CrossMark

Using Administrative Data to Count and Classify Households with Local Applications

Gillian Harper [1] · Les Mayhew [1]

Abstract Households rather than individuals are being increasingly used for research and to target and evaluate public policy. As a result accurate and timely household level statistics have become an increasing necessity especially at local level. However, official sources of information on households are fragmented with significant gaps and inaccuracies that limit their usefulness. This paper reviews present statistical arrangements and then describes a new approach to data collection and household classification which combine various local administrative sources. An intermediate step is the creation of local population counts which are converted into household types and these methods are described in two companion papers previously published in this journal. The utility and advantages of the approach are demonstrated using the example of the six Olympic London Boroughs for whom the data collection was undertaken in 2011 and the analysis subsequently.

Keywords Administrative data · Household classification · Household statistics · Local applications

Introduction

Why Analyse Households?

Households are fundamental economic units of production and consumption in which goods and tasks are shared for mutual benefit. Important examples of productive

---

✉ Les Mayhew
    lesmayhew@googlemail.com

    Gillian Harper
    gill.harper@geocreate.co.uk

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

household activities include cooking, cleaning, childcare, care for older people and education (Van der Heyden et al. 2003; Eurostat 2009).

The Office for National Statistics (ONS), which is the UK's largest independent producer of official statistics and is responsible for the census in England and Wales, was among the first to measure and value unpaid goods and services produced by households and others have since followed in their path (ONS 2000; Holloway et al. 2002).

Statistics Finland, for example, found that GDP is increased by 40 % and household consumption by 60 % when its value is included in the National Accounts (Varjonen and Aalto 2006). More recently, the ONS estimated the contribution to GDP of raising children alone to be worth 23 % (Fender et al. 2013).

However, the use of household level information has been overlooked in many potentially useful applications, in part because of difficulties of measurement and definition but also the problem of assigning attributes such as income to households as opposed to individuals. The root of the problem, it can be argued, is the difficulty of classifying households systematically among their many variants.

The ONS defines a household as one person living alone, or a group of people (not necessarily related) living at the same address who share cooking facilities and share a living room, sitting room or dining area. It can consist of more than one family or no families in the case of a group of unrelated people. A dwelling by contrast is a self-contained unit of accommodation with its own front door potentially containing more than one household. [1]

In practice, there is a cost in meeting the strict conditions of such a tightly worded definition and experience shows that many users will prefer something that is less than ideal rather than nothing at all. Partly for these reasons, household data tends to be produced by agencies or companies for specific purposes and so, overall, the system is somewhat fragmented and incoherent.

There are other arguments why more should be done to put households at the forefront of research. Households are the basic units for transactions such as paying utility bills, property taxation and for rubbish collection and so tend to have a commercial or proprietary basis, and households are also the basis for measuring poverty in society in the UK. [2]

Estimates of the demand for house building rely heavily on household forecasts, which depend in turn on preferred living arrangements (e.g., people living alone as opposed to family units with or without children). Household attributes such as age, gender, occupancy etc. are useful predictors of the need for local services but are hard to source (Bowling 1991; Ohwaki et al. 2009; Larsson et al. 2006; Ulker 2008).

From a health perspective, there is interest in the protective value of living in different types of households (e.g., see Marmot 2010). For example, Vaupel (2010) considered that only about 25 % of the variation in adult life spans is attributable to genetic differences, noting that "older people are healthier when they live in insulated housing, wear appropriate clothing, eat appetizing food and enjoy their days."

---

[1] See ONS: https://www.gov.uk/definitions-of-general-housing-terms
[2] The standard measurement is known as 'Households Below Average Income' (see http://research.dwp.gov.uk/asd/index.php?page=hbai and www.poverty.org.uk/techinical/hbai.shtml).

The importance of households is established in other social policy domains. Examples include childcare (Eurostat 2009), accessing GPs, nursing care and hospital admissions (Van der Heyden et al. 2003), childhood immunisation (Bronte-Tinkew and Dejong 2005; House and Keeling 2009), exposure to smoking (King et al. 2009), and alcohol and marijuana use (Wagner et al. 2008).

At the political level, socially excluded or otherwise economically challenged households have been a recent focus of attention because of their high social costs and demands on public money (HM Government 2011). All of the above arguments suggest that it is timely to look again at how household statistics can be improved and hence this is the primary focus of this paper.

Our particular perspective is from the standpoint of using administrative rather than official data sources. We illustrate our approach at local level where we believe the opportunities for change are greatest. Our methods do not rely on any one single data source but rather several which are combined systematically. The methods and results presented are a sequel to two previous papers by the authors in this journal (Harper and Mayhew 2012a and b).

Our approach is bottom up in that we use locally available administrative data that we link at a person and address level. This flexibility means that it can be manipulated to suit different definitions and types of household as used by other agencies in the UK or overseas including Eurostat or OECD. Because it goes further than the current system of household classification used by the Department of Communities and Local Government (DCLG) in England, it can also be tailored to local circumstances.

Present Arrangements and the Case for Change

The expanding demand for household statistics reflected in the above is only partly being met by official statistical sources, which tend to be disparate, lacking in consistency, only available in certain geographies and variable in periodicity. In England, the ONS (Office for National Statistics) and DCLG are the responsible agencies for providing official national statistics on households and their equivalents elsewhere in the UK.

To date our work has focussed on the 'local authority' unit, which is equivalent to the concept of a 'municipality' as used by Eurostat. A local authority is responsible for planning and providing services such as housing, education, social care, roads, libraries and rubbish collection. It raises taxes through a levy on properties and receives grants from central government. The term is interchangeable and essentially equivalent to other frequently used terms such as 'borough', 'council', or 'district'.

Data collection begins with the decennial census, the latest of which was in 2011 with households used as the primary unit of enumeration (Baffour et al. 2013). The lowest geography for which data are available at is Output Areas with between 40 and 125 households in each. With the census as a baseline, DCLG then generally provide two-yearly projections at local authority district level and indicative figures of future numbers by household type if past demographic trends were to continue.

Producing household statistics is split into two main stages. The first is the production of ONS local authority based population projections by sex and single year of age, using assumptions about births, deaths and migration. The second stage combines this with information on household composition from Censuses to estimate the proportions

of households by local authority area and household type (Department for Communities and Local Government, November 2010).

DCLG are required to provide consistent national and regional projections (Department for Communities and Local Government 2008, 2010b) in order to allow for comparisons. The results are in effect statistical projections in which household types are fixed and inflexible and not actual counts of households.[3] However, in the view of users, the rapidly changing population in some areas reduces their accuracy and value and figures are not easy to reconcile with other data (Department for Communities and Local Government 2008).

Nevertheless, DCLG forecasts are extensively used by government departments and local authorities in preparation of development plans (Department for Communities and Local Government 2010b), and in the assessment of future housing need by house builders and utility providers as well as, for example, the Cambridge Centre for Housing and Planning Research, the Town and Country Planning Association, and the Joseph Rowntree Foundation (Holmans 2012).

More detailed information about households (e.g., their income and spending patterns) is captured in surveys and used to inform social research and policy development at a national level. The UK Data Service disseminates many of the UK large-scale survey datasets that are available for households such as the Labour Force Survey, the Family Expenditure Survey, and the General Lifestyle Survey which ran from 1971 to 2012.[4] However, they are difficult to use for other purposes and cannot be easily linked to other data at local level without the use of imputation.

In addition to the above sources, commercial geo-demographic products are available at the household level but users must pay. These provide consumer and lifestyle typologies of households rather than an enumeration of household demographics, and are reliant on census, survey and estimated data and so also use imputation to a large degree. In summary therefore, if we take all the different sources of data on households available, the central problems are a lack of coherence among the different sources, coupled with complicated and sometimes opaque methodologies.

The gap that we address in this paper is at local authority level, although our approach is both generalisable and scaleable to other geographies. Local authorities require timely and granular information on population, housing stock, housing costs, tenancy and a host of other variables at sub-local authority level to help inform and review current policy and services. The problem is that the relevant data are distributed among a range of council administrative systems that exist in silos (e.g., property registers, and local Council Tax records).

By being unable to access or link these data can create real problems for users. For example, a recent House of Commons Select Committee report said that 'local government would really like more frequent data, if that was possible' (House of Commons 2014). However, because support for households and families is largely provided through local authorities, their work is being severely hampered by the lack of evidence for social investment initiatives or calibrating interventions (HM Government 2011; Harper 2002; Voas and Williamson 2001).

---

[3] Although DCLG household types have changed since previous projections following user consultation
[4] See: http://ukdataservice.ac.uk/get-data/key-data.aspx#/tab-uk-surveys

Aside from this, the demand for better local data has been growing apace with the introduction of new legislation and financial pressures for local authorities to become more efficient. Examples include the Localism Act (2011) which gives local authorities more freedoms including responsibility for their own Local Development Framework (e.g., Leeds City Council 2011) and the Health and Social Care Act (2012), which resulted in the creation of Health and Wellbeing Boards to guide local commissioners of services across the NHS.

Given all the above, the inevitable conclusion is that the present state of household statistics is highly unsatisfactory. Available information is too aggregated, inflexible, and out of date or modelled and cannot be easily linked to other data domains such as education, social care or the private rented sector for effective local policy and planning. It is for these reasons that this paper puts forward a different basis for collecting and maintaining household statistics using locally available administrative sources (Harper and Mayhew 2012a and b).

Our paper is also timely because it coincides with wider moves to utilise administrative sources for the future production of official population statistics. In particular the aim is to expand the role of administrative data to replace or supplement the Census as first set out as part of the 'Beyond 2011' programme and now being implemented under the 'Census Transformation Programme'. The currently recommended system is for a predominantly online census in 2021 supplemented by further use of administrative and survey data (ONS, March 2014).

Aims of Paper

In this paper, we concentrate on local authority areas for reasons previously set out but also because they capture and are able to provide the required administrative data sets. The proof that a gap in official sources exists is evidenced by the many different studies we have been commissioned to carry out by local authorities and reference is made to some of these. Nevertheless, the approach is by no means perfect as there are some shortcomings that cannot be easily filled and these are also identified.

An important advantage of administrative data is that it is possible to add attributes to households that are not available in official data. This includes for example information on Council Tax, low income households, environmental health and education but also the usage of local services such as libraries or social care. Data are not necessarily limited to local administrative data sources; for example, survey-derived attitudinal data such as health, diet, and household spending patterns can also be considered subject to availability (examples can be provided on request).

The examples in this paper draw upon work undertaken for the six London Olympic boroughs which hosted the Olympic Games in 2012 between 2011 and 2014 (for example see Mayhew et al. 2011). The local authorities concerned are Barking and Dagenham, Greenwich, Hackney, Newham, Tower Hamlets and Waltham Forest with a combined population of 1.5 m and 0.6 m households. The resultant databases are in use by each local authority and the combined database used by the Greater London Authority.

The specific aims of the paper are to:

(a) Describe a system for producing flexible classifications and enumerations of household types using locally collected administrative data at address level

(b) Provide worked examples including a short case study on child poverty for informing local policy and decision-making

(c) Inform the wider statistics community by comparing our methods and results with official figures and to highlight differences as appropriate

The rest of the paper is structured as follows. The next section describes the core methodology for converting administrative data into household types and the attachment of attributes. The following section describes an application based on a case study of child poverty in the London Borough of Hackney, one of the six Olympic boroughs. The next section compares our results with official data sources and assesses any differences. A final section discusses the findings and concludes.

Creating Household Statistics from Administrative Data

Background to Demographic Counts

We use as our base the population estimations created by Mayhew Harper Associates during 2011 for the six Olympic boroughs[5] as at 27th March 2011. Full details of the methodology can be found in Harper and Mayhew (2012a, b). This database contains the age and gender of residents for every address within the local authority areas (i.e., taxable entities on the Council Tax register and entries on the Local Land and Property Gazetteer).

Our approach implicitly assumes that households and addresses are one and the same, i.e., they appear as separate entries on address registers and correspond to taxable units for Council Tax purposes. The reason for making this point clear is that administrative sources are address based, so that the case where there is more than one household per address this would not necessarily be identifiable.

Where there is more than one address per individual, we use the most recent and delete the others, but we cannot verify second addresses if they fall outside the boundary areas. It is possible to live at more than one address e.g., when parents have joint custody of children. The 2011 Census deals with this by recording the usual residence and the second address of children in this situation and we seek to do the same.

Such limitations are not easily overcome and are features shared with other sources of household statistics. For example official statistics often struggle with the identification of HMOs (Houses in Multiple Occupation) or communal establishments. We are usually able to distinguish these from private and social households and also distinguish between residential and non-residential uses. However, we cannot necessarily differentiate between married or unmarried households as the data are not sufficiently complete or accurate.

Examples of the main communal establishments are prisons, care/nursing homes and educational establishments (e.g., student halls of residence). A residual category

---

includes hotels, hostels, boarding houses, guest-houses, hospitals, sheltered accommodation, children's homes, psychiatric homes/hospitals and defence establishments. Generally speaking using administrative sources, it is possible to identify such establishments from local registers, property gazetteers and other sources.

Alternative Household Classification Systems Using Administrative Data

Local policy makers and planners need to be able to identify differences between household types e.g., pensioner households or three-generation households to support effective policy and decision making. As we shall show, the DCLG types are restrictive and unhelpful, although it is probable that DCLG would maintain that other types of households could be recreated if there was a demand for them. However, such decisions are not in local authority control and so lead to inflexibility.

Figure 1 summarises the stages in the process that starts with a list of the administrative systems that provide the starting point for the creation of the person level database in which all subjects are de-identified and which ends in applications in specific policy domains. In the next sections, we explain our method of household enumeration which follows directly after the population enumeration stage as shown in the diagram.

The process aggregates person level data by age, sex or other attributes into one of eight core household types. In this standard classification, households are broken down into eight higher level categories A-H (see Fig. 1). The next stage in the process is to link to each household data pertaining to particular attributes of households such as household size, tenure, tax band, benefit status etc.

From this it is relatively simple to develop sub-types of households for addressing specific issues of interest. These typically fall into policy domains such as health, housing, education, the local economy etc. Such information can be used for planning local services, drawing up strategic plans including Joint Strategic Needs Assessments
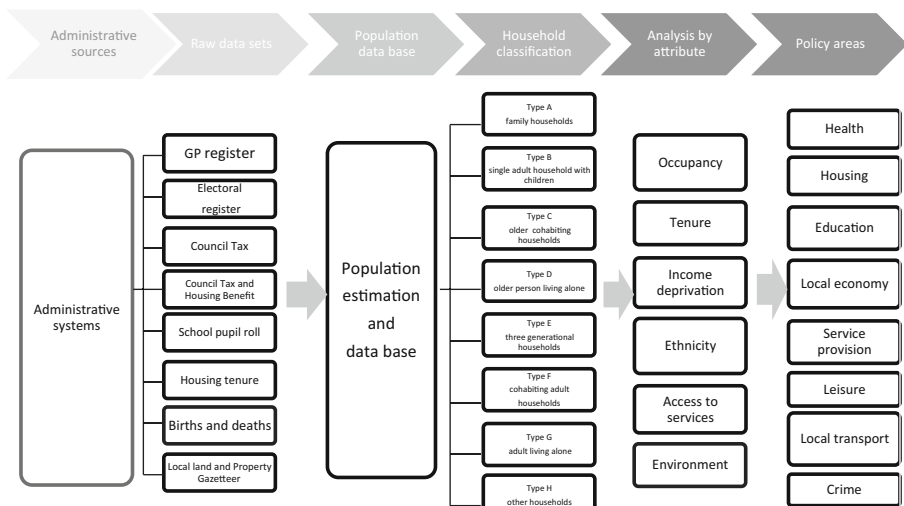

Fig. 1 Stages in the production of person and household level data and policy domains supported

(JSNAs) and so on. However, the same information can also be considered for use in a multitude of other different applications.

Take the example of older households, these are much more likely to frequent local shops, doctors, libraries, post offices and day centres than other households and so the whereabouts of these households can be used to inform public providers of these services as appropriate. Other examples include the identification of households suffering isolation or neglect, or houses in disrepair which may be a danger to health or an encouragement for anti-social behaviour.

In consideration of the possible applications a key point to note is that we restrict ourselves to statistical uses of the data. In other words we are not concerned with operational uses which rely on the identification of households or the names of people living in them in order to support some form of Council action. This use of the data is not covered by data protection legislation for which different legal considerations apply but it is covered for statistical uses.

Enumerating Household Types

We start with the premise that people can be sorted into household types according to their age and the number of occupants. Based on these two variables, we show that it is possible to define as many categories and sub-categories of household as we wish in a single consistent framework. Initially, we define eight household types based on the definitions in Fig. 2 which we call the 'eight standard types'. Using age and size of households as descriptors we can divide each type into their constituent age groups as shown in Table 1 with examples of each.

The methodology is flexible with regard to the number of age groups to be included. To keep to description and presentation manageable we use just three here: Group 1 children (0–19), group 2 working age adults (20–64), and group 3 older adults (65+). Row one is a Type A is a family household with two children and two or more adults (the additional adults could be an older sibling, friend or relative, or someone temporarily resident at an address); row two a Type B household with one adult and one child

Table 1 Specific examples of households defined by size and age group (Key: O indicates a person)

Type	Age group 1	Age group 2	Age group 3	House-hold size	Description
A	OO	OO		4	Couple household with two children
B	O	O		2	Single adult household with one child
C		O	O	2	Older couple household with one person aged 65+
D			O	1	Older person living alone
E	O	OO	O	4	3-generational household with one child, couple and an older person
F		OOO		3	Cohabiting adult household
G		O		1	Adult living alone
H	OO		OO	4	Split generation household
H'	OOOO			4	Young household (e.g., students, teenage parents)

Type A Family household with children — 2 adults and 1+ children; 2+ adults and 1+ children

Type B Single adult with children — 1 adult and 1+ children

Type C Older cohabiting couple — Couple household with at least one person aged 65+ and no children

Type D Older person living alone — Person aged 65+ living alone

Type E Three generational household — At least one person aged 0-19, one person aged 20-64, and one person aged 65+

Type F Cohabiting adult household — Adult household with 2+ adults aged 20-64

Type G Adult living alone — One adult aged 20-64 living alone

Type H Other households — Skipped generational households with older person and children but no adults; Household in which nobody is aged 20+ e.g. student household, teenage couple household; Anomalous households e.g. household with only children and no confirmed adult
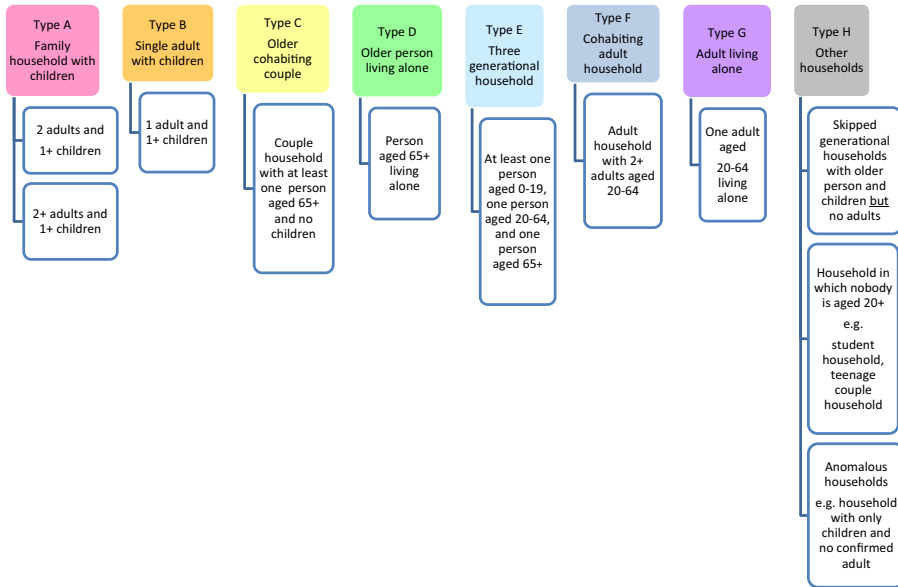
Fig. 2 The eight standard household types derived from administrative data

and so on. Gender differences can also be included as further sub-types and these are also discussed.

Of the examples shown, Type H households are easily the smallest in number and tend to fall into several heterogeneous sub-categories. Occupants could be young people (possibly students), or are from a split generation (e.g., a household in which children live with their grandparents). Type H may contain what are described as 'concealed' households i.e., separate households within a single address. It can also include anomalous cases where the administrative data has identified children as living at an address but no adult; these cases may be genuine or anomalous due to missing data.

Although, as previously noted, local administrative data cannot easily determine whether a couple household is a married household, divorced or whether people are related in some other way (other than by sharing a surname which is not always reliable or sufficient), for typical uses of household level data it is rarely essential to know this. Conversely, the ability to specify both age widths and household size offers scope to study various attributes of households in far greater detail.

First we need to be able to enumerate all possible combinations by age and size of household in order to analyse their relative occurrence in the population as well as their attributes. It can be shown that the equation for the number of possible combinations N of households with r age categories and up to n people is given by:

$$N = \frac{1}{0!} + \frac{r}{1!} + \frac{r(r+1)}{2!} + \frac{r(r+1)(r+2)}{3!} + \cdots + \frac{r(r+1)(r+2)\ldots.(r+n-1)}{n!}$$

Where n is the number of occupants per household $(0, 1, 2, 3, 4\ldots n)$ and r is the number of age categories $(1, 2, 3, 4\ldots.r)$. Each term inside the brackets multiplied by r gives the number of households with $0, 1, 2, 3, 4\ldots n$ people where zero indicates the 'void' case (i.e., an empty property). Table 2 enumerates the number of possible household types for

Table 2 Possible combinations of household demographic types based on size and age (see text for details of the highlighted cells)

Number of age categories (r)	Number of people per household, n						
	0	1	2	3	4	5	6
1	1	1	1	1	1	1	1
2	1	2	3	4	5	6	7
3	1	3	6	10	15	21	28
4	1	4	10	20	35	56	84
5	1	5	15	35	70	126	210
6	1	6	21	56	126	252	462

up to 6 age categories and 6 occupants. The inclusion of void households, the first term in the equation, is retained in order to derive an empty property rate for an area.

For any given value of r and n the sum of the terms gives the total possible combinations of household types. For example, there are a total of $1+3+6+10+15=35$ combinations of household types with 3 age categories and up to 4 people if the void case is included. This is highlighted in row three of Table 2 which adds to 35. Note that this is the same as for the number of combinations for 4 age groups and up to 4 people in the cell below and for 5 age categories and 3 occupants, similarly highlighted.

This result in turn is the same as the number of household types with 2 people and 5 age categories ($1+3+6+10+15=35$) which is highlighted in column three. In general therefore it can be seen that: $\sum_n N_{rn} = N_{r+1,n.}$ and $\sum_r N_{rn} = N_{r,n+1.}$

A standard result and important simplification in combinatorial mathematics is that:

$$N = \frac{(n+r-1)!}{(r-1)!\, n!}$$

Where r is a row in Table 2 and n is a column, for which $n+r \geq 1$ and $r \geq 1$. For example, for N_{44} this is $\frac{7!}{3!4!} = 35$ which is the same as the previous result as previously highlighted. As can be seen the accounting framework that is the result can expand rapidly which means that the number of categories can soon become unwieldy. Examples will be given shortly in which different sub-sets are selected and analysed in more digestible form.

The inclusion of gender to identify same sex households can also be considered although this leads inevitably to even more variants, but may be relevant in specific applications. However, this possibility simplifies if we are only concerned with the gender mix of a household and not with gender mix within an age group or level of occupancy.

For example any household can be labelled single sex (*M* or *F*, or of mixed gender, *m*). Occasionally there may be data gaps and the gender of one of more people at an address cannot be sourced in which case an 'unknown' category may be included. In practice the number of cases of gender 'unknown' is small and so it is convenient to combine the 'mixed and unknown categories' without much information loss.

Table 3 Mapping household demographic combinations on to the eight standard types, A to H for the case of three age groups and up to four occupants

Case	Age group 0–19 (A)	Age group 20–64 (B)	Age group 65+ (C)	Household occupancy (A+B+C)	Standard household type
1	0	0	0	0	void
2	0	0	1	1	D
3	0	1	0	1	G
4	1	0	0	1	H
5	0	1	1	2	C
6	0	0	2	2	C
7	0	2	0	2	F
8	1	1	0	2	B
9	1	0	1	2	H
10	2	0	0	2	H
11	0	2	1	3	C
12	0	1	2	3	C
13	0	0	3	3	C
14	0	3	0	3	F
15	1	2	0	3	A
16	1	1	1	3	E
17	1	0	2	3	H
18	2	1	0	3	B
19	2	0	1	3	H
20	3	0	0	3	H
21	0	3	1	4	C
22	0	2	2	4	C
23	0	1	3	4	C
24	0	0	4	4	C
25	0	4	0	4	F
26	1	3	0	4	A
27	1	2	1	4	E
28	1	1	2	4	E
29	1	0	3	4	H
30	2	2	0	4	A
31	2	1	1	4	E
32	2	0	2	4	H
33	3	1	0	4	B
34	3	0	1	4	H
35	4	0	0	4	H

In cases where gender is included, the number of household combinations must be scaled by a factor of 3 except for people living alone in which case the scale factor is 2.

If the aim is to consider both occupancy *and* gender mix then the possible combinations is further increased. For example if occupancy is three, then the possible combinations are MMM, MMF, MFF, and FFF.

Mapping Household Counts on to Standard Types

All possible combinations of households conveniently map on to the eight standard types A to H as previously defined in Table 1. Proceeding with the example above based on three age groups and occupancy levels of up to four per household, Table 3 shows how this mapping works (similar tables can be produced for other combinations of age and occupancy).

This example produces 35 mutually exclusive household types including the void case and is chosen simply because it is compact enough to include in a small table, albeit it is not exhaustive (i.e., it excludes cases where occupancy is greater than 4 persons). This example gives rise to three variants of Type A family households, three Type B, 9 Type C and so on.

Although it is seen that the most occurring standard type is Type H of which there are 10 variants, in practice they only account for less than 2 % of all households. If voids are excluded, typically the most numerous household types, accounting for around 96 % of the total, are Types A, B, C, D, F and G. Type E 3- generational households also account for less than 2 % of the total and so are similar to Type H.

Examples of Household Enumeration

An administrative data-derived population count is arranged such that each person is represented as a row in a de-identified database to which other attributes can be linked relating to the individual or to the household. Users of this approach will be particularly interested in examples that would not be reproducible using official sources but are nevertheless deemed useful.

These will depend on their availability in other datasets used. This could include information about the services accessed by individuals or households (e.g., schools attended); or it could involve commonly required attributes such as size and tenure as already suggested. An especially important example is benefit status: in the UK households may qualify for financial support to pay their rent or reduce Council Tax bills. Eligibility is based on income and savings and so we use this as a proxy for a low income household.

The example in Table 4 is designed to provide boroughs with a picture of low-income households. It covers all six Olympic boroughs and is simply a summation of the population and household types split by gender, average occupancy, low income status, tenure and tax band. Tenancy figures indicate the size of the council and social rented sectors; tax band information is included on relative housing wealth by household type, with bands A to C being a proxy for relatively low value housing.[6]

It confirms that the most numerous household types are Types A and G which are single working age adult households or cohabiting working age adult households.

---

[6] In the UK residential properties are banded by value into eight categories from A (lowest value) to H (highest value) with bands A to C being a proxy for low value housing and D to H for higher value housing. The convention is to italicise tax bands in order to distnguish them from household types.

Table 4 Summary table showing a breakdown of households across the six Olympic Boroughs and selected key attributes

Household type	Frequency of household type	Population	Average occupancy per household	% male only	% female only	% mixed or unknown	% housing units tax banded A-C	% of households on benefits	% of social housing units
A	125,856	597,670	4.7	0.8	2.0	97.2	65.4	38.9	34.2
B	57,209	160,366	2.8	6.8	27.2	66.0	74.5	56.2	43.6
C	40,596	100,702	2.5	6.0	6.9	87.1	59.6	41.8	32.4
D	42,114	42,114	1.0	36.4	63.4	0.2	76.9	58.3	48.0
E	11,893	71,027	6.0	0.6	3.2	96.2	56.1	55.9	35.0
F	108,798	274,558	2.5	13.4	10.2	76.5	62.1	21.3	25.6
G	188,610	188,610	1.0	52.7	40.2	7.1	68.8	23.5	26.7
H	8797	19,715	2.2	27.8	32.6	39.6	57.2	41.4	30.9
Total	583,873	1,454,762	2.5	23.9	23.7	52.4	66.8	34.3	31.9

However, the greater proportion of the population lives in Type A family households and Type F cohabiting adult households. Numerically, the smallest standard types are Type E 3-generational households and Type H households. Types B, D, E and H are more likely to be benefit households; Types B, D and E are more likely to live in social housing; and Types B, D and G more likely to live in lower value properties.

Table 4 also shows that gender mix by household type is quite intuitive but it is not always possible for data reasons to identify gender, so that in 2.5 % of households gender is 'unknown'. For simplicity this has been subsumed into the 'mixed and unknown column'. From the table we can infer that females are nearly twice as likely to be the sole survivors in older type D households and are more common in single parent households (e.g. the case of a female parent or guardian and at least one female child).

Differences in the average size, occupancy and age of households can be demonstrated in various ways. Figure 3 is a scatter-gram showing average occupancy versus average age at output area level across all six boroughs. It shows that each household type forms a characteristic cluster in these two dimensions; only Type H does not show any clustering tendency.

Table 5, on which Fig. 3 is based, shows that Type A family households typically contain four or five persons, including children, with an average age of 25 years and occupancy of 4.7 persons; Type B single parent households are about 6 years younger and range in size from two to three persons and an occupancy of 2.8 persons; Type C older cohabiting households range in size between two or three persons with average age of 62 years and an average occupancy of 2.5 adults.

Type D older single person households average 77 years and are the dominant type of older household at the oldest ages; Type E three-generational households have an average age of 36 years and occupancy of 5.8 persons; Type F households are cohabiting adult households with an average age of 40 years and occupancy of 2.8 persons; Type G single occupancy adult households have an average age of 42 years. Type H is the least homogenous type of all but only account for 1.5 % of all households.
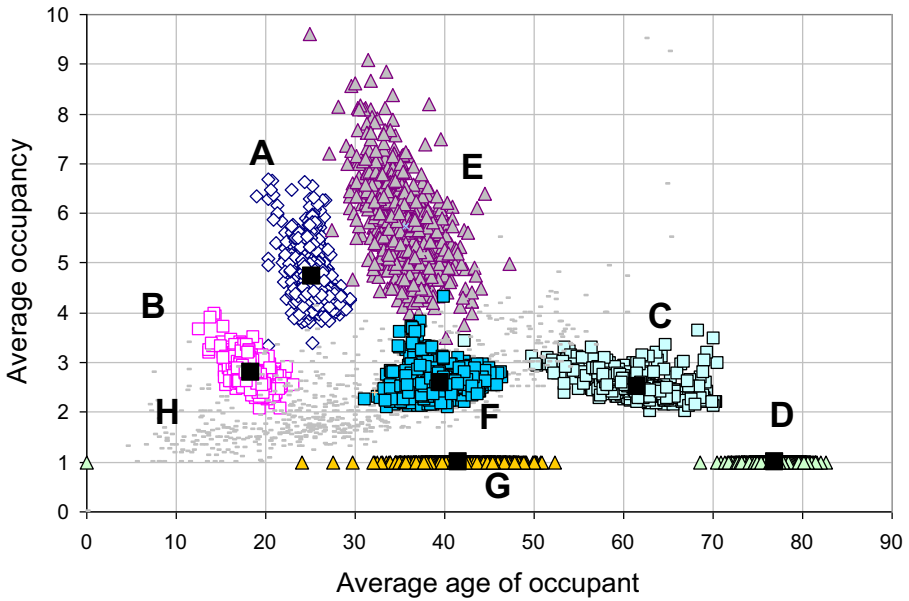
Fig. 3 Scatter-gram of household types showing occupancy versus average age by output area

A Case Study: Child Poverty in Hackney

Local authorities are interested in enumerating the number of child households for a range of purposes; for example, the concept of 'child yield' is frequently used to predict the demand for housing and school places. Such information is also extremely valuable to health providers and social services to identify vulnerable families and health needs.

In this short case study, we enumerate, map and analyse children living in households by tenure and benefit status and compare access to children's centres in the borough. We choose as our case study the London Borough of Hackney, one of the six

Table 5 Average household age and occupancy

Household type	Average age	Standard deviation	Average occupancy	Standard deviation	As % of all households
A	25.2	1.4	4.7	0.5	21.6
B	18.5	1.3	2.8	0.2	9.8
C	61.5	3.8	2.5	0.2	7.0
D	76.9	3.3	1.0	0.0	7.2
E	35.6	2.8	5.8	0.9	2.0
F	39.6	2.8	2.6	0.3	18.6
G	41.6	3.5	1.0	0.0	32.3
H	32.1	12.1	2.4	1.4	1.5

Olympic boroughs.[7] In its state of the borough report in 2013, it records that about 37 % of all children in Hackney are affected by child poverty, according to the standard national child poverty measure and is the third highest rate in London.[8]

Poverty varies spatially within the borough and impacts different communities unequally. Hackney has a very diverse population with at least 14 nationalities each having over 1000 members. It is also home to the Charedi population, a major Jewish orthodox sect with a population of around 18,000.

One of the things we are able to do is to identify households by ethnicity and in one case by religion. This is based on an extensive data set based on self-declared ethnicity derived from the School Pupil Census which we use to probabilistically assign ethnic status to people and households.

Hackney Council believes it is useful to build up a picture of different communities, whether defined by socio-economic criteria such as age, gender, ethnicity or different religious affiliations, in order to understand better their size and distribution and to design services that better meet their needs and expectations equally and fairly.

Working closely with the Charedi community, we used the Shomer Shabbas, a register of Charedi heads of households of Jewish orthodoxy, to estimate the population. Using the highly distinctive names therein, we estimated the probability of people with these names being Charedi, extending our search to include the whole population, not only those on the register.

In parallel, we also identified two other communities for analysis, namely Turkish and Bangladeshi households. Each community forms a distinctive group in terms of child yield, tenure and benefit status and like the Charedi are easy to identify. In comparison the Charedi community is highly clustered towards the north of the borough, but the other two communities are more widespread.

Table 6 enumerates the whole population and each community by number of children, tenure and benefit status. Our definition of a 'child' is any one age 19 or under for these purposes (later we focus on the 0–5 s). As can be seen the table shows quite different experiences in each community according to each of the attributes: benefit status (a proxy for low income[9]) and social housing (a proxy for supported housing).

For households with at least one child the table incorporates two household types: Types A family households and B single adult households with children. Households that have no children or are empty ('void') are included for completeness. Comparing columns it can be seen that child yield in each of the communities is much higher than for the whole borough.

In the case of the Charedi community household size is especially large with 53 households having 10 children and 31 more than 10. Also we see that Charedi households are far less likely to live in social housing than either Turkish or Bangladeshi households.

---

[7] http://www.hackney.gov.uk/Assets/Documents/estimating-and-profiling-the-population-of-hackney.pdf

[8] See 'State of the Borough Report 2013; Section 2 Child Poverty and Family Well-being'. http://www.hackney.gov.uk/Assets/Documents/Reduce-Child-Poverty-and-improve-Family-Well-being.pdf

[9] Hackney bases its measure of child poverty on the proportion of children living in families in receipt of out of work benefits or tax credits with reported income less than 60 % of median income. Our measure based on locally administered means tested benefits gives a very close approximation to this.

Table 6 Summary table showing a breakdown of households in Hackney by the number of children, housing tenure and benefit status according to three different communities

No. of children in household aged 0-19	All				Turkish				Bangladeshi				Charedi			
	No. of house-holds	% social hous-ing	% on benefits	% of all house-holds	No. of house-holds	% social hous-ing	% on benefits	% of all house-holds	No. of house-holds	% social hous-ing	% on benefits	% of all house-holds	No. of house-holds	% social hous-ing	% on benefits	% of all house-holds
void	5975	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
0	73,310	45.9	33.5	66.3	1943	66.2	70.8	45.3	511	56.6	54.6	35.1	1424	19.0	35.2	36.0
1	13,733	60.2	48.5	12.4	926	80.9	79.5	21.6	227	74.9	74.4	15.6	550	16.5	61.1	13.9
2	9635	63.9	53.5	8.7	925	85.3	86.4	21.6	256	81.3	76.6	17.6	473	18.0	67.9	12.0
3	4432	71.0	61.7	4.0	370	87.3	87.8	8.6	215	85.1	84.7	14.8	364	19.5	76.4	9.2
4	1832	70.7	66.3	1.7	89	82.0	75.3	2.1	140	87.1	82.9	9.6	319	18.2	69.6	8.1
5	736	62.4	70.8	0.7	29	75.9	89.7	0.7	68	77.9	88.2	4.7	219	19.6	65.8	5.5
6	394	51.8	72.1	0.4	6	83.3	83.3	0.1	19	68.4	84.2	1.3	178	23.0	69.1	4.5
7	242	38.8	72.3	0.2	2	50.0	100.0	0.0	11	45.5	72.7	0.8	150	19.3	67.3	3.8
8	139	30.9	73.4	0.1	1	100.0	100.0	0.0	6	83.3	100.0	0.4	108	21.3	72.2	2.7
9	99	24.2	65.7	0.1	0	0.0	0.0	0.0	1	100.0	100.0	0.1	89	22.5	66.3	2.2
10	59	22.0	66.1	0.1	0	0.0	0.0	0.0	0	0.0	0.0	0.0	53	18.9	66.0	1.3
>10	41	22.0	17.1	0.0	0	0.0	0.0	0.0	2	50.0	50.0	0.1	31	16.1	58.1	0.8
Total/average	110,627	48.3	37.5	100.0	4291	75.7	77.7	100.0	1456	72.1	71.0	100.0	3958	18.8	56.0	100.0

Table 7 Odds of income deprivation by risk factor including 95 % confidence intervals (CI)

Risk factor	Odds ratio	Lower CI	Upper CI
Any child 0–19	2.6	2.50	2.70
Single adult 20+	1.3	1.26	1.33
At least one person age 65+	3.4	3.30	3.60
At least one person aged 20–64	0.9	0.87	0.98
Living in social housing	4.2	4.10	4.40

The table shows that the percentage of households on benefits is 33.5 % in a childless household rising to 48.5 % in one-child households and steadily increasing to 73.4 % in 8-child households. This percentage varies considerably between the three communities, but in general the greater the number of children the more likelihood a household will qualify for financial assistance.

How does this compare with other low income households? A useful finding is that the risk of any household being on low income can be boiled down to a small number of risk factors. Using logistic regression it can be shown that a household is 2.6 times more likely to be on means tested benefits if there is any child aged 0–19; and 3.4 times more likely if there is an older person aged 65+ (for further information on logistic regression see e.g., Altman 1999).

Table 7 summarises the five main risk factors and their influence on income poverty: Four relate to ages of occupants and one to housing tenure and together they statistically explain 87 % of the variation in benefit households. However, they also have the special property that they can be used in combination to reproduce each of the eight standard household types. For example, a Type A household must have at least one-child age 0–19 and two adults aged 20–64, but if it has only a single adult and at least one child then it is a Type B household.

The odds in Table 7 are multiplicative so that for example a single adult Type B household would be $2.6 \times 1.3 \times 4.2 = 14.2$ times more likely to be on benefits than a household with none of these risk factors. Extending this further, a Type C older household must have at least one person aged 65+, but if that person lives alone then it is a Type D one person older household. In contrast, a Type E 3-generational household must have at least one child, an older person and a working age adult.

Access to Children's Centres in Hackney

It is generally accepted that having access to affordable, good quality childcare has a bearing on parental decisions when they are in the process of returning to or entering work which can help lift families out of poverty. For many years there has been a national programme of Sure Start children's centres to target those in greatest need of support and for which responsibility for their running has since been devolved to local authorities. [10]

---

[10] https://www.gov.uk/sure-start-childrens-centres-local-authorities-duties

In this section we evaluate to what extent childcare and other needs are being met in Hackney based on the existing network of centres. Under present rules children up to 5 years old who are resident in the borough are eligible to attend one of 22 such centres located in the borough. However, their attendance at these centres is subject to strict criteria including evidence of residence in Hackney and also proof of household income.

Plainly it is important that the centres should be accessible to those in greatest need and so we mapped all households likely to qualify on these criteria and also the centres. We would expect a typical average catchment radius of 0.5 km for this number of centres and size of local authority; we term this radius 'pram pushing distance' and it is equivalent to a 6 to 10 min walk time.

We split all households with children under 5 years old into groups identifiable by being in one of the three communities above and also whether on means tested benefits or not. We then mapped the results and tabulated how many households meeting these criteria had access to none, 1, 2 or 3+ centres according to households in each community: Turkish, Bangladeshi or Charedi.

Based on the 15 k households with children aged under 5, 8.2 k were low income households. Of these, 26 % of all households had no access to one or more children's centres within pram pushing distance whereas 74 % did. We also found that benefit households had slightly better access than non-benefit households which is what one would expect, albeit by only an unexpectedly small margin (only 68 versus 66 %).

Figure 4 is a map of children's centres and of all households in Hackney that meet both the benefit criterion and have at least one child under 5 years old. Each household is colour-coded according whether there are 0, 1, 2 or 3+ centres within 500 m. It can be seen that children's centres are widespread throughout the borough, but also that many households fall outside the pram pushing criterion (see black symbols).
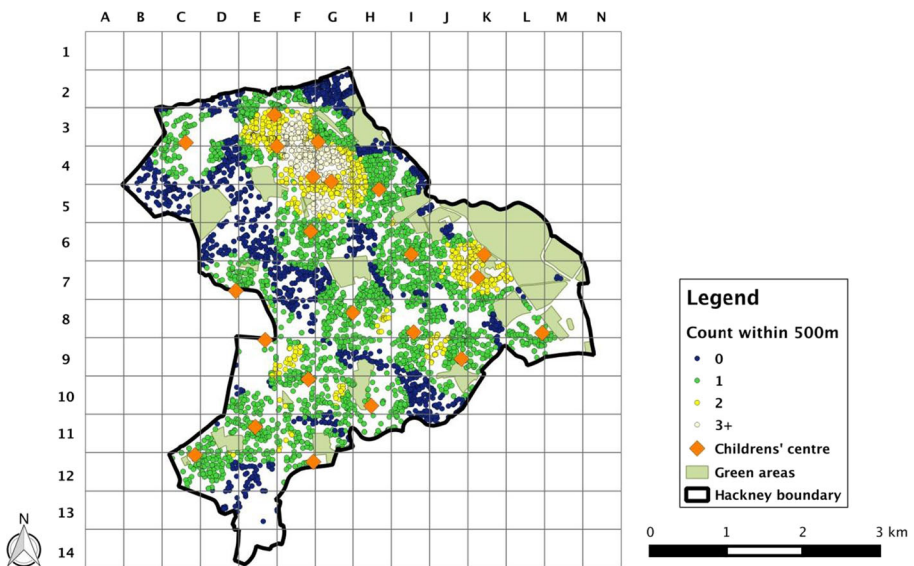

Fig. 4 Map showing the locations of households on benefits with children aged<5 that are outside pram pushing distance from the nearest children's centre

The clear impression is that the map shows several large gaps in the network – but also areas with access to 3 or more centres. The area of greatest choice is in the north of the borough between cells E3 and G5. This area is strongly identified with the Charedi community but because the Turkish and Bangladeshi communities do not experience the same degree of co-location, their access is much more variable by comparison.

Further analysis shows that 46 % of Charedi households have a choice of two or more centres within pram pushing distance as compared with only 13 % of Turkish and Bangladeshi households. Clearly local authorities do not set out to create unequal access to public services but we would argue the quality of the data they use often mean that decisions are too broad brush relative to the objectives they seek to achieve.

Our main conclusion therefore is that children's centres are widely dispersed in this borough but their planning could have benefited from better fine tuning. Although it is not possible to reverse the clock by reconfiguring existing centres, it cannot be ruled out that some centres may be forced to close due to budget constraints and so this is also a another possible use of the data.

This kind of analysis can be used to ensure services are located equitably but other factors are important too. For example, although the Charedi community appears to be very well served geographically, it tends to make parallel arrangements for its own children's needs based on their religious beliefs. Hence, the issues are even more complex but it is precisely for dealing with these issues that our methodology is well suited.

Access to children's centres is based on residence and so there is a further question of boundary effects when we are dealing with other services that are located just outside the local area. We did not cover cases here because of the strict residence eligibility conditions governing this service, but they can be easily addressed by working with services in neighbouring areas or by analysing several boroughs together especially for services with trans-boundary catchment areas.

Administrative Counts Versus Official Household Statistics

Previous sections have sought to explain the use of local administrative data to enumerate and classify households starting with the raw administrative data. How to present and use the information in digestible form for different purposes was set out in a simple accounting framework using examples and tested using a case study. A key question is how much confidence can users have in administrative approaches to the enumeration of households?

In this section we compare our findings with official figures produced by the ONS, DCLG and GLA (Greater London Authority). Whilst we do not expect to find an exact correspondence, it is useful nonetheless to identify reasons for any differences. From our experience of reconciling the administrative approach with official sources, we expect to meet two potential problems. One is the different basis used to count populations (i.e., Census versus administrative sources); the second is translating administrative data into exact copies of officially used households definitions.

One obvious and insurmountable difference is that Census data are for a point in time and updated only 10-yearly, whereas administrative data are constantly being updated. For this reason official household statistics covering intervening years will

Table 8 Comparison based on total number of households by local authority using administrative, DCLG, GLA and ONS Census data

Households ('000 s)						Difference % in households			
Local authority	Admin 2011 (A)	GLA 2011[a] (B)	DCLG 2011[b] (C)	DCLG 2011[c] (D)	Census 2011[d] (E)	Admin - GLA	Admin- DCLG (C)	Admin- DCLG (D)	Admin - Census
Barking and Dagen-ham	70.5	72.3	69.3	70.1	69.7	−2.5	1.7	0.6	1.1
Greenwich	101.6	106.7	99.5	101.4	101.0	−4.8	2.1	0.2	0.6
Hackney	103.0	98.1	92.1	102.1	101.7	5.0	11.8	0.9	1.3
Newham	104.0	103.2	91.8	102.3	101.5	0.8	13.3	1.7	2.5
Tower Hamlets	101.2	100.6	98.0	102.1	101.3	0.6	3.3	−0.9	−0.1
Waltham Forest	97.9	95.1	92.0	97.4	96.9	2.9	6.4	0.5	1.0
Total	578.2	576.0	542.7	575.4	572.1	0.4	6.6	0.5	1.1

[a] copyright © Greater London Authority, 2011

[b] copyright © CLG, 2010

[c] copyright © DCLG, 2013

[d] copyright © ONS, 2012

tend to be based on a complicated mix of fact and imputation in which potential errors are impacted by timing differences in population counts and the assumptions used regarding household formation.

As well as DCLG, the GLA (Greater London Authority) also produces its own household projections for London boroughs using housing development trajectories based on the Strategic Housing Land Availability Assessment (SHLAA). The GLA use the same household definitions as DCLG but a key difference is that they use their own population estimates as a basis. However, the availability of GLA data affords the opportunity to benchmark our household counts with both sources. The version available at the time of this analysis is the 2009 round SHLAA based projections.[11]

Our results for the year 2011 are shown in Tables 8 and 9 for each of the six Olympic boroughs. Table 8 compares household counts based on administrative sources with comparative figures created by DCLG in 2010 (the version available at the time of the analysis), an updated DCLG version made available in 2013,[12] Census 2011, and the GLA 2009 round SHLAA based projections, all *without* communal establishments.

The results show that for the whole region administrative household counts based on administrative data are 0.4 % higher than GLA at 578 k but 6.6 % higher than the 2010 version of the DCLG figures. The addition of the more recent DCLG 2013 figures substantially reduces this difference down to 0.5 %. This demonstrates that the DCLG

---

[11] The most recent GLA data available at the time of writing is based on 2013 SHLAA data

[12] As well as subsequent interim results since 2013, an update incorporating full Census 2011 information has been postponed by DCLG until late 2015

Table 9 Comparison of vacant dwellings by local authority using administrative, DCLG, GLA and ONS Census data

Local authority	Total available dwellings ('000 s) Admin 2011	Vacant properties ('000 s)					Vacant properties %				
		Admin 2011 (A)	GLA 2011[a] (B)	DCLG 2011[b] (C)	DCLG 2011[c] (D)	Census 2011[d] (E)	Admin 2011 (A)	GLA 2011[a] (B)	DCLG 2011[b] (C)	DCLG 2011[c] (D)	Census 2011[d] (E)
Barking and Dagenham	72.9	2.4	0.6	3.6	2.8	3.2	3.3	0.8	4.9	3.8	4.4
Greenwich	112.8	11.2	6.1	13.3	11.4	11.8	9.9	5.4	11.8	10.1	10.5
Hackney	107.9	4.9	9.8	15.8	5.8	6.2	4.5	9.1	14.6	5.4	5.7
Newham	108.9	4.9	5.7	17.1	6.6	7.4	4.5	5.2	15.7	6.1	6.8
Tower Hamlets	115.7	14.5	15.1	17.7	13.6	14.4	12.5	13.1	15.3	11.8	12.4
Waltham Forest	103.1	5.2	8.0	11.1	5.7	6.2	5.0	7.8	10.8	5.5	6.0
Total	621.3	43.0	45.3	78.6	45.9	49.2	6.9	7.3	12.7	7.4	7.9

[a] copyright © Greater London Authority, 2011

[b] copyright © CLG, 2010

[c] copyright © DCLG, 2013

[d] copyright © ONS, 2012

figures have fallen more in line with the administrative data results, and that the Census figures are also very close – as is to be expected because the 2013 DCLG figures draw on the Census results.

It also demonstrates the change from using the Mid-Year estimates as the original population base for DCLG 2010 figures that were known to have undercounts for these areas, to the Census 2011 population estimates in the DCLG 2013 version. This is hence demonstration of the inconsistencies than can occur and which are reinforced even within the same sources over time.

Within the Olympic boroughs the differences vary by local authority but two that particularly stand out are between the administrative and original DCLG household counts for Hackney and Newham. Again, it is noteworthy that these are greatly reduced in the post-Census 2013 DCLG figures.

Another comparable measure is estimates of the number of vacant dwellings. For the administrative data method we define the vacant dwelling rate as the percentage of the total number of residential addresses on the LLPG (Local Land and Property Gazetteer) that are not occupied, having first removed all communal establishments from the LLPG for consistency. We use this particular gazetteer for our population estimations because it is provided and used by the local authorities themselves.

Table 9 shows the differences in vacant property rates between the five sources. These range in value from 0.8 % in Barking and Dagenham using GLA definitions to as high as 15.7 % in Newham using the original DCLG definitions. One reason why rates between the local authorities vary to this extent is due to the exceptionally active regeneration in the Docklands area of east London, affecting mainly Tower Hamlets where there are large numbers of new apartments which were unoccupied at the time.

Another reason is traceable to the lower ONS population counts on which the original 2010 DCLG household counts are based. Without these, in Table 9, across the whole area the total vacancy rate only varies by 1 % across the sources, if the original DCLG estimates are excluded. For example, our work in the six Olympic boroughs produced an administrative-based population count of 1.46 m, which is 0.8 % higher than the GLA's but nearly 11 % higher than the equivalent count published by the ONS at that time.

However, this is not an artefact of when data were produced but a systemic problem which can be traced back in time. The London Borough of Newham is a particularly good example of this. At the time of our work in March 2011, the published ONS population for Newham was 240 k compared with our own figure of 299 k. Following revisions to their methodology, the ONS released new figures in November 2011 in which Newham's population had increased from 240 to 272 k.

The final Census 2011 population estimate for Newham is 308 k, a figure created from Census surveys and a number of subsequent adjustments. This may be compared with figures published by the GLA which increased its own estimate for Newham from 268 to 296 k in June 2011, a figure that was partly informed by our own work. The discrepancies between ONS, GLA and administrative sources and also within ONS sources are illustrative of how figures can quickly get out of kilter in areas of high in-migration and regeneration, as the case in Newham but also in neighbouring boroughs.

Table 10 The government household typology scheme

Household type	Description
One person households	Male
	Female
One family and no others[a]	Couple[b]: No dependent[c] children
	Couple: 1 dependent child
	Couple: 2 dependent children
	Couple: 3+ dependent children
	Lone parent: 1 dependent child
	Lone parent: 2 dependent children
	Lone parent: 3+ dependent children
A couple and one or more other adults[d]	No dependent children
	1 dependent child
	2 dependent children
	3+ dependent children
Lone parent and one or more other adults	1 dependent child
	2 dependent children
	3+ dependent children
Other households[e]	See notes

[a] Households with dependent children and no non-dependent children

[b] 'Couple households' are either married or cohabiting

[c] A dependent child is a person in a household aged 0 to 15 (whether or not in a family) or a person aged 16 to 18 who is a full time student in a family with parent(s)

[d] In these categories, the other adults may include another couple and/or another lone parent and/ or a non-dependent child

[e] The 'Other households' category above is an aggregation of five categories from the original Census table C1092 supplied by ONS

Comparison of Household Types Using Official Figures

The household typology based on the government's own published methodology is shown in Table 10 (taken from Department of Communities and Local Government, November 2010). On the face of it, there is no reason why the previously identified discrepancy between sources should impact unduly on our own household typology as long as definitions are comparable even if the total quantum of households differs. As can be seen in Fig. 2, the scope of our own typology is richer in detail due to the accounting framework we have created.

As illustration of the differences we focus again on the London Borough of Hackney. This is because it has one of the largest differences in household counts based on our figures and 2010 DCLG's and so provides a rigorous test. We acknowledge that any generalisations concluded from one local authority may not necessarily apply to other types of local authority area or nationally (see check list at Appendix A which provides a summary of quality issues relevant to our research).

Table 11 Household type counts in London Borough of Hackney. Differences compared to the administrative data count are given

Households ('000 s)					Difference % in households		
Household type	Admin 2011 (A)	DCLG 2011[a] (B)	DCLG 2011[b] (C)	Census 2011[c] (D)	Difference (A–B)	Difference (A–C)	Difference (A–D)
Family household (couple)	20.5	18.2	24.8	28.3	2.30	−4.30	−7.80
Family household (lone parent)	8.6	7.7	8.3	10.8	0.90	0.30	−2.20
Family household with other adults & dependent children	10.0	8.2	9.5	4.2	1.80	0.50	5.80
One person household	47.2	43.5	36.5	35.6	3.70	10.70	11.60
Other households	16.7	14.5	23	22.8	2.20	−6.30	−6.10
Total	103.0	92.1	102.1	101.7	10.90	0.90	1.30

[a] copyright © DCLG, 2010

[b] copyright © DCLG, 2013

[c] copyright © ONS, 2012

For comparison purposes, we recreated DCLG household types using administrative sources. DCLG definitions are quite demanding in terms of their specificity, so reproducing these figures is likely to provide a robust test. The first issue to consider is the definition of a dependent child. For DCLG purposes, this is a person in a household aged 0–15 (whether or not in a family) or a person aged 16 to 18 who is a full time student (in a family with parents). As was seen, our primary classification uses age 19 and under and so the first task was to alter our definition of a younger person.

A practical problem was to identify children in full time education (FTE). Our principal data source was the school pupil census, in which persons are flagged if they are aged 16–18 and attended a school in the borough or a neighbouring borough. They are also flagged if they are on Connexions[13] data in which young people are flagged as 'FTE' if they are in full time education. In practice it could not be determined how many of those not registered as FTE attended private schools or other state schools in neighbouring boroughs.

Using administrative data we were able to re-create the 'one person household' category, identified by DCLG as persons aged 16 or over living on their own: however, anyone aged less than 16 living on their own (an extremely rare case and probably anomalous) are considered to be in the 'other category'.[14] The DCLG category 'one family and no others' was also able to be re-created. This includes mixed sex couple

---

[13] Connexions was a UK governmental information, advice, guidance and support service for young people aged 13 to 19 (up to 25 for young people with learning difficulties and/or disabilities), created in 2000 following the Learning and Skills Act. Its work has since been reorganised and re-focused by the Coalition government.

[14] In reality, children under 16 will not be living on their own, but in these cases the resident adults were unable to be captured using administrative data. These cases are small in number.

households aged 19 and over with or without dependent children or households with only one adult aged 19 and over with dependent children.

The DCLG categories of 'couple' or 'lone parent' with one or more other adults' were combined for simplicity into 'family households with other adults and dependent children' where these households have dependent children. Note that in evaluating the 2010 DCLG classification system, the definition of a 'couple' only includes mixed sex, married or cohabiting couples, as defined by the ONS. In other words, DCLG household types tend to be narrower in scope by their omission of same sex households, and also omit three-generational households.

The 2011 Census household types can be grouped in the same way as DCLG categories for comparison purposes, although now same sex couples are considered a 'family' household. Notwithstanding these definitional subtleties, a crude comparison of the results using DCLG 2010 (column B) and 2013 figures (column C), Census 2011 (column D) and administrative sources (column A) is shown in Table 11. As can be seen, the totals are generally quite close although the administrative total is marginally higher than the official sources at that time.

Two key findings are that administrative sources enumerate far fewer 'couple family' households and 'other households' than do 2013 DCLG and the Census, but more one-person households and family households with more than two adults. Such differences are due mainly to definitional issues particularly the boundary line between children and adults; however, there are discrepancies between DCLG and the Census, again exposing a lack of consistency between sources.

Reasons for Differences Between Sources

Other factors have come to light following a recent discussion of the results from the 2012-based projections of households in England as compared with the Census. This review concluded that differing definitions of 'couples' was one of the key issues, but it is also maintained there had been a too slow a reaction to changes in migration (BSPS, May 2015).

Another source, The UNECE (2011), criticises census estimates of single parent households, noting that there can be large differences between one-person households, defined on the basis of a 'housekeeping unit', or a 'dwelling'. This could be another reason for the discrepancy in census counts that use the former definition, and the administrative counts that use the latter which is also reflected in Table 11.

Other factors can be speculated for the differences seen such as the effects of the recession and housing crash at the time of the Census which may have suppressed household counts (e.g., DCLG/RSS meeting 2013; McDonald and Williams 2014). All of the above suggests that definitional changes as well as differences in methodology continue to be a serious problem for DCLG.

We understand DCLG is now reviewing the methodology with the aim to make it 'simpler, and more transparent'. It is especially telling, that even after all this work, the most recent DCLG figures are still labelled as provisional because they 'do not yet fully

incorporate Census 2011 data' (BSPS, May 2015). Overall, the picture therefore remains complex and unsatisfactory.

McDonald and Williams (2014) suggest that it is time for local authorities to consider their own situation carefully and their statistical needs, a view with which we would strongly concur. The root of the problem is not only the lack of coherence in how household statistics are produced, but also a lack of granularity and flexibility over definitions and therefore outputs.

Considering all the difficulties it is re-assuring that our analysis gives us greater confidence that administrative sources are more likely to provide a long-term solution to these issues than continually tweaking present arrangements. We believe this is a strong argument for adopting the approach described in this paper, which gives users more control, greater flexibility and better timeliness.

Discussion

This paper has identified a gap in the availability, quality and functionality of household statistics at local level. As well as a confusing diversity in sources, household statistics in general suffer from over-aggregation, a lack of flexibility and coherence, coupled with in inability to link to other data except at output area level. Using administrative data, we are able to provide a current enumeration and typology of households with flexible definitions and geography that supports linkage.

We observed that there were a considerable number of administrative data assets available to local authorities for these purposes. This availability creates the conditions for local authorities to develop their own local systems for meeting local needs provided this work is carried out in a data-secure environment, if not alone then in consortia. This would fit with the grain of locally devolved powers and the roles and responsibilities of newly created Health and Wellbeing Boards under the 2012 Health and Social Care Act.

Our approach has been refined in numerous studies in which accuracy, timeliness and specificity of detail were important considerations. This includes the six Olympic borough study referred to in this paper but also in other parts of London and England. So far we have not attempted to create household projections from administrative data, but we have started to consider household turnover and changes in household type between administrative snapshots.

The current momentum in the UK is to increase the use of administrative data in population analysis and censuses as confirmed by the 2014 ONS announcement (ONS 2014). Administrative data present different problems for users as compared with censuses especially concerning their coverage and reliability (e.g., see ADT 2012; Zhang 2012). Our view is that their potential is not fully realisable unless they are jointly analysed within a systematic rule based framework. These issues and how to address them are further discussed in Harper and Mayhew 2012a, b.

It is useful nevertheless to refresh ourselves on how these issues may propagate through the process of producing household statistics, and if the outputs are sensitive to these. Our administrative data based population and household estimation relies on the linkage of multiple administrative data sources, making it necessary to assess uncertainty and error with great care. This is because no one source captures the whole

population, and so rules and assumptions are needed to deal with conflicting information, duplicates, and over or under coverage.

In the absence of comparable methodologies it is difficult to know if more efficient methods can be devised, but a checklist of quality control issues is contained at Appendix A which may be cross-referred with the analysis and approach taken in this paper. Our methodology is not based on a sampling procedure and as such does not support the use of confidence intervals.[15] We use external comparators as much as possible to back up our results, and the fact that official statistics have fallen into line and validated our results especially after the 2011 Census helps to vindicate this.

There still remain several unsolved issues. For example boundary issues arise where a person has two addresses but in two separate authorities and so some double counting is possible. In theory these issues are reconcilable at national level but not if they have addresses in different countries; in other words no system is perfect. We also mentioned in passing the difficulties of identifying the marital status of households. In a perfect world this would require us to link records on marriage and divorce which could pose formidable technical and other challenges.

In terms of transferability, equivalent administrative datasets to those found in the UK would not necessarily be available in exactly the same form elsewhere, although for countries with registration systems instead of censuses such as Finland it should be possible to recreate our typology without difficulty. For countries still using censuses it should also be possible to recreate our typology but probably not at the same level of geographical granularity or with the same flexibility.

To conclude, this paper proposes a different basis for classifying and maintaining household statistics using locally available information sources. The key advantages are that data can be produced on a timely basis in any geography and can be linked to other attributes. This of course requires not only a good knowledge of local data sources but also how to access and exploit them. In our case, access to the data forming the basis for this paper was approved for use by the then local PCTs (Primary Care Trusts) and local councils and underpinned by legally enforceable data sharing protocols including non-disclosure to third parties.

A - Measuring Statistical Quality of Population Estimation and Household Counts from Administrative Data Method

ONS provide a checklist of quality measures and indicators for use when measuring and reporting on the quality of statistical outputs. They also record the dimension of quality being measured in each case, using the six European Statistical Service (ESS) Dimensions of Quality developed by Eurostat. [16] In the following table our methodology is compared against each of these standards.

---

[15] Note that DCLG are in a similar position, stating that their 'projection methodology does not enable calculations of probability, standard errors or confidence intervals

[16] Based on ONS Guidelines for Measuring Statistical Quality http://www.ons.gov.uk/ons/guide-method/ method-quality/quality/guidelines-for-measuring-statistical-quality/index.html

Relevance: The degree to which the statistical product meets user needs for both coverage and content.

The administrative data population estimation is an estimate of a local authority's population at a snapshot in time derived from the use and linkage of core local authority administrative datasets and a rule-based system applied to establish who are current residents.

Each person in this count is a separate database entity and assigned to a property address. The summary of the demographic profile of each property address is the basis for the household counts and classification typologies.

The method was developed in response to a need from local authorities to have an alternative source of population statistics that did not rely on survey methods and that was more timely and quicker to produce.

The outputs are used by local authorities who have requested the service to quantify their current population and to inform commissioning, service planning and policy.

The output is created bespoke for a local authority, and is therefore highly relevant to their local context.

Accuracy: The closeness between an estimated result and the (unknown) true value

The methodology does not enable calculations of probability, standard errors or confidence intervals and therefore these cannot be calculated.

Outputs are compared to available benchmarks to ensure results are sensible.

The methodology uses current data on actual existing residents rather than projections or survey adjustments and imputation. It accounts for no one dataset having complete population coverage by joining separate datasets together to maximize coverage, and for inflation and duplication.

The methodology is inevitably dependent on the accuracy of the input data and is vulnerable to the known issues associated with administrative datasets (noted elsewhere). It is also dependent on the accuracy of the assumptions used in the rule-based system.

No rounding is used in the output.

The outputs are representative of that snapshot in time and may date quickly if the population is in a high state of change.

Timeliness and Punctuality: Timeliness refers to the lapse of time between publication and the period to which the data refer.

The population estimation is carried out at a snapshot in time and results are available typically 2 to 3 months after that time. This is possible because existing data assets are used and no new information is required to be gathered.

Accessibility and Clarity: Accessibility is the ease with which users are able to access the data. It also relates to the format(s) in which the data are available and the availability of supporting information.

Outputs are handed over to the relevant local authority staff with full training, description and metadata. These are available only to nominated and approved staff at the individual and household level due to the potentially identifying nature of the data. Staff can provide aggregate outputs for others if required.

This output is in database format for ease of use in statistical systems.

Outputs are owned and managed by the local authority only and are not otherwise publicly available unless in published form.

Comparability: The degree to which data can be compared over time and domain.

Over time, where a local authority has commissioned population estimations more than once, outputs are comparable in that the same datasets (for that snapshot in time), methodology and assumptions are used each time. Input datasets are kept consistent as much as possible but are vulnerable to known change issues associated with administrative datasets (noted elsewhere).

Geographically, outputs between the local authorities that have outputs available are comparable in the same way as over time, mentioned previously. It has not been quantified if the datasets and assumptions are biased by different types of local area, therefore caution should be used in this respect.

Outputs are not available for every local authority area in England and Wales and are therefore not nationally comparable.

Coherence: The degree to which data that are derived from different sources or methods, but which refer to the same phenomenon, are similar.

Outputs are compared to available benchmarks to ensure results are sensible.

Population estimations from other sources may differ due to different methodologies, assumptions and definitions used and care should be taken when making comparisons.

References

ADT (2012). The UK administrative data research network: improving access for research and policy. Report from the Administrative Data Taskforce. http://www.esrc.ac.uk/_images/ADT-Improving-Access-for-Research-and-Policy_tcm8-24462.pdf

Altman, D. G. (1999). *Statistics for Medical Research*. London: Chapman & Hall.

Baffour, B., King, T., & Valente, P. (2013). The modern census: evolution, examples and evaluation. *International Statistical Review, 81*(3), 407–425.

Bowling, A. (1991). Social support and social networks: their relationship to the successful and unsuccessful survival of elderly people in the community: an analysis of concepts and a review of the evidence. *Family Practice, 8*(1), 68–83.

Bronte-Tinkew, J., & Dejong, G. (2005). Do household structure and household economic resources predict childhood immunization? Evidence from Jamaica and Trinidad and Tobago. *Population Research and Policy Review, 24*, 27–57.

BSPS meeting (18th May 2015). The 2012-based household projections for England: methodological issues. LSE, London. http://www.lse.ac.uk/socialPolicy/BSPS/dayMeetings/Home.aspx.

DCLG/RSS meeting (2013). *DCLG Household projections*. Royal Statistical Society, London.

Department for Communities and Local Government (2008). *Options for the future of the household projection model*. London: Communities and Local Government. (Available from http://www.communities.gov.uk/documents/housing/pdf/housingprojectionmodel.pdf)

Department for Communities and Local Government (2010b). *Consultation on proposed changes to the national statistics on household projections*. London: Communities and Local Government. (Available from http://www.communities.gov.uk/documents/housing/pdf/1487114.pdf).

Department for Communities and Local Government (2010). *Updating the Department for Communities and Local Government's household projections to a 2008 base – methodology*. London: Communities and Local Government. (Available from http://www.communities.gov.uk/documents/statistics/pdf/1780350).

Department for Communities and Local Government (2013). *2011-based Interim Household Projections Quality Report*. London: Communities and Local Government.

Eurostat Statistical Books (2009). Reconciliation between work, private and family life in the European Union, section 4.1.4. Luxembourg: Office for Official Publications of the European Communities.

Fender, V., Foster, R., Khan, A., Punt, S. & Carolan, G. (2013). *Household satellite accounts, valuing informal childcare in the UK*. Office for National Statistics.

Harper, G. (2002). Using surfaces to inform local policy – modelling deprivation in Brent. Unpublished MSc dissertation.

Harper, G., & Mayhew, L. (2012a). Using administrative data to count local populations. *Applied Spatial Analysis and Policy, 5*(2), 97–122.

Harper, G., & Mayhew, L. (2012b). Applications of population counts based on administrative data at local level. *Applied Spatial Analysis and Policy, 5*(3), 183–209.

HM Government (2011). *Growing the social investment market: A vision and strategy*. London: Cabinet Office. (Available from http://www.parliament.uk/deposits/depositedpapers/2011/DEP2011-0271.pdf).

Holloway, Short, & Tamplin (2002). *Household satellite account (Experimental) Methodology*. London: Office for National Statistics.

Holmans, A. E. (2012). *Interim revised estimates of future demand and need in England in 2006 – 2026 with 2008-based demography*. Cambridge Centre for Housing and Planning Research. (Available from http://www.cchpr.landecon.cam.ac.uk/Downloads/Future%20demand&need%20WEB%20COPY.pdf).

House of Commons Public Administration Select Committee (2014). *Too soon to scrap the census. Fifteenth report of session 2013–14.* House of Commons London: The Stationery Office Limited.

House, T., & Keeling, M. J. (2009). Household structure and infectious disease transmission. *Epidemiology and Infection, 137*(5), 654–661.

King, K., Martinenko, M., Bergman, M., Liu, Y.-H., Winickoff, J. P., & Weitzman, M. (2009). Family composition and children's exposure to adult smokers in their homes. *Pediatrics, 123*(4), 559–564.

Larsson, K., Thorslund, M., & Kareholt, I. (2006). Are public care and services for older people targeted according to need? Applying the behavioural model on longitudinal data of a Swedish urban older population. *European Journal of Ageing, 3*(1), 22–33.

Leeds City Council (2011). *Leeds strategic housing market assessment update*. Leeds City Council. (Available from http://www.leeds.gov.uk/files/Internet2007/2011/27/final%20leeds%20shma%2027-05-11(1).pdf).

Marmot, M. (2010). The Marmot review - fair society healthy lives. Strategic review of health inequalities in England post-2010. (Available from http://www.instituteofhealthequity.org/).

Mayhew, L., Harper, G., & Waples, S. (2011). Counting Hackney's population using administrative data – an analysis of change between 2007 and 2011. (Available from http://www.hackney.gov.uk/Assets/Documents/estimating-and-profiling-the-population-of-hackney.pdf).

McDonald, N. & Williams P. (2014). Planning for housing in England: Understanding recent changes in household formation rates and their implications for planning for housing in England' RTPI Research Report no.1.

Ohwaki, K., Hashimoto, H., Sato, M., Tamiya, N., & Yano, E. (2009). Predictors of continuity in home care for the elderly under public long-term care insurance in Japan. *Aging Clinical Experimental Research, 21*(4–5), 323–328.

ONS. (2000). *Household Satellite Account (Experimental)*. London: Office for National Statistics.

ONS (March 2014). The census and future provision of population statistics in England and Wales: Recommendation from the National Statistician and Chief Executive of the UK Statistics Authority.

Ulker, A. (2008). Household structure and consumption insurance of the elderly. *Journal of Population Economics, 21*(2), 373–394.

UNECE (2011). Measurement of different emerging forms of households and families. Prepared by the Conference of European Statisticians Task Force on Families and Households.

Van der Heyden, J. H. A., Demarest, S., Tafforeau, J., & Van Oyen, H. (2003). Socio-economic differences in the utilisation of health services in Belgium. *Health Policy, 65*(2), 153–165.

Varjonen, J. & Aalto, K. (2006). *Household production and consumption in Finland 2001: Household satellite account*. Statistics Finland and the National Consumer Research Centre.

Vaupel, J. (2010). Biodemography of human ageing. *Nature, 464*(25), 536–541.

Voas, D., & Williamson, P. (2001). The diversity of diversity: a critique of geo-demographic classification. *Area, 33*(1), 63–76.

Wagner, K. D., Ritt-Olson, A., Soto, D. W., & Unger, J. B. (2008). Variation in family structure among urban adolescents and its effects on drug use. *Substance Use and Misuse, 43*(7), 936–951.

Zhang, L. (2012). Topics of statistical theory for register-based statistics and data integration. *Statistica Neerlandica, 66*(1), 41–63.