Using local administrative data to count populations: the UK case

'Life After the Census'
University of Ulster 9th May 2012

Gill Harper

nkm and GeoIntel Ltd

gill.harper@geointel.co.uk

Outline

- Background
- Limitations of Census
- Administrative data as an alternative
- *nkm* methodology to count populations
- Examples
- Benefits of administrative data
- Challenges of administrative data
- Progress of use of administrative data in GB

Background

- In the 2001 Census, response rates fell below acceptable response rates for many local authorities

- As a result some expressed doubts about the accuracy of their population counts and challenged ONS

- Since 2001 there have been significant intervening changes (population influxes, fertility, regeneration)

- When converted into revenue allocations, undercounts can be worth tens of millions of pounds in shortfalls to LAs and PCTs over 10-year cycle

Response rates in 2001 by London borough
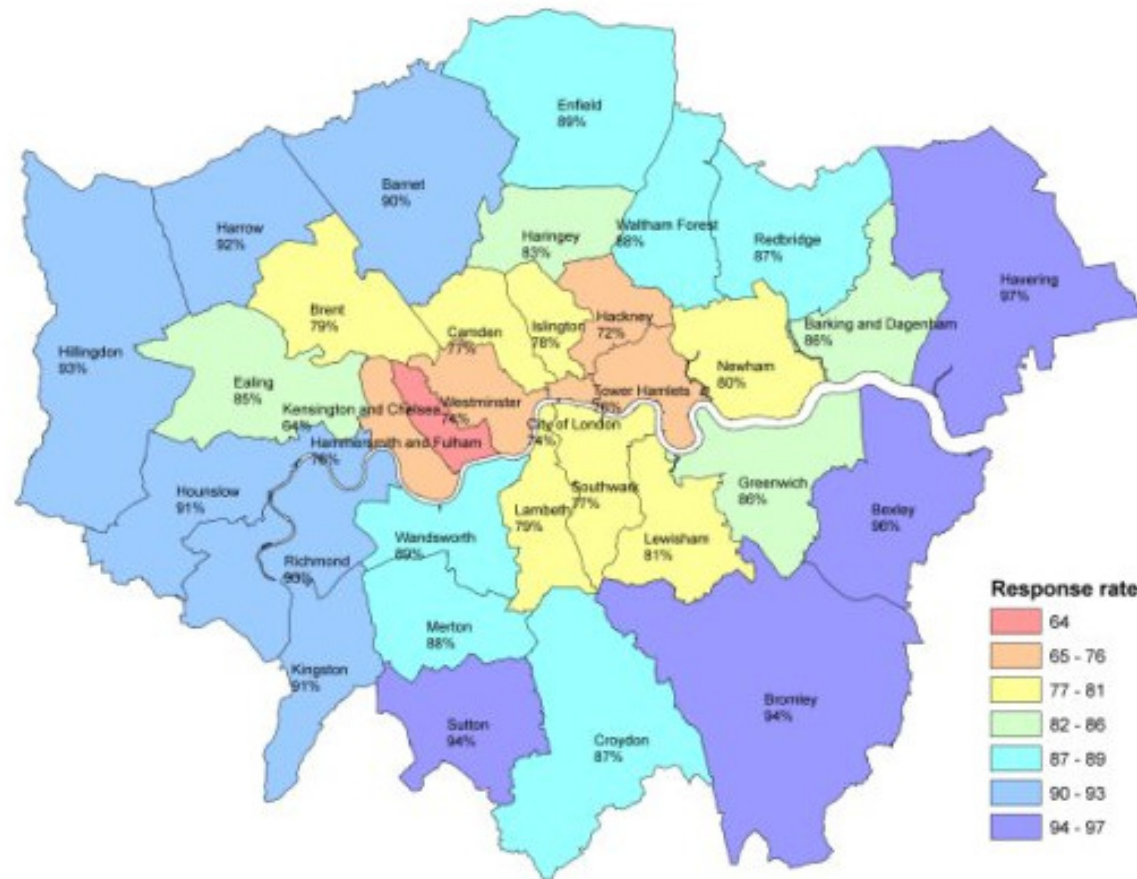
Census 2011

- The 2011 UK Census is estimated to cost in excess of £500m
- In spite improvements, there are concerns about potential response rates in certain households (HMOs, migrant households, households with 7+ persons)
- As part of the quality assurance process, ONS considered locally derived evidence
- Future format is uncertain but is likely to rely more heavily on administrative data
- Initial 2011 outputs in July 2012

Limitations

Even with good response rates:

- Census baseline is only renewed every ten years
- But data up to 12 years out of date by time *new* data are available
- Geographical units are inflexible and/or inappropriate
- Inflexible definitions of data items (e.g. age, ethnicity)
- MYEs built on shaky baseline because of population fluxes
- Linkage to administrative data or surveys impractical or error prone (ecological fallacy, MAUP)

Administrative data as an alternative

- There is much routinely collected administrative data at household or individual level:
  - GP Register
  - Council Tax Register
  - Electoral Register
  - Benefits Register
  - School Census
  - Births and Deaths
  - Housing Waiting List
  - Social Services

Data sharing agreements and concept of a 'virtual group'

PCT

'Virtual Group'

Police

Local authority
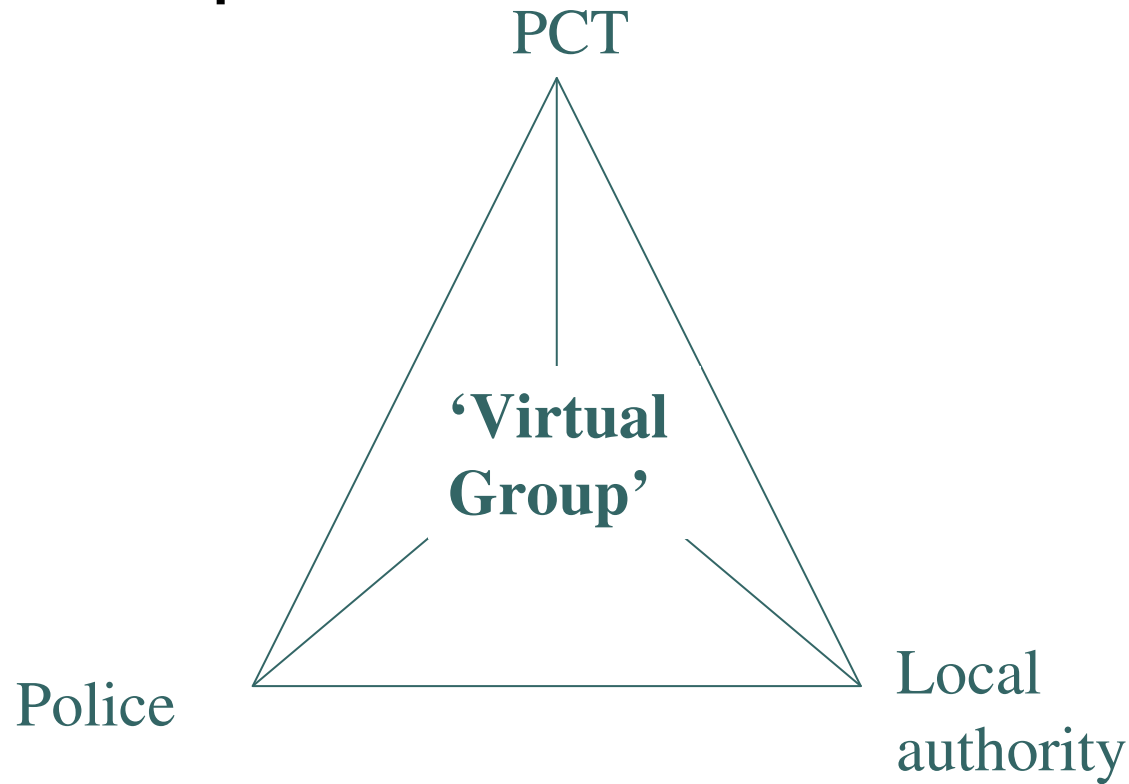
A virtual group is an analytical hub with membership from different agencies that acts as a secure haven for data and is bound by strict rules about data confidentiality, data protection legislation, and protocols with data suppliers

nkm

Successful implementation

- Feasibility proven: *nkm* developed a system to exploit administrative data to count local populations
- Secure data sharing and analysis
- Implemented for numerous local authorities
- Carried out for 6 Olympic boroughs March 2011 to benchmark with Census results and QA
- By end of June 2011 we had provided population breakdowns by age, sex and ethnicity
- By end of July databases transferred to each LA and used locally
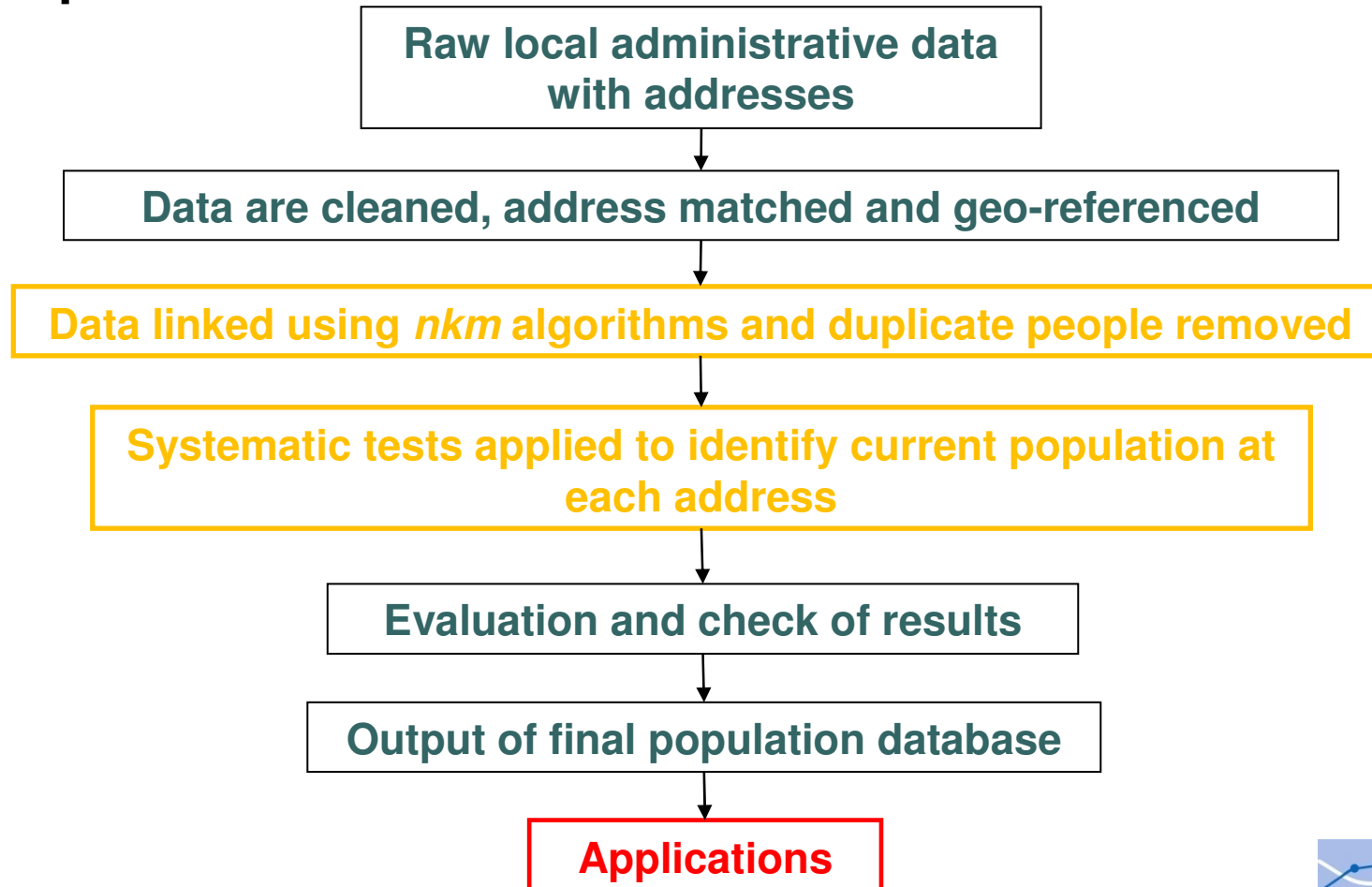
Procedure

- Data are addressed matched to the local property gazetteer
- Duplicates are removed
- Births added and deaths removed
- Tests are applied to ensure a person is the latest at an address
- Output is a minimum 'confirmed population'

Procedure

Raw local administrative data
with addresses

↓

Data are cleaned, address matched and geo-referenced

↓

Data linked using *nkm* algorithms and duplicate people removed

↓

Systematic tests applied to identify current population at
each address

↓

Evaluation and check of results

↓

Output of final population database

↓

Applications

Record linkage

- No consistent unique identifier in GB
  - NI number
  - NHS number
- Record linkage required
- Algorithms compare person identification fields: forename, surname, gender, date of birth
- Sophisticated rules

Methodology

- Systematic and rule-based
- Stages represented in truth tables
- Boolean algebra to test a logical expression as T or F:

Venn element	a	b	c	decision	comment
0	0	0	0	R	not on any data set
1	1	0	0	R	on the GP register only
2	0	0	1	R	empty property
3	0	1	0	R	on other data set only
4	1	0	1	A	on GP and address register
5	1	1	0	R	on GP register and other data set
6	0	1	1	A	on other data set and on address register
7	1	1	1	A	on GP register and other data set and address register

A = accept R = reject
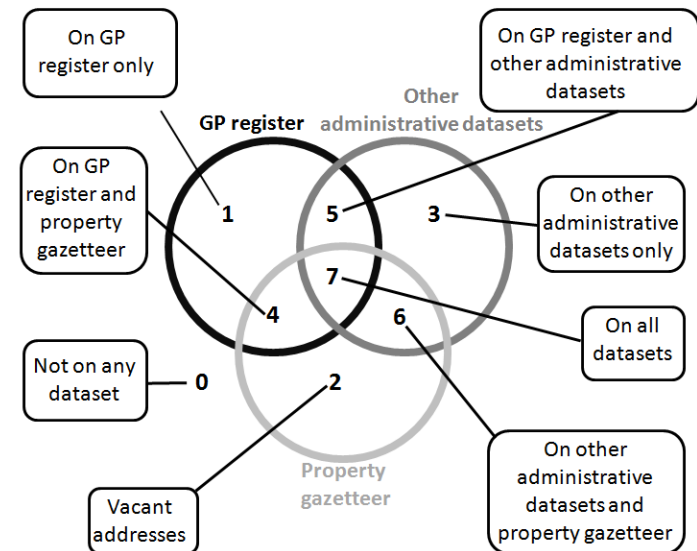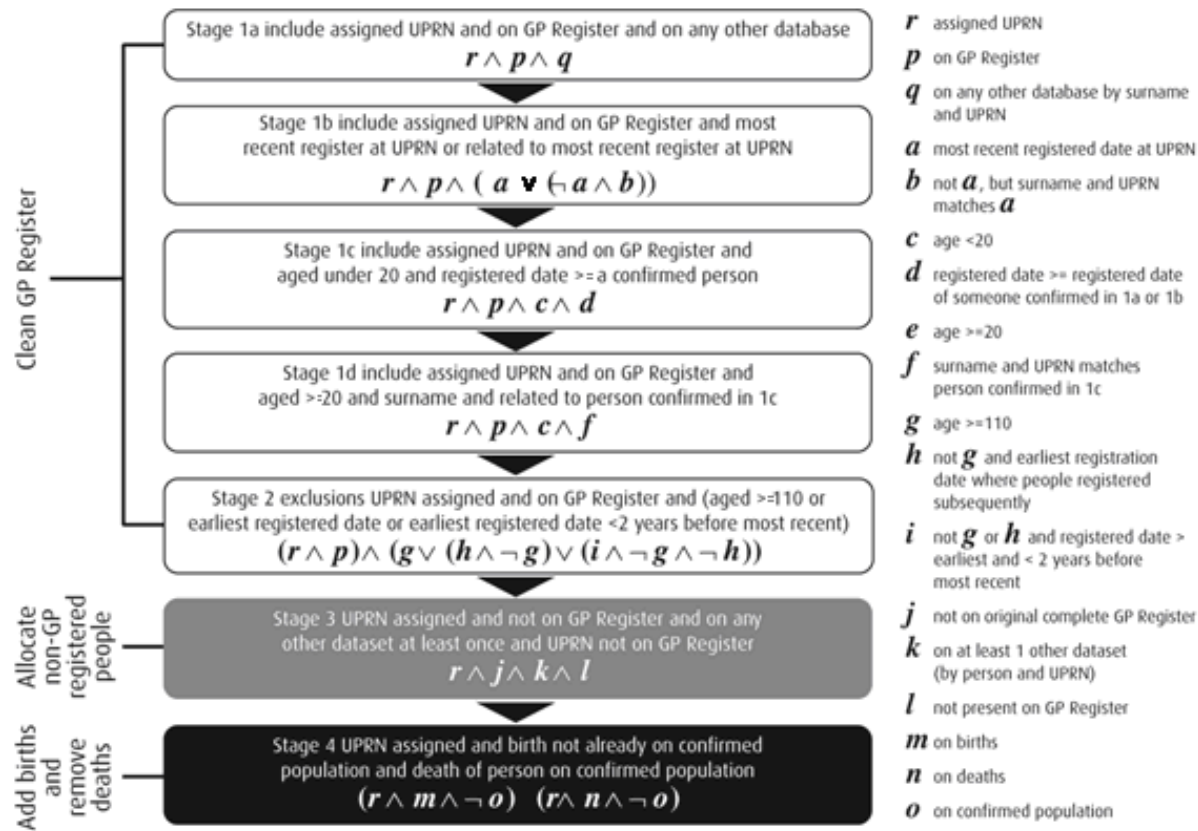
a: on GP register

b: on any other database
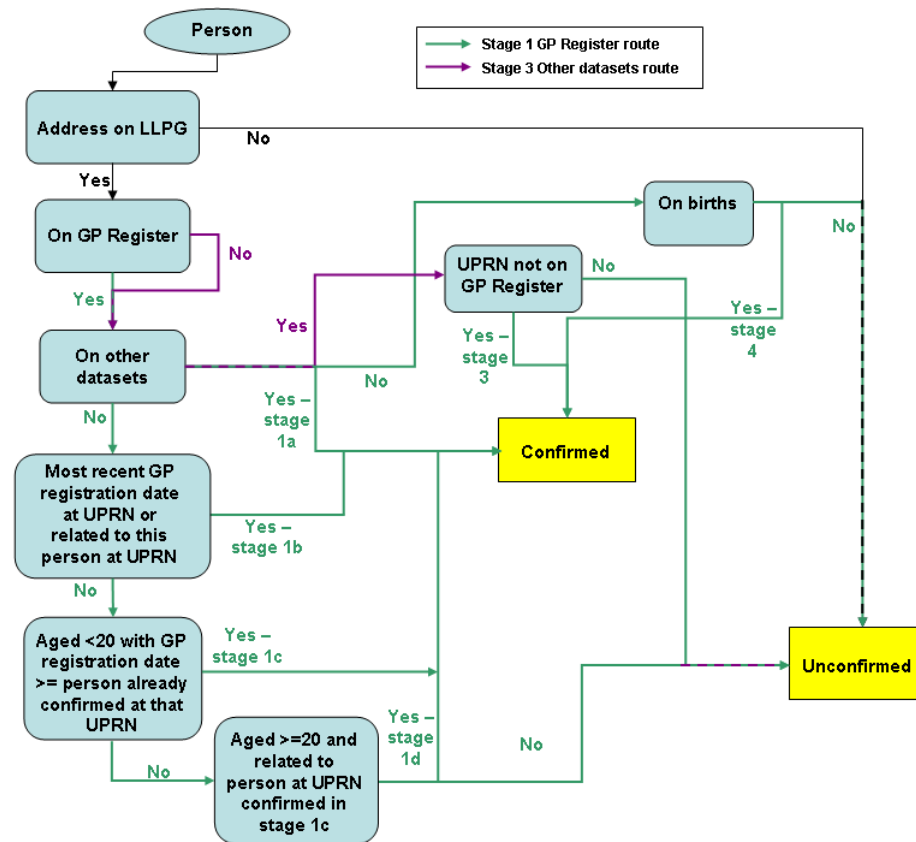
c: assigned a UPRN (living at a recognised address)

Methodology

- Summary of stages as truth tables:

Methodology

- Pathway to confirmation:

Output population table

ID	Age	Gender	UPRN	Easting	Northing	Ward_name	LSOA	Occupancy	Social_housing	Council_Tax_band
1	32	F	000100056303	543912	184140	Abbey	E01000010	4	0	B
2	0	M	000100030905	548510	187902	Whalebone	E01000112	9	0	D
3	35	M	000100056959	543667	184531	Abbey	E01000008	2	1	A
4	2	M	000100056959	543667	184531	Abbey	E01000008	2	1	A
5	4	M	000100008596	547432	185360	Mayesbrook	E01000077	4	0	C
6	1	F	010002170633	544361	183983	Abbey	E01000010	3	0	C
7	37	M	010002170633	544361	183983	Abbey	E01000010	3	0	C
8	1	M	000100055498	544495	184681	Abbey	E01000009	5	0	C
9	26	F	000100070114	544646	183242	Gascoigne	E01000049	2	0	B
10	4	M	000100013136	548887	188168	Whalebone	E01000110	4	0	D
11	5	F	000100013136	548887	188168	Whalebone	E01000110	4	0	D
12	30	F	000100013136	548887	188168	Whalebone	E01000110	4	0	D
13	29	F	010002168540	549054	186182	Heath	E01000062	9	0	D
14	2	F	000100062296	544744	183419	Gascoigne	E01000048	3	1	B
15	26	F	000100062296	544744	183419	Gascoigne	E01000048	3	1	B
16	0	M	000100002519	546758	185839	Becontree	E01000021	7	0	C
17	33	M	000100059212	544575	184133	Abbey	E01000010	6	0	D
18	1	M	000100059212	544575	184133	Abbey	E01000010	6	0	D
19	30	F	000100059212	544575	184133	Abbey	E01000010	6	0	D
20	11	F	010002171822	549620	184891	Alibon	E01000015	5	0	D
21	43	M	000100046878	550035	184690	Village	E01000108	1	0	C
22	2	M	000100028883	549158	185600	Alibon	E01000016	3	0	C
23	4	M	000100005894	545659	185198	Longbridge	E01000067	7	0	F
24	0	F	000100088966	547247	183009	Thames	E01000095	3	0	C

Example – Barking & Dagenham

age group	administrative population at 30/9/2008	ONS* 2008 MYE (old)	ONS** 2008 MYE (revised)	GLA*** 2008 (revised)
Total	171,851	168,853	171,600	171,976

*   published 2009

**   published in 2010

*** published 2010

Example – Newham 2011

age groups	nkm	nkm adjusted 2011	nkm adjusted 2007	GLA 2011	ONS MYE 2010
0-4	26,124	26,124	24,152	25,835	25,800
5-9	21,841	21,841	19,500	18,268	17,800
10-14	19,889	19,889	19,278	15,892	14,000
15-19	20,031	20,031	18,492	15,347	15,200
20-24	25,753	25,753	32,348	23,242	22,000
25-29	31,692	36,458	25,912	34,027	24,400
30-34	28,719	31,530	22,901	30,096	18,600
35-39	22,913	22,913	21,246	22,726	18,000
40-44	20,790	20,790	18,834	18,341	18,400
45-49	17,759	17,759	16,073	14,780	15,400
50-54	14,607	14,607	12,240	12,522	13,200
55-59	11,036	11,036	9,707	9,621	9,700
60-64	8,756	8,756	7,475	7,976	7,800
65-69	6,111	6,111	6,436	5,828	5,600
70-74	5,629	5,629	5,238	5,258	5,300
75-79	4,050	4,050	3,984	3,954	3,800
80-84	2,701	2,701	2,606	2,610	2,500
85+	2,139	2,937	3,669	2,530	2,700
age/unknown	8,376				
Total	298,916	298,916	270,091	268,854	240,200
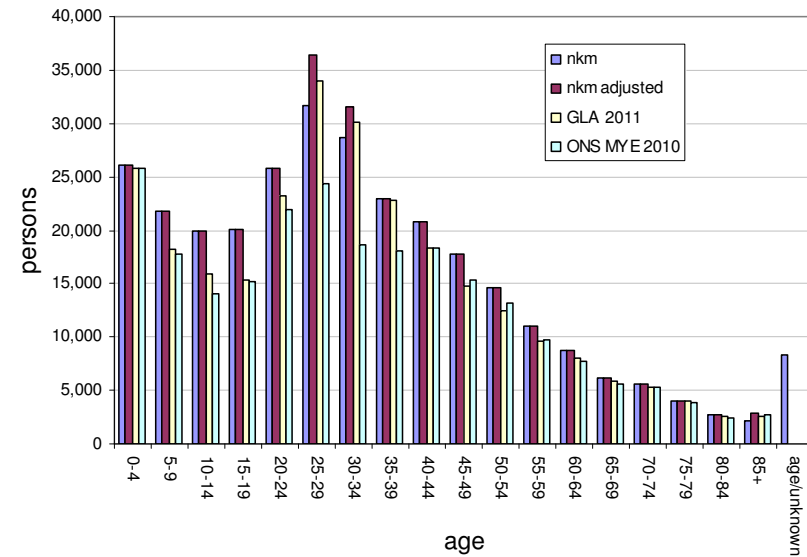
age group	2011-2007	% change
0-9	4,313	9.9
10-19	2,150	5.7
20-64	22,867	13.7
64+	-505	-2.3
total	28,825	10.7

The benefits of using administrative data

- Quicker turnaround
- Reduces burden on respondents
- Can be repeated frequently
- Reduced cost
- Outputs at the individual and household level including age and sex
- Population intelligence to support local decision making
- Easily combine with surveys
- Flexible and granular output for more responsive analytical services

The challenges in using administrative data

- Administrative registers are changing e.g. Academies, CCGs, Health and Wellbeing Boards

- Data access and lack of national legal framework

- No consistent unique identifier

- National implementation

Progress in UK

- *nkm* has successfully estimated local populations using administrative data
- ONS used this as local evidence in some cases
- ONS MYEs utilising administrative data – counts became closer to *nkm* estimates
- ADLS
- ONS considering *nkm* method

Beyond 2011

- The 2011 Census will cost £480m but excludes various interim costs
- Coalition announcement that the 2011 Census would be the last
- ONS to report on alternatives to Census by 2014
- Broadly 3 administrative options: central government data, local government data, commercial sources

END

gill.harper@geointel.co.uk

www.nkm.org.uk

www.geointel.co.uk

References:

Harper, G. and L. Mayhew (2011) Using administrative data to count local populations. Journal of Applied Spatial Analysis and Policy DOI: 10.1007/s12061-011-9063-y

Harper, G. and L. Mayhew (2011) Applications of population counts based on administrative data at local level. . Journal of Applied Spatial Analysis and Policy DOI 10.1007/s12061-011-9062-z

Harper, G. and L. Mayhew (2012) Re-thinking households - Using administrative data to count and classify households with some applications. Research paper, Cass Business School, Faculty of Actuarial Science and Insurance