Making Use of Local Administrative Data For Population Estimates and Service Planning

UPTAP Leeds March 2009

Les Mayhew

(lesmayhew@googlemail.com)

Gillian Harper

(harpergill@googlemail.com)

Mayhew Harper Associates Ltd.

Outline

- Limitations of traditional population statistics
- Local needs and challenges
- Administrative data as an alternative
- Methodology
- Application in service planning and delivery
- Risk ladder theory
- Overview

FAQs

- What is the population of my community, council or PCT area?
- What is the IMD for this housing estate?
- How many single parents live in social housing and are on benefits?
- How many nurseries are there within pram pushing distance of households with young children?
- Are services accessible to those that need them and how much unmet demand is out there?
- Who needs to have face to face contact and where should face-to-face caller centres be located?
- Are there special groups that need more personalised services and how many are there (e.g. older people, single parent households, ethnic groups)?

Limitations of Official Population Statistics

- Decennial Census
- Disseminated 24 months later
- Output Area is smallest unit
- Units are inflexible and/or inappropriate
- Data aggregation
- Pre-determined cross-referencing
- False correlation
- 2001 address and response problems
- Not particularly good at identifying special groups and therefore at answering complex questions

Local Needs and Challenges

- Rapidly changing populations
- Better information on migration
- Under-counting reduces monetary allocations
- Resources may be misallocated
- Spatial diversity
- Customer segmentation
- Need small area level evidence base

Political context

- Treasury Sub-Committee 2008 recognised weaknesses of current Census
- Current MYE are not fit for purpose

"National policies need to be informed by good quality local statistics"

An Alternative – Administrative Data

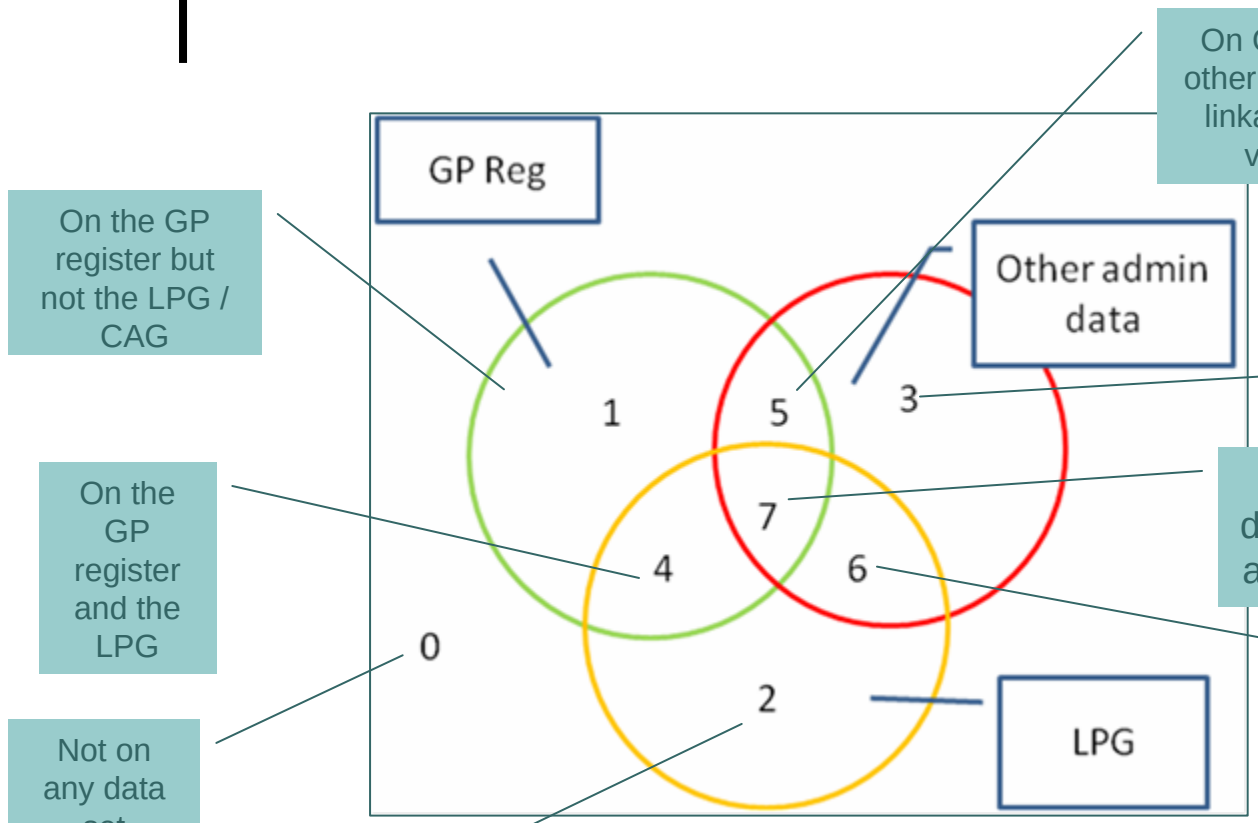
- From an existing data linking technique
- Routinely collected administrative data
- Household or individual level
- Flexible boundaries
- Up-to-date and repeatable

- GP Register
- Council Tax Register
- Electoral Register
- Benefits Register
- School Census
- Births and Deaths

Methodology

- Data records linked together by address after standardisation to a property gazetteer
- GP Register is base
- All records for each address cross-referenced and assessed for who is current
- Sequential logical assumptions used to include or exclude people

Principles



On GP register and other data sets but no linkable address or valid address

On the GP register but not the LPG / CAG

On other data bases but no valid address

On 1+ data sets and LPG

On the GP register and the LPG

Not on GP register but on 1+ data sets and the LPG/CAG

Not on any data set

Vacant addresses

Concept of a ‘confirmed minimum population’

Principles ~ concept of a truth table

ABC	number of people	decision A/R	confirmed	unconfirmed	comments
000	0	R	0	0	empty set
001	50	R	0	50	empty property
010	30	R	0	30	no valid address
011	200	A	200	0	confirmed
100	10	R	0	10	no valid address
101	80	A	80	0	confirmed
110	70	R	0	70	no valid address
111	100	A	100	0	confirmed
total	540		380	160	

A	B	C	|	(A | B)			&	C	
0	0	0	|	0	0	0	**0**	0	R
0	0	1	|	0	0	0	**0**	1	R
0	1	0	|	0	1	1	**0**	0	R
0	1	1	|	0	1	1	**1**	1	A
1	0	0	|	1	1	0	**0**	0	R
1	0	1	|	1	1	0	**1**	1	A
1	1	0	|	1	1	1	**0**	0	R
1	1	1	|	1	1	1	**1**	1	A

A - accept

R - reject

A	assigned a UPRN (living at recognised address)
B	on the GP register
C	on any other data base by surname and UPRN

Algorithm for estimating population from administrative data

- Actual algorithm is based on 18 different variables and at least 7 data sets
- Process is divided into 4 stages with stage 1 having 4 sub-stages
- Each stage and sub-stage generates ‘truth tables’ which are used to build up the population
- Symbolic logic is used to define each stage so that whole process can be represented in compact mathematical form
- First two stages involve the GP register and last two stages other data sets
- Final output is a set of records containing ‘confirmed’ population, geo-coordinates and demographic characteristics to which other data may be appended

Truth table for stage 1b

This example is based on the truth table for stage 1c in which *r* and *p* are data sets and where *a* and *b* are filters

r p a b	r&p&(a|(~a&b))
0 0 0 0	0 0 0 0 0 0 1 0 0 0
0 0 0 1	0 0 0 0 0 1 1 0 1 1
0 0 1 0	0 0 0 0 1 1 0 1 0 0
0 0 1 1	0 0 0 0 1 1 0 1 0 1
0 1 0 0	0 0 1 0 0 0 1 0 0 0
0 1 0 1	0 0 1 0 0 1 1 0 1 1
0 1 1 0	0 0 1 0 1 1 0 1 0 0
0 1 1 1	0 0 1 0 1 1 0 1 0 1
1 0 0 0	1 0 0 0 0 0 1 0 0 0
1 0 0 1	1 0 0 0 0 1 1 0 1 1
1 0 1 0	1 0 0 0 1 1 0 1 0 0
1 0 1 1	1 0 0 0 1 1 0 1 0 1
1 1 0 0	1 1 1 0 0 0 1 0 0 0
1 1 0 1	1 1 1 1 0 1 1 0 1 1
1 1 1 0	1 1 1 1 1 1 0 1 0 0
1 1 1 1	1 1 1 1 1 1 0 1 0 1

Stage 1b: person has a UPRN

& is on GP register & is most recent registered at UPRN

or is related to most recent registered at UPRN

$$r \wedge p \wedge (a \vee (\neg a \wedge b))$$

Residuals (unconfirmed cases)

Confirmed cases

Stages

- Decide snapshot and data time windows
- Clean and geo-reference all data sets
- GP register as base
- Start process of confirming people at each address according to rules of algorithm
- Each category has a set of rules and weights
- Add births and remove deaths
- Assess high occupancy and vacancy rates

Validation

- Reasonability checks
- Comparable estimates and trends
- ‘Ground truthing’
- Look at relevant national statistics (e.g. child benefit counts)
- Take more than one snapshot

Requirements

- Requires understanding of the scope of each dataset
- Originally collected for different purposes
- Requires partnership work and data sharing
- Creates a ‘minimum confirmed population’
- Each person has an age and gender and is geo-referenced

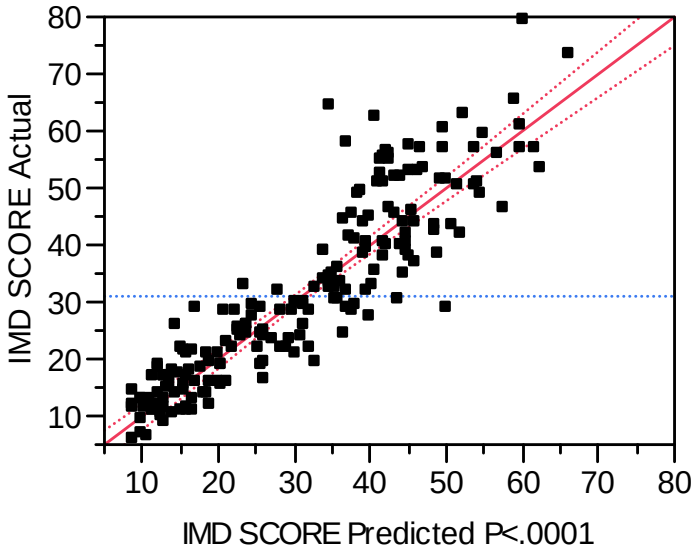
Service Planning and Delivery

- Population is linked to a wealth of socio-economic and health information from source datasets by address
- Segment the population and profile any user-defined area or subject
- Identify gaps in need and small populations at risk
- Impossible with aggregated official statistics

Service Planning and Delivery 1 – IMD in Non-standard Areas

- IMD only available down to SOA level
- LAs need to know levels of deprivation to any geography (buffer areas, high streets, split geographies)
- *nkm* allows users to estimate a consistent IMD to any area or shape
- Method works by modelling the association between IMD at SOA level and *nkm* variables at a household level

Service Planning and Delivery 1 – IMD in Non-standard Areas



This model is based on 8 variables derived from combinations of 3 risk factors:

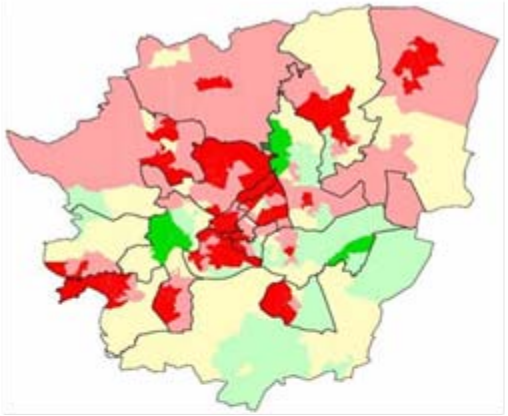
•A –households with at least 1 person 75+ *or* a single parent *or* 3+ children under 20

•B –households that are in council tax band A (i.e. lowest value housing)

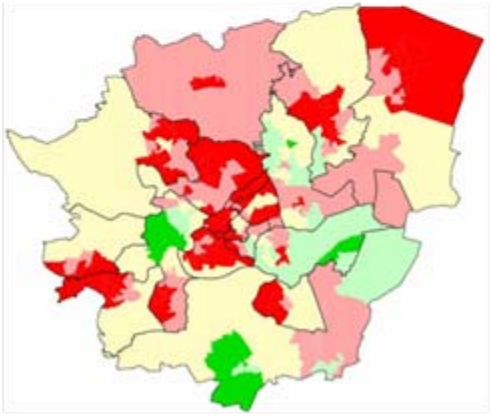
•C –housing rented from the local authority (i.e. Council housing)

Service Planning and Delivery 1 – IMD in Non-standard Areas

- use linear regression to fit household variables to IMD
- then able to measure deprivation at any geographical unit
- reveals pockets of deprivation previously disguised within SOAs
- to track the effect of local initiatives over short timescales



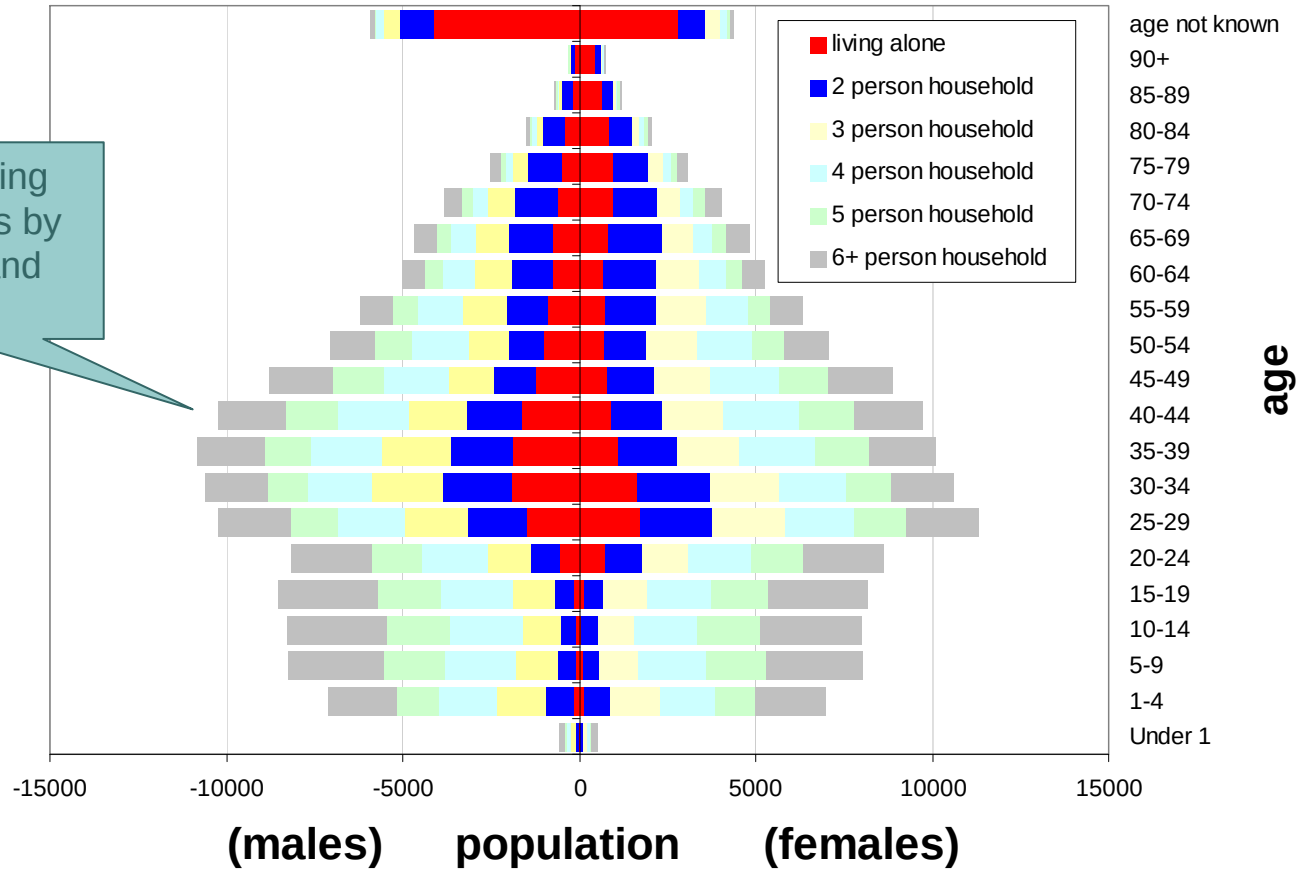
IMD 2007



Modelled IMD

Service Planning and Delivery 2 – Partitioning Populations

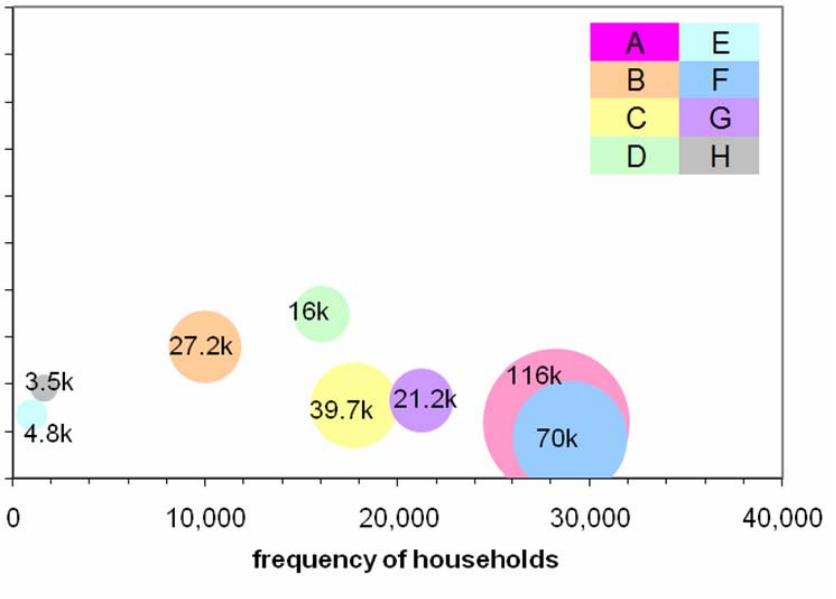
Indicative living arrangements by age group and gender



Service Planning and Delivery 3 – Household Classification

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

First tier household classification based on household demography and 81 sub-types

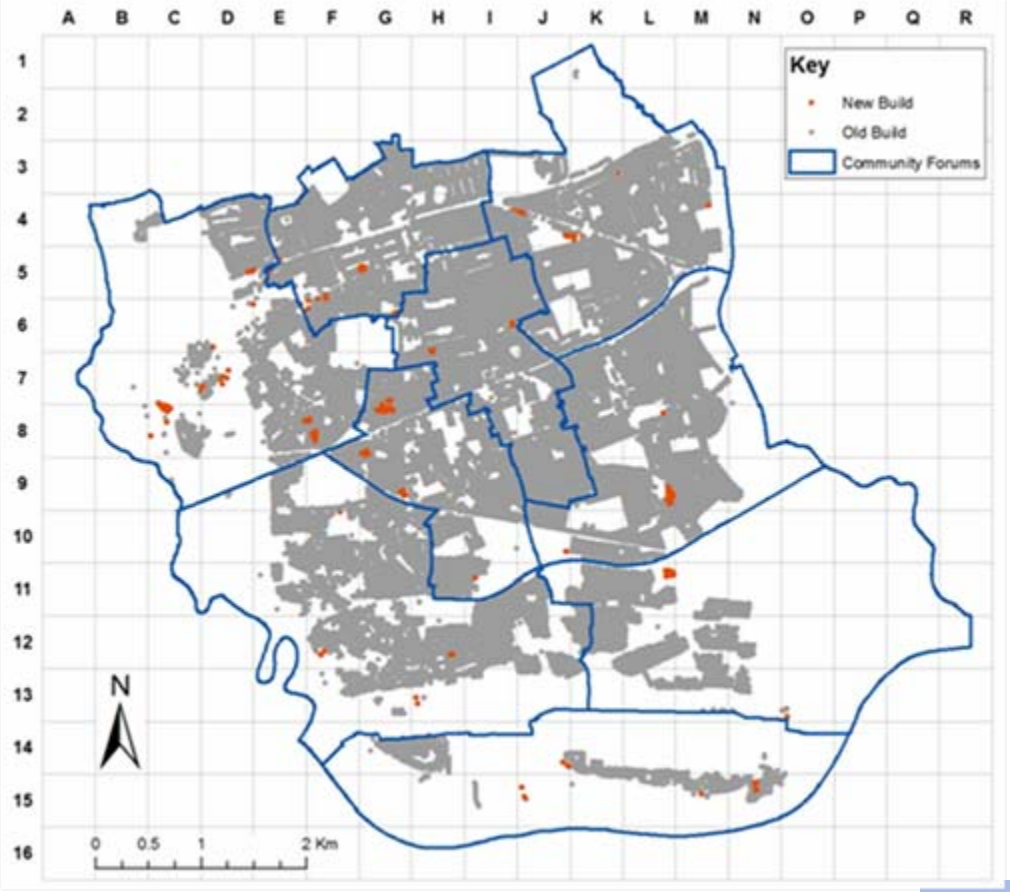


- The percentage of households living in local authority housing by household type and frequency and population size

- Type D households, older people living alone, have the highest percentage in local authority housing, at 34.8%.

Service Planning and Delivery 4 - Regeneration

- Are 'new builds' being occupied by the better off?
- What are the characteristics of new occupants?



Service Planning and Delivery 4 - Regeneration

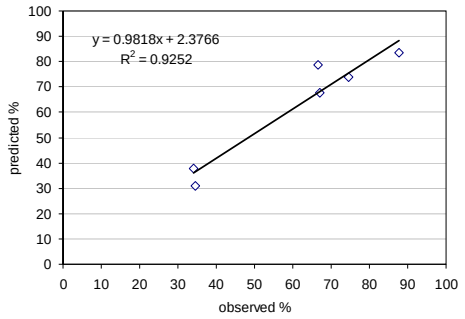
- Segmenting occupants of new and old builds:

row	Frequency of UPRNs	new build	adults no children	private tenure	% not on benefits	lower CI %	upper CI %
1	1896	Y	Y	Y	87.7	86.1	89.2
2	37994		Y	Y	74.6	74.1	75.0
3	38815			Y	67.1	66.6	67.6
4	666	Y		Y	66.5	62.8	70.1
5	8112				34.7	33.6	35.7
6	8864		Y		34.2	33.2	35.2
total	96347	2562	48754	79371	64.7	64.4	65.0

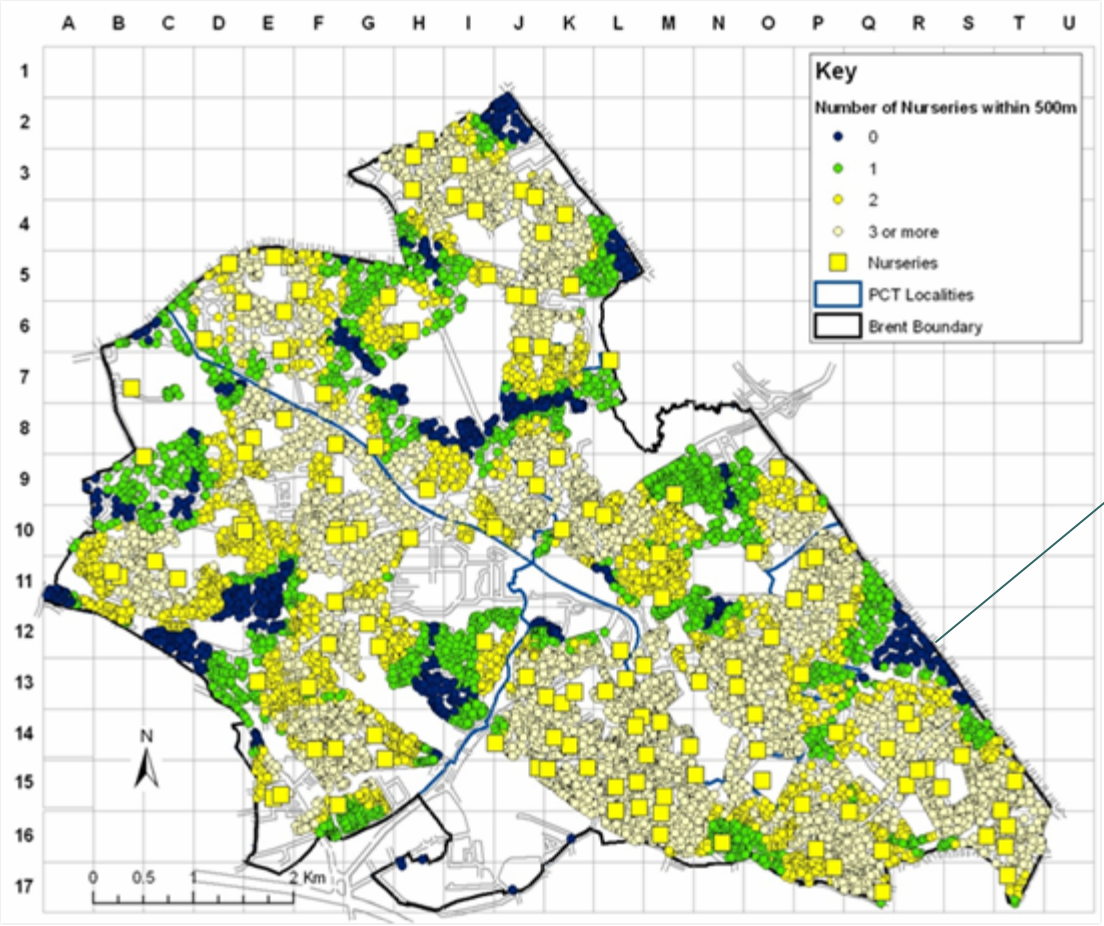
Risk ladder

The odds of ***not*** being on benefits increase:

- 1.77 times if the household is in a new build
- 1.36 times if there are adults and no children
- 4.7 times if the new build is privately owned



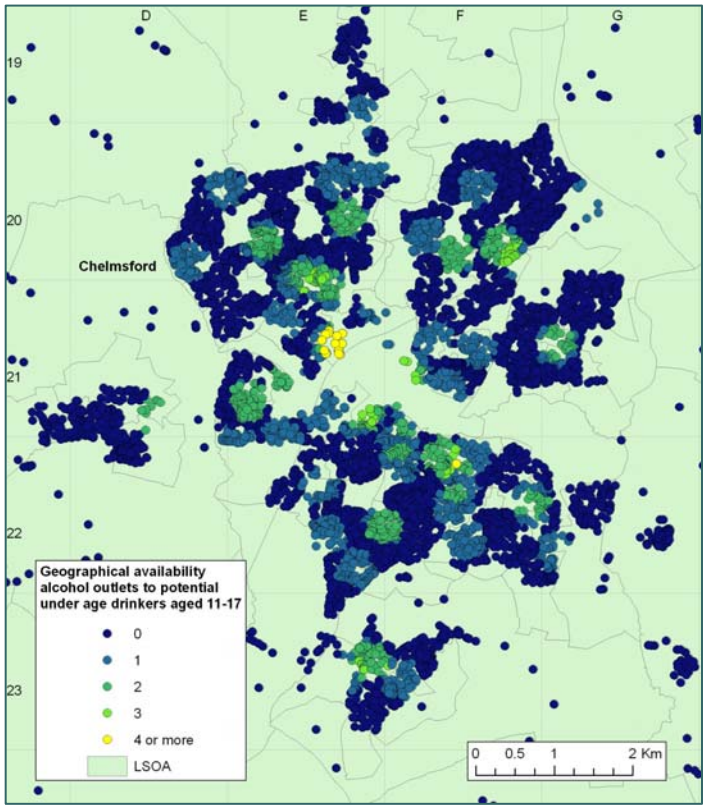
Service Planning and Delivery
5 – Access to Services



How many nurseries are there within pram pushing distance of where children live?

There no nurseries in 'pram pushing' distance in areas shaded in black

6- Public health ~ underage drinking

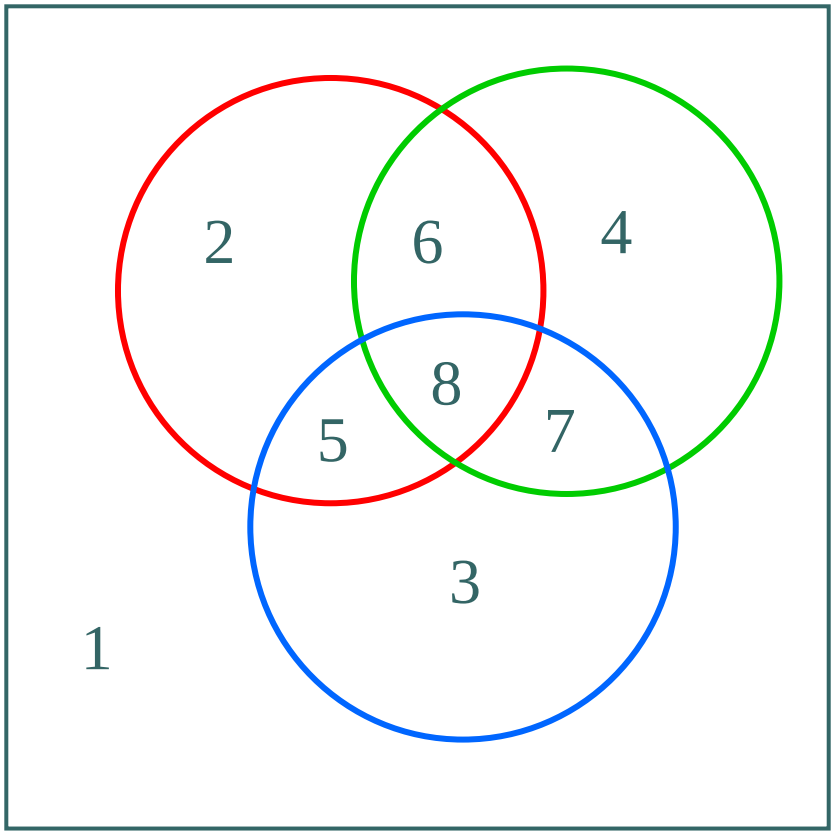


category	frequency	1+ children aged 11-17 at address	single adult household	social housing	% of households within 250m of an alcohol outlet	lower CI%	upper CI%
1	7,266		Y	Y	52.3	51.2	53.5
2	5,815			Y	44.9	43.6	46.2
3	1,989	Y		Y	44.6	42.4	46.9
4	929	Y	Y	Y	43.9	40.7	47.2
5	31,242		Y		40.8	40.3	41.4
6	3,151	Y	Y		36.6	34.9	38.3
7	69,772				33.7	33.3	34.0
8	17,860	Y			30.3	29.6	31.0
total	138,024	23929	42588	15999	36.6	36.4	36.9

Risk ladder

Odds of living near an alcohol outlet increase 1.3 times if living in social housing and 1.6 times if a single adult household

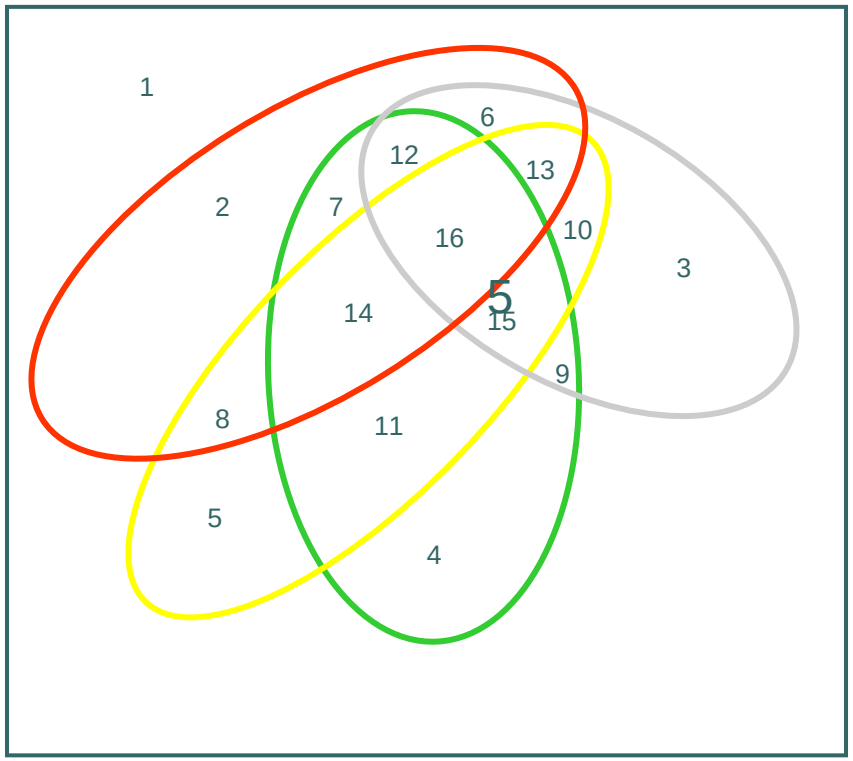
Diagrammatic interpretation of risk ladders using Venn diagrams – 3 dimensions



	RBG
1	000
2	100
3	010
4	001
5	110
6	101
7	011
8	111

Risk factor combinations in binary barcode form

4-dimensional Venn diagram showing factor combinations



	RBGY
1	0000
2	1000
3	0100
4	0010
5	0001
6	1100
7	1010
8	1001
9	0110
10	0101
11	0011
12	1110
13	1101
14	1011
15	0111
16	1111

Specific Applications

- Population estimation
- Strategic needs assessments
- Access to local services
- Regeneration
- Well being and life expectancy
- Environment, transport and housing
- Deprivation
- Child care sufficiency
- Children's services
- Community safety
- Older peoples services
- Chronic disease management
- Educational attainment
- Policy evaluation

Overview

- Innovative technique, underexploited data
- A more granular and flexible evidence base
- Improves planning and delivery at the small area level
- Change can be monitored more frequently
- Explores directly relationship between population characteristics and outcomes
- Feed into parallel investigation of using administrative data for future Census

END