latter type of area classification are rural–urban classifications, which are especially relevant with respect to capturing 'rurality' effects on health outcomes, a variable which even current deprivation indices have thus far been found slightly wanting as discussed in the section on deprivation indices. The latest rural–urban classifications for the UK are the ONS Rural and Urban Area Classification 2004 for England and Wales, the NISRA Urban–Rural Classification 2005 and the Scottish Executive's Urban–Rural Classification 2003–04 (Bibby and Shepherd, 2004; Northern Ireland Statistics and Research Agency, 2005; Scottish Executive, 2008). Again, as with the National Statistics 2001 Area Classification, there are geographic tiers to the rural–urban classifications. An example of the use of a rural–urban classification in health care research is Jordan *et al.* (2004a) a study which included analysis of accessibility to GP services in southwest England.

Neighbourhood statistics

A recent development in the availability of socio-demographic data in the UK is the provision of Neighbourhood Statistics by the national statistics agencies covering all geographical scales in the UK from the lowest socio-demographic units to the national level. The neighbourhood statistics datasets include a range of information, including those covered in the census and the Indices of Deprivation, which characterise both population and environmental characteristics of areas. The provision of a range of socio-demographic indicators in a single dataset customisable to a given locality is a great boon and is especially valuable for integrating primary health care provision within urban and rural planning, a topic returned to later in this review.

Boundary data

Administrative, census and postcode locations are fundamental to mapping health to both visualise geographic patterns and analyse spatial clustering. The geographic hierarchy of boundaries within the UK is complex, with differences of detail between the countries in question and within them. Essentially, boundaries are available for government administrative units at every level down to census OAs. However, this neighbourhood statistics geography is intentionally static to allow the production of non-disclosive and comparable data series, while the units of primary health care organisation change repeatedly and their boundaries are not necessarily aligned with the hierarchy built on OAs as its foundation. These OAs were, wherever possible, built from whole unit postcodes (the smallest element in the postcode hierarchy) and represent consistent-sized populations of around 300 persons. They were intended as flexible statistical spatial blocks rather than analysis units in their own right (Majeed *et al.*, 1995; Scrivener and Lloyd, 1995; Martin, 2004). To aid this goal further, a lower tier of super output areas, LSOAs ('Datazones' in Scotland) have been produced by aggregating 2001 OAs to produce areas with a mean population size of 1500 and a middle tier (MSOAs, Middle Super Output Area) with a mean population size of 7500. Digital boundary data are freely available for 2001 OAs and ONS are proposing to retain a stable OA geography as far as possible for the 2011 census.

Primary Health Care Research & Development 2010; **11**: 287–300

In addition to data published for these geographical areas, there is extensive use of the postcode as a geographical reference and Royal Mail, Ordnance Survey and ONS work together to produce the Office for National Statistics Postcode Directory (<http://www.ons.gov.uk/about-statistics/geography/products/geog-products-postcode/nspd/index.html>), a listing of all postcodes giving their Ordnance Survey grid reference and their membership of a very large number of geographical areas, including all levels of the neighbourhood statistics geography and also current and past primary health care organisations.

A present, though perhaps only a short-term, problem with boundary datasets as applied to primary health care contexts, is that the current PCT boundaries have not been published and only pre-2005 boundaries are available except for London where the boundaries are unchanged. This situation is a result of the number of PCTs in England having been reduced from 303 to 152 in number during 2006, as part of a move to make PCT boundaries coterminous with those of local authorities to facilitate integrated urban and rural planning. Those needing to use current ones outside of London are obliged to construct the boundaries themselves, though considering the fluctuations in PCT boundaries over the last decade it is our recommendation to perform analyses at the local authority level to permit the study of change over time. A separate limitation that the reader may also have noted that socio-demographic boundaries in general are inflexible and additionally do not reflect any 'real world' physical boundaries. While for most planning and evaluation purposes this is not a significant issue, those requiring more refined, complex analyses which take into account the problems posed by inflexible boundaries the use of spatial smoothing techniques is paramount. A discussion of the application of spatial statistics in health geography is beyond the scope of this data sources review and instead readers are directed towards Elliott *et al.* (2000) and Cromley and McLafferty (2002).

Linking datasets

Combining clinical data with those of other types (eg, census, geodemographic) not only enables services planning and research but also potentially

creates new, value-added datasets. For instance by combining geographic data on lifestyles and behaviours with those on health, geographical-based health profiles and healthcare needs assessment datasets can be produced (Dedman *et al.*, 2006). An example of this is the locality health profiles produced by the Association of Public Health Observatories (Pencheon, 2008). Of key relevance to linking existing datasets is the National Statistics Postcode Directory, where a postcode or recognised area code is available in each dataset the directory can be used to link records directly or to reweight data from one area unit to another. Researchers within UK academic settings can take advantage of the GeoConvert facility to undertake such linkage (Mimas GeoConvert at <http://geoconvert.mimas.ac.uk/>). Where direct lookup in this way is not possible, the researcher must perform GIS analysis to determine the best match between the available data.

The socio-demographic data sources discussed in this review can also be linked with transport and health datasets to identify localities with both high health care needs and limited geographical access to service provision, and indeed much of this work has focused on primary health care (eg, Luo and Wang, 2003; Guagliardo, 2004; Luo, 2004; Wang and Luo, 2005; Mobley *et al.*, 2006). Further to this forecasting future demand for primary health care services in areas currently undergoing neighbourhood regeneration is increasingly becoming an integrated part of planning wider social infrastructures especially in spearhead groups (Blackman, 2006) and relates directly to the information toolkits at the disposal of health care planners. It is in this respect that neighbourhood statistics are invaluable. A current example of this is the London Thames Gateway Social Infrastructure Framework (LTG-SIF). This urban regeneration programme covers much of east London is funded by the London Development Agency and local PCTs, and is coordinated by the National Health Service London Healthy Urban Development Unit in part aims to effectively plan future health services in the area by working closely with local urban planners to model the implications of new housing developments for health care demand and supply (National Health Service London Healthy Urban Development Unit, 2006a; 2006b). In data terms

this means integrating population growth models based on census information with morbidity trends, deprivation scores and transport datasets in order to optimally plan new educational centres, green spaces, GP clinics and other facilities in an integrated framework to produce sustainable communities.

A simple example of a data linkage between socio-demographic and health care data in a GIS environment, using spatial regression, is provided in Figure 1, illustrating the strength of the association between limiting long-term illness prevalence and NS-SEC class 8 (never worked and long-term unemployed; see Table 2) of the NS-SEC for LSOAs across southwest England, with PCT boundaries overlain. Notice the markedly higher regression residuals in largely rural Cornwall in comparison to highly metropolitan Greater London, demonstrating the weaker association in rural areas between long-term morbidity and a key indicator of area socio-economic composition. There are not only clear implications for relying solely upon socio-economic indicators for making resource allocation decisions for rural areas, as discussed in the section on deprivation indices. The figure also shows the value of looking at socio-demographic data sources from a geographical perspective.

Nevertheless, making use of the full range of datasets discussed in this review is not always straightforward. For instance geodemographic datasets derived from consumer questionnaires and administrative sources at present cannot match the population coverage and quality assurance levels achieved by the decennial censuses. The substantial increase in information available for small areas is a welcome development; however, this trend in turn raises questions about statistical robustness and interpretation. For example consideration of local factors, such as the presence of nursing homes in some areas, must be taken into account in the interpretation of geographic patterns. Additionally, the limitations of aggregated, socio-demographic datasets can in some circumstances compound the variable data quality of clinical and health care datasets; hence thoughtful use of these sources, especially when linking them with health-related data, is imperative.

Currency of data also plays a role in limiting the effectiveness of these datasets. While geographical

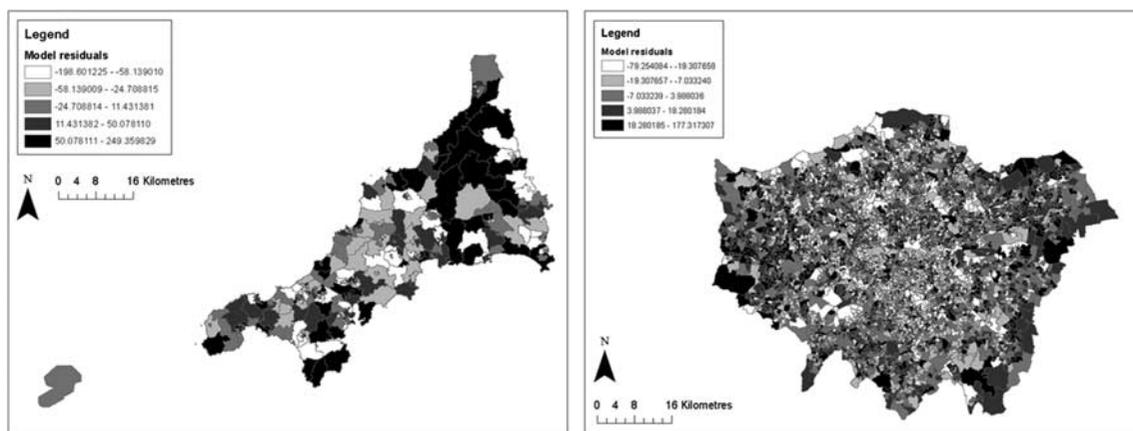


Figure 1 Illustrating the weaker relationship between socio-economic indicators and health outcomes for rural areas when compared to urban areas, using the example of residuals from a spatial regression model of limiting long-term illness and unemployment applied to data from Cornwall and Greater London. *Source:* 2001 census data provided with the support of the Economic and Social Research Council and Joint Information Systems Committee. Census output is Crown copyright and is reproduced with the permission of the Controller of Her Majesty's Stationery Office and the Queen's Printer for Scotland.

planning and evaluation of primary health care does not require real-time data unlike say emergency response services, nevertheless the fact that there is significant variation in how current the socio-demographic data sources discussed in this review are makes analysis more complex. For example, census data which, even with annual mid-year estimates become increasingly inaccurate in terms of demographic structure the further from the year of the census one goes. Localities with a significant turnover of population are especially problematic for monitoring locality health profiles and planning health care services (Hennell, 2004) particularly with respect to forecasting future demand for health care. By contrast with the census situation, some variables in the current indices of deprivation are updated annually. As a result there is inevitably a certain, perhaps unquantifiable, amount of temporal mismatch between the different datasets. This complicates small area analyses especially considering that the census provides unrivalled coverage and quality assurance in comparison to other datasets such as the frequently updated geodemographic classifications available commercially.

A key issue in terms of monitoring equity in health care is that the identification of spatial variation in the need for health services in rural areas is more complex given the available data

sources than for urban settings. While a new generation of flexible deprivation indices, geodemographic sources and the NS-SEC collectively provide a more effective information toolkit for monitoring and planning rural health care, other variables such as access to proximate out-of-hours services, distance to healthcare centres and public transport infrastructure often play a role in the effectiveness of health services delivery (Baird *et al.*, 2000). This situation is made more complex by the fact that although there is an access to services domain in each of the Indices of Deprivation, it should be noted that the access domain in each of these indices uses straight line distances rather than road network distances and it has been noted that the former only has moderate correlation with the latter in many localities (Jordan *et al.*, 2004a; 2004b; Niggebrugge *et al.*, 2005), a problem for monitoring inequalities in health care accessibility.

It should be noted that the aggregated data sources discussed, while well suited for resource allocation purposes are not always fit for purpose in other health-related analyses, for example aetiological studies. We stress that household-level or individual-level data need to be used in conjunction with aggregate, contextual datasets for researching associations in health and behavioural outcomes in individuals and groups

(eg, Nieuwenhuijsen, 2000). Obtaining data on individuals from healthcare records, questionnaires and/or interviews requires prior ethical approval but can provide valuable individual-level information to complement aggregated datasets, for example Blackman *et al.* (2001) and Day (2008).

In the longer term, the 2008 Treasury Subcommittee of Inquiry on Counting the population has recommended that following 2011 ONS move towards the replacement of the conventional census by data sourced from administrative records (House of Commons Treasury Committee, 2008). Such development is currently at a very early stage, but would have major implications for the entire data landscape presented in this review and all involved in primary health care research and development should seek to engage in the forthcoming debate. In terms of the more immediate future there is currently a consultation programme for Census 2011, with many issues relevant to the health care sector being covered, for example debate regarding proposed new census questions for inclusion such as that on multi-residency (Cabinet Office, 2008). Of notable interest to primary health care researchers and developers will be the consultation over whether to collect data on numbers of legal short-term migrants, due to take place in April and May 2010. This potential short-term migrant dataset could have great value for services planning; however, there will inevitably be question marks over the accuracy of such a dataset. Further details on how to become engaged in the Census 2011 consultation can be obtained by visiting the ONS website (<http://www.ons.gov.uk/census/2011-census/consultations/index.html>). The consultation process is now underway and will continue up until autumn and winter 2010.

Conclusions

Developing routine, published health and geographic data sources to study the distribution of socioeconomic and health inequalities may assist local and population-based health intervention strategies to be more efficiently targeted. The advent of OAs/SOAs, a new generation of deprivation measures, the NS-SEC and progress in GIS, have delivered a powerful suite of data, tools, methods and an associated body of literature

for primary health care research and development. Geodemographic profiling in particular is likely to play an increasing role in local strategies.

Acknowledgement

The authors thank Allan Baker of the Office for National Statistics (ONS) Mortality Statistics Team for valuable comments on drafts of this review paper.

Declarations

Funding: Dr Saxena holds an NIHR postdoctoral fellowship.

Competing interests: The authors declared no competing interests.

Ethical approval: The authors declared that ethical approval is not required.

References

- Abbas, J., Carlin, H., Cunningham, A., Dedman, D. and McVey, D. 2009: *Geodemographic segmentation*. York, UK: Association of Public Health Observatories [Technical Briefing 5].
- Adams, J. and White, M. 2006: Removing the health domain from the Index of Multiple Deprivation 2004 – effect on measured inequalities in census measure of health. *Journal of Public Health* 28, 379–83.
- Asthana, S., Halliday, J. and Brigham, P. 2002: *Rural deprivation and service need: a review of the literature and an assessment of indicators for rural service planning*. Bristol, Avon, UK: South West Public Health Observatory.
- Aveyard, P., Manaseki, S. and Chambers, J. 2002: The relationship between mean birth weight and birth weight using the Townsend deprivation score and the Super Profile classification system. *Public Health* 116, 308–14.
- Baird, A.G., Donnelly, C.M., Miscampell, N.T. and Wemyss, H.D. 2000: Centralisation of cancer services in rural areas has disadvantages. *British Medical Journal* 320, 717.
- Barnett, S., Roderick, P., Martin, D. and Diamond, I. 2001: A multilevel analysis of the effects of rurality and social deprivation on premature limiting long term illness. *Journal of Epidemiology and Community Health* 55, 44–51.
- Bates, A. 2006: Methodology used for producing ONS small area population estimates. *Population Trends* 125, 30–46.
- Bibby, P. and Shepherd, J. 2004: Developing a new classification of urban and rural areas for policy purposes: the Methodology. Retrieved 19 April 2009 from <http://www.statistics.gov.uk/geography/nrudp.asp>

- Birkin, M.** and **Clarke, G.** 1998: GIS, geodemographics and spatial modelling: an example within the UK financial services industry. *Journal of Housing Research* 9, 87–112.
- Birkin, M., Clarke, G.** and **Douglas, L.** 2002: Optimising retail mergers and acquisitions geographically. *Progress in Planning* 58, 229–318.
- Birkin, M., Clarke, G., Gibson, P., Dewhurst, R.** and **Bobby, J.** 2005: Re-evaluating funding mechanisms for Primary Care Trusts: a new model of secondary care usage incorporating lifestyle data. *Environment and Planning C* 23, 317–36.
- Blackman, T.** 2006: *Placing health: neighbourhood renewal, health improvement and complexity*. Bristol, Avon, UK: Policy Press.
- Blackman, T., Harvey, J., Lawrence, M.** and **Simon, A.** 2001: Neighbourhood renewal and health: evidence from a local case study. *Health and Place* 7, 93–103.
- Bottle, A., Millett, C., Xie, Y., Saxena, S., Wachter, R.M.** and **Majeed, A.** 2008: Quality of primary care and hospital admissions for diabetes mellitus in England. *Journal of Ambul Care Management* 31, 226–38.
- Bowling, A., Barber, J., Morris, R.** and **Ebrahim, S.** 2006: Do perceptions of neighbourhood environment influence health? Baseline findings from a British Survey of Aging. *Journal of Epidemiology and Community Health* 60, 476–83.
- Brown, P.J.B.** and **Batey, P.W.J.** 1994: *Design and construction of a geodemographic targeting system: super profiles 1994*. Liverpool, UK: University of Liverpool, Urban Research and Policy Evaluation Laboratory.
- Cabinet Office.** 2008: *Helping to shape tomorrow: the 2011 Census of Population and Housing in England and Wales*. London, UK: The Stationery Office.
- Carstairs, V.** 2000: Socio-economic factors at area level and their relationship with health. In Elliott, P., Wakefield, J., Best, N. and Briggs, D., editors., *Spatial epidemiology: methods and applications*. Oxford, UK: Oxford University Press, 51–67, Ch. 4.
- Congdon, P.** 2006: Estimating diabetes prevalence by small area in England. *Journal of Public Health* 28, 71–81.
- Cox, J.** 1998: Poverty in rural areas is more hidden but no less real than in urban areas. *British Medical Journal* 316, 722–30.
- Cromley, E.K.** and **McLafferty, S.L.** 2002: *GIS and public health*. New York, USA: The Guilford Press.
- Damiani, M., Propper, C.** and **Dixon, J.** 2005: Mapping choice in the NHS: cross-sectional study of routinely collected data. *British Medical Journal* 330, 284–89.
- Day, R.** 2008: Local environments and older people's health: dimensions from a comparative qualitative study in Scotland. *Health and Place* 14, 299–312.
- Dedman, D., Jones, A., Tocque, K.** and **Bellis, M.A.** 2006: *Population targeting: tools for social marketing*. Liverpool, UK: Northwest Public Health Observatory.
- Department of Health.** 2004: *Choosing health: making healthier choices easier*. London, UK: DH.
- Primary Health Care Research & Development* 2010; **11**: 287–300
- Dixie, J.** and **Dorling, D.** 2002: New questions for the 2001 census. In Rees, P., Martin, D. and Williamson, P., editors. *The census data system*. Chichester, West Sussex, UK: Wiley, 283–94, Ch. 20.
- Doran, T., Drever, F.** and **Whitehead, M.** 2006: Health underachievement and overachievement in English local authorities. *Journal of Epidemiology and Community Health* 60, 686–93.
- Elliott, P., Wakefield, J., Best, N.** and **Briggs, D.** 2000: *Spatial epidemiology: methods and applications*. Oxford, UK: Oxford University Press.
- Field, K.** 2000: Measuring the need for primary health care: an index of relative disadvantage. *Applied Geography* 20, 305–32.
- Gatrell, A.C.** 2002: *Geographies of health: an introduction*. Oxford, UK: Blackwell.
- General Register Office for Scotland.** 2007: *2006 census test evaluation report*. Edinburgh, UK: GROS.
- General Register Office for Scotland.** 2008: *Mid-2007 Population Estimates Scotland*. Edinburgh, UK: GROS.
- Gnani, S.** and **Majeed, A.** 2006: *A user's guide to data collected in primary care in England*. Cambridge: Eastern Region Public Health Observatory.
- Guagliardo, M.F.** 2004: Spatial accessibility of primary care: concepts, methods and challenges. *International Journal of Health Geographics* 3, 3–15.
- Harris, R., Sleight, P.** and **Webber, R.** 2005: *Geodemographics: GIS and neighbourhood targeting*. Chichester, West Sussex, UK: Wiley.
- Hennell, T.** 2004: Have we learned to count yet? A health services perspective on the 2001 census. *Area* 36, 124–35.
- House of Commons Treasury Committee.** 2008: *Counting the population: eleventh report of session 2007–2008 Volume 1*. London, UK: The Stationery Office.
- House of Commons Treasury, Sub-committee of Inquiry on Counting the Population.** 2008: Retrieved January 2009 from <http://www.publications.parliament.uk/pa/cm200708/cmselect/cmtreasy/183/183.pdf>
- Jenkins, C.** and **Campbell, J.** 1996: Catchment areas in general practice and their relation to size and quality of practice and deprivation: a descriptive study in one London borough. *British Medical Journal* 313, 1189–92.
- Jordan, H., Roderick, P.** and **Martin, D.** 2004b: The Index of Multiple Deprivation 2000 and accessibility effects on health. *Journal of Epidemiology and Community Health* 58, 250–57.
- Jordan, H., Roderick, P., Martin, D.** and **Barnett, S.** 2004a: Distance, rurality and the need for care: access to services in South West England. *International Journal of Health Geographics* 3, 21–29.
- Kalthenthaler, E., Maheswaran, R.** and **Beverley, C.** 2004: Population-based health indexes: a systematic review. *Health Policy* 68, 245–55.
- Longley, P.** 2005: Geographic information systems: a renaissance of geodemographics for public service delivery. *Progress in Human Geography* 29, 57–63.

- Luo, W.** 2004: Using a GIS-based floating catchment method to assess areas with shortage of physicians. *Health and Place* 10, 1–11.
- Luo, W.** and **Wang, F.** 2003: Measures of spatial accessibility to health care in a GIS environment: synthesis and a case study in the Chicago region. *Environment and Planning B: Planning and Design* 30, 865–84.
- Lynch, J.W., Kaplan, G.A., Pamuk, E.R., Cohen, R.D., Heck, K.E., Balfour, J.L. and Yen, I.H.** 1998: Income inequality and mortality in metropolitan areas of the United States. *American Journal of Public Health* 88, 1074–80.
- Majeed, F.A., Cook, D.G., Poloniecki, J. and Martin, D.** 1995: Using data from the 1991 census. *British Medical Journal* 310, 1511–514.
- Martin, D.** 2002: Geography for the 2001 census in England and Wales. *Population Trends* 108, 7–15.
- Martin, D.** 2004: Neighbourhoods and area statistics in the post-2001 census era. *Area* 36, 136–45.
- Martin, D.** 2006: Last of the censuses? The future of small area population data. *Transactions of the Institute of British Geographers* 31, 6–18.
- Martin, D.** 2007: Census present and future. *Journal of the Royal Statistical Society A* 170, 263–66.
- Millett, C., Zelenyanski, C., Binysh, K., Lancaster, J. and Majeed, A.** 2005: Population mobility: characteristics of people registering with general practices. *Public Health* 119, 632–38.
- Mimas GeoConvert.** Retrieved 22 March 2009 from <http://geoconvert.mimas.ac.uk/>
- Mobley, L.R., Root, E., Anselin, L., Lozano-Gracia, N. and Koschinsky, J.** 2006: Spatial analysis of elderly access to primary care services. *International Journal of Health Geographics* 5, 19–35.
- Morgan, O. and Baker, A.** 2006: Measuring deprivation in England and Wales using 2001 Carstairs scores. *Health Statistics Quarterly* 31, 28–33.
- Muggli, E.E., McCloskey, D. and Halliday, J.L.** 2000: Health behaviour modeling for prenatal diagnosis in Australia: a geodemographic framework for health service utilisation and policy development. *BMC Health Services Research* 6, 109–19.
- National Health Service London Healthy Urban Development Unit.** 2006a: *London Thames Gateway Social Infrastructure Needs Assessment*. London, UK: HUDU.
- National Health Service London Healthy Urban Development Unit.** 2006b: *London Thames Gateway Social Infrastructure Framework: a toolkit to guide decision making*. London, UK: HUDU.
- Neighbourhood Renewal Unit.** 2004: *The English Indices of Deprivation 2004: Summary (Revised)*. London, UK: Office of the Deputy Prime Minister.
- Nieuwenhuijsen, M.J.** 2000: Personal exposure monitoring in environmental epidemiology. In Elliott, P., Wakefield, J., Best, N. and Briggs, D., editors, *Spatial epidemiology: methods and applications*. Oxford, UK: Oxford University Press, 360–74, Ch. 20.
- Niggebrugge, A., Haynes, R., Jones, A., Lovett, A. and Harvey, I.** 2005: The Index of Multiple Deprivation 2000 accessibility domain: a useful indicator for public health? *Social Science and Medicine* 60, 2743–753.
- Northern Ireland Statistics and Research Agency.** 2005: *Report of the inter-departmental urban–rural classification group: statistical classification and delineation of settlements*. Belfast, County Antrim, UK: NISRA.
- Office of Population Censuses and Surveys.** 1994: *Under-coverage in Great Britain*. London, UK: OPCS [1991 Census user guide 58].
- Office for National Statistics.** 2008: *Helping to shape tomorrow: the 2011 census of population and housing in England and Wales*. London, UK: The Stationery Office.
- Office for National Statistics Postcode Directory.** Retrieved 22 March 2009 from <http://www.ons.gov.uk/about-statistics/geography/products/geog-products-postcode/nspd/index.html>
- Office for National Statistics 2011 Census Consultations.** Retrieved 19 April 2009 from <http://www.ons.gov.uk/census/2011-census/consultations/index.html>
- Pencheon, D.** 2008: *The good indicators guide: understanding how to use and choose indicators*. Coventry, West Midlands, UK: NHS Institute for Innovation and Improvement.
- Powell, J., Tapp, A., Orme, J. and Farr, M.** 2007: Primary care professionals and social marketing of health in neighbourhoods: a case study approach to identify, target and communicate with ‘at risk’ populations. *Primary Health Care Research and Development* 8, 22–35.
- Rose, D. and O’Reilly, K.** 1998: *The ESRC review of government social classifications: final report*. London, UK: The Stationery Office.
- Sanmartin, C., Ross, N.A., Tremblay, S., Wolfson, M., Dunn, J.R. and Lynch, J.** 2003: Labour market income inequality and mortality in North American metropolitan areas. *Journal of Epidemiol Community Health* 57, 792–97.
- Saxena, S., Car, J., Eldred, D., Soljak, M. and Majeed, A.** 2007: Practice size, caseload, deprivation and quality of care of patients with coronary heart disease, hypertension and stroke in primary care: national cross-sectional study. *BMC Health Services Research* 7, 96–104.
- Scottish Executive.** 2008: *Urban–rural classification 2007–2008*. Edinburgh, UK: Office of the Chief Statistician.
- Scrivener, G. and Lloyd, D.C.E.F.** 1995: Allocating census data to general practice populations: implications for study of prescribing variation at practice level. *British Medical Journal* 311, 163–65.
- Senior, M.** 2002: Deprivation indices. In Rees, P., Martin, D. and Williamson, P., editors. *The census data system*. Chichester, West Sussex, UK: Wiley, 123–38, Ch. 9.
- Sigfrid, L.A., Turner, C., Crook, D. and Ray, S.** 2006: Using the UK primary care Quality and Outcomes Framework to audit health care equity: preliminary data on diabetes management. *Journal of Public Health* 28, 221–25.

- Steele, F., Brown, J. and Chambers, R.A.** 2002: Controlled donor imputation system for a one-number census. *Journal of the Royal Statistical Society Series A* 165, 495–522.
- Strong, M., Maheswaran, R. and Radford, J.** 2006: Socio-economic deprivation, coronary heart disease prevalence and quality of care: a practice-level analysis in Rotherham using data from the new UK general practitioner Quality and Outcomes Framework. *Journal of Public Health* 28, 39–42.
- Subramanian, S.V. and Kawachi, I.** 2006: Whose health is affected by income inequality? A multilevel interaction analysis of contemporaneous and lagged effects of state income inequality on individual self-rated health in the United States. *Health and Place* 12, 141–56.
- Thiru, K., Hassey, A. and Sullivan, F.** 2003: Systematic review of scope and quality of electronic patient record data in primary care. *British Medical Journal* 326, 1070–74.
- Twigg, L., Moon, G. and Jones, K.** 2000: Predicting small-area health-related behaviour: a comparison of smoking and drinking indicators. *Social Science and Medicine* 50, 1109–20.
- Vickers, D. and Rees, P.** 2007: Creating the UK National Statistics 2001 output area classification. *Journal of the Royal Statistical Society Series A* 170, 379–404.
- Voas, D. and Williamson, P.** 2001: The diversity of diversity: a critique of geodemographic classification. *Area* 33, 63–76.
- Wang, F. and Luo, W.** 2005: Assessing spatial and nonspatial for healthcare access: towards an integrated approach to defining healthcare professional shortage areas. *Health and Place* 11, 131–46.
- Webber, R.** 2004: *Neighbourhood inequalities in the patterns of admissions and their application to the targeting of health promotion campaigns*. London, UK: CASA, University College London.