

Harper, G. & Mayhew, L. (2012). Re-thinking households - using administrative data to count and classify households with some application (Report No. Actuarial Research Paper No. 198). London, UK: Faculty of Actuarial Science & Insurance, City University London.



**CITY UNIVERSITY  
LONDON**

[City Research Online](#)

**Original citation:** Harper, G. & Mayhew, L. (2012). Re-thinking households - using administrative data to count and classify households with some application (Report No. Actuarial Research Paper No. 198). London, UK: Faculty of Actuarial Science & Insurance, City University London.

**Permanent City Research Online URL:** <http://openaccess.city.ac.uk/2329/>

#### **Copyright & reuse**

City University London has developed City Research Online so that its users may access the research outputs of City University London's staff. Copyright © and Moral Rights for this paper are retained by the individual author(s) and/ or other copyright holders. All material in City Research Online is checked for eligibility for copyright before being made available in the live archive. URLs from City Research Online may be freely distributed and linked to from other web pages.

#### **Versions of research**

The version in City Research Online may differ from the final published version. Users are advised to check the Permanent City Research Online URL above for the status of the paper.

#### **Enquiries**

If you have any enquiries about any aspect of City Research Online, or if you wish to make contact with the author(s) of this paper, please email the team at [publications@city.ac.uk](mailto:publications@city.ac.uk).



**Cass Business School**  
CITY UNIVERSITY LONDON

**Faculty of Actuarial  
Science and Insurance**

**Actuarial Research Paper  
No. 198**

**Re-thinking households – Using  
administrative data to count and  
classify households with some  
applications**

**Gillian Harper  
Les Mayhew**

**May 2012**

ISBN 978-1-905752-32-4

Cass Business School  
106 Bunhill Row  
London EC1Y 8TZ  
Tel +44 (0)20 7040 8470  
[www.cass.city.ac.uk](http://www.cass.city.ac.uk)

“Any opinions expressed in this paper are my/our own and not necessarily those of my/our employer or anyone else I/we have discussed them with. You must not copy this paper or quote it without my/our permission”.

## **Re-thinking households - Using administrative data to count and classify households with some applications**

**Gillian Harper and Les Mayhew**

**Key words: administrative data; local; household classification; household statistics; policy**

### **Summary:**

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, present 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 based on local administrative sources. The result is a more integrated and flexible system. The utility and advantages are demonstrated using recent examples from the six Olympic London Boroughs.

### **1.1 Introduction**

#### *Why analyse households?*

Historically households were recognised as an important unit for economic and sociological analysis with concepts such as ‘breadwinners’ and head of household’. Much of this household analysis was predicated on the assumption that males’ roles were economically and socially more important than females’. However even without this gender bias there are sound reasons for looking at households:

- (a) Households are self contained economic units of production and consumption: income and expenditure are often shared. For example, the link between households and consumption patterns has resulted in classification systems that are widely used in retail marketing and increasingly in social marketing (Webber and Farr, 2000). Statistics Finland, for example, found that GDP is increased by 40% and household consumption by 60% when household production is included in the National Accounts (Varjonen and Aalto 2006).
- (b) Specifically for the public sector, households are transactional units for the purposes of paying utility bills, property taxation and for rubbish collection. The household unit is used for measuring poverty. 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/technical/hbai.shtml](http://www.poverty.org.uk/technical/hbai.shtml)).
- (c) Household characteristics such as occupancy, average resident age, tenure status, and gender composition are often good predictors of social outcomes and the take up of services including health, social care and education (Bowling, 1991; Ohwaki et al. 2009; Larsson et al. 2006; Ulker, 2008).

## *Re-thinking households using administrative data*

- (d) The Marmot Review (Marmot, 2010) emphasises how poor health arises from inequalities in society in the conditions in which people are born, work, and age. Solutions, the review argues, include supporting families and lifting households out of poverty.
- (e) Household characteristics are significant determinants of life expectancy. Vaupel (2010), remarking on why there has been such an expansion in life expectancy, considers that only about 25% of the variation in adult life spans could be attributed 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”.
- (f) The social and economic importance of households is established in numerous other social domains. Examples of studies that report outcomes by household composition are childcare (Eurostat, 2009), accessing GPs, nursing care and hospital admissions (Van der Heyden et al. 2003), childhood immunisation (Bronte-Tinkew and Dejong, 2005; House et al. 2009), exposure to smoking (King et al. 2009), and alcohol and marijuana use (Wagner et al. 2008).
- (g) Socially excluded or otherwise challenged households that require multiple interventions from different agencies have been the focus of attention among both the current and previous governments mainly because high social costs they incur. There is now considerable interest in intervention models which earn a return for investors if certain measurable outcomes are met e.g. a household is taken out of welfare benefits (HM Government 2011).

### *Sources of data*

There are several disparate and unconnected sources of data on households in the United Kingdom. These include surveys which often have a secondary purpose in terms of underpinning household classifications but their main purpose is for research and policy development at a national level. The Economic and Social Data Service (ESDS) disseminates many of the UK large-scale survey datasets that are available for households such as the General Household Survey, the Labour Force Survey, and Family Expenditure Survey (<http://www.esds.ac.uk/government/surveys/>).

Presently, the main official source of statistics on households in England is the Department for Communities and Local Government (DCLG). Their projections are linked to the latest Office for National Statistics sub-national population projections and are available at district (i.e. Local Authority level). They provide the evidence base for the assessment of future housing requirements and are used by DCLG, other government departments, the National Housing and Planning Advice Unit, regional planning bodies, and local authorities.

In spite of the above, a major barrier for users is a lack of timely and accurate information at a sub-local authority level. There are commercial systems that are able to provide household level data. However, because they also rely to a large degree on census and synthetically estimated data they suffer from similar problems with accuracy and timeliness. Because support for households and families is largely provided through local authorities and the voluntary sector, their work is being

## *Re-thinking households using administrative data*

severely hampered by the lack of evidence for social investment initiatives or calibrating interventions (HM Government, 2011; Harper, 2002; Voas and Williamson, 2001).

Given the above, it is hard to avoid the conclusion that the present state of household statistics is unsatisfactory. Available information tends to be inflexible in terms of one or more attributes and, crucially, the data is imputed or modelled in some way using out of date information. In a DCLG consultation in 2008, for example, users said that the rapidly changing population in some areas reduced their accuracy and value and figures were not easy to reconcile with other data (Department for Communities and Local Government, 2008).

For all the above reasons, this paper puts forward a different basis for collecting and maintaining household statistics using locally available information sources. The key advantages of this approach is that the data are at household level, can be produced on a timely basis and can be linked to other data particularly information about services supplied. This does however require knowledge of local data sources and how to access and exploit them.

The approach is effective because it creates one set of base data for all users rather than the silo-based and fragmented approach that currently exists. It is flexible because it allows data to be produced to any level of geography, cheaper because it does not involve commercial licenses and efficient because it gives analysts, service providers and policy makers new and more accurate tools to aid decision making, and dynamic because users can create their own household classification system to suit different applications.

### **1.2 Aims of paper**

In this paper, we consider the use of administrative sources of data for defining and quantifying households rather than existing sources for the reasons given. In the absence of access to national administrative sources, we focus our attention on local areas using examples of areas with populations of around 300k but the methodology we develop is both general and flexible and can be scaled up to any size of area.

A problem is that whilst administrative data sets and registers at the household level may be a viable source for capturing the population, the data need to be linked and analysed systematically before they can be used for statistical purposes. Harper and Mayhew (2011a and b) describe a methodology for combining local administrative data sets to create population counts using a formal system of logic based on 'truth tables' and also provide examples of applications. We build on that research in this paper.

Local authorities and health trusts hold a wealth of such data on their local populations that can have added value by linking them together and using them in this way. The data base which is the end product of this process is not identical to the census, but it produces core demographic data by individual and household that in practical terms can be linked to a wide range of other administrative data sets. The examples in this paper are drawn from the six London host Olympic boroughs and are

## *Re-thinking households using administrative data*

based on work undertaken by Mayhew Harper Associates during 2011 (see for example Mayhew et al. 2011).

Because, unlike the census outputs, each person is geo-referenced, the data can be flexed according to geography and other dimensions. The form of aggregation considered in this paper is into households; however, because households exist in a multitude of forms, some of which are more common than others, it is necessary to establish a framework that can be adapted to suit specific issues.

Our overall aim is to enhance evidence for policy-making and our objectives are fourfold:

- (a) To compare and contrast administrative and official sources of statistical information on households
- (b) To examine the scope for defining more flexible classifications of households that rely on locally collected administrative data and are more up-to-date
- (c) To consider how such definitions can be extended to include attributes such as housing tenure and income deprivation
- (d) To show how administrative data can be used to evaluate and inform local policy and decision-making

In particular, we use administrative data as a basis for a framework in which it is possible to define, enumerate and analyse different household structures. The framework is inclusive in terms of the whole population but flexible in terms of dimensions such as size, occupancy and age. We show how different household demographic types can be linked to attributes such as income deprivation and housing tenure and therefore useful in policy analysis. The examples however are not exhaustive and it is possible to create as many attributes as there are data to support them.

### **2.1 Definitions and current sources of data on households**

The 2011 Census 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, dining room or kitchen’ (Cabinet Office, 2008). However, this definition has changed in every Census since 1971, mainly in minor ways, but it also changed more significantly in 2001 (Grundy et al. 2010).

The Joseph Rowntree Foundation defines a household as a single person or group of people living at the same address who share common housekeeping or a living room (Palmer et al. 2006). A ‘dwelling’ on the other hand is a self-contained unit of accommodation (including the basic facilities of kitchen, bathroom and toilet) which has its own front door. There is a distinction because a dwelling can be occupied by a single household or by a number of households.

The current statistical system in England which broadly adheres to these definitions is based on sub-national household projections which are published every two years to

## *Re-thinking households using administrative data*

maintain consistency with the national population projections. They are termed 'projections' because in reality they are based on information that is pieced together using the ten yearly census, population estimates for intervening years, and survey sources such as the Labour Force Survey.

Producing household statistics is split into two main stages coordinated by DCLG (Department for Communities and Local Government, November 2010)). The first is ONS (Office for National Statistics) 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 the 1991 and 2001 population Censuses to estimate the proportions of households by local authority area, household type etc.

Our main concerns with current arrangements may be summarised as follows:

- (a) Household statistics are not actual figures but in reality only sophisticated estimates;
- (b) The level of geography available is restrictive – some counts only apply nationally and it is not possible to go below local authority level;
- (c) Users have little control over either definitions of households or choice of geography;
- (d) The data cannot be linked to other attributes such as deprivation, ethnicity and so forth.

Commercial household classifications are available at a cost to users. The typologies used tend to be colourful in their descriptions and do not enumerate household demographics. However, they are also highly reliant on the Census.

Our argument in this paper is that administrative data can potentially address most of these points and offer significant advantages but they are not perfect substitutes. For example, it is possible to derive the exact age and gender of current residents at every residential property address in an area with obvious advantages in a range of applications. Based on separate research (not included in this paper), we also demonstrate how administrative data can be used to assign ethnicity to households.

On the other hand administrative data does not generally allow identification of family or marital relationships, nor do they tell us whether household members share cooking and other facilities to allow distinguishing between a household and a dwelling. Instead, we make the simplifying assumption that people at the same residential address constitute a 'household'. HMOs (Houses in Multiple Occupation) are identified as separate entities using administrative data and therefore as separate households rather than be lost within one dwelling address.

This may not be as big an information loss as it first appears. For example, information about marital status is obtained from the Census which is only conducted 10-yearly and so is largely out of date, and for most normal applications an address or 'dwelling' based definition of households is a sufficient basis for analysis. This fits with the grain of how information is recorded in administrative systems, whether it is for council tax purposes, benefits or utility bills and so definitional units are consistent and verifiable.



## *Re-thinking households using administrative data*

The current household typology based on the government's own published methodology is shown in Table 1 (from Department of Communities and Local Government, November 2010). In the next section we compare household counts based on these definitions with our own methodology to see how well they align. We argue that our approach is able to replicate these but is also capable of a much richer variety of household types that are suited to a range of different purposes to any level of geography. In this way, we argue that the information gain so derived significantly outweighs any information loss.

**Table 1: The government household typology scheme**

Household type	Description
One person households	Male Female
One family and no others†	Couple‡: No dependent§ 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§§	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*	See notes
† Households with dependent children and no non-dependent children	
‡ 'Couple households' are either married or cohabiting	
§ 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)	
§§ In these categories, the other adults may include another couple and/or another lone parent and/ or a non-dependent child	
* The 'Other households' category above is an aggregation of five categories from the original Census table C1092 supplied by ONS	

Official household statistics distinguish between communal institutions (i.e. establishments providing managed residential accommodation) and private households as separate categories. The main communal establishments include 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.

Identifying communal establishments is not a straight forward process and DCLG identify these in a number of ways (Department for Communities and Local Government, November 2010, Appendix 2). Using administrative sources, the process of identifying actual communal establishments from current local registers and property gazetteers tends to be simplified. In the next section, we compare DCLG household counts with administrative counts, but for the purposes of comparing our figures with DCLG's we first remove communal establishments from the administrative data.

## **2.2 Comparison of counts using official versus administrative definitions**

We now investigate the extent to which household counts based on administrative data replicate DCLG household counts. Whilst we do not expect to find an exact correspondence, it is useful nonetheless to identify reasons for similarities and differences. From previous experience, we expect to encounter two potential problems. One is the different basis used to count populations (i.e. the Census is survey based and the other administration based); the second is the process of translating administrative data into exact copies of DCLG household definitions.

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 2009 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 fact that a third source is also available affords the opportunity to compare and benchmark our household counts with two sources rather than one.

Our results for the year 2011 are shown in Table 2 for each of the six Olympic boroughs. The table compares household counts based on administrative sources *without* communal establishments with comparative figures from DCLG and GLA also *without* communal establishments. The results show that for the whole region administrative household counts are 0.4% higher than GLA at 578k but 6.6% higher than DCLG. Within the region the differences vary by local authority but two differences that particularly stand out are between the administrative and DCLG household counts for Hackney and Newham.

*Re-thinking households using administrative data*

**Table 2: Comparison based on total number of households by local authority using administrative, DCLG and GLA data in 000s.**

Local authority	Households ('000s)			difference % in households		Total available dwellings ('000s)	vacant properties ('000s)			vacant properties %		
	admin 2011 (A)	GLA 2011† (B)	DCLG 2011‡ (C)	admin - GLA	admin- DCLG	admin 2011	admin 2011 (A)	GLA 2011† (B)	DCLG 2011‡ (C)	admin 2011 (A)	GLA 2011† (B)	DCLG 2011‡ (C)
Barking and Dagenham	70.5	72.3	69.3	-2.5	1.7	72.9	2.4	0.6	3.6	3.3	0.8	4.9
Greenwich	101.6	106.7	99.5	-4.8	2.1	112.8	11.2	6.1	13.3	9.9	5.4	11.8
Hackney	103	98.1	92.1	5	11.8	107.9	4.9	9.8	15.8	4.5	9.1	14.6
Newham	104	103.2	91.8	0.8	13.3	108.9	4.9	5.7	17.1	4.5	5.2	15.7
Tower Hamlets	101.2	100.6	98	0.6	3.3	115.7	14.5	15.1	17.7	12.5	13.1	15.3
Waltham Forest	97.9	95.1	92	2.9	6.4	103.1	5.2	8	11.1	5	7.8	10.8
<b>Total</b>	<b>578.2</b>	<b>576</b>	<b>542.7</b>	<b>0.4</b>	<b>6.6</b>	<b>621.3</b>	<b>43</b>	<b>45.3</b>	<b>78.6</b>	<b>6.9</b>	<b>7.3</b>	<b>12.7</b>

† copyright © Greater London Authority, 2011

‡ copyright © CLG, 2010

## *Re-thinking households using administrative data*

Another key measure is estimates of the number of vacant properties. We define the vacant property rate as the number of households divided by the total number of residential addresses on the LLPG (Local Land and Property Gazetteer) having first removed all community establishments from the LLPG for consistency. Table 2 shows the consequential differences in vacant property rates between the three sources. These range in value from 0.8% in Barking and Dagenham using GLA definitions to as high as 15.7% in Newham using DCLG definitions.

There are two key reasons why rates may vary to this extent and in some cases appear exceptionally high. One is due to the exceptionally active regeneration in the City fringe and Docklands areas affecting mainly Tower Hamlets where there are large numbers of apartments that are as yet unoccupied. The other can be primarily traced to the lower ONS population counts on which DCLG household counts are based.

Our work in the six Boroughs produced an administrative-based population count of 1.46m, which is 0.8% higher than the GLA's but nearly 11% higher than the equivalent count published by the ONS. 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 240k compared with our own figure of 299k. Following revisions to their methodology, the ONS released new figures in November 2011 in which Newham's population had increased from 240k to 272k. This may be compared with figures published by the GLA which increased its own estimate for Newham from 268k to 296k in June 2011, a figure that was partly informed by our own work which by then had been made available.

The discrepancies between ONS, GLA and administrative 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. In the approach described in the following sections we link households to population counts in a very clear and precise way which affords much flexibility over household definitions and also greater control over population counts including their validation.

### *Split by household type*

The second issue to discuss is whether the above discrepancy feeds into and therefore distorts the breakdown of households by type. On the face of it, there is no reason why this should impact unduly as long as definitions are comparable even if the total quantum of households differs. Since local authorities are separate legally constituted decision making units we considered it more appropriate to analyse one borough in detail rather than all at once.

We decided on Hackney because it not only has one of the largest difference in household counts based on our figures and DCLG's, but it also provides an excellent case study for testing and evaluating different household typologies because it is an inner city borough with a more complicated household structure than a typical suburban borough. It also allows us to bring together more factors into the mix such as tenure and deprivation than is possible using DCLG data.

## *Re-thinking households using administrative data*

For the moment our purpose is to recreate DCLG household types using administrative sources. As was shown in Table 1, DCLG definitions are quite demanding in terms of their specificity, so reproducing these figures is likely to provide a robust test. The first critical issue to consider is the definition of a dependent child. For DCLG proposes, 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). However, by far the most immediate practical problem is to identify children in full time education (FTE) using administrative data.

Our key data sources were the school pupil census, in which persons are separately flagged if they are aged 16-18 and attended a school in Hackney or a neighbouring borough. They are also flagged if they are on Connexions data in which young people are as flagged as 'FTE' if they are in full time education (Connexions is the national information and advice service for young people that tracks young people's intended destination from school year 8 until they are 20). However, this process is not exhaustive since there were some children receiving education in the independent sector or education in establishments located in boroughs for which we had no information.

One person households are identified as persons aged 16 or over living on their own: anyone aged less than 16 living on their own are considered to be in the category 'other' households. The category 'one family and no others' includes mixed sex couple households aged 19 and over with or without dependent children or households with only one adult aged 19 and over with dependent children.

The categories of 'couple or lone parent with one or more other adults' are combined into one category 'family households with other adults and dependent children' where these households have dependent children. In evaluating the DCLG classification system, the definition of a 'couple' only includes mixed sex, married or cohabiting couples, as defined by the ONS and marriage in particular may be hard to verify using administrative data except by shared surname. By the same token, same sex couples would not be considered a 'family' household in the DCLG classification.

A comparison of the results using DCLG figures and administrative sources is shown in Table 3. Although, as was expected, absolute numbers are lower for DCLG as compared with administrative sources for the reasons given above, the percentage breakdown turn out to be very similar for each household type. This gives us confidence that administrative sources are able to replicate the basic split in households if not the total quantum of households.

As we shall argue, policy makers and planners need to be able to identify key differences between household types e.g. pensioner households or three-generation households, and on this basis the DCLG types described in this section 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.

*Re-thinking households using administrative data*

**Table 3: Comparison of percentage of households in the London Borough of Hackney belonging to each DCLG household type using DCLG and administrative data**

Household type	Administrative sources 2011 (000s) (A)	DCLG 2011 (000s) (B)	difference (000s) (C)	% of households by type (admin) (C)	% of households by type (DCLG) (D)	% difference (C-D)
Family household(couple)	20.5	18.2	2.3	19.9	19.8	0.1
Family household (lone parent)	8.6	7.7	0.9	8.3	8.4	-0.1
Family household with other adults& dependent children	10	8.2	1.8	9.7	8.9	0.8
One person household	47.2	43.5	3.7	45.8	47.2	-1.4
Other households	16.7	14.5	2.2	16.2	15.7	0.5
Total	103	92.1	10.9	100	100	0

To summarise, we find that local administrative data is able to replicate the household types in the standard DCLG classification in Table 1, suggesting that both approaches can arrive at similar results; however, the underpinning population on which government figures are based are too aggregate and out of date before they are published. In the next sections we present the case for a different approach for counting and classifying households which uses only administrative data and which deals effectively with many of these drawbacks.

### **2.3 Alternative household classification systems using administrative data**

In what follows, our objective is to develop a new framework for counting and analysing households which is based on detailed demographic data that is extractable from locally available administrative systems. Since this data is captured in real time, household statistics, unlike DCLG figures, are synchronised snapshots at a point in time rather than projections.

We start with a brief review of the stages in the production process beginning from administrative sources of data to the creation and enumeration of household types, and finally examples of uses to inform policy and decision-making in local areas. The first stage in this approach is the prior creation of an administrative population count.

This may be conceptualised as a data base in which every person on the data base is allocated to an address or more exactly a UPRN (Unique Property Reference Number corresponding to an address and an entry on the Local Land and Property Gazetteer). The number of people and their ages then form the basis for the household classification.

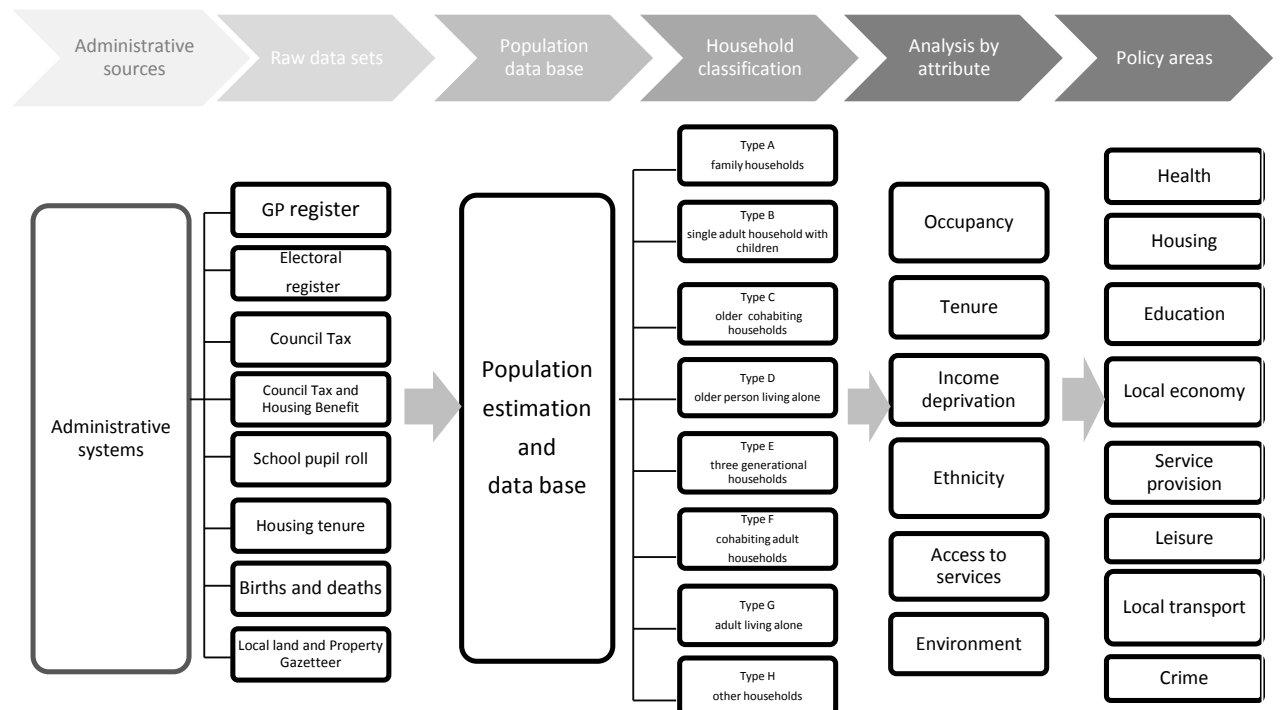
Figure 1 summarises the stages in the process that starts with a list of the administrative systems that provide the starting point for the creation of the person level data base. A technical description of this process is beyond the scope of this paper but full details may be found in Harper and Mayhew (2011a and b) and this was the method used in the Six Olympic Boroughs study of 2011. Reference 2011b, for example, provides examples of applications of the methodology to which the results set out in this paper can be considered complementary.

Since our interest in this paper is in households, we focus on the processes for aggregating person level data by age, sex or other attributes into suitable household types. In our standard classification, we define eight basic types which are defined below. The stage after this is to append to each household data pertaining to particular attributes of households such as household size, housing tenure, or benefit status. The objective is to demonstrate how easy it is to develop sub-types of households for addressing specific issues of interest.

Note that that whilst these are common enough attributes, households can also be characterised in a host of other ways. For example, they can be categorised on the basis of their levels of access to services and amenities in an area and so on. This could include something as elementary as proximity to a bus stop or proximity to a basket of local services. A case in point might be older households which are more likely to use local shops, health care, libraries, post offices and so on.

Other examples that meet a specific need in local authorities include identifying workless households, children households with poor school attendance or households suffering isolation or neglect, all of which involve local authority services and responsibilities to a degree. In these cases a distinction is drawn between statistical uses of data and case management which rely on personal identification. Other attributes we have considered and developed approaches for include ethnicity identification although a discussion of the techniques involved is outside the scope of this paper.

**Figure 1: Stages in the production of person and household level data and the policy domains supported**



### 3. Enumerating household types

In this section, we lay the foundations for enumerating and analysing different household types. We start by assuming 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 both categories and sub-categories of household in a single consistent framework which can be used for multiple purposes.

Initially, we define eight household types based on the definitions in Figure 2 which we call the ‘standard types’. Using age and size of households as two descriptors we can divide each type into their constituent age groups as shown in Table 4 which shows examples of each. We let age group 1 represent children, age group 2 working age adults, and age group 3 older adults.

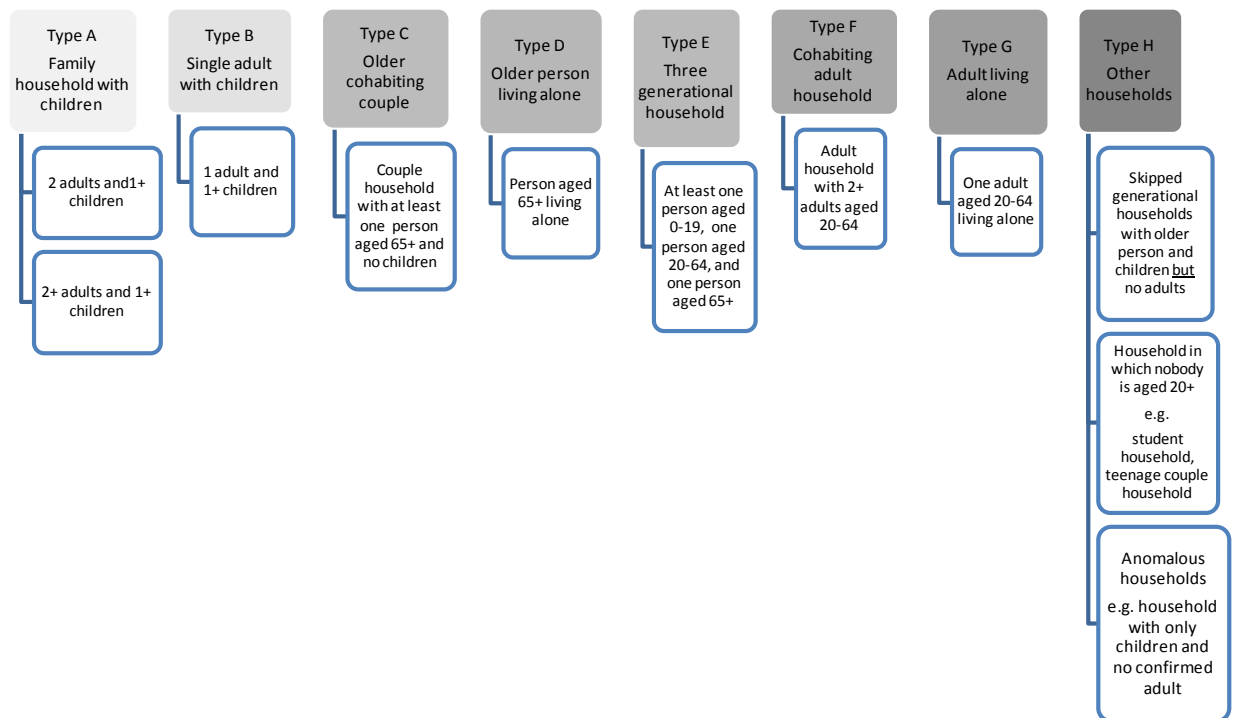


## *Re-thinking households using administrative data*

Row one is a Type A couple household with two children and two or more adults (the additional adults could be an older sibling, friend or relative, lodgers or someone temporarily resident at an address ); row two a Type B household with one adult and one child and so on. Of the examples shown, Type H households in the last two rows are the least homogenous and can be split into several sub-categories as required. In these cases the occupants are all young people (possibly students), or are from a split generation (e.g. a household in which children live with their grandparents). Note that in this classification scheme, couples may be mixed sex or same sex which can be separated out into sub-categories as required.

Type H is easily the smallest group among the standard types and may include a few households with young children and no adults. These are cases where the administrative data has identified children as living at an address but no teenage parent or adult. The most probable explanation in this small number of cases is that the adults are unregistered with a GP and do not appear on other data bases.

**Figure 2: Stages in the production of person and household level data and the policy domains supported**



**Table 4: 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	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 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 parent)

Although administrative data cannot determine whether a couple household is a married household 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 not essential to know this. Conversely, the ability to specify both age widths and household size offers scope to study various attributes of households in 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:

$$N = 1 + \frac{r}{1!} + \frac{(r+1)r}{2!} + \frac{(r+2)(r+1)r}{3!} + \frac{(r+3)(r+2)(r+1)r}{4!} + \dots + \frac{(r+k-1)(r+k-2)\dots(r+1)r}{k!} \quad (1)$$

Where  $n$  is the number of occupants (1, 2, 3, 4... $n$ ) and  $r$  is the number of age categories (1, 2, 3, 4... $r$ ). Each term inside the brackets multiplied by  $r$  gives the number of households with 1, 2, 3, 4... $n$  people plus one further combination for the 'void' case where an address is not occupied and therefore undefined. The inclusion of void households is retained in resultant set of household accounts 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 3+6+10+15=34 combinations of household types with 3 age categories and 4 people or 35 if the void case is included. This is shown in Table 5 row 3, which enumerates all possible combinations for up to six age categories and six occupants.

**Table 5: Possible combinations of household demographic types based on size and age**

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

The matrix in Table 5 is symmetrical about the diagonal so that there are the same number of combinations with  $j$  age categories and  $m$  occupants as there is with  $m$  age categories and  $j$  age categories. We can also note that each cell is the sum of the row above plus one.

For example, consider row three for up to 4 occupants the possible combinations are:

$$1 + \sum_{n=1}^{n=4} N_{3i} = N_{44} = 1 + 3 + 6 + 10 + 15 = 35 \quad (2)$$

It can also be shown that the value for any particular cell is also given by:

$$N = \frac{(n+r-1)!}{(r-1)!n!} \quad (3)$$

For example, for  $N_{44}$  this is  $\frac{7!}{3!4!} = 35$  which is the same as the previous result.

The inclusion of gender can also be considered although this leads inevitably to more variants which can become unwieldy in general use but may be highly relevant in specific applications.

If we are only concerned with the gender mix of a household and not with gender mix within an age group, a household with 2 occupants will have 3 possible gender combinations (MM, MF, FF) and one with 3 occupants four gender combinations (MMM, MMF, MFF, FFF) and so on. This is reflected in Table 6 which shows the possible combinations of household demographic types based on size and age and gender.

**Table 6: Possible combinations of household demographic types based on size and age taking gender into account**

Number of people per household ( <i>n</i> )						
number of age categories ( <i>r</i> )	1	2	3	4	5	6
1	2	3	4	5	6	7
2	4	9	16	25	36	49
3	6	18	40	75	126	196
4	8	30	80	175	336	588
5	10	45	140	350	756	1470
6	12	63	224	630	1512	3234

#### **4. Classifying and counting households**

Because by using administrative data age categories and occupants can be specified exactly, this provides considerable scope for analysing specific features and attributes of households. This can be illustrated using the example of a base case with three age categories and up to four occupants which has the convenient feature that it can be tabulated on one page.

In most studies to date, we have used the following age ranges 0-19, 20-64 and 65+. These are fairly broad and yet descriptively rich categories which can be fine-tuned and sub-divided into smaller age groups as required (e.g. 0-4, 5-11, 12-16, 17-19). In practice, we seek to enumerate the number of each household type in a population by specifying both the age categories and occupancy and so tables can quickly grow to a large size.

It can be postulated (and would save on analytical resources if true) that the actual distribution of household sizes and types could be replicated on the basis of a few known demographic facts, such as the sizes of each age group. If household structures were predictable or at least partly predictable from the demographic structure alone, it would be a huge simplification in many areas of activity particularly in housing, education and social services departments.

The problem is essentially one of transforming demographic data into the most household structures according to a set of rules based on a small number of demographic variables and then comparing the results with an exogenous distribution with known properties and outcomes. The aim is ambitious since much will depend on the population structure of the area being investigated (rural or urban for example), cultural and economic factors.

One way to proceed is to consider whether household patterns are the outcome of a random process (i.e. observed household patterns are simply the result of chance). Such approaches have been commonplace in geography or the natural sciences for many years and are used to determine whether processes generating a particular distribution match reality (e.g. Lewis, 1977; Rogers, 1974). If it is purely random then it means that household composition could be predicted on the basis of population demography alone.

Intuitively, a purely random process is likely to be restrictive because it requires that any household has an equal probability of containing a person and that a person at an

address lives there independently of any other person at that address. The contrary argument is that there are societal processes such as the desire to live in families or the need to couple adults with children which cause patterns to deviate from the expected.

However, the amount by which a random process deviates from actual household patterns is itself informative and may have some predictive value. Suppose that the total population is  $P$  and the number of households is  $N$ , the average household size is  $P/N$  or  $\lambda$ . The expected number of households of size  $n$  is given by the Poisson distribution and the expected age composition for any level of occupancy within a sub-type is described by the multinomial distribution.

Combining both distributions we obtain the following expression for the expected number of households of size  $n$  which are split into age categories  $n_1, n_2, n_3$  such that

$$n_1 + n_2 + n_3 = n$$

This is given by:

$$Np(n_1, n_2, n_3, \lambda) = N \frac{\lambda^n e^{-\lambda}}{n_1! n_2! n_3!} p_1^{n_1} p_2^{n_2} p_3^{n_3} \quad (4)$$

Where  $\lambda = \frac{P}{N}$

$n$  = household size

$n_i$  = the number of occupants in age category  $i$

$p_i$  = the proportion of the total population in age category  $i$

Continuing with our illustration, we use data from the London Borough of Hackney, one of the six London Olympic boroughs, with a population of 237,646 people at a snapshot date of 27/03/2011. Of the total population, 27.1% are aged between 0-19, 64.9% between 20-64 and 8% are age 65+. The total number of available registered addresses is 110,627 including community establishments and average occupancy of 2.148 persons per household.

For purposes of exemplification, Table 7 enumerates all possible combinations of household of up to four people in size based on the discussion in section 3 which we call the ‘base case’. We limit the example to occupancy levels of four simply to contain the size of the table so that it fits to a page.

It shows that there are 35 possible sub-types of household including the void case in row 1 (this is an empty property). The columns show the number of people in each age category, followed by household size in ascending order. A further column shows the corresponding household type based on the definitions in Figure 1 (A to H).

*Re-thinking households using administrative data*

**Table 7: Base case enumerating household type and sub types with random assignment of people to households based on household size of up to 4 people**

Case	age group 0-19	age group 20-64	age group 65+	household occupancy	household type	expected number of households	expected number of people
1	0	0	0	0	void	12,910	-
2	0	0	1	1	D	2,217	2,217
3	0	1	0	1	G	17,996	17,996
4	1	0	0	1	H	7,520	7,520
5	0	1	1	2	C	3,090	6,180
6	0	0	2	2	C	190	381
7	0	2	0	2	F	12,543	25,085
8	1	1	0	2	B	10,483	20,966
9	1	0	1	2	H	1,291	2,582
10	2	0	0	2	H	2,190	4,381
11	0	2	1	3	C	2,154	6,461
12	0	1	2	3	C	265	796
13	0	0	3	3	C	11	33
14	0	3	0	3	F	5,828	17,484
15	1	2	0	3	A	7,306	21,919
16	1	1	1	3	E	1,800	5,400
17	1	0	2	3	H	111	333
18	2	1	0	3	B	3,053	9,160
19	2	0	1	3	H	376	1,128
20	3	0	0	3	H	425	1,276
21	0	3	1	4	C	1,001	4,003
22	0	2	2	4	C	185	740
23	0	1	3	4	C	15	61
24	0	0	4	4	C	0	2
25	0	4	0	4	F	2,031	8,124
26	1	3	0	4	A	3,395	13,579
27	1	2	1	4	E	1,255	5,018
28	1	1	2	4	E	155	618
29	1	0	3	4	H	6	25
30	2	2	0	4	A	2,128	8,512
31	2	1	1	4	E	524	2,097
32	2	0	2	4	H	32	129
33	3	1	0	4	B	593	2,371
34	3	0	1	4	H	73	292
35	4	0	0	4	H	62	248
<b>Total</b>	<b>42779</b>	<b>75798</b>	<b>14751</b>			<b>103,214</b>	<b>197,114</b>
					households >4 people	7,413	40,532
					<b>Grand total</b>	<b>110,627</b>	<b>237,646</b>

## *Re-thinking households using administrative data*

With up to 4 occupants there are 3 distinct types of A and B households, 9 of type C, 4 of type E, and 10 of type H. Type C and Type D households are 1-person households of which by definition there is only one distinct type of each. Two final columns show the total number of households and population by household sub-type based on these data using the formula above.

Because the example only allows for occupancy of up to four people per household, it is necessary to add a residual category for households with more than four people as shown at the foot of the table. Among other things, the table highlights the fact that certain household sub-types are likely to be more common than others on the assumption of a purely random process.

For example, older households (types C and D) in which there are no children and at least one occupant aged 65+ are among the least numerous. This is because older people only comprise 8% of the total population and so are statistically more likely to live alone or in couples than in three-generational settings or households with only older people and children.

### **4.1 Analysis according to the eight standard types**

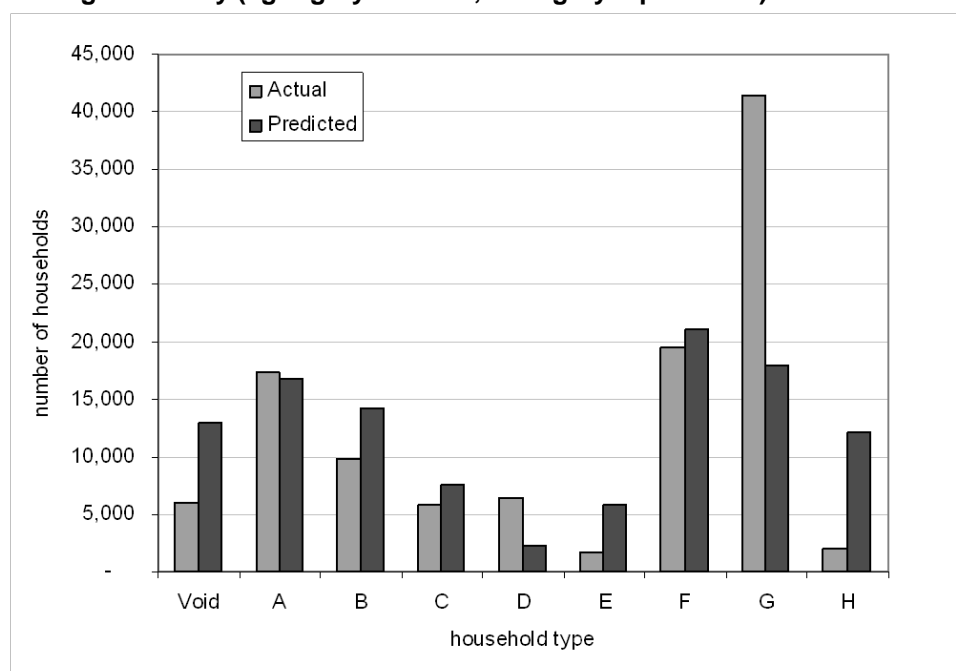
The potential utility of this table can be more readily understood if it is summarised by aggregating the information across household types. Table 8 and Figure 3 compare the expected number of households and population summarised by the eight standard household types: A, B, C etc. with the actual numbers of each type of household so that we can compare what actually occurs with a hypothesised random process.

The results show that many more adults aged 20-64 in Type G households live alone than would be the case if household formation were a purely random process. In addition, the actual number of Type B single parent households and population are substantially less than the expected number as are the number of Type E and Type H

**Table 8: The actual and expected number of households and people by household type based on data from the London Borough of Hackney**

Household type	actual households (A)	expected households (B)	Diff A-B	actual population	expected population	actual occupancy	expected occupancy
Void	5,975	12,910	-6,935	0	0	0.0	0.0
A	17,406	16,704	702	81,617	65,149	4.7	3.9
B	9,859	14,226	-4,367	28,074	32,996	2.8	2.3
C	5,885	7,531	-1,646	15,564	21,979	2.6	2.9
D	6,467	2,217	4,250	6,467	2,217	1.0	1.0
E	1,714	5,793	-4,079	9,255	24,681	5.4	4.3
F	19,527	21,131	-1,604	47,093	54,541	2.4	2.6
G	41,422	17,996	23,426	41,422	17,996	1.0	1.0
H	1,948	12,118	-10,170	2,764	18,068	1.4	1.5
>10	424	7	417	5,390	74	12.7	9.9
Total	110,627	110,633	-6	237,646	237,700	2.1	2.1

**Figure 3: The actual and expected number of households by type in the London Borough Hackney (light grey = actual, dark grey = predicted)**



Key to Figure 3

Category	Description
A	family households with dependent children
B	single adult households with dependent children
C	older cohabiting households
D	older person living alone
E	three generational households
F	cohabiting adult households no children
G	single adult households
H	other households

households. A third observation is that many more properties would be predicted to be empty than is actually the case.

We might speculate therefore that adults much prefer to live together as couples if they have children but express a desire to live alone if they do not and assuming they can afford to do so. Alternative living arrangements such as single parent households, 3- generational households, and to some extent co-habiting adult households are all consistently fewer than would be expected by comparison which may relate to selection processes and societal norms (for example, three generational households were once much more common than they are today and single person households much less common).

Another insight is that the much smaller than expected number of empty properties can be considered as indicative of the influence of the housing market which prefers occupied to unoccupied properties. This does not mean however that the housing market is seamlessly efficient as there are entry barriers due to high prices and supply side shortages in some housing categories which tend to impact on large families. We also find for example that the empty property rate in social housing is lower than in the private housing market.



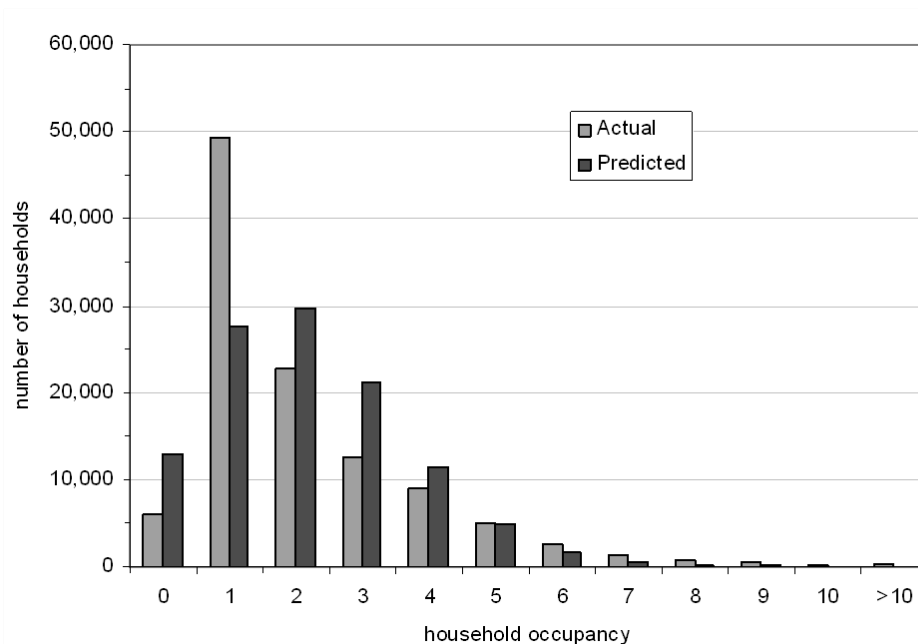
## *Re-thinking households using administrative data*

The extent to which the actual household demographics deviate from a random process may be extended in other ways. Table 9 and Figure 4 show the actual and expected household size based on occupancy from zero to ten people. They show that the general tendency is for a greater number of households to be smaller in size than expected, the main reasons for which are firstly that fewer households than expected are empty and secondly because more people than expected tend to live alone.

**Table 9: Household size and occupancy by tenure and occupancy status in the London Borough of Hackney 2011**

Number of occupants	number of households (A)	expected households (B)	diff A-B	% social housing (actual)	% benefit households (actual)
void	5,975	12,910	-6,935	0.0	0.0
1	49,284	27,733	21,551	43.0	32.6
2	22,803	29,787	-6,984	51.3	37.7
3	12,592	21,329	-8,737	60.1	46.6
4	8,972	11,455	-2,483	63.6	50.2
5	5,107	4,921	186	69.5	56.0
6	2,608	1,762	846	70.1	60.2
7	1,371	541	830	65.9	63.6
8	753	145	608	60.7	64.5
9	482	35	447	50.4	62.0
10	256	7	249	42.6	64.5
>10	424	7	417	0.2	60.1
Total	110,627	110,633	-6	48.1	37.3

**Figure 4: The actual and expected number of households by occupancy in the London Borough of Hackney (light grey = actual, dark grey = predicted)**



## *Re-thinking households using administrative data*

The size of household also affects the split between private and social housing tenure and entitlement to means tested benefits. Nearly 50% of households live in social housing in Hackney. This rate increases from 43% in single person households, rises to 70% with occupancy levels of six before falling again to 43% with occupancy levels of 10.

Means tested financial support for housing and council tax costs is paid locally through the welfare benefits system for households on low income and few assets. The table shows that the percentage of benefits households increases steadily from 32.5% in single person households before levelling at between 62% and 65% in 8+ person households.

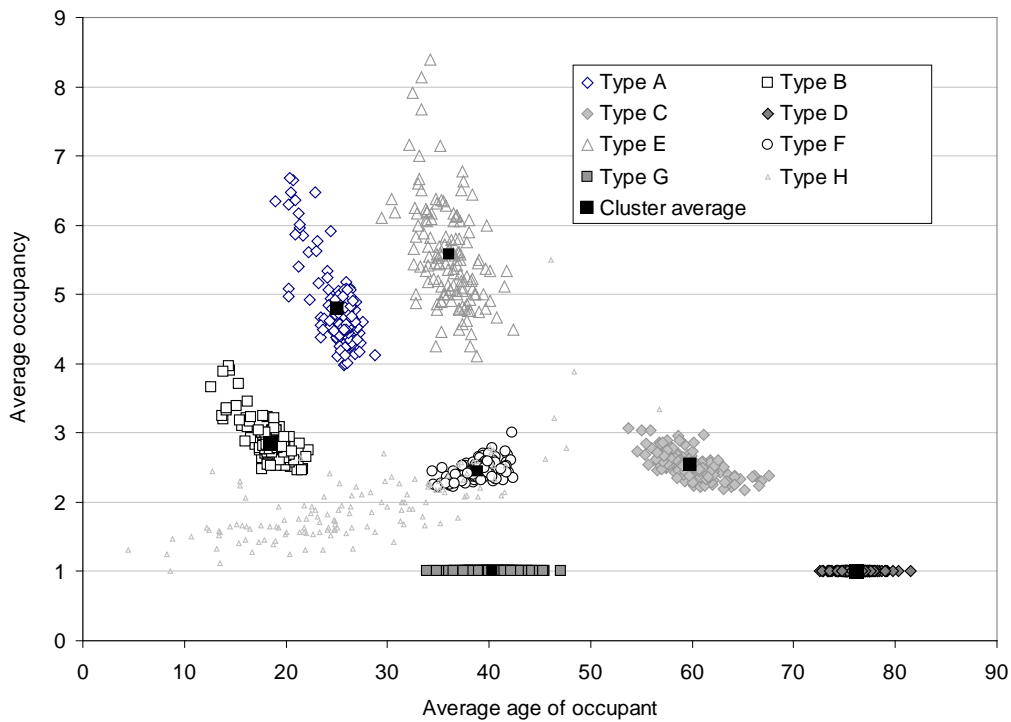
Within this broad analysis of household types we also find that adult men and women are more likely to live in same sex than mixed households. Of households with at least one adult man or one woman present, 30% are male only and 35% female only and 35% are mixed. However a much lower proportion are same sex couple-only households. Of all households, two-person male only adult households account for 3.3% of the total, and two-person female households for 4.4%.

### **4.2 Age and occupancy by standard household type**

Although we find that statistical distributions are not a short cut to estimating household types, they can serve a purpose as in the illustration given above. Households can also be predicted from aggregate data given more limited information such as the size of housing stock and it is possible for example to construct regression models which can do this job, although accuracy tends to be much better in some household categories than others.

A key advantage of our approach is that we have full demographic data on each household and so we can analyse households with considerable confidence and clarity as well as link other data (e.g. tenure, and benefit status). Characteristic differences between the eight standard types in terms of age and occupancy are observable from Figure 5. This is a scattergram showing average occupancy versus average age per household at output area level. Each household type forms a characteristic cluster in these two dimensions and so form clear categories; only Type H which is a residual category does not show any clustering tendency.

Figure 5: Household types by average occupancy and age at output area level



In more detail:

- (a) Type A family households typically contain four or five persons, including children, with an average age of 25 years and occupancy of 4.8 persons
- (b) Type B single parent households range in size from two to three persons with an average age of 19 years and occupancy of 2.8 persons
- (c) Type C older cohabiting households range in size between two or three persons with average age of 60 years and occupancy of 2.5 persons
- (d) Type D older single person older households have an average age of 76 years upwards and become the dominant type of older household at the oldest ages
- (e) Type E three generational households range in occupancy levels from four to eight people with an average age of 36 years and occupancy of 5.8 persons
- (f) Type F households are cohabiting adult households with an average age of 39 years and occupancy of 2.4 persons
- (g) Type G single occupancy adult households range have an average age of 40 years

## 4.2 Households with children

On a day to day basis local authorities are interested in households that are more likely to be users of council services. Two demographic groups that rely on the council services to a greater degree are children and older people through the housing, education, and social care and benefits systems. The following is an example based on a decomposition of households with children to show the relationship between the number of children per household, housing tenure and income deprivation.

Table 10 shows the frequency of households by number of children and the percentage that is social housing tenure or in receipt of means tested benefits. Of all occupied residential properties, it shows that only 30% of households have any children living in them. Of these 27.8k have fewer than 4 children and 3.2k has 4 or more. It is noteworthy that the effect of having more children is to increase the probability of being in a low income household or living in social housing.

**Table 10: Number of households based on number of children in household age 0-19 in Hackney**

Number of children in household age 0-19	number of households	% social housing	% on benefits
void	5975	0	0
0	73,684	45.7	33.3
1	13,732	60.2	48.5
2	9,630	63.9	53.5
3	4,420	71.1	61.7
4	1,803	71.2	66.5
5	702	63.8	71.2
6	355	53.8	73.2
7	188	38.8	77.1
8	84	32.1	79.8
9	12	33.3	83.3
10	1	100.0	100.0
>10	41	9.0	24.0
Grand total	110,627	48.3	37.5

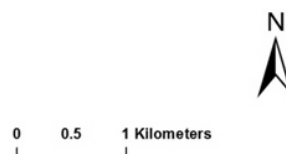
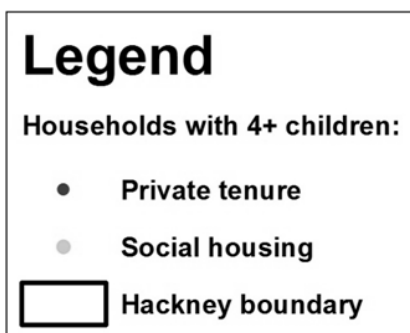
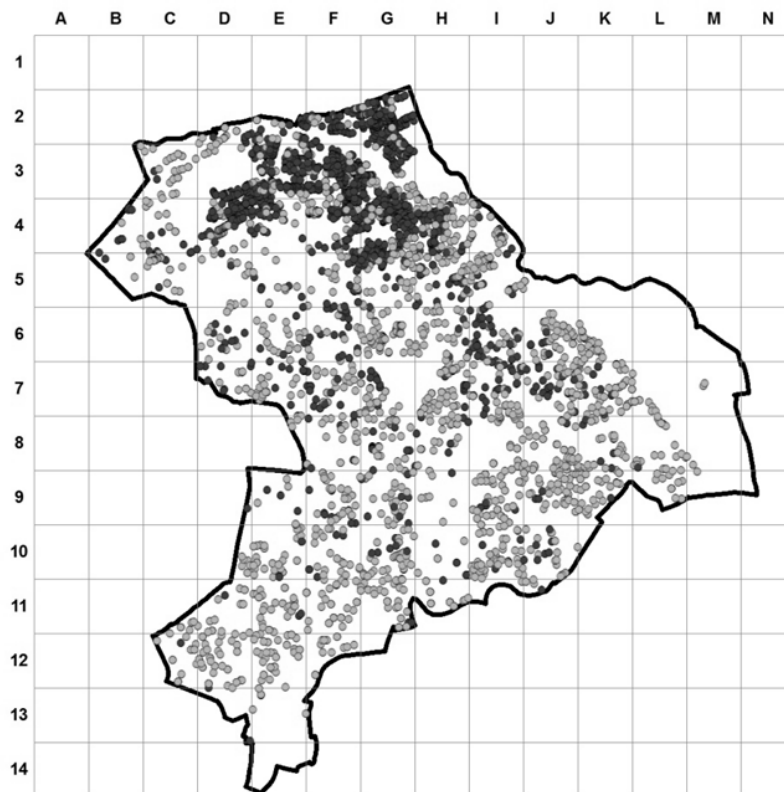
In Hackney, for example, 48.3% of all households are in social tenure. The table shows that the chances of living in social housing increases from 45.7% with no children to 71.2% with four children after which the percentage declines. Table 10 also shows that percentage of households on benefits is 33.3% in a childless household rising to 48.5% in one-child households and steadily increasing to 83% in 9-child households.

Because in our data base households are geo-referenced, we are able to analyse any potential geographical patterns among sub-categories of children households. Figure 6 for example, is a map of the locations of households in Hackney with 4+ children of which there are 3,186 examples in the data. Superimposed on the map is a grid which works like a spreadsheet and can be used to identify particular cells or localities.

## Re-thinking households using administrative data

In this example each household has been colour-coded according to whether tenure is private or social housing. We observe a clear clustering of households with 4+ children above row 5 and between columns D to H. We know from other work that this clustering is related to the concentration of Charedi (Ultra Orthodox Jewish) households which tend to have larger than average numbers of children. The map is hence another demonstration of the granularity of the information and the multiple uses to which it can be put.

**Figure 6: Map of households in Hackney with 4+ children (● =private tenure household,○ =social housing household)**



### **4.3 Minimum data sets for identifying household types including applications**

As has been argued a key advantage of administrative data is that it is possible to link other attributes to household classifications. This could include information about health, education, crime or access to services and public transport. One can look at educational outcomes, access to and uptake of health services, households with unmet or complex needs, household churn and so on and so there is no doubt that a rich tapestry of detailed local information can thus be created.

We have determined that our ability to predict and therefore estimate household structures from aggregate information is limited. An important question therefore is what is the minimum administrative data sets that would be needed to produce household counts of the eight standard types? Such information could be built into policy on welfare and housing especially if it is of sufficient granularity but not overly complex to measure or define.

Where available we would be able to identify single parent households and older households as being not only among the more vulnerable groups but also gaps or similarities with other household types. In essence we thus seek a set of markers of identifiers with which to flag each household in such a way they these can be used in combination to produce the eight standard types.

Furthermore these markers, or risk factors, would be predictive in terms of their ability to identify some attribute or other of those households. For purposes of illustration we choose income deprivation as this attribute and we use benefit status as our proxy for defining it. They are: presence of a child aged 0-19, an adult aged 20+, at least one person aged 65+, and another aged 20-64. Since housing tenure and income deprivation are also co-related we define a fifth risk factor as 'living in social housing'.

Of these factors the first four demographic factors are sufficient to transform demographic data into household counts of the eight standard types. The method is general and does not have to rely on administrative counts of the population and could be produced during the early stages of processing of census data in order to produce compact statistical counts at sub-local authority level. The fifth factor is added since our focus in this case is on income deprivation and inequality among households.

Returning to the problem at hand, the valid combinations of these five risk factors create 22 valid and mutually exclusive combinations each of which can be enumerated and assigned to one of the eight standard household types. Table 12 is based on 104,228 households of up to 10 people and shows the outcome of this analysis. For consistency with previous tables, it excludes 424 households with more than ten people and 5,975 empty properties as can be verified. It can also be noted that if the table is summed by the standard types we would obtain the same breakdown as in Table 8 column 2.

**Table 11: Breakdown of household types by risk factor combination and risk of income deprivation**

Case	household type	frequency	any child age 0-19	single adult age 20+	at least one person age 65+	at least one person aged 20-64	living in social housing	Observed % of households on benefits
1	H	76	Y	Y	Y		Y	88.2
2	D	4,979		Y	Y		Y	80.2
3	E	1,139	Y		Y	Y	Y	72.4
4	C	3,073			Y	Y	Y	71
5	B	6,482	Y	Y		Y	Y	70.8
6	H	14	Y		Y		Y	64.3
7	C	813			Y		Y	63.2
8	A	11,170	Y			Y	Y	60
9	H	705	Y				Y	52.5
10	E	575	Y		Y	Y		49
11	D	1,488		Y	Y			48.8
12	G	15,700		Y		Y	Y	46.7
13	H	28	Y	Y	Y			46.4
14	B	3,377	Y	Y		Y		45.1
15	F	9,094				Y	Y	43
16	H	10	Y		Y			40
17	C	1,575			Y	Y		35.9
18	A	6,236	Y			Y		33.1
19	C	424			Y			25.5
20	H	1,115	Y					24.8
21	F	10,433				Y		16.1
22	G	25,722		Y		Y		13.9
Total		104,228	30,927	57,852	14,194	94,576	53,245	39.6

The rows in the table are arranged such that the most probable sub-group in terms of income deprivation is row 1. Of the 76 households in this category 88.2% are households on benefits whereas in row 22 of the 25,722 Type G households only 13.9% are on benefits. This compares with 39.6% of households for Hackney as a whole. Hence this is a wide spectrum of households that fall into greater or lesser income deprivation if benefits status is used for measurement purposes.

The column totals represent the number of households to which the given risk factor applies so for example 30,927 households have at least one child living in them and 53,245 households are social tenure. The estimated value of each risk factor for predicting income deprivation can be obtained using logistic regression techniques. We set up the following model in which we regress the logit of living in a household on benefits against each of the grouped risk factors as follows.

$$\hat{L}_i = \ln \left[ \frac{\hat{r}_i}{1 - \hat{r}_i} \right] = \beta_0 + \sum_{m=1}^{m=M} \beta_m x_{mi} + u_i \quad (5)$$

## Re-thinking households using administrative data

where  $i$  is the  $i^{th}$  factor combination,  $i=1,22$  and  $M$  is the number factors in the model,  $M=5$ , and  $L_i$  is the sample log-odds in category  $i$  of a reported DV occurrence.

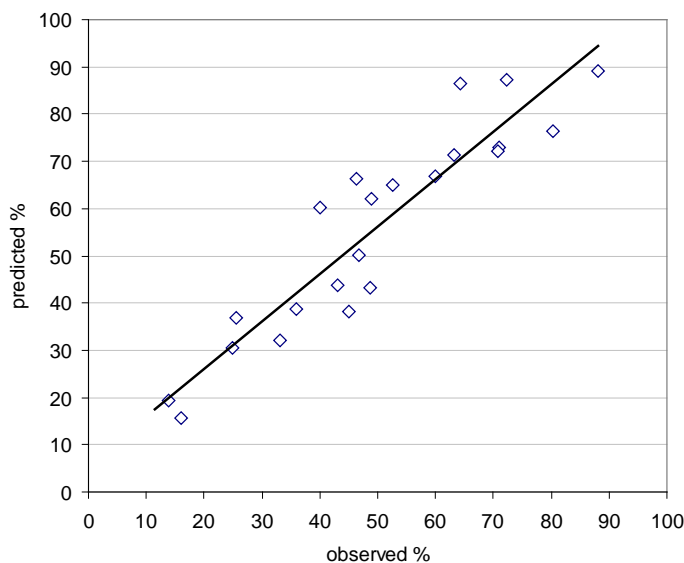
This results of the regression given in Table 13. They show that the odds of living in a household on benefits increase 4.3 times if living in social housing, 3.4 times if a person resident is age 65 or over, 2.6 times if there are children living in the household and 1.3 times if there is a single adult aged 20 or over. If there is an adult age 20 to 64 the odds are reduced by 0.9 which is related to the fact that this is peak working age. All the estimates are significantly different from zero at the 95% level of probability.

**Table 13: odds of income deprivation by risk factor including 95% confidence intervals**

Risk factor	parameter estimate	odds ratio	lower CI	upper CI
Any child 0-19	0.95	2.6	2.5	2.7
Single adult 20+	0.26	1.3	1.26	1.33
At least one person age 65+	1.24	3.4	3.3	3.6
At least one person aged 20-64	-0.08	0.9	0.87	0.98
Living in social housing	1.4	4.2	4.1	4.4
Constant	-1.77			

The odds are multiplicative so that a household with an older person living alone in social housing is  $1.3 \times 3.4 \times 4.2 = 18.6$  times more likely to be in receipt of benefits than a household with none of these risk factors based on this example. Figure 7 shows the predicted percentage of households versus the observed percentage. The fitted equation has a value of  $R^2$  of 0.86 and the slope of the equation is 1.00 indicating a very model fit.

**Figure 7: Expected percentage of households on benefits based on risk factors versus actual percentage by risk group ( $R^2=0.8557$ )**



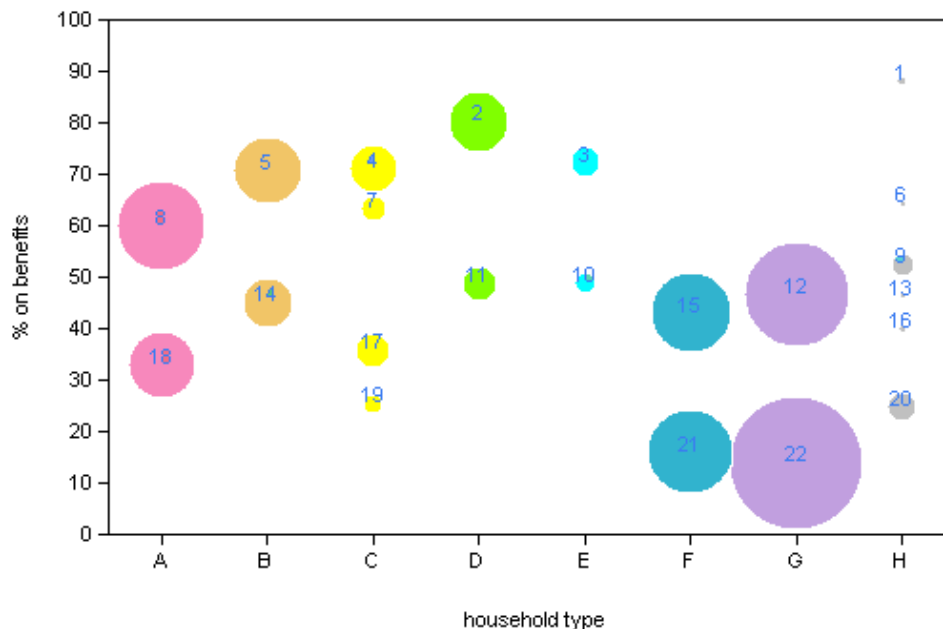


## Re-thinking households using administrative data

The significant correlation with social housing is observable in the bubble chart in Figure 8. On the vertical axis is the percentage of households in receipt of means tested benefits and on the horizontal axis is the standard household type. Each numbered colour coded bubble corresponds with a row in Table 12 in which the size of the bubbles is proportional to the number of households in that sub-category.

For any given household type the higher bubble corresponds to households in social tenure and the lower bubble in private tenure so the correlation between social housing and income deprivation is quite stark in this borough. Although not shown here these results can be mapped in various ways to show more and less affluent neighbourhoods, income poverty among the different standard types, or income poverty in household with or without children etc.

**Figure 8: The relationship between the percentage of households on benefits and household type by risk category**



### 5. Implications and critique

This paper has identified a significant gap in the availability and quality of household statistics at local level. Despite the diversity in sources of information available, this information is only available in the form of national surveys or data down to district level. For methodological reasons and lags in reporting, we have shown that figures have a tendency to be distorted especially in areas of high population growth and churn.

The various sources of information on households, surveys etc, inform one another to a degree for classificatory or imputation processes but they do not constitute an integrated system and therefore a proper basis for household accounts that can be used at local level or for other than a narrow range of purposes. This limits their usefulness across a broader set of policy domains of which several examples were given.

## *Re-thinking households using administrative data*

On the other hand DCLG district level household projections fulfil a purpose in that they provide consistent national and regional projections (Department for Communities and Local Government, 2008; March 2010b). Revisions to the DCLG methodology and outputs have also resulted in improvements including projections for the number of households with dependent children, based on revised ONS population and migration estimates (Department for Communities and Local Government, 2008; March 2010a; March 2010b; August 2010; November 2010).

They are extensively used by government departments, the National Housing and Planning Advice Unit, regional planning bodies and local authorities (Department for Communities and Local Government, March 2010b), and in work done on future housing need by among others the Cambridge Centre for Housing and Planning Research for Shelter, the Town and Country Planning Association, and the Joseph Rowntree Foundation (Holmans, 2012).

However DCLG projections are clearly not ideal, as demonstrated by the GLA's use of alternative population evidence for London for their household projections, from the higher administrative data counts in Table 2, and from general literature on dissatisfaction with ONS mid-year population estimates (House of Commons Treasury Committee, 2008). In addition, the statistics are in reality forecast counts and not actual counts of households, whilst definitional types are fixed and inflexible.

Surveys by contrast fulfil a need at national level and are widely exploited in social research and frequently form the evidence bases in policy domains such as welfare policy or child care. However, again they are not something that can easily be used for other purposes especially by local users since they are not geographically disaggregate or timely enough to be representative, and nor can they be easily linked to other data.

The gaps that this paper exposes are, to a limited extent, being filled by one method or another as evidenced by the commercial systems that are available such as Mosaic. Unlike DCLG data, these are available at a sub-district level and are therefore an improvement. However, because they also rely to a large degree on government census data their accuracy is also questionable and are lacking in detail as regards demography and occupancy.

One important catalyst for change is the Localism Act (2011) which gives local authorities more freedoms including responsibility for their own Local Development Framework. This often means finding evidence beyond or even contrary to the DCLG figures to be more locally representative (e.g. Leeds City Council, 2011).

In addition there are new demands required for statistics for evidence required under the Localism Act and the Health and Social Care Act (2012). The latter will bring into being Health and Wellbeing Boards. These will be a forum for local commissioners across the NHS, public health and social care, elected representatives, and others with responsibility to draw up Health and Well-Being strategies and plans.

As we have demonstrated, by using administrative data we are able to provide a more complete current enumeration of households, flexible definitions, and linkage to other important household level variables from the bottom-up. This is likely to be more

## *Re-thinking households using administrative data*

locally relevant and therefore better able to support these new duties on local authorities in areas ranging from health care provision to social and leisure services.

Our data have been used in a number of studies in which accuracy timeliness and specificity are important. This include work on complex families, overcrowding, health assessments, access to and uptake of services including unmet need, educational outcomes, and a range of health studies including costs of secondary care by household type, access to child care and so on.

In this paper we have not attempted to create future household projections from administrative data, but this may be considered. A basic requirement is to recreate the data at intervals and to identify trends within the population and household types. For example, work has already been conducted on analysing population change and household formation in the six Olympic boroughs, with encouraging results.

### **6. Conclusions**

Household level statistics about the size, type and demography of households are seen as increasingly important in a range of applications. As well as forming the backbone of evidence for assessing future housing need, household level data is useful for designing services, predicting consumption patterns and generally informing policy in areas such as welfare benefits, childcare and taxation.

The premise of this research is that current sources of information used in each of these areas are inflexible, fragmented and piecemeal. A key consequence of this is that the availability of information especially at local level is limited. This is impacting on local authorities' ability to conduct its business efficiently and effectively in terms of policy development and service delivery.

With new duties placed on local authorities, for example as a result of the Localism and Health and Social Care Acts, this is likely to highlight further the present unsatisfactory position and lead to demands for better data. At present it seems unlikely that this gap can be filled by either the ONS or other government departments because of a lack of resources and other priorities.

This is unfortunate as is it is local authorities that are most likely to benefit from better household data. On the other hand this paper has shown that the considerable administrative data sets available to local authorities could be put to much better use. It has described how these assets can be used first to provide detailed population estimates and secondly how these estimates can be used to convert population into households.

To this end a new methodology was described for doing so and a fully worked case study was presented based on data from one London borough, Hackney, with a population of almost 238k and nearly 111k households (including community establishments). The same methodology has been applied to other local authorities over a period of 10 years including all six of the London Olympic Boroughs of which Hackney is one.

## *Re-thinking households using administrative data*

Our results showed that our methodology could replicate the existing DCLG typology accurately on a proportionate basis. The main issue for DCLG to address is that the population data on which their household methodology is based undercounts the number of households in the six London boroughs as compared with our figures and the Greater London Authority's.

We find that this weakness is due primarily to the ONS demographic estimates which are part of DCLG's household methodology. This is apparent not only from direct comparisons with our own estimates and GLA's but also from the considerably higher empty property rates when household counts are compared with the number of dwellings.

However, even if DCLG's estimates had been more accurate and closer to ours there would still be problems with the lack of flexibility over definition and geography and in the capacity to link with other important household variables. In contrast, our method of creating household counts and types is able to offer flexibility in these aspects, and draws upon actual current population counts rather than estimated trends, thereby reflecting a more accurate population.

The approach adopted demonstrates that individual households can form the building block of any geographical area of interest and be linked to any other household level variable, and so this also bypasses the need for expensive surveys to a degree. Surveys seeking opinions or personal habits such as smoking or drinking behaviours are a different matter of course. Even in these cases the results can be linked to administrative data as appropriate although by definition it would only include a sample of all households.

At this point the methodology cannot currently verify the marital relationships between residents of a household, and nor has this aspect been evaluated in terms of accuracy or consistency against national household figures. This would require us to link records on marriage and divorce which, even if available electronically, would not contain the current address or personal identifiers to enable this and would pose formidable legal and practical challenges.

It is debateable whether this gap is an important omission or not when compared against the other advantages that are available, especially an ability, demonstrated in the Olympic Boroughs study, to link other local data sources including health and social services data, leisure services and many other sources to provide a single statistical platform for use locally.

To conclude, this paper proposes a different basis for collecting and maintaining household statistics using local available information sources. The key advantages are that data are at household level, can be produced on a timely basis and can be linked to other data particularly information about services supplied. This does however require knowledge of local data sources and how to access and exploit them. Finally, use of the data forming the basis for this paper was approved for use by the local PCTs (Primary Care Trusts) and underpinned by a legally enforceable data sharing protocol.

## References

- 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, 68-83.
- Bronte-Tinkew, J. and 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.
- Cabinet Office (2008) Helping to shape tomorrow: The 2011 census of population and housing in England and Wales. HM Stationery Office. (Available from <http://www.official-documents.gov.uk/document/cm75/7513/7513.pdf>).
- Department for Communities and Local Government (December 2008) Options for the future of the household projection model. Communities and Local Government, London. (Available from <http://www.communities.gov.uk/documents/housing/pdf/housingprojectionmodel.pdf>)
- Department for Communities and Local Government (March 2010a) Testing methodological changes to the household projections model. Communities and Local Government, London. (Available from <http://www.communities.gov.uk/documents/housing/pdf/1487127>).
- Department for Communities and Local Government (March 2010b) Consultation on proposed changes to the national statistics on household projections. Communities and Local Government, London. (Available from <http://www.communities.gov.uk/documents/housing/pdf/1487114.pdf>).
- Department for Communities and Local Government (August 2010) Proposed changes to the National Statistics publication on household projections: summary of responses to consultation. Communities and Local Government, London. (Available from <http://www.communities.gov.uk/documents/housing/pdf/1665808.pdf>).
- Department for Communities and Local Government (November 2010) Updating the Department for Communities and Local Government's household projections to a 2008 base – methodology. Communities and Local Government, London. (Available from <http://www.communities.gov.uk/documents/statistics/pdf/1780350>).
- 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.
- Grundy, E., Stuchbury, R. and Young, H. (Spring 2010) Households and families: Implications of changing census definitions for analyses using the ONS Longitudinal Study' *Population Trends*, 139, 64-69.
- Harper, G. and Mayhew, L. (2011a) Using administrative data to count local populations. *Applied Spatial Analysis and Policy*, April, 1-26.

## *Re-thinking households using administrative data*

Harper, G. and Mayhew, L. (2011b) Applications of population counts based on administrative data at local level. *Applied Spatial Analysis and Policy*, March, 1-27.

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

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

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 Treasury Committee. (14th May 2008) Counting the population. Eleventh report of session 2007-8, 1, House of Commons London: The Stationery Office Limited.

House, T. and 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. and Weitzman, M. (2009) Family composition and children's exposure to adult smokers in their homes. *Pediatrics*, 123 (4), 559-564.

Larsson, K., Thorslund, M. and 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 (May 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](http://www.leeds.gov.uk/files/Internet2007/2011/27/final%20leeds%20shma%2027-05-11(1).pdf)).

Lewis, P. (1977) *Maps and Statistics*. Methuen, London.

Mayhew, L., Harper, G. and Waples, S. (July 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>).

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/>).

## *Re-thinking households using administrative data*

Ohwaki, K., Hashimoto, H., Sato, M., Tamiya, N. and 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.

Palmer, G., Kenway, P. and Wilcox, S. (2006) Housing and neighbourhoods monitor. Joseph Rowntree Foundation. (Available from <http://www.jrf.org.uk/sites/files/jrf/1800-housing-neighbourhoods-demand.pdf>).

Rogers, A. (1974) *Statistical Analysis of Spatial Dispersion*. Pion: London.

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

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

Varjonen, J. and K. Aalto (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. and Williamson, P. (2001) The diversity of diversity: a critique of geodemographic classification. *Area*, 33(1), 63-76.

Wagner, K.D., Ritt-Olson, A., Soto, D.W. and 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.

Webber, R. and Farr, M. (2000) MOSAIC – from an area classification system to household classification. Cathie Marsh Centre for Census and Survey Research conference: The Census of Population 2000 and Beyond, 22-23 June 2000. (Available from <http://www.ccsr.ac.uk/conference/papers.htm>).

### **Web pages:**

<http://research.dwp.gov.uk/asd/index.php?page=hbai>.

<http://www.poverty.org.uk/summary/households.shtml>

<http://www.esds.ac.uk/government/surveys/>

## FACULTY OF ACTUARIAL SCIENCE AND INSURANCE

### Actuarial Research Papers since 2001

<b>Report Number</b>	<b>Date</b>	<b>Publication Title</b>	<b>Author</b>
160.	January 2005	Mortality Reduction Factors Incorporating Cohort Effects. ISBN 1 90161584 7	Arthur E. Renshaw Steven Haberman
161.	February 2005	The Management of De-Cumulation Risks in a Defined Contribution Environment. ISBN 1 901615 85 5.	Russell J. Gerrard Steven Haberman Elena Vigna
162.	May 2005	The IASB Insurance Project for Life Insurance Contracts: Impart on Reserving Methods and Solvency Requirements. ISBN 1-901615 86 3.	Laura Ballotta Giorgia Esposito Steven Haberman
163.	September 2005	Asymptotic and Numerical Analysis of the Optimal Investment Strategy for an Insurer. ISBN 1-901615-88-X	Paul Emms Steven Haberman
164.	October 2005.	Modelling the Joint Distribution of Competing Risks Survival Times using Copula Functions. ISBN 1-901615-89-8	Vladimir Kaishev Dimitrina S, Dimitrova Steven Haberman
165.	November 2005.	Excess of Loss Reinsurance Under Joint Survival Optimality. ISBN1-901615-90-1	Vladimir K. Kaishev Dimitrina S. Dimitrova
166.	November 2005.	Lee-Carter Goes Risk-Neutral. An Application to the Italian Annuity Market. ISBN 1-901615-91-X	Enrico Biffis Michel Denuit
167.	November 2005	Lee-Carter Mortality Forecasting: Application to the Italian Population. ISBN 1-901615-93-6	Steven Haberman Maria Russolillo
168.	February 2006	The Probationary Period as a Screening Device: Competitive Markets. ISBN 1-901615-95-2	Jaap Spreeuw Martin Karlsson
169.	February 2006	Types of Dependence and Time-dependent Association between Two Lifetimes in Single Parameter Copula Models. ISBN 1-901615-96-0	Jaap Spreeuw
170.	April 2006	Modelling Stochastic Bivariate Mortality ISBN 1-901615-97-9	Elisa Luciano Jaap Spreeuw Elena Vigna.
171.	February 2006	Optimal Strategies for Pricing General Insurance. ISBN 1901615-98-7	Paul Emms Steve Haberman Irene Savoulli
172.	February 2006	Dynamic Pricing of General Insurance in a Competitive Market. ISBN1-901615-99-5	Paul Emms
173.	February 2006	Pricing General Insurance with Constraints. ISBN 1-905752-00-8	Paul Emms
174.	May 2006	Investigating the Market Potential for Customised Long Term Care Insurance Products. ISBN 1-905752-01-6	Martin Karlsson Les Mayhew Ben Rickayzen



<b>Report Number</b>	<b>Date</b>	<b>Publication Title</b>	<b>Author</b>
175.	December 2006	Pricing and Capital Requirements for With Profit Contracts: Modelling Considerations. ISBN 1-905752-04-0	Laura Ballotta
176.	December 2006	Modelling the Fair Value of Annuities Contracts: The Impact of Interest Rate Risk and Mortality Risk. ISBN 1-905752-05-9	Laura Ballotta Giorgia Esposito Steven Haberman
177.	December 2006	Using Queuing Theory to Analyse Completion Times in Accident and Emergency Departments in the Light of the Government 4-hour Target. ISBN 978-1-905752-06-5	Les Mayhew David Smith
178.	April 2007	In Sickness and in Health? Dynamics of Health and Cohabitation in the United Kingdom. ISBN 978-1-905752-07-2	Martin Karlsson Les Mayhew Ben Rickayzen
179.	April 2007	GeD Spline Estimation of Multivariate Archimedean Copulas. ISBN 978-1-905752-08-9	Dimitrina Dimitrova Vladimir Kaishev Spiridon Penev
180.	May 2007	An Analysis of Disability-linked Annuities. ISBN 978-1-905752-09-6	Ben Rickayzen
181.	May 2007	On Simulation-based Approaches to Risk Measurement in Mortality with Specific Reference to Poisson Lee-Carter Modelling. ISBN 978-1-905752-10-2	Arthur Renshaw Steven Haberman
182.	July 2007	High Dimensional Modelling and Simulation with Asymmetric Normal Mixtures. ISBN 978-1-905752-11-9	Andreas Tsanakas Andrew Smith
183.	August 2007	Intertemporal Dynamic Asset Allocation for Defined Contribution Pension Schemes. ISBN 978-1-905752-12-6	David Blake Douglas Wright Yumeng Zhang
184.	October 2007	To split or not to split: Capital allocation with convex risk measures. ISBN 978-1-905752-13-3	Andreas Tsanakas
185.	April 2008	On Some Mixture Distribution and Their Extreme Behaviour. ISBN 978-1-905752-14-0	Vladimir Kaishev Jae Hoon Jho
186.	October 2008	Optimal Funding and Investment Strategies in Defined Contribution Pension Plans under Epstein-Zin Utility. ISBN 978-1-905752-15-7	David Blake Douglas Wright Yumeng Zhang
187.	May 2008	Mortality Risk and the Valuation of Annuities with Guaranteed Minimum Death Benefit Options: Application to the Italian Population. ISBN 978-1-905752-16-4	Steven Haberman Gabiella Piscopo
188.	January 2009	The Market Potential for Privately Financed Long Term Care Products in the UK. ISBN 978-1-905752-19-5	Leslie Mayhew
189.	June 2009	Whither Human Survival and Longevity or the Shape of things to Come. ISBN 978-1-905752-21-8	Leslie Mayhew David Smith
190.	October 2009	ilc: A Collection of R Functions for Fitting a Class of Lee Carter Mortality Models using Iterative fitting Algorithms* ISBN 978-1-905752-22-5	Zoltan Butt Steven Haberman
191.	October 2009	Decomposition of Disease and Disability Life Expectancies in England, 1992-2004. ISBN 978-1-905752-23-2	Domenica Rasulo Leslie Mayhew Ben Rickayzen

<b>Report Number</b>	<b>Date</b>	<b>Publication Title</b>	<b>Author</b>
192.	October 2009	Exploration of a Novel Bootstrap Technique for Estimating the Distribution of Outstanding Claims Reserves in General Insurance. ISBN 978-1-905752-24-9	Robert Cowell
193.	January 2010	Surplus Analysis for Variable Annuities with a GMDB Option. ISBN 978-1-905752-25-6	Steven Haberman Gabriella Piscopo
194.	January 2010	UK State Pension Reform in a Public Choice Framework. ISBN 978-1-905752-26-3	Philip Booth
195.	June 2010	Stochastic processes induced by Dirichlet (B-) splines: modelling multivariate asset price dynamics. ISBN 978-1-905752-28-7	Vladimir Kaishev
196.	August 2010	A comparative Study of Parametric Mortality Projection Models. ISBN 978-1-905752-29-4	Zoltan Butt Steven Haberman
197.	April 2012	Are the Dimensions of Private Information More Multiple than Expected? Information Asymmetries in the Market of Supplementary Private Health Insurance in England* ISBN 978-1-905752-31-7	Martin Karlsson Florian Klohn Ben Rickayzen
198.	May 2012	Re-thinking households – Using administrative data to count and classify households with some applications. ISBN 978-1-905752-32-4	Gillian Harper Les Mayhew

### **Statistical Research Papers**

1.	December 1995.	Some Results on the Derivatives of Matrix Functions. ISBN 1 874 770 83 2	P. Sebastiani
2.	March 1996	Coherent Criteria for Optimal Experimental Design. ISBN 1 874 770 86 7	A.P. Dawid P. Sebastiani
3.	March 1996	Maximum Entropy Sampling and Optimal Bayesian Experimental Design. ISBN 1 874 770 87 5	P. Sebastiani H.P. Wynn
4.	May 1996	A Note on D-optimal Designs for a Logistic Regression Model. ISBN 1 874 770 92 1	P. Sebastiani R. Settini
5.	August 1996	First-order Optimal Designs for Non Linear Models. ISBN 1 874 770 95 6	P. Sebastiani R. Settini
6.	September 1996	A Business Process Approach to Maintenance: Measurement, Decision and Control. ISBN 1 874 770 96 4	Martin J. Newby
7.	September 1996.	Moments and Generating Functions for the Absorption Distribution and its Negative Binomial Analogue. ISBN 1 874 770 97 2	Martin J. Newby
8.	November 1996.	Mixture Reduction via Predictive Scores. ISBN 1 874 770 98 0	Robert G. Cowell.
9.	March 1997.	Robust Parameter Learning in Bayesian Networks with Missing Data. ISBN 1 901615 00 6	P. Sebastiani M. Ramoni
10.	March 1997.	Guidelines for Corrective Replacement Based on Low Stochastic Structure Assumptions. ISBN 1 901615 01 4.	M.J. Newby F.P.A. Coolen
11.	March 1997	Approximations for the Absorption Distribution and its Negative Binomial Analogue. ISBN 1 901615 02 2	Martin J. Newby

<b>Report Number</b>	<b>Date</b>	<b>Publication Title</b>	<b>Author</b>
12.	June 1997	The Use of Exogenous Knowledge to Learn Bayesian Networks from Incomplete Databases. ISBN 1 901615 10 3	M. Ramoni P. Sebastiani
13.	June 1997	Learning Bayesian Networks from Incomplete Databases. ISBN 1 901615 11 1	M. Ramoni P. Sebastiani
14.	June 1997	Risk Based Optimal Designs. ISBN 1 901615 13 8	P. Sebastiani H.P. Wynn
15.	June 1997.	Sampling without Replacement in Junction Trees. ISBN 1 901615 14 6	Robert G. Cowell
16.	July 1997	Optimal Overhaul Intervals with Imperfect Inspection and Repair. ISBN 1 901615 15 4	Richard A. Dagg Martin J. Newby
17.	October 1997	Bayesian Experimental Design and Shannon Information. ISBN 1 901615 17 0	P. Sebastiani. H.P. Wynn
18.	November 1997.	A Characterisation of Phase Type Distributions. ISBN 1 901615 18 9	Linda C. Wolstenholme
19.	December 1997	A Comparison of Models for Probability of Detection (POD) Curves. ISBN 1 901615 21 9	Wolstenholme L.C
20.	February 1999.	Parameter Learning from Incomplete Data Using Maximum Entropy I: Principles. ISBN 1 901615 37 5	Robert G. Cowell
21.	November 1999	Parameter Learning from Incomplete Data Using Maximum Entropy II: Application to Bayesian Networks. ISBN 1 901615 40 5	Robert G. Cowell
22.	March 2001	FINEX : Forensic Identification by Network Expert Systems. ISBN 1 901615 60X	Robert G. Cowell
23.	March 2001.	Wren Learning Bayesian Networks from Data, using Conditional Independence Tests is Equivalent to a Scoring Metric ISBN 1 901615 61 8	Robert G Cowell
24.	August 2004	Automatic, Computer Aided Geometric Design of Free-Knot, Regression Splines. ISBN 1-901615-81-2	Vladimir K Kaishev, Dimitrina S. Dimitrova, Steven Haberman Richard J. Verrall
25.	December 2004	Identification and Separation of DNA Mixtures Using Peak Area Information. ISBN 1-901615-82-0	R.G. Cowell S.L. Lauritzen J Mortera,
26.	November 2005.	The Quest for a Donor : Probability Based Methods Offer Help. ISBN 1-90161592-8	P.F. Mostad T. Egeland., R.G. Cowell V. Bosnes Ø. Braaten
27.	February 2006	Identification and Separation of DNA Mixtures Using Peak Area Information. (Updated Version of Research Report Number 25). ISBN 1-901615-94-4	R.G. Cowell S.L. Lauritzen J Mortera,

<b>Report Number</b>	<b>Date</b>	<b>Publication Title</b>	<b>Author</b>
28.	October 2006	Geometrically Designed, Variable Knot Regression Splines : Asymptotics and Inference. ISBN 1-905752-02-4	Vladimir K Kaishev Dimitrina S.Dimitrova Steven Haberman Richard J. Verrall
29.	October 2006	Geometrically Designed, Variable Knot Regression Splines : Variation Diminishing Optimality of Knots. ISBN 1-905752-03-2	Vladimir K Kaishev Dimitrina S.Dimitrova Steven Haberman Richard J. Verrall
30.	November 2008	Scheduling Reentrant Jobs on Parallel Machines with a Remote Server. ISBN 978-1-905752-18-8	Konstantin Chakhlevitch Celia Glass
31.	November 2009	Probabilistic Expert Systems for Handling Artifacts in Complex DNA Mixtures. ISBN 978-1-905752-27-0	R.G.Cowell S.L. lauritzen J. Mortera
32.	April 2012	Detecting a Family Trio in a DNA Mixture. ISBN 978-1-905752-30-0	Robert G. Cowell

***Papers can be downloaded from***

***<http://www.cass.city.ac.uk/facact/research/publications.html>***

# **Faculty of Actuarial Science and Insurance**

## **Cass Business School**

---

Copyright 2012 © Faculty of Actuarial Science and Insurance,  
Cass Business School  
106 Bunhill Row,  
London EC1Y 8TZ.

ISBN 978-1-905752-32-4