

Applications of Population Counts Based on Administrative Data at Local Level

Gillian Harper · Les Mayhew

Received: 17 December 2010 / Accepted: 24 February 2011

© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract The House of Commons Treasury Committee (2008) identified the three main purposes of population statistics as resource allocation, population ratios and local planning and intelligence, and described current population estimates as ‘unfit for purpose’. This paper considers the limitations of current population statistics in the context of these purposes and how using administrative data available at local authority level can overcome them. A sequel to a companion paper in this journal dealing with the population counting methodology, it considers how administrative data can be captured and organised at a local level for the benefit of local service providers and communities. A detailed example based on one London local authority shows the results of this approach in applications related to the three purposes given. A concluding section reflects on the wider applicability of the approach.

Keywords Local population counts · Census limitations · Key purposes of population data · Use of alternative administrative sources · Application example

Introduction

On the face of it, why we need to count populations seems a question hardly worth debating. After all, the first UK census was in 1841 and it has continued every 10 years except for 1941.¹ Debate today is much more about how accurate and timely the

¹With the exception of 1966 where a trial interim 5 year census was carried out using a short form for every household and a long form for a sample of households

L. Mayhew

Faculty of Actuarial Science and Insurance, Cass Business School, City University, London, UK

G. Harper · L. Mayhew (✉)

Mayhew Harper Associates Ltd, London, UK

e-mail: lesmayhew@googlemail.com

G. Harper

e-mail: harpergill@gmail.com

URL: www.nkm.org.uk

counts need to be. The answer depends entirely on one's point of view and the purpose for which the information is needed. How many lifeboats should an ocean liner be equipped with needs a precise answer for obvious safety reasons; for retailers, used to uncertainty, population is arguably secondary as compared with market share and profits. For public service organisations, identifying the 'right' population depends on, for example, whether it is people who will contribute or require resources, whether it is a day-time, night-time, temporary or long-term population.

What is clear today is that there is a demand for ever timelier and more detailed population data to satisfy a growing thirst for population intelligence. We have reached a point in which most service organisations know in detail who their customers are and where they live, but are much less sure about how many others there are like them in the wider population and what other services they use or need. However, part of the thrust for more detailed and accurate population statistics is political in origin—for example to strengthen local democracy, encourage joint working across public sector boundaries, and generally encourage greater information sharing. This is partly in the name of greater efficiency and effectiveness but also, ultimately, to serve wider social objectives such as reducing health inequalities, promoting social cohesion and protecting the vulnerable and so forth.

The House of Commons Treasury Committee (May 2008) was damning in its assessment of the UK system for counting the population, which it described as 'unfit for purpose'. It usefully set out what it considered the three main purposes of population statistics. They were to:

1. Allocate resources based on the distribution of the central Government grants to countries of the United Kingdom, to local authorities, health care providers, police and other services at local level
2. Provide denominators to construct ratios such as the number of crimes committed per head of population, unemployment rates and so on, and hence to evaluate policy at a national level (but also at a local level)
3. Plan, deliver and evaluate services at a local level taking into account need and demand.

This is not to refute that there are many other purposes besides. However, in any of these cases, there can be no doubt that an inaccurate population figure could skew resources, by giving a false picture about an area. In a companion paper in this volume (Harper and Mayhew 2011), we have set out an alternative methodology for counting populations using routinely collected administrative data which we define as data not primarily collected for statistical purposes (Vale 2006). Although the greater use of administrative data has been talked about for years (e.g. see Ericksen and Kadane 1986; Brackstone 1987; Steffey and Bradburn 1994; Penneck 2007; Keohane 2008), there has been remarkably little progress in the UK in implementing the necessary changes to statistical systems, or in exploiting the potential of these alternative sources for a combination of reasons. In addition, very little of this debate has percolated down to the local level which arguably is the level of government that stands to benefit most from more accurate and detailed population counts.

In this paper, we show how administrative data collected at a local level can be used to overcome the significant weaknesses under the current arrangements identified by the Treasury Committee. "[Limitations of Official Population Statistics](#)"

briefly considers the limitations of presently available population statistics in each of the three core purposes identified; “[Data Structures Using Administrative Data Sources](#)” considers how administrative data could be structured for statistical purposes to overcome these limitations; “[Application Example](#)” provides a worked example using actual data that covers aspects of the main purposes of population statistics. Our focus is on the local level (local authority or below) although a national perspective is introduced where relevant; for example, it is sometimes necessary to understand the process from a national perspective (e.g. in the case of resource allocation) in order to understand the local ramifications. The wider adoption of our approach at other levels of government is critically appraised in a concluding section which considers what has been achieved and future directions for development.

Limitations of Official Population Statistics

Resource Allocation

Resource allocation may be defined as the process by which the public sector allocates resources to activities and areas based on specified objectives in circumstances in which the process cannot be entrusted to market forces (Barr 2004). Resource allocation formulae typically need to capture, directly or indirectly, the need or demand for a service per unit of population, in which population is the scaling factor which must be combined with the unit cost of a service or the total quantum of resource to be distributed to territorial units, administrative areas or service delivery organisations. The particular formula used will depend on the population served and the underlying policy objectives which are integral to achieving wider social objectives—for example, to reduce health inequalities, to combat crime or raise educational standards (Marmot 2010).

In spite of their increasing complexity and the number of other variables built into such formulae, good population estimates are crucial elements of the process for distributing resources fairly or equitably—whether it is school places, police on the beat, or health budgets. The NHS was one of the first public sector organisations to use a formulaic approach to resource allocation based on a weighted population for distributing health care budgets down to regional level and below (RAWP 1976; for review of history see, Bevan 2009). However, within this broad canvas of applications and approaches, a general distinction can be drawn between area-based funding models, and those which allocate resources to organisations that deliver specific services such as schools.²

Area-based funding differs because it allocates block grants to geographically bounded administrative units such as local authorities which in turn use the funding to deliver a range of services. The Local Government Finance Settlement is a good example of the use of population estimates, but also of how other factors such as deprivation and specific local circumstances are taken into account. There is in addition an emerging trend towards funding models which allocate resources direct

² A further category would be allocations at a household level such as educational vouchers, benefits or budgets for personal care all of which are eligibility based.

to delivery organisations (e.g. schools, primary care). Through the need to use more detailed and timelier population data, these processes have progressively exposed weaknesses and anomalies in population statistics. Such data limitations may partly explain why formulaic approaches to sub-local authority resource allocation processes are uncommon, with much more regard being given to local judgement and politics.

The problems with population data are essentially of two kinds. The first is its inflexibility; for example, geographical units are often unsuited to the applications that depend on it, and specific variables such as age bands are in fixed formats, making it difficult to identify the demand for services based on non-standard age groups. However, more fundamental is the poor quality of the data itself which can be traced to a combination of low response rates during the previous census in 2001, especially in inner city areas, and subsequent population fluxes through migration (Simpson 2007; Simpson and Brown 2008). The consequential undercounting of population has meant that some areas have effectively experienced nearly a decade of underfunding since the last census in 2001 (e.g. see Mitchell et al. 2002; Dorling 2007; LGA 2007). Lawrence et al. (2007), working in Brent, a suburb of London, found for example that population undercounting potentially equated to a loss in revenue of an estimated £40 m per year for the primary care trust (Lawrence et al. 2007).

Use as Denominators

The second class of applications, ratios or related indicators based on rates, has more to do with providing a societal barometer or dashboard of indicators for economic management, policy evaluation or other applications. Ratios or rates are used in numerous contexts (employment, health, crime, education etc.), and have been used increasingly by governments and agencies for setting local targets for deliverers of related services or holding local authorities to account. They are usually expressed as percentages or rates per thousand of the population that are exposed to a particular outcome or risk such as unemployment or disease, either incidence (new cases) or prevalence (all cases). Population denominators are also needed for calculating life expectancy which is widely used to measure health inequalities.

In public health applications, attention is properly directed toward the ascertainment of accurate numerators (e.g. the number of MMR vaccinations in children, women who are breast screened). Unlike numerators which tend to rely on administrative counts through case control and reporting systems, denominators are arguably as great a source of inaccuracy. Issues arising include statistical imprecision of population counts, appropriate choice of administrative boundary, breaks in time series, a lack of contemporaneousness, or an inability to measure the population at risk due to lack of specificity in the data (e.g. in terms of age, sex, ethnicity, housing).

In epidemiological studies, good reporting systems are obviously essential for counting the numerators, but good denominators are needed also for measuring vulnerable sub-groups such as ethnic minorities or recent arrivals to the country (e.g. see Roderick and Connelly 1992; Hayward et al. 2010). The NHS in England, for example, recommends different levels of medical provision based on TB incidence

rates and so accurate data are crucial in order to calibrate appropriate levels of medical need in an area to combat this socially corrosive disease (NHS 2007). However, alternatives to population denominators can be considered when population information is unavailable or unreliable, for example the use of satellite imagery, although clearly this suggestion would not be appropriate in the TB case or many other applications of a similar nature (Viel and Tran 2009).

Newcastle City Council in the UK, for example, argued for the use of residential properties as the main denominator when creating neighbourhood rates (e.g. crime rates per 1,000 properties, rather than persons). The number of residential properties is available from the council's business and residential property gazetteer. Among the advantages claimed is the high quality of the data, that it is regularly updated and reflects changes on the ground (e.g. new builds, demolitions, and conversions), and that residential properties are identifiable separately from business properties. In addition, the council has control over the data so the denominator matches the period of the data for the numerator.³ However this may not be useful where the subject of interest is people rather than properties or households. Our approach also uses property data, but a key difference is that we link property data to administrative data so that we can construct population as well as property ratios in the denominators.

There is also potential for unwelcome interactions between a false numerator and an inaccurate denominator to produce perverse results. A national indicator used by the Government in England until recently provides local measures of the rate of hospital admissions for alcohol-related harm for every 100,000 members of the population. It uses the concept of 'attributable fraction' which is assigned to patients entering hospital based on how much of their condition may be related to alcohol consumption. The calculation is a function of relative risk estimates and population drinking estimates, and therefore relies on the accuracy of population estimates of alcohol consumption and the availability and quality of the relative risk estimates reported in the epidemiological literature (Jones et al. 2008). As a consequence, it is likely that the method overstates the alcohol harm in some areas and understates the harm in others.

In summary, the danger of using misleading ratios is potentially exacerbated where ratios are used as management targets and result in resources being redirected. The lesson of the last 10 years is that management ratios need to be defined and used with caution including where there is scope for error in both numerator and denominator. It is noteworthy that the new Coalition Government has rescinded the use of targets as a means of control over public funded services and organisations and so the consequences of uncritical applications of ratios are less than previously, although the reasons are primarily political and not data driven. Nevertheless, ratios remain one of the few means of comparing one area or organisational unit with another and so the more that can be done to improve the data on which they are based the better the outcomes are likely to be.

Use in Delivery of Local Services

The third class of applications concerns the design and delivery of local services. Arguably, this is the most challenging of applications as it is much harder to

³ <http://www.newcastle.gov.uk/core.nsf/a/nnispop> [date accessed: November 2010]

fudge the data. Local authorities in the UK are responsible for supplying local public services such as schools, libraries, public leisure facilities, collecting Council Tax taxes, maintaining electoral registers and managing local public facilities and infrastructure. They are expected to work in partnership with the police, emergency services and health care providers. Each has its own information systems which capture many features of local areas, including the built environment, frequently employing GIS (Geographic Information Systems); however, these are not linked together into a unified system and it is not unusual for different departments of local authorities to use unharmonised data including different population data sources.

In most cases, these systems capture data only on users and not on the population as a whole i.e. people that do not use the service as well as those that do; however, complete information about a population is normally required to identify gaps, undertake needs assessments, or to identify hard to reach groups such as older people living alone. Because management of public services is predominantly carried out at local level, population statistics must be capable of supporting this role. 'With local government in a key 'place-shaping' leadership role, it is vital that every opportunity is taken to refine and improve the available information used to gauge crucial decisions' (Keohane 2008, p.4). Similar arguments can be set out for health and police services.

For previously stated reasons, population statistics are a long way off meeting this requirement. Local authorities need far more flexible information than is available so that they can answer questions such as: what are the population and deprivation levels for a given housing estate? How many single parents live in social housing and are on benefits? How many nurseries are there within pram pushing distance of households with young children? Who needs to have face to face contact with local services? Are there vulnerable groups that need more personalised services and how many are there (e.g. older people, single parent households, and ethnic groups)? In the following sections, we consider the necessary changes to the system of collecting population counts needed to meet these challenges.

Geography as a Barrier

We have seen that the drive for more detailed local information puts heavier reliance on local population estimates. At a national level the percentage error in a population estimate is believed to be small, but broken down into small spatial units at the level at which service providers require accurate information, errors are magnified (see Harper and Mayhew 2011). Pre-determined spatial units, ranging from electoral wards down to Census output areas, may not correspond to the areas that users are interested in, which may be housing estates, brown field development sites, town centres etc. In this section we shall argue that users need greater control over the data to link and merge different sources and also to have greater control over geography (i.e. the spatial building blocks on which decision-making is founded).

A good example of the problems caused by geographical inflexibilities of population data is the Sure Start programme for young children. Sure Start is the previous Government's programme (1997 to 2010) to improve the development

prospects for young children by co-ordinating and streamlining services for this age group. In evaluating the impact of Sure Start to see if it had met its objectives, it was found necessary to adjust or apportion data that did not match pre-determined boundaries (Frost and Harper 2007). As Harper (2002) noted, apportionment techniques are the only option for users that require non-standard breakdowns of data which are far less accurate as a result. It is therefore arguable whether it will ever be possible to conduct robust evaluations of government initiatives as long as this arrangement persists.

History tells us that statistical boundaries of administrative areas are subject to alteration making it impossible to create an accurate picture of change through time. So one can argue a key requirement for any new system of population statistics is that it must be flexible so that the population of any area may be determined swiftly, accurately and simply. However, even if the estimates were accurate and boundaries are unchanged, there would still be problems unless data are collected in ways that can be flexed to deal with boundary change. Having a single stable geography to suit all needs is arguably unrealistic. Moreover, it is noteworthy that attempts to change definitions and initiate new collections tend to become bureaucratic and unwieldy for a range of reasons.

The inflexibility of administrative geography is also associated with analytical problems of measurement and interpretation. There are two main effects which we now discuss in order to press home the case for change. The first of these is known as the 'ecological fallacy' (e.g. see Greenland and Robins 1994; Openshaw 1984a). This is based on an error in the interpretation of statistical data, in which inferences about the nature of specific individuals are based solely upon aggregate statistics collected for the group in a geographical defined area to which those individuals belong—in other words projecting on to the individual, generalizations that apply to a population. In extreme cases, this may have unfortunate consequences in terms of false attribution of causality as well as association, for example, a high crime area may have a high number of single parent households, but single parents are not the cause of crime.

The second class of effects is known as the modifiable areal unit problem (or MAUP). This is a source of statistical bias that affects statistical hypotheses by causing the correlation, or association, between two variables to vary widely (first noted by Gehlke and Biehl 1934; see also Openshaw and Taylor 1981). It arises when point-based measures such as the number of people at an address are aggregated into districts or zones so that summary values (e.g., totals, rates, proportions) are heavily influenced by choice of boundary. As Openshaw (1984b, p3) laments, "the areal units (zonal objects) used in many geographical studies are arbitrary, modifiable, and subject to the whims and fancies of whoever is doing, or did, the aggregating."

The problem is that we do not know to what extent either the ecological fallacy or MAUP bias decision-making processes unless we have more flexible data. The obvious solution is to create data that can be aggregated into spatial units of any shape or size through using geo-referenced point data at a household or person level. This approach would give users flexibility to choose their own boundaries, add new data from their own sources by person or household linking; it would have the effect of improving accuracy and timeliness and overcome the necessity for apportionment;

the potential for ecological fallacy could be minimised by using individual level data (Tranmer and Steel 1998; Mayhew 2002); and potential MAUP problems could be investigated using thorough sensitivity analysis.

Use of Administrative Data in Practice

We have argued that knowing the demographic and household characteristics of local populations is a key requirement for policy purposes. Better data help to improve local services and decision making in various ways by improvements to quality, fairness, and value for money. Different sub-groups of the population have different needs and risks; for example young children, older people and single-parent households (Coleman and Schofield 1986). The current arrangements for collecting information about the population may have served the UK well in the past, but with today's more exacting applications we therefore fully concur with the Treasury Sub-Committee that they leave much to be desired.

As Freedman et al. (2008) argue, local level analysis is critical for local level decisions and policy in which service providers require information not only about costs and volumes of the services delivered, but also about need (how much input is required based on individual requirements), and risk (e.g. what is the probability of an adverse event such as a fall leading to injury with and without the provision of social services). Since the alternatives set out in this paper are based on locally owned data sources, the cost of compiling the data would be easily afforded by the main users. The barriers to producing better population data are, it is argued, not technical but organisational and bureaucratic.

A key requirement is to have data sharing arrangements between data owners in which data are processed in a secure environment or 'safe haven' by a small unit comprising skilled analysts who would be legally bound by data confidentiality. Their role is to geo-reference and link data at a household or individual level, and tabulate and anonymise for statistical purposes. A linked set of administrative data covering local key data sets would provide a platform for more responsive analytical services using better quality data, and more timely population intelligence to support council and local health services. When combined, the data can add greatly to what is known about a population's characteristics. For example, the Annual School Census contains much more detailed ethnic categories than those used by many authorities based on the decennial census.

There are also cost savings because of less reliance on external data sources, product licences etc., less duplication of analysts across organisations through scale economies, and enablement of data sharing with partners e.g. at district level and among health providers. The wider benefits of better data management across partner organisations are more efficient services through joined up working and, ultimately, better outcomes for local people. Figure 1 sets out how a system based on administrative sources would dovetail with present arrangements in which locally available data would be processed and used to support local decision-making as well as providing information at higher levels of government.

In this scheme, data provided by the local community (box C) to service providers (Q) is stored in administrative systems (A). Such systems currently provide data to government departments by way of (B) and in turn these data are processed by

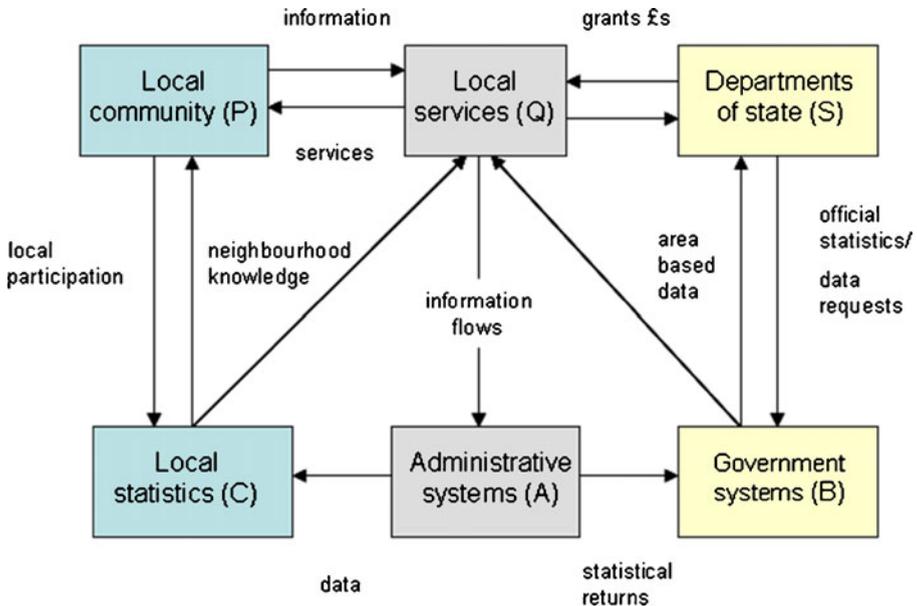


Fig. 1 The flow of administrative data at national and local level

departments of state (S) and used to create and evaluate policy and allocate funds accordingly. Statistics are fed back to local areas principally in the form of geographically aggregated data in appropriate administrative units (e.g. local authorities and below). The new feature of our approach is that (A) would be used to generate local statistics and intelligence through (C); (C) in turn would be used to create ‘neighbourhood knowledge’ that would be fed back to local services (to enable them to perform better) and to the local community (to enable them to participate in decisions about local service as appropriate).

Integral to any new system would be locally available data which, as well as supplying central government with its information requirements, would be exploited and used directly at source. In the next section, we compare the structure of current official population statistics with what can be made available using local administrative data. We do not go into detail here about sources of local administrative data, which are numerous, but instead concentrate on data structures and differences with present arrangements (however, see companion paper for examples).

Data Structures Using Administrative Data Sources

The main characteristic of data structures typically provided in official population figures is that they are based on territorial units, whereas in the method described here they are provided at an individual or household level. The main units used by statisticians are Electoral Wards, Census Output Areas, Super Output Areas and Postal Districts, Postal Sectors or individual postcodes. Hence, whereas a typical local authority may have thirty or so wards and each unit postcode (the smallest

administrative unit) could have up to eighty addresses,⁴ the administrative alternative would have one row per individual up to the population of an area, perhaps 250,000 records or more. The information contained about each individual would be coded in such a way as to facilitate statistical analysis but would not be cluttered with the myriad of other administrative information held on source data bases.

The broad structure of current official population data published by the Office for National Statistics (ONS) is shown in Table 1. Variables of interest are specified by the content of the forms used in the Census and published as pre-determined tables. Each variable is an aggregate, such as the number of people aged 85+. Other official data include data sets based on national administrative sources, but these are not fully integrated with ONS data or necessarily harmonised geographically, or by reporting period, or available contemporaneously. Examples include hospital admissions data, National Insurance registrations and social security data. Some administrative data measure stocks at a point in time and others flows (e.g. migration); time periods are variable (e.g. financial year, calendar year, school year); and publication dates are variable.

The data structure based on administrative data methods described here is shown in Table 2 and this is the basis for the approach we have adopted in the application example in “[Application Example](#)” below. Each row is an individual and each column a variable of interest. The nearest equivalent under present population data arrangements would be the census Sample of Anonymised Records (SARs), but as well as being out of date, these do not contain information that would routinely enable other data to be linked. Typically in the administrative data method there would be an area code which could represent an existing administrative unit or one designed by the user that the individual resides in. This would be followed by demographic information and include for example exact age and gender; all records in addition would be anonymised.

Subsequent columns would contain variables derived from a range of administrative sources. The data held in administrative data bases essentially comprise four types: (a) categorical or fixed variables such as gender or ethnicity or date of birth; (b) event variables such as the date of a visit or transaction or a birth or death; (c) flow variables comprising the date a service or payments began and ended; (d) the quantum of service provide (e.g. hours of care, meals provided, childcare sessions, day attendance at a care centre and the costs thereof). Not all of this detail is strictly needed for statistical purposes but much will depend on the requirements of the user, with different users having access to different fields as deemed appropriate by local policy and legal requirements.

Neither does the level of detail have to be onerous from a data collection or data definition perspective. Often all that is needed is binary information, for example whether a person lives in social housing or not, categorical data such as household type, and numerical data such as the total number living in a household. Beyond these basic measures a host of other variables may be added from a range of sources such as whether a person lives in a household on benefits (a proxy for low income) and Council Tax band (a proxy for housing

⁴ <http://www.royalmail.com/portal/rm/content1?catId=400044&mediaId=9200078#3400054>

Table 1 Typical structure of currently available official population data

area	area code ID	population	age group	variable 1	variable 2	etc.
1						
2	<e.g. SOA>	<number of units>				
3						
.						
.						
n						

wealth). While these variables are not identical to census outputs, they offer valuable individual and household level socio-economic information that is both detailed and precise. Exact comparisons with the range of information available in the census are impossible, since it will depend on how many administrative data sets and therefore variables are included. However, there are some obvious examples such as religion, place of work which are variables included in the census but not in any comprehensive local administrative data source. Conversely, administrative data is far more comprehensive in areas such as such as benefit status, education, crime, housing, health care and so on.

The *x* and *y* columns in Table 2 are geographical references so that populations or combinations of variables can be analysed and mapped geographically. These are ascertained by linking a person’s address to the Local Land and Property Gazetteer (or equivalent) and extracting the Easting (*x*) and Northing (*y*). Usually access to *x* and *y* co-ordinates are limited to those that work in a GIS environment and who are usually the custodians of the data. Other users of the data may have more restricted access depending on the context and sensitivity of particular variables or pieces of information (e.g. certain crime or health data). The important point is that the master data base is flexible and users’ access can be designed and tailored appropriately to their needs.

The final data set contains additional variables that are constructed from the base data either for individuals or households, such as the number of co-residents in a household. For instance, for some purposes households are a more appropriate unit

Table 2 Typical structure of data bases using administrative data sources

person	area code ID	age	gender	<i>x</i>	<i>y</i>	variable 1	variable 2	etc.
1								
2	<e.g. SOA>							
3								
.	.							
.	.							
n								

of analysis than individuals, in which case household level variables are devised by counting or summarising variables by the property reference number assigned to each individual. For example, if five individuals on the database share the same property reference number, it is inferred that the occupancy of that property is five. To meet this need, we have developed an eight-fold household classification scheme based on individual household demography which is described in the next section. This uses descriptors such as family households with dependent children, older cohabiting households, 3-generational households and so on. These can be broken down into more detailed sub-types as required or users can even create their own classification instead.

There are many other non-core data sets that can be used to enhance the data base. In local authorities these include information from service users of adult social care, libraries, educational data and data on children and families. To these can be added a range of NHS and other data sets, each of which contain much information of potential value. These include data on community health services, or hospital admissions, although special arrangements are usually needed to gain access to some of these sources depending on the application. Similar considerations apply to more sensitive information such as crime data or personal medical records; however, a discussion of the details pertaining to data access and related issues in these cases is outside the scope of this paper.

The other main source of local data, which should be mentioned in passing, is survey information i.e. information obtained from specially commissioned surveys. The advantage of surveys is that they collect precisely those data that cannot be obtained from administrative sources. Examples include qualitative and attitudinal data of one kind or another such as willingness to give up smoking or optimism about the future or satisfaction with local services. For example, a health and well-being survey commissioned by one health authority sought information on a person's self-evaluated state of health, drinking habits, income, cohabitation arrangements and so forth. Although usually based on a small sample of residents, such data can be linked to administrative data and used to impute and infer social and other characteristics of whole populations; however, a description of the methods and assumptions involved are also outside the scope of this paper.

We may generalise these statements by adding that virtually any data set could be appended provided they can be linked accurately to individual records by means of a shared identifier. For instance, some commercial data sets such as loyalty card customer data which may provide valuable information on shopping habits and expenditure patterns. In general, the most useful data would therefore be sources that had the potential to fill gaps and were known to be of high quality; however, the most important barrier to obtaining access to such data sets for statistical purposes is likely to be their commercial confidentiality. Finally, although we have stressed that it is desirable that administrative data should be linkable at a person level, it is possible to contemplate versions of data sets in which linkage takes place at a household (address level) or at higher levels (e.g. output areas). We now turn to a worked example to show how administrative data can be used in actual applications, and explain why such applications would not be possible if users had to rely on existing population data.

Application Example

Following the lead provided by the Treasury Committee, we use this section of the paper to demonstrate examples of applications using administrative data in the three main purposes suggested in their report: (a) to allocate resources; (b) provide denominators to construct ratios such as the number of crimes committed; and (c) plan, deliver and evaluate services at a local level taking into account need and demand. Whilst we cannot provide an example of resource allocation based on the whole country, we can demonstrate a case study that exemplifies each of these purposes at a local level (i.e. typically populations up to 300 k). The principles involved are no different from those at a regional or national level, although of course the analysis will be more disaggregate as a result and therefore more relevant to local decision-makers.

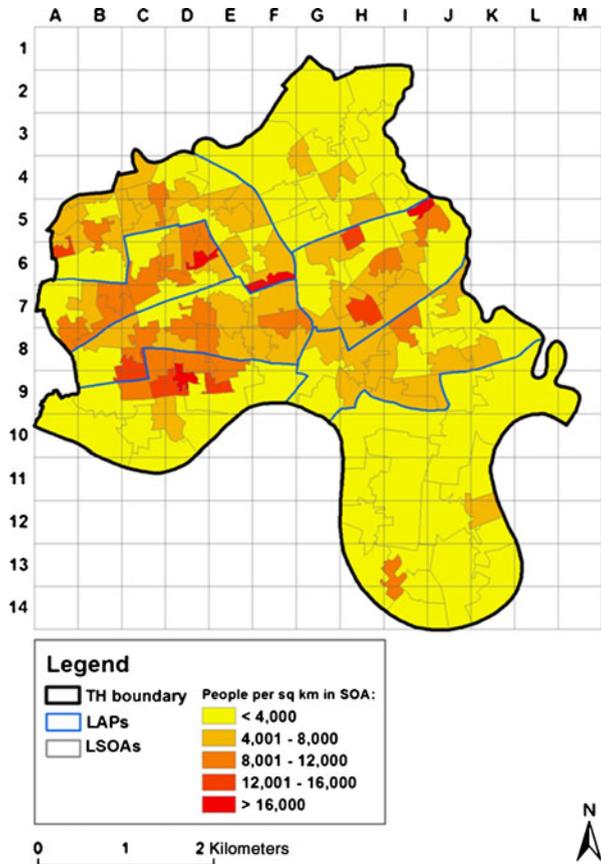
The key point is that data structures enabled by the use of administrative data allow common methodological approaches, regardless of geographical scale and across sub-populations whether at a household or some other spatial level for any given purpose. In addition, through their greater flexibility they help to minimise the danger of MAUP issues or problems of false correlation. We will illustrate the methodology by referring to a health related study conducted in the London Borough of Tower Hamlets concerning the take-up of free NHS eye tests among older people. Eye testing is simply one of a host of public services that requires people to attend a location to receive a service and so the principles are general even if the details of each service differ (e.g. whether the service is discretionary such as an eye test or compulsory such as education for 5 to 16 year olds). Such distinctions do not affect what follows although it may influence the technical procedures of how resources are allocated.

Background to Case Study

Tower Hamlets is a densely populated inner London borough, located to the east of the City of London financial centre and bordering the river Thames including Canary Wharf to the south. The borough is ethnically diverse, but also diverse in terms of income and wealth, being on average one of the most deprived boroughs in the country. Using local administrative data sources, we estimated the population in 2009 to be 234,828 people, living in 100,995 dwellings (Mayhew and Harper 2010). The biggest ethnic group are the Bangladeshi community who account for about 32.1% of the population, white British and other white 30.8%, and the rest a mix of Black, other Asian and mixed origins. Figure 2, for example, is a density map of the Bangladeshi population by Lower Super Output Area (LSOA), which has been constructed from administrative sources and is provided as one of many possible illustrations of the descriptive detail attainable.⁵ However, it also turns out that the Bangladeshi population takes

⁵ Ethnicity is assigned to individual records in the data base using a combination of school census records, hospital admissions data and a purpose developed surname recognition algorithm (not discussed here). For this reason the assignment is probabilistic on a scale of 0 to 1 and population counts by ethnic group are constructed on this basis. A full discussion of the methodology is outside the scope of this paper.

Fig. 2 Density of Bangladeshi population in Tower Hamlets by Lower Super Output Area (LSOA) (Contains Ordnance Survey data © Crown copyright and database right 2010, and data sourced from London Borough of Tower Hamlets) Note: LAPs are Local Area Partnerships



up eye tests more than other groups and why this is so is also of keen interest to health commissioners.

By way of further background, Table 3 shows the household structure, population and benefit status of all households in the local authority based on the same administrative data and methodology. This household structure is demographically defined with regard to the age of the occupants of households and is distilled from 81 different sub-types.⁶ The table shows that income deprivation is particularly concentrated in household types A to E (see key beneath table), which account for 60% of the population and 36% of households. Around 58% of households in categories A to E receive means tested benefits as compared with an average of 32% of all households, and 55% of households in categories A to E are in social housing.

⁶ This classification is not to be confused with commercial products such as MOSIAC or ACORN which also classify households into types. Note that these products rely on a combination of census data and other data sources and use multivariate techniques to produce synthetic rather than demographically based characterisations.

Take-up Rates of Free Eye Tests Under the NHS

We wished to focus on the older population and its access to free eye tests under the NHS.⁷ We considered the take-up of eye tests in this group relative to their geographical access to eye testing centres. Lastly, we asked the question by how much eye test take-up would increase if the geographical access to eye tests could be improved by allowing further centres to be opened through enabling GP practices to offer their premises as locations for free sight tests. The map in Fig. 2 is relevant because it happens that the Bangladeshi population tends to live closer to eye testing centres than other sub-groups and this appears to give rise to a higher take-up of this service. The ‘LAPs’ in the legend are Local Area Partnership areas which are local authority subdivisions based on aggregations of wards. For some purposes these are the fundamental local unit of resource allocation.

In the UK the National Health Service provides help with the cost of glasses based on age, health or income. Eye tests provided by optometrists are either free under the NHS or must be paid for privately. Entitlement to a free eye test is granted where a person is under 16 (under 18 if in full time education), or aged 60 or over. Free eye tests are also available to people diagnosed with diabetes or glaucoma, or who are advised that they are at risk of glaucoma, who are registered as blind or partially sighted or who are being treated in hospital for an eye condition or who are being prescribed contact lenses. Entitlement is also extended to people in receipt of certain social security benefits, or people aged 40+ whose immediate relatives have been diagnosed with diabetes or glaucoma. Older people are the largest users of this service, but also the most likely not to be tested if the service is inaccessible.

Based on NHS data, Tower Hamlets has the lowest take-up of free eye tests anywhere in London among older people. In 2007–2008, take-up among the 60+ population was 17.8% as compared with a London average of 37.5%, although by 2008–2009 this situation had improved somewhat. However, the true figure may be lower because of known problems with population estimates for this borough. The local primary care trust (the commissioners of these services) were concerned to understand why take up was so low and what could be done to improve the situation. The study ranged widely into areas of epidemiology, aspects of service provision and so forth; but one strand of enquiry concerned the level of geographical access to sight testing centres.

Geographical Access to Free Eye Tests

Although a densely populated borough, ease of travel is variable and the locations of eye test centres tended to be skewed towards well established commercial areas in the middle and west of the borough. Figure 3 is a map showing the locations of eye testing centres and the locations of all households with a person aged 60+ living there. Each household has been colour coded according to whether there are 0, 1, 2, or 3 or more centres within a 500 m radius (10 min walk time) of each household

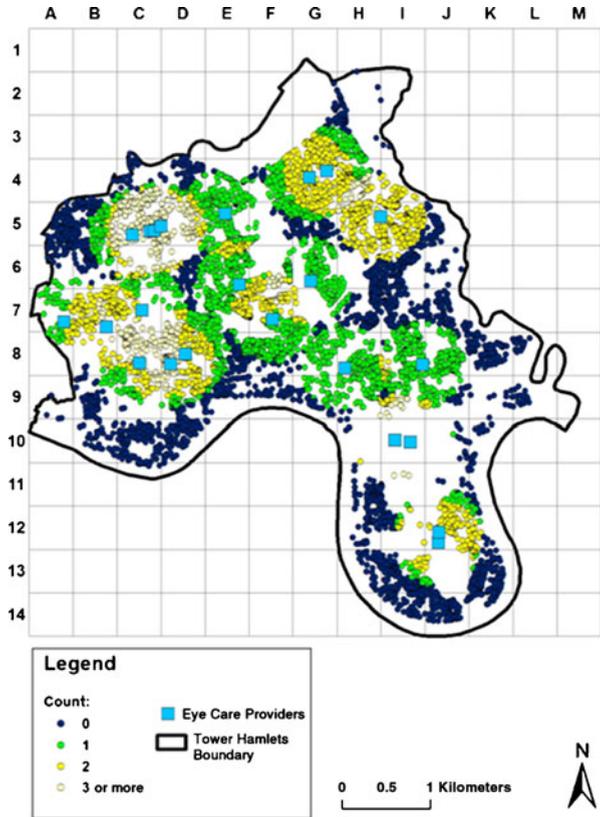
⁷ This section draws in small part on a project undertaken in conjunction with PHAST (Public Health Action Support Team) on behalf of Tower Hamlets PCT. Care Needs Assessment: Eye Health, findings and recommendations. Lead author M. Simons, 2009. See also www.nkm.org.uk

Table 3 Household structure, population, tenure and benefit status

Household type	Frequency by household type	Population by household type	Number of households on benefits	% of all households	Social housing household type	% social housing	Category	Description
A	17,337	87,635	9,234	53.3	9,321	53.8	A	Family households with dependent children
B	6,819	19,311	3,806	55.8	3,420	50.2	B	Single adult households with dependent children
C	4,530	11,158	2,570	56.7	2,384	52.6	C	Older cohabiting households ⁽¹⁾
D	5,389	5,389	3,539	65.7	3,565	66.2	D	Older person living alone
E	2,530	16,696	2,032	80.3	1,466	57.9	E	Three generational households ⁽²⁾
F	21,703	50,699	3,259	15	4,071	18.8	F	Cohabiting adult households no children
G	41,305	41,305	6,940	16.8	8,128	19.7	G	Single adult households
H	1,382	2,635	503	36.4	470	34	H	Other households
Total	100,995	234,828	31,883	31.6	32,825	32.5		

(1) At least one resident must be aged 65+; (2) At least one resident must be aged under 20, and one aged 65+

Fig. 3 Geographical access to eye testing centres based on 10-minute walk time or 500 m. Round symbols indicate locations of households with one or more persons aged 60+ (Contains Ordnance Survey data © Crown copyright and database right 2010, and data sourced from London Borough of Tower Hamlets)



(across the border centres are excluded). The lightest coloured symbols have most access and the darkest symbols are homes with least access. Those with least access are spread throughout the borough but especially along a strip bordering the River Thames to the south and in patches elsewhere (e.g. cells A5 and B5 and I6 and I7).

We started by examining whether older people are more or less disadvantaged in terms of access than other sub-groups of the population. Under the National Assistance Act 1948 local authorities have a statutory duty to maintain a register of people living in their area who are visually impaired. A person does not have to register as blind or partially sighted, but if they do they may be entitled to certain benefits and services. We analysed the register and found that a person was nearly 13 times more likely to be registered if they were aged 60+.

Using the population data base, we segmented the whole population into one of 16 mutually exclusive groups with each group sharing similar attributes. In this case it included whether a person is aged 60+, living alone, is *not* Bangladeshi, and private housing tenure. We call these risk factors because they act as markers whose influence can be quantified using regression techniques. Note that it would be possible to define other types of risk factors which could include for example a range of clinical risk factors (see for example Alder et al. 2005); however, our purpose here was to look at differentials in access.

Table 4 shows our results in which those groups with the least access are ranked first and with greatest access last. The numbers in each category are given in column one, and the levels of access and 95% confidence intervals in the final three columns. The intermediate columns indicate the presence or absence of an attribute by the symbol ‘Y’ and column totals give the total population, the population aged 60+, the numbers in private tenure etc. The results show that levels of access range from 20.5% living more than 500 m from an eye test centre in the best case (row 16) to 46% in row one (worst case). The Tower Hamlets average is 37.1%. This form of tabulation, known as a ‘risk ladder’, is only possible using linked data. The results show that confidence intervals are acceptably tight around the central estimate and that they capture access differentials succinctly.

Using logistic regression techniques, we ascertained which particular groups were the most disadvantaged. We found that a person was 1.2 times more likely to live more than 500 m from a centre if living in private tenure, 1.1 times more likely if living alone, 1.6 times more likely if *not* Bangladeshi (all significantly different from one at the 95% level of confidence). It followed that the more disadvantaged groups were likely to be non-Bangladeshi, those living in private tenure, and living alone.

Those aged 60+ were only 0.74 times as likely to live further than 500 m from the nearest centres (all coefficients significantly different from 1 at 95% level of confidence). This suggested that older people had better access than younger people but not as good as the Bangladeshi population which tended to be located nearer to

Table 4 Table segmenting the population of Tower Hamlets by access to eye test centres according to the given risk factors

Case	Number in category	Aged 60+	Not in social housing	Living alone	Not Bangladeshi	% >500 m from nearest eye test centre	Lower 95% CI %	Upper 95% CI %
1	32131		Y	Y	Y	46.0	45.5	46.6
2	68535		Y		Y	43.3	43.0	43.7
3	5178	Y	Y		Y	38.2	36.9	39.6
4	35345				Y	36.1	35.6	36.6
5	2451	Y	Y	Y	Y	34.8	32.9	36.7
6	6556			Y	Y	33.6	32.4	34.7
7	895			Y		32.9	29.8	36.1
8	4962	Y			Y	31.8	30.5	33.1
9	43746					30.4	30.0	30.9
10	956		Y	Y		30.2	27.3	33.2
11	1875	Y				29.3	27.2	31.4
12	26433		Y			27.6	27.1	28.2
13	4270	Y		Y	Y	27.3	26.0	28.7
14	1248	Y	Y			24.3	21.9	26.8
15	178	Y		Y		23.2	17.2	30.2
16	69	Y	Y	Y		20.5	11.7	32.0
total	234828	20231	137001	47506	159427	37.1	36.9	37.3

eye testing sites. The 60+ age group with the poorest access tended therefore to be people living alone in private tenure and *not* Bangladeshi; the reason why this may be important is that it provides a potential illustration of the inverse care law. This states that the availability of good medical care tends to vary inversely with the need for it in the population served (Tudor-Hart, 1971) especially if it can be shown that people that live farther away from a source use the service less.

Impact of Geographical Access on Take-up Rates

To understand how geographical access might affect the take-up of eye tests in the 60+ age group, we analysed 14,000 administrative forms filled in by optometrists after an eye test has taken place. These forms contained a range of other useful information including the presence of certain eye conditions such as glaucoma, and so we were able identify key risk groups. Our analysis showed that males, older people and Bangladeshis were significantly more likely to be diagnosed with glaucoma than other groups, and so older people were clearly one of the high risk groups.

We found that the ratio of people with diabetes receiving free eye tests to all those receiving eye tests was 8.1%, but this rose to 20.6% in the highest risk group (older males, and Bangladeshis). This compares with independent estimates for Tower Hamlets as a whole of 5%; however, it is not known whether the 3% margin of difference is a mixture of self selection or other effects (i.e. people having eye tests are more likely to have an eye condition). Overall, we found that the proportion of older people tested was twice the proportion of people aged under 16 tested, which in turn was twice the number of working age adults tested (i.e. 4:2:1).

Concentrating on the 60+ population, Fig. 4 shows the percentage take-up of free eye tests based on their distance from the nearest eye testing centre. It shows that around 35% of those living next to an optometrist will receive an eye test in a given year, but this then falls to around 25% at 500 m and to below 10% after one kilometre. Although take-up in this age range is greater overall because needs are greater, it was noteworthy that the amount of attrition (i.e. the fall off in take-up with distance) was higher than in other age groups, and is also higher than in the Bangladeshi population. To put this in perspective, a 60+ person living nearby an eye testing centre would be tested once every 3 years on average but this would slip to 5 or more years or longer if they lived further away.

Evaluation of an Alternative Service Configuration

In answer to the question what could be done to improve access, it is the role of service commissioners to consider the best arrangements for delivering health services. One suggestion was to use local GP practices. We therefore estimated what take-up would be likely to occur in the older group if an optometrist were to perform eye tests in existing GP practices, the argument being that GPs are more numerous and more evenly spread in the borough. In other words, would re-allocating resources to more convenient locations incentivise take-up in this high risk group? In doing so, we presumed that GP practices would be able to make space available and an optometrist would be able to travel between locations (a mobile service exists for

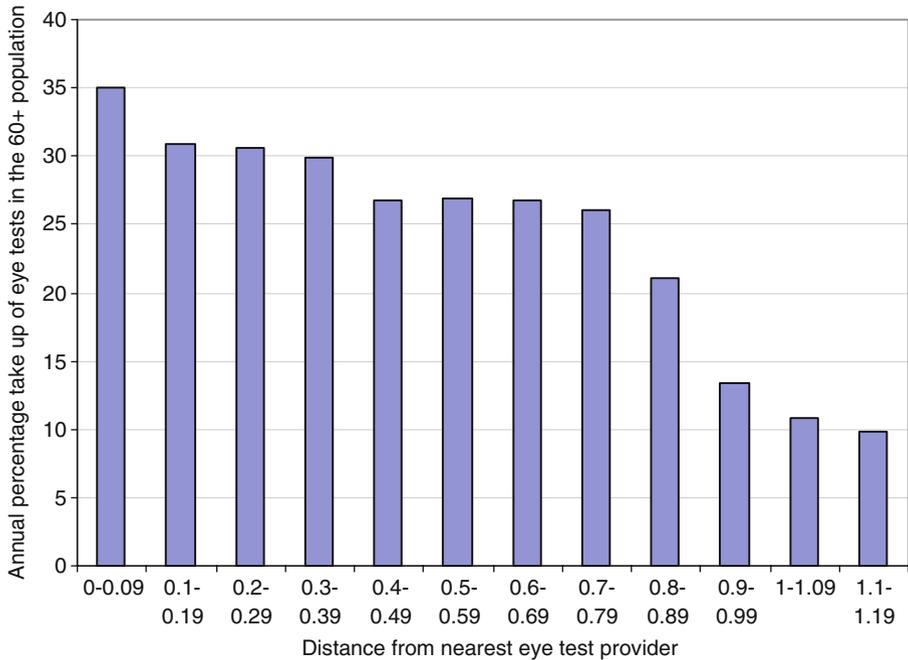


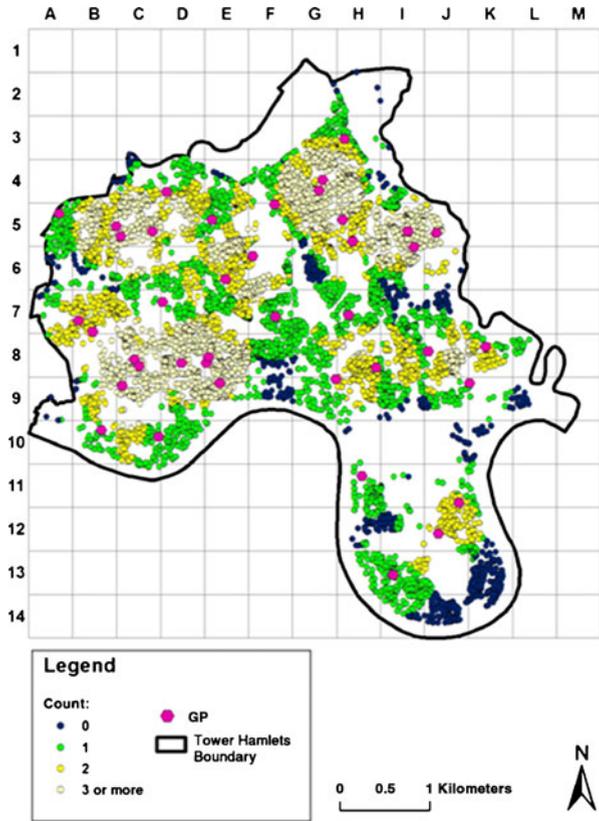
Fig. 4 Free eye test take-up in the 60+ population based on distance from nearest eye test centre

care homes but service levels are currently low). Figure 5 shows the geographic effects on access were this to occur on the assumption that travel behaviour would react to distance effects in the same way. As is seen, there would be a far greater equity of access throughout the borough as a consequence, but what would be the effect on take-up?

Figure 6 shows the predicted level of take-up following the hypothetical reassignment of the service to GP practice surgeries. It shows that access would be improved in the 0 to 500 m distance range and that the numbers having to travel more than 500 m would fall substantially. Overall we found that a re-configuration would improve take-up in the borough in the 60+ age group by 8% based on this argument and the rate of overall take-up by 2%. This would have the effect of improving the borough's position within London by a few places, but it would not be sufficient to lift it up to the London average. However, this predicted effect is predicated on the assumption that there would be no other accompanying changes. One such behavioural change arising from the opportunities of co-location would be that older people would seek eye tests on routine visits for clinical check ups at their GPs rather than having to make separate trips to different locations.

Thus, we can argue that the 8% improvement in take-up would be the minimum uplift attainable. Note that this analysis does not take account of the costs of re-configuration of the service or the willingness of optometrists to travel between sites. The cost implications however are likely to be fairly small relative to the benefits and since this is a geographically small borough it is maintained that optometrists would not be especially inconvenienced especially as there is already a small mobile service provided.

Fig. 5 Geographical access to GP practices based on 10-minute walk time or 500 m. Round symbols indicate locations of households with one or more persons aged 60+ (Contains Ordnance Survey data © Crown copyright and database right 2010, and data sourced from London Borough of Tower Hamlets)



What has been achieved using administrative data in this example? Firstly, evidence has been provided that, based on administrative data, distance attenuates the take-up of discretionary local services such as eye tests; secondly, that take-up is affected by demographic and socio-economic factors as well as medical need; thirdly, providing services in GP practice locations would go some way towards correcting the low take-up in vulnerable groups and estimates of the effect were provided. Similar findings might be expected from consideration of many other kinds of services; but what are the resource implications in this case?

Implications for Resource Allocation

NHS commissioners of these and other services must decide how services will be delivered and a number of funding mechanisms can be envisaged. One is that general practices would pay optometrists to visit their surgeries and so the question arises how GP budgets should be recompensed. This borough is sub-divided into eight Local Area Partnerships (LAPs), each comprising between 25,000 to 37,000 people. As previously noted, the LAPs are used for allocating resources for some services. We will illustrate what theoretical difference three simplistic funding formulae would make to the primary care budgets of each LAP for this service: (A)

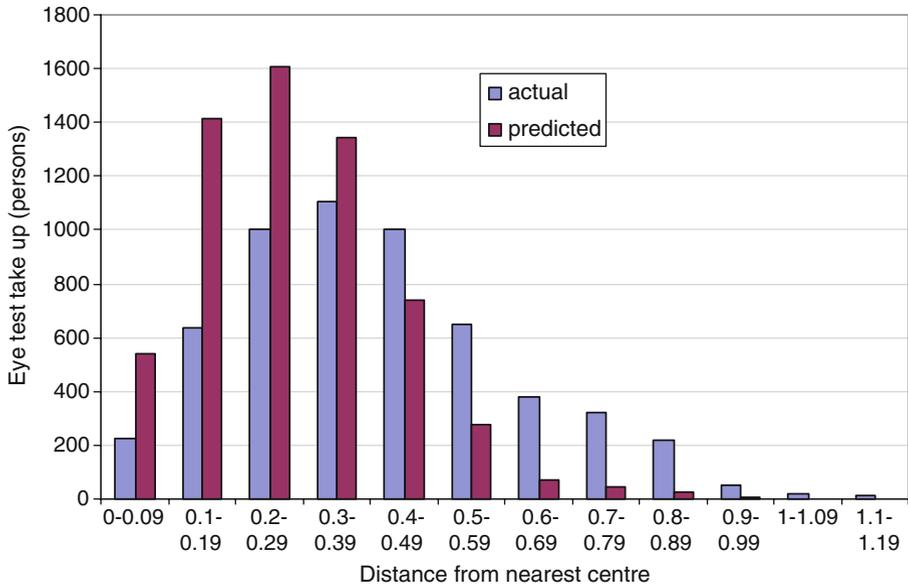


Fig. 6 Predicted change in eye test take-up in the 60+ population following re-configuration

based on resident population; (B) based on the population served if each resident used their nearest GP; (C) based on the population served if each 60+ person sought to have an eye test at their nearest GP.

Table 5 shows the results. Under scenario (A) for example 15.9% of any budget would be allocated to LAP 1 and 9.7% to LAP 2; under scenario B 16.1% would be allocated to LAP 1 and 10% to LAP 2 and so on. The percentage difference in allocation between A and B varies from -0.8% to $+0.9\%$, so overall a range of 1.7% . Under scenario C, which is based on the 60+ population, the range of variation is considerably higher, from -3.4% to $+2.4\%$. Hence, choosing different population bases will lead to different allocation outcomes; in this case, it may be argued that this allocation would have the merit of raising take up in a vulnerable group, which, as was seen, is most likely to be deterred by having to travel.

Discussion

The above example has shown how administrative data can aid local providers of a key service and enable the commissioning of better services through reconfiguration. We have not addressed in detail how resource allocation formula would work in other circumstances or for other services as there are different possibilities that would need to be worked through (for example a peripatetic service for the housebound). Much would also depend on the funding mechanisms and the budget holders who would be responsible for specific services, in this case GPs. However, assuming that primary care is the unit responsible for the delivery of a range of services, this illustration shows the extra evidence that administrative data can contribute to this category of decision-making. In the context of government plans to

Table 5 Alternative resource allocation scenarios

LAP area	Total population	% of population by LAP (A)	% of population allocated to LAP based on assignment to nearest GP (B)	% of 60+ population allocated to LAP based on assignment to nearest GP (C)	Difference (B-A)%	Based on 60+ population only	Difference (C-A)%
1	37386	15.9	16.1	18.1	0.1	18.7	2.2
2	22813	9.7	10.0	9.6	0.3	8.3	-0.1
3	30048	12.8	12.2	11.9	-0.6	13.1	-0.9
4	26352	11.2	10.4	11.0	-0.8	11.0	-0.2
5	25065	10.7	10.6	13.1	-0.1	13.3	2.4
6	26406	11.2	12.1	11.5	0.9	10.7	0.3
7	30175	12.8	12.6	12.6	-0.2	13.4	-0.2
8	36583	15.6	15.9	12.1	0.3	11.5	-3.4
total	234828	100.0	100.0	100.0	0	100.0	0

give clusters of GPs a much bigger role in commissioning services, the availability of robust evidence at the local level is essential.

A discussion of all possible methods of resource allocation is outside the scope of the present paper; however, there is a well established literature on location-allocation techniques. Within the literature, a broad distinction could be drawn between methods that rely solely on the number of registrants with GP practices and those which took account of the proximity or accessibility of patients to a practice location or locations or a hybrid of both (since they would give different results). In this case we have chosen to use where people live rather than where they are registered to avoid a possible circularity of aim (i.e. people registered with a GP, because it is the only one available).

Could currently available population data have provided a similarly detailed analysis? The user would be able to associate eye tests with locations of residence and hence the geographical distribution of take-up at a population level. However, demographic data would only have been available at output area level but not necessarily in disaggregated age categories. Because data are spatially aggregated, it would not have been possible to calibrate the level of take-up attrition with distance with sufficient accuracy and indeed none may have been found; in addition MAUP issues would also have arisen so the results would have been biased by the geography used.

The second problem is that the data would not have been able to distinguish between ethnicity, housing tenure or household demography, so that calibration of the influence of individual risk factors such as these would have been ruled out and yet these were found to be significant influences on take-up. The possibility of ecological fallacy would also have arisen e.g. low take-up is the result of deprivation and not old age *and* deprivation. Finally, the accuracy of the base data would be questionable since it would be reliant on mid-year estimates which in turn are based on a census baseline that was over 8 years old. To

conclude, it is hard to see how such a re-configuration of resources could have been evaluated or justified except through anecdotal evidence and trial and error unless administrative data had been used.

Conclusions

This paper is intended to contribute to the highly topical debate on how the use of administrative data can replace or improve the current sources of data on population especially at local levels. It has done so by pointing out the significant deficiencies and disadvantages of present arrangements and showing how administrative data could be captured, structured and used in more useful ways. A worked example has been included as evidence that the approach is both practical and achievable.

The first of the three main purposes of population data is to allocate resources to local authorities, health care commissioners and providers, police and other services, and subsequently to areas and services within each territorial entity. The current system of information on population arguably does a reasonable job down to territorial level, albeit the data are flawed through being out of date, and are based on ineffective collection methods with high levels of imputation.

At sub-local authority scale, inflexibilities and inaccuracies are magnified with the result that figures could skew decision making by creating a false evidential picture on the ground. One direct consequence is that denominators used to construct ratios, the second given main purpose of population statistics, such as the number of crimes committed, unemployment rates or new TB cases per head of population will be wrong in the most hard to count areas, with a range of possible consequences.

The approach adopted in this paper is shown to work well at a local level and has several key advantages over the alternatives, including more granularity, greater flexibility and timelier data. This is especially true as far as the given main purpose of population statistics is concerned, namely local planning and intelligence. By being able to link data at person and household level reduces possible concerns about the modifiable areal unit problem and ecological fallacy issues. Working with this level of granularity gives one much greater control over definitions, geography, time windows and analytical methods.

The three main purposes of population estimates have been stated as resource allocation of central funds; to provide denominators; and to aid the effective planning of services and delivery at the local level. The first two relate mainly to capturing accurate population counts, and at the very least, counts that are more accurate than those presently available from national statistics. The proposed methodology meets these criteria because it was originally developed in response to requests from local authorities who perceived there to be a discrepancy between official estimates of their populations and the actual population they believed they had which in turn impacted on their central government revenue allocations.

Implementing the methodology at a national level has not yet been attempted but can be considered. At first glance, it would be a matter of carrying out the administrative population count for each of the local authorities in England and Wales, and combining them into national coverage. This is certainly feasible given the universal coverage of the National Land and Property Gazetteer (NLPG), and

component data sets described in a sequel paper in this journal. Consideration would need to be given to people not on standard local data sets or for various reasons treated differently in terms of administrative data. These include the armed forces, prison populations and students in higher education. However, this is no different to present arrangements under the census. Similarly, other external independent data sources could provide information on private school pupils, or for example private GP patients.

Based on our experiences of using administrative data, we believe that the bureaucratic issues are probably more of a barrier to implementation than the technical issues. Both this paper and Harper and Mayhew (2011) have shown how technical issues can be resolved, such as data linking, how information may be structured, and finally how the information may be used to address each of the key areas of application outlined in the introduction. Our experience of working in different locations is that the bureaucratic impediments to a wider adoption of these techniques are mainly the result of confusion between the uses of data for personal and research purposes.

In legal terms, the advice of the Information Commissioner is that “Section 33” of the 1998 Data Protection Act provides that personal data may be processed for research purposes notwithstanding the requirement of the second data protection principle providing that a number of conditions are satisfied. These are: no substantial damage or distress is likely to be caused to any data subject; personal data will not be processed in order to support decisions about particular individuals; personal data will not be disclosed (except to a researcher) in a form which identifies living individuals. The data forming the basis for the worked example described in this paper was approved for use by the local PCT and underpinned by a legally enforceable data sharing protocol.

Acknowledgements This paper is based on research for the ESRC UPTAP (Understanding Population Trends and Processes) programme. We acknowledge contributions from Sam Waples of Birkbeck College, London, and thank Richard Verrall of Cass Business School, David Lawrence of the London School of Hygiene and Tropical Medicine and John Eversley of ppre CIC for their comments and support.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Alder, J., Mayhew, L., Moody, S., Morris, R., & Shah, R. (2005). The chronic disease burden: An analysis of health risks and health care usage. Cass Business School, ISBN: 1-901615 - 87-1.
- Barr, N. (2004). *The Economics of the Welfare State*. 4th edition. Oxford University Press.
- Bevan, G. (2009). The search for a proportionate care law by formula funding in the English NHS. *Financial Accountability and Management*, 25, 391–410.
- Brackstone, G. (1987). Issues in the use of administrative records for statistical purposes. *Survey Methodology*, 13(1), 29–43.
- Coleman, D., & Schofield, R. (Eds.). (1986). *The state of population theory—forward from Malthus*. Oxford: Blackwell.
- Dorling, D. (2007). How many of us are there and where are we? A simple independent validation of the 2001 Census and its revisions. *Environment and Planning A*, 39(5), 1024–1044.
- Ericksen, E., & Kadane, J. (1986). Using administrative lists to estimate census omissions. *Journal of Official Statistics*, 2(4), 397–414.

- Freedman, M., Lane, J., & Roemer, M. (2008). New approaches to creating data for economic geographers. *Journal of Official Statistics*, 24(1), 133–156.
- Frost, M., & Harper, G. (2007). The challenge of profiling communities. In J. Belsky, J. Barnes, & E. C. Melhuish (Eds.), *The national evaluation of sure start: Does area-based early intervention work?* (pp. 45–61). Bristol: Policy Press.
- Gehlke, C., & Biehl, H. (1934). Certain effects of grouping upon the size of the correlation coefficient in census tract material. *Journal of the American Statistical Association Supplement*, 29, 169–170.
- Greenland, S., & Robins, J. (1994). Invited commentary: ecological studies—biases, misconceptions, and counter examples. *American Journal of Epidemiology*, 139, 742–760.
- Harper, G. (2002). Using surfaces to inform local policy – modelling deprivation in Brent. MSc Dissertation, University College, London.
- Harper, G., & Mayhew, L. (2011). Using administrative data to count populations 1: Data sets and methodology. *Applied Spatial Analysis and Policy* (in review).
- Hayward, J., Murray, D., Iny, I., Jarrett, J., Lonergan, K., Pillas, D., & Seager, S. (2010). *London TB Service Review and Health Needs Assessment*. http://www.csl.nhs.uk/Publications/Documents/Tuberculosis/PHAST_London_TB_report_1.pdf
- House of Commons Treasury Committee. (14th May 2008). Counting the population. *Eleventh Report of Session, 2007–8*, (1).
- Jones, L., Bellis, A., Dedman, D., Sumnall, H., & Tocque, K. (2008). Alcohol attributable fractions for England, alcohol attributable mortality and hospital admissions. Centre for Public Health, Liverpool John Moores University, ISBN: 978-1-906591-34-2.
- Keohane, N. (2008). Local counts – the future of the Census. Report. New Local Government Network. <http://www.nlgn.org.uk/public/wp-content/uploads/local-counts.pdf>
- LGA. (2007). Estimating the scale and impacts of migration at the local level. Report. Local Government Association. <http://www.lga.gov.uk/lga/aio/109536>
- Lawrence, D., Patel, I., & Sievright, M. (2007). Brent primary care trust shortfall in revenue allocations for 2006/7 and 2007/8. Report. Brent PCT.
- Marmot, M. (2010). Fair society, healthy lives, The marmot review. Report. www.ucl.ac.uk/marmotreview. ISBN 978–0–9564870–0–1
- Mayhew, L. D. (2002). The neighbourhood health economy: A systematic approach to the examination of health and social risks at neighbourhood level. Actuarial Research Paper 144. Cass Business School, City University, London.
- Mayhew, L., & Harper, G. (2010). Counting the population of Tower Hamlets – A London borough in transition. Report. http://www.towerhamlets.gov.uk/lgs/901-950/916_borough_statistics/population_growth.aspx
- Mitchell, R., Dorling, D., Martin, D., & Simpson, L. (2002). Bringing the missing million home: correcting the 1991 small area statistics for undercount. *Environment and Planning A*, 34(6), 1021–1035.
- NHS. (2007). Tuberculosis prevention and treatment: A toolkit for planning, commissioning and delivering high-quality services in England. Report. Department of Health. http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_075638.pdf
- Openshaw, S. (1984a). Ecological fallacies and the analysis of areal census data. *Environment and Planning A*, 16(1), 17–31.
- Openshaw, S. (1984b). The modifiable areal unit problem. *Concepts and Techniques in Modern Geography*, 38, 41.
- Openshaw, S., & Taylor, P. J. (1981). The modifiable areal unit problem. In N. Wrigley & R. J. Bennett (Eds.), *Quantitative geography: A British view* (pp. 60–69). London: Routledge.
- Penneck, S. (2007). The future of using administrative data sources for statistical purposes. *Proceedings of the Third International Conference on Establishment Surveys, Survey Methods for Businesses, Farms, and Institutions*. Montreal, Quebec. June 18–21.
- RAWP. (1976). Sharing resources for health in England. Report. Resource Allocation Working Party. HMSO: London.
- Roderick, P., & Connelly, J. (1992). The problems of monitoring tuberculosis in an inner-city health district: Integrated information is required. *Public Health*, 106(3), 193–201.
- Simpson, L. (2007). Fixing the population: from census to population estimate. *Environment and Planning A*, 39(5), 1045–1057.
- Simpson, L., & Brown, M. (2008). Census fieldwork in the UK: the bedrock for a decade of social analysis. *Environment and Planning A*, 40(9), 2132–2148.
- Steffey, D., & Bradburn, N. (1994). *Counting people in the information age*. Washington: National Academy Press.

- Tudor-Hart, J. T. (1971). The Inverse Care Law. *Lancet*, I, 405–12.
- Tranmer, M., & Steel, D. G. (1998). Using census data to investigate the causes of the ecological fallacy. *Environment and Planning A*, 30(5), 817–831.
- Vale, S. (2006). The use of administrative sources for economic statistics – an overview. Regional Workshop on Use of Administrative Data in Economic Statistics, Moscow. http://unstats.un.org/unsd/economic_stat/Web/PDF/Overview%20-%20Vale.pdf
- Viel, J. F., & Tran, A. (2009). Estimating denominators: satellite-based population estimates at a fine spatial resolution in a European urban area. *Epidemiology*, 20(2), 214–222.