Working Paper 12/03

Exploring the potential of microdata from a large commercial survey for the analysis of demographic and lifestyle characteristics of internal migration in Great Britain

Michael Thomas
Myles Gould
and John Stillwell

School of Geography, University of Leeds, Leeds, LS2 9JT, UK
Email gymjt@leeds.ac.uk

Paper presented at "Innovative Perspectives on Population Mobility: Mobility, Immobility and Well-being" a conference organized under the auspices of the Population Geography Research Group (RGS-IBG) and the ESRC Centre for Population Change, University of St Andrews, 2-3 July, 2012

July 2012

Abstract

Spatial analysts of internal migration in Britain have typically sourced data from population censuses or administrative registers because national survey data are usually restricted by sample size. Censuses provide extensive demographic and socio-economic attributes for migrants that are reliable and comprehensive but are only decadal. Administrative sources provide inter-censal data on a continuous quarterly or annual basis but the range of attributes tends to be very limited. As part of the 'Beyond 2011' programme, national statistical agencies across Britain are currently attempting to identify and evaluate alternative sources of migration data from administrative registers and social surveys in order to establish robust methods of annual population estimation as well as methods for providing disaggregated population counts for service providers.

This paper considers Acxiom's Research Opinion Poll (ROP) commercial data, hitherto unused for the analysis of internal migration in GB. It is a very large and rich national lifestyle survey collected annually for postcoded individuals (n≈750,000 per annum). The ROP contains individual level demographic and socio-economic, as well as indicators of area satisfaction. It is large enough to analyse/model 'movers' and 'stayers' and determine geographical patterns of residential mobility at relatively detailed spatial scales. Before presenting some preliminary empirical results of characteristics and patterns, the paper will outline some data quality issues and report an initial validation exercise involving comparison of the ROP data with census, administrative and/or survey data from alternative sources. Finally, the paper will outline potential modelling framework(s) for further analysis of this rich data set.

Keywords

Internal migration; annual survey data; lifestyles; data quality; validation

Contents

List of Figures

List of Tables

1 Introduction

Acxiom's Research Opinion Poll (ROP) is a large annual lifestyle survey carried out across Great Britain (GB). During the mid-2000s, the survey contained a series of questions about migration, providing a source of data with considerable potential for research that examines the lifestyle characteristics of those who had moved address in the past, those who intended to move in the future and those who had not experienced a residential move. This paper outlines some of the data quality issues associated with the survey and reports some initial validation exercises involving the comparison of Acxiom's ROP microdata with census, administrative and survey data from alternative sources. The paper is divided into four further main sections (2-5). Section 2 is largely concerned with providing an overview of what are common sources of migration data, however, a particular focus is on situating the ROP within the wider context of social survey sources in GB. Section 3 reports the findings of the initial validation exercises and is itself divided into four sub-sections that contain the validation exercises and analyses that have been undertaken. These sub-sections are arranged so as to cover the validation of aggregate, micro and spatial dimensions of the ROP respectively. Section 4 moves beyond the validation exercise and explores some of the ROP's unique attributes in terms of both variables available and the spatial detail. Section 5 offers a potential framework for analysis. There is then brief conclusion in Section 6.

2 Migration data sources

2.1 Conventional sources

There is no single comprehensive source of data that can cover all the disparate requirements of those interested in internal migration. Instead, researchers and policy makers find themselves in a situation where, in order to satisfy their requirements, they must utilise a variety of sources characterised by sharp contrasts in coverage, detail (of both data and geography) and accuracy. Spatial analysts of internal migration in both GB and farther afield have typically sourced data from censuses or administrative registers because national survey data are usually restricted by sample size (Nam *et al.*, 1990; Rees and Kupiszewski, 1999; Stillwell *et al.*, 2011). Indeed, the IMAGE project[1], which aims to identify forms of data on internal migration collected by individual countries around the world, indicates that 91% of the 179 countries in the project data inventory draw their internal migration statistics

[1] Internal Migration Around the GlobE (http://www.gpem.uq.edu.au/image)

primarily from a census (Bell *et al.*, 2012). A further 25% employed data from administrative and/or population registers with 66% utilising surveys, as shown in Table 1.

Table 1: Primary sources of internal migration statistics

		Data sources			
Region	**Countries**	**Census**	**Admin/Register**	**Survey**	**Multiple sources**
Africa	50	47	0	41	38
Asia	40	37	14	24	26
Europe	42	32	28	35	35
Latin America	31	31	0	14	31
North America	3	3	2	2	3
Oceania	13	13	1	2	3
TOTAL	179	163	45	118	136

Source: IMAGE's global inventory of internal migration sources

In the UK context, censuses provide extensive demographic and socio-economic attributes for migrants that are reliable but are only decadal. Given their near comprehensive coverage, their relative reliability and the rich detail of their demographic and socioeconomic variables, censuses are currently considered as the optimum points of reference for those interested in local population statistics (Raymer *et al*., 2012) and in small area demographic analysis in particular.

Registers and administrative sources provide inter-censal migration data on a continuous quarterly or annual basis but the range of attributes tends to be very limited. The Office for National Statistics (ONS) employs a combination of proxy administrative sources from which migration estimates are drawn and used subsequently in the estimation of mid-year populations. For England and Wales, two sets of internal migration estimates are produced by the ONS using administrative sources (Rees *et al*., 2009). The first is based on National Health Service Central Register (NHSCR) data and captures events of migration between health authorities (HAs). The second is based on the Patient Register Data System (PRDS) which measures transitions in the NHS patient data. The PRDS data are available from 1999 and are produced at the local authority district level which means that, with the potential to aggregate into counties, they are consistent with the geographical units used in the 2001 and 2011 Censuses respectively (Smith *et al*., 2010; Raymer *et al*., 2012). Data from an additional administrative source, the Higher Education Statistics Agency (HESA), is used to adjust the estimates and control for students in higher education, a subgroup with unique flow patterns (ONS, 2010a). There are comparable registration systems for patients used to estimate

internal migration flows in Scotland (Community Health Index data, CHI) and Northern Ireland (Central Health Index, NI-CHI). Beyond this, the cross-border moves between England and Wales and Scotland and Northern Ireland are provided to ONS for inclusion in the published estimates by the respective national statistics agencies (ONS, 2011a) although cross-border flows between districts comprising the home countries remain unavailable (Lomax *et al*., 2011). As noted above, such data sources are extremely useful for those interested in measuring up-to-date migration counts and flows; however, at the same time, they have not been designed explicitly for the purpose of capturing migration, they suffer from a distinct lack of demographic and socio-economic detail in most cases and are, in some cases, only relevant to distinct sub-sections of society such as school children (School Census) and HE students (HESA).

Surveys, on the other hand, represent timely sources of highly detailed socio-demographic data, but are typically characterised by relatively small sample sizes which restrict the potential for reliable analysis at more detailed geographical levels. As Poston and Bouvier (2010: 34) argue, *"[b]y administering surveys to carefully selected random samples of the larger populations, demographers are better able to uncover underlying patterns of demographic behaviour than is possible with materials from censuses and registration systems"*. However, as stated, this level of attribute detail is commonly constrained by the level of spatial detail and coverage included in surveys.

As part of the 'Beyond 2011' programme, national statistical agencies across the UK are currently attempting to identify and evaluate alternative sources of migration data from administrative registers and social surveys in order to establish robust methods of annual population estimation as well as methods for providing disaggregated population counts for service providers (ONS, 2011b). Indeed, going forward, it is far from certain that the traditional census will remain as the primary point of reference for small area demographic and socio-economic statistics in the UK.

With this context in mind, this paper is focused on evaluating and validating a source of migration data hitherto unused for migration analysis that comes from the commercial sector. Acxiom's ROP is a very large national lifestyle survey, providing a wealth of information about respondents on an annual basis that includes individual level demographic and socio-economic characteristics (e.g. age, occupation, education, household income, health problems) and well as area level wellbeing (e.g. neighbourhood satisfaction).

2.2 The Acxiom ROP and its value as a source of migration data

Acxiom is an international commercial company and a global leader in interactive multichannel marketing services. Acxiom's main product, and the product that is of most interest here, is the annual Research Opinion Poll (ROP) survey. According to Thompson *et al*. (2010: 7-8), the primary aim of the ROP is to *"gather detailed and up-to-date information on consumer spending habits, preferences, socio-demographic information and the respondents' geographic locations"*. The ROP is essentially a paper-based survey that is delivered through direct mail twice a year, in September and January, with the results of the two surveys being combined to form the annual data file, i.e. the surveys from September 2010 and January 2011 combine to form the results for 2011. Acxiom offers additional products that are themselves derived from the ROP. One such derived product is the annually produced 'Aggregated Data' which, after a process of weighting and manipulation, is argued to be fully representative of the GB population at district and lower super output area (LSOA) levels (Thompson *et al.*, 2010).

One of the key benefits of the ROP survey lies in its sheer size; the microdataset for 2009 involved over 1,000,000 household responses. While the exact details are not disclosed by Acxiom, it is clear that this huge return is based on a great deal of planning by Acxiom's data acquisition team who employ a number of sources of addresses so as to ensure a geographically even, and demographically robust, response (Rees *et al*., 2009). The success of their approach is cited by Thompson *et al*. (2010: 13) who highlight the fact that for the 2009 ROP, *"only 0.4% of all Middle Super Output Areas (MSOAs) across [Great Britain] did not return a response"*. Given the size of the sample population, the Acxiom ROP microdata represent a unique and potentially important source of data for social science research.

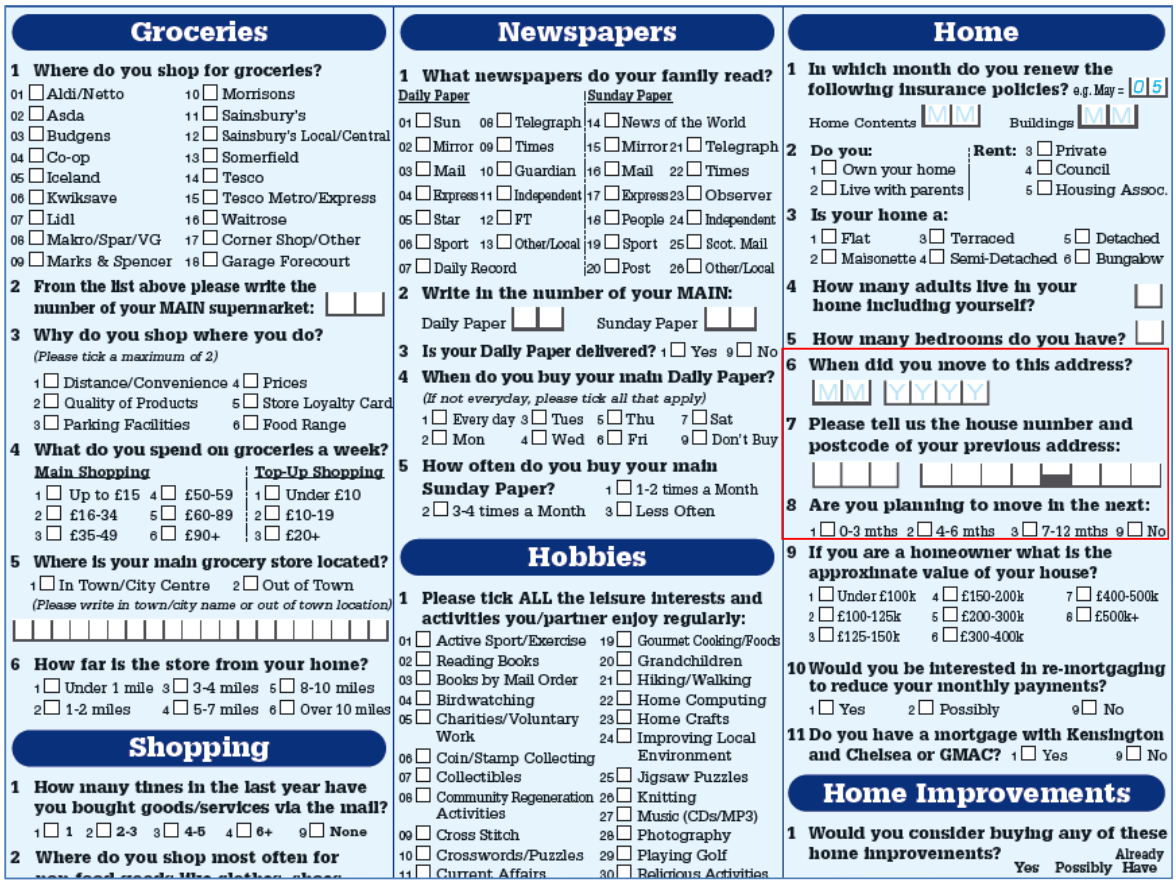
However, beyond the size of the survey in terms of numbers of respondents, the ROP also offers a great deal in terms of attribute detail. For example, in 2009, there were 130 questions asked, allowing for over 1,000 possible answers, covering 26 broad topics including: groceries; shopping; your local area; environment; outgoings; occupation; home; leisure; education; health; and even including a section on the 'credit crunch'. The questions asked can be broken down into two broad categories: core questions and sponsored questions. The core questions are repeated from survey to survey and cover such characteristics as respondents' address, age, sex, income, occupation, tenure and grocery spend. Given the

continuous nature of the core questions, it is possible to generate data appropriate for consistent time-series analysis from 2004 (Thompson *et al.*, 2010). There are also sponsored questions that, as one would expect, are included in the survey having been paid for by clients. The now defunct regional development agency, Yorkshire Forward, for instance, over a number of years sponsored a number of questions ranging in topic from specific questions on Yorkshire, through to more general issues relevant to the environment and tenure.

For the years 2005, 2006 and 2007, the ROP surveys prove particularly useful for those interested in residential moves within GB because the following questions were included about respondents' migration behaviour:

- 'When did you move to this address? (month and year)';
- 'Please tell us the house number and postcode of your previous address';
- 'Are you planning to move in the next: 0-3 months; 4-6 months; 7-12 months; No?'

The 'Home' section of the 2007 ROP questionnaire is shown in Figure 1.



Groceries

1 Where do you shop for groceries?

01 ☐ Aldi/Netto
02 ☐ Asda
03 ☐ Budgens
04 ☐ Co-op
05 ☐ Iceland
06 ☐ Kwiksave
07 ☐ Lidl
08 ☐ Makro/Spar/VG
09 ☐ Marks & Spencer
10 ☐ Morrisons
11 ☐ Sainsbury's
12 ☐ Sainsbury's Local/Central
13 ☐ Somerfield
14 ☐ Tesco
15 ☐ Tesco Metro/Express
16 ☐ Waitrose
17 ☐ Corner Shop/Other
18 ☐ Garage Forecourt

2 From the list above please write the number of your MAIN supermarket: ☐☐

3 Why do you shop where you do?
(Please tick a maximum of 2)

1 ☐ Distance/Convenience
2 ☐ Quality of Products
3 ☐ Parking Facilities
4 ☐ Prices
5 ☐ Store Loyalty Card
6 ☐ Food Range

4 What do you spend on groceries a week?

Main Shopping		Top-Up Shopping	
1 ☐ Up to £15	4 ☐ £50-59	1 ☐ Under £10	
2 ☐ £16-34	5 ☐ £60-89	2 ☐ £10-19	
3 ☐ £35-49	6 ☐ £90+	3 ☐ £20+	

5 Where is your main grocery store located?
1 ☐ In Town/City Centre 2 ☐ Out of Town
(Please write in town/city name or out of town location)

6 How far is the store from your home?
1 ☐ Under 1 mile 3 ☐ 3-4 miles 5 ☐ 8-10 miles
2 ☐ 1-2 miles 4 ☐ 5-7 miles 6 ☐ Over 10 miles

Shopping

1 How many times in the last year have you bought goods/services via the mail?
1 ☐ 1 2 ☐ 2-3 3 ☐ 4-5 4 ☐ 6+ 9 ☐ None

2 Where do you shop most often for

Newspapers

1 What newspapers do your family read?

Daily Paper			Sunday Paper	
01 ☐ Sun	08 ☐ Telegraph		14 ☐ News of the World	
02 ☐ Mirror	09 ☐ Times		15 ☐ Mirror	21 ☐ Telegraph
03 ☐ Mail	10 ☐ Guardian		16 ☐ Mail	22 ☐ Times
04 ☐ Express	11 ☐ Independent		17 ☐ Express	23 ☐ Observer
05 ☐ Star	12 ☐ FT		18 ☐ People	24 ☐ Independent
06 ☐ Sport	13 ☐ Other/Local		19 ☐ Sport	25 ☐ Scot. Mail
07 ☐ Daily Record			20 ☐ Post	26 ☐ Other/Local

2 Write in the number of your MAIN:
Daily Paper ☐☐ Sunday Paper ☐☐

3 Is your Daily Paper delivered? 1 ☐ Yes 9 ☐ No

4 When do you buy your main Daily Paper?
(If not everyday, please tick all that apply)
1 ☐ Every day 3 ☐ Tues 5 ☐ Thu 7 ☐ Sat
2 ☐ Mon 4 ☐ Wed 6 ☐ Fri 9 ☐ Don't Buy

5 How often do you buy your main Sunday Paper?
1 ☐ 1-2 times a Month
2 ☐ 3-4 times a Month 3 ☐ Less Often

Hobbies

1 Please tick ALL the leisure interests and activities you/partner enjoy regularly:

01 ☐ Active Sport/Exercise
02 ☐ Reading Books
03 ☐ Books by Mail Order
04 ☐ Birdwatching
05 ☐ Charities/Voluntary Work
06 ☐ Coin/Stamp Collecting
07 ☐ Collectibles
08 ☐ Community Regeneration Activities
09 ☐ Cross Stitch
10 ☐ Crosswords/Puzzles
11 ☐ Current Affairs
19 ☐ Gourmet Cooking/Foods
20 ☐ Grandchildren
21 ☐ Hiking/Walking
22 ☐ Home Computing
23 ☐ Home Crafts
24 ☐ Improving Local Environment
25 ☐ Jigsaw Puzzles
26 ☐ Knitting
27 ☐ Music (CDs/MP3)
28 ☐ Photography
29 ☐ Playing Golf
30 ☐ Religious Activities

Home

1 In which month do you renew the following insurance policies? e.g. May = 05
Home Contents ☐☐ Buildings ☐☐

2 Do you:
1 ☐ Own your home
2 ☐ Live with parents
Rent: 3 ☐ Private 4 ☐ Council 5 ☐ Housing Assoc.

3 Is your home a:
1 ☐ Flat 3 ☐ Terraced 5 ☐ Detached
2 ☐ Maisonette 4 ☐ Semi-Detached 6 ☐ Bungalow

4 How many adults live in your home including yourself? ☐

5 How many bedrooms do you have? ☐

6 When did you move to this address?
☐☐ ☐☐☐☐

7 Please tell us the house number and postcode of your previous address:
☐☐☐ ☐☐☐☐☐☐☐

8 Are you planning to move in the next:
1 ☐ 0-3 mths 2 ☐ 4-6 mths 3 ☐ 7-12 mths 9 ☐ No

9 If you are a homeowner what is the approximate value of your house?
1 ☐ Under £100k 4 ☐ £150-200k 7 ☐ £400-500k
2 ☐ £100-125k 5 ☐ £200-300k 8 ☐ £500k+
3 ☐ £125-150k 6 ☐ £300-400k

10 Would you be interested in re-mortgaging to reduce your monthly payments?
1 ☐ Yes 2 ☐ Possibly 9 ☐ No

11 Do you have a mortgage with Kensington and Chelsea or GMAC? 1 ☐ Yes 9 ☐ No

Home Improvements

1 Would you consider buying any of these home improvements?

	Yes	Possibly	Already Have

Figure 1: Section of January 2007 Acxiom ROP with migration questions (Source: Acxiom n.d.)

When these questions are combined with each respondent's 'current address' (house number and postcode), the unique value of this data source for migration analysis becomes apparent. After all, as Rees *et al.* (2009: 110) make clear, surveys of this type *"provide data on the origin, destination and timing of the last move together with some indications of future migration propensity"*. Not only does the precise geo-referencing of cases allow the researcher to generate any geography of their choice, with the timing of the move, the potential exists for the integration of time within the research, for instance through the analysis of variations between cohorts of movers (e.g. are some cohorts more/less 'successful' in their moves than others?). Moreover, with the ROP surveys being the size that they are, they do offer the unique potential, at least when compared to other survey sources, for use in a comprehensive geographical study of migration below district level (Table 5). Table 2 represents the variables that have been collected from Acxiom for use within this research project.

Table 2: Overview of Acxiom ROP variables available for use in this project

Variable	Jan-05	Jan-07	Sep-07
Postcode	✓	✓	✓
Gender	✓	✓	✓
D.O.B (respondent & partner)	✓	✓	✓
Marital Status	✓	✓	✓
Ethnic or national background	✓	✓	
Occupation (respondent & partner)	✓	✓	✓
Income (combined annual)	✓	✓	✓
Approximate value of house		✓	✓
Qualifications (respondent & partner)	✓	✓	✓
Number of cars in household	✓	✓	✓
Make of car (main car)	✓		
Model & type (main car)	✓		
First four characters of car registration (main car)	✓		
Is the car owned (main car)	✓		
Respondent/partner suffer from (see below):			
1 hearing aid; 2 hearing difficulties; 3 mobility problems; 4 slight loss of bladder control	✓		
1 mobility problems; 2 arthritis/ rheumatism; 3 hearing difficulties; 4 hearing aid; 5 diabetes; 6 slight loss of bladder control		✓	
1 mobility problems; 2 arthritis/ rheumatism; 3 hearing difficulties; 4 hearing aid; 5 slight loss of bladder control; 6 concerns about urine flow; 7 back pain; 8 osteoporosis			✓
Own or rent residence	✓	✓	✓
Residence type	✓	✓	✓
Number of people living at home	✓		✓
Month moved to the address	✓	✓	✓
Year moved to the address	✓	✓	✓
Number and postcode of last address	✓	✓	✓
Planning to move in future	✓	✓	✓
Like the neighbourhood they live in	✓	✓	✓
Neighbourhood (improved/ same/ got worse/ recently moved)		✓	✓

2.3 Overview of alternative survey sources

Despite the ROP's specific strengths, surveys are typically rich in variable detail but lacking in geographical detail and coverage. In order to demonstrate this, and allow for comparison with Acxiom's ROP, we provide an overview of some selected survey sources with particular attention being paid to the questions they ask about migration and consequently their usefulness for the analysis of internal migration in GB. The selection of the following three survey sources for discussion is based on the opinion that they are the most comparable surveys to the ROP in terms of sample size and variable detail and, furthermore, they appear to be the best equipped to answer questions relevant to the study of migration in the UK.

2.3.1 Integrated Household Survey (IHS)

The Integrated Household Survey (IHS) is a continuous composite survey with annual updates being released quarterly on a rolling basis (April 2009 to March 2010, July 2009 to June 2010, October 2009 to September 2010, January 2010 to December 2010, and so on). According to the ONS (2011c: 2), *"[t]he aim of the IHS is to produce high-level estimates for particular themes to a higher precision and lower geographic level than current Office for National Statistics (ONS) social surveys"*. Indeed, for the first IHS data release (April 2009 to March 2010), a sample size of approximately 450,000 individuals was achieved. Quite uniquely, the IHS has been formed from the merging of a number of existing government surveys. All of the component surveys contain a number of similar questions that form what are the 'core' questions of the IHS, covering themes including: economic activity, education, health and disability, identity and income (ONS, 2011c; Walthery, 2011). There are roughly 100 'core' questions within the IHS. As outlined in Table 3, the component surveys for the April 2009 to March 2010 IHS include: the General Lifestyle Survey (GLF), the Living Cost and Food Survey (LCF), the Opinions Survey (OPN), the English Housing Survey (EHS), the Labour Force Survey/Annual Population Survey (LFS/APS), and the Life Opportunities Survey (LOS). However, since the first data release, there have been a number of component changes to the IHS. In January 2010, the OPN survey was removed, a decision based on the requirement to reduce the size of the OPN survey by removing the IHS 'core' questions (ONS, 2010b). Further to this, both the LOS and EHS were also removed in April 2011. The LOS removal was based on a change in its sampling methodology making it inappropriate for inclusion while the EHS removal was based on funding restrictions. Finally, the last data from the GLF were contributed in December 2011 (Jones, 2011). However, as Walthery (2011: 3) notes, while the contributing surveys have been reduced in number over recent

months, *"it is expected that the composition of the IHS will be flexible with some surveys leaving the IHS and others entering each year"*.

Table 3: IHS (April 2009-March 2010) component surveys and their respective sizes

Component Survey	Observations (Individuals)	Percentage of IHS sample
Annual Population Survey (APS)	334,206	74
English Housing Survey (EHS)	40,753	9
Life Opportunities Survey (LOS)	23,368	5
Opinions Survey (OPN)	20,981	5
General Lifestyle Survey (GLF)	18,033	4
Living Cost and Food Survey (LCF)	11,989	3
Total (IHS)	*449,330*	*100*

Source: ONS (2011c)

Given that the IHS is a new composite survey, made up of separate surveys with their own specific designs, a certain level of care is required when attempting to use the data it provides for research. Indeed, Walthery (2011: 7) warns that *"it is not recommended to produce tables or estimates of the data without using weights, given the heterogeneity of possible source or error within each variable"*. In terms of geographical coverage, the IHS includes the whole of the UK including Northern Ireland. However, it should be noted that not all component surveys cover the UK. For instance, of the April 2009 to March 2010 components, only the LFS and associated APS cover the whole of the UK – the GLF, LCF, LOS and OPN cover GB and the EHS only covers England. The highest geographic detail available to standard users of the IHS is Government Office Regions (GORs) although, if the user is granted 'special licence' access, a range of further geographies is available including Unitary Authorities/Local Authority Districts (UA/LAD), Counties, NUTS2 and NUTS3 regions (ONS, 2011c).

With regard to migration, the IHS provides two key questions from which internal flow data can be derived: 'place of residence 3 months ago (UA/LAD)' and 'place of residence one year ago (UA/LAD)'. However, unfortunately, the data derived from these questions are currently only available on the ONS internal research datasets. The best option available publically, via the Economic and Social Data Service (ESDS), is data derived from 'place of residence 3 months ago (UK/somewhere else)', which is itself restricted to 'special license' access. Within the ESDS End User dataset, it is possible to gather data on period at current address. Table 4 indicates the nature of the migration questions in each of the major social surveys carried out in the UK, including those discussed in this section.

Table 4: Overview of social survey migration questions

Survey	Migration Questions and Access Restrictions
Integrated Household Survey	Period at current address; place of residence 3 months ago (UK or somewhere else) (Special License); place of residence 3 months ago (range of codes as UA/LAD) (ONS internal only); place of residence one year ago (range of codes as UA/LAD) (ONS internal only).
Labour Force Survey	Period at current address; region of residence 3 months ago and 1 year ago (Special License: UA/LAD).
Annual Population Survey	Period at current address (Special License: region of residence 3 months and 1 year ago).
General Lifestyle Survey	Period at current address (can be linked through IHS to migration questions in 'core' module).
Family Resources Survey	How long have you lived at the address? (0-12 months / 1 year to more than 20 years).
English Housing Survey	Period at current address (can be linked through IHS to migration questions in 'core' module).
Life Opportunities Survey	Period at current address, and ability to track longitudinally (can be linked through IHS to migration questions in 'core' module).
Living Costs and Food Survey	Period at current address (can be linked through IHS to migration questions in 'core' module).
Acxiom Lifestyle Survey	When did you move to this address (Month and Year); the number and postcode of previous address; planning to move in the next (up to 12 months). Sponsored questions asked: Jan 2005, 2006, 2007 and September 2007.
British Household Panel Survey	Moved in past year; future intention to move; move into residential home, and ability to track moves longitudinally.
Understanding Society	Lived at address whole life; moved to address (month, year); prefers to move house; expects to move house; expects to move in next year, and ability to track moves longitudinally.

2.3.2 Labour Force Survey (LFS)/Annual Population Survey (APS)

The LFS is a continuous quarterly survey with a sample population (April to June 2011) of around 100,300 individuals in 42,900 households (ONS, 2011d: ii). As the ONS (2002: 550) states, the primary purpose of the LFS should be recognised as *"the prompt publication of key aggregate, whole economy indicators for the integrated assessment of labour market conditions"*. In its current form, the LFS employs a rotational sampling design, whereby respondents are interviewed five times at 13-week intervals and once they complete wave 5 they drop out and new respondents take their places. The broad socio-economic and demographic categories included in the LFS have remained roughly the same since the major

format changes of 1992. Categories in the LFS include, for example, ethnic group, gender, age, religion, education and training, income, health and employment type/location/hours worked. Paying specific attention to its use for migration analysis, the LFS household dataset includes a question on region of residence three months ago and one year ago, making it possible to formulate flow matrices by cross-tabulating previous residence with the region of usual residence. Furthermore, as Champion *et al*. (1998) have noted, the LFS also includes information on the respondents' labour market position one year prior, an interesting characteristic for those interested in analysing the causes and effects of migration.

Beyond this, the LFS also has strengths and limitations more particular to its use for the analysis of internal migration. Perhaps its key strength lies in the fact that it offers quarterly migration data that can be disaggregated further by a large number of socio-economic and demographic characteristics (individual level and household level). The quarterly continuous production of such data means that the LFS proves a useful source for researchers interested in examining migration and population change between census dates, and by characteristics not included in the patient register datasets (Owen and Green, 1992). A major limitation is its relatively small sample size, when compared against data sources like the NHSCR/PRDS and the census, which makes it a less satisfactory source for those interested in estimating migration counts and flows within the UK. Furthermore, the relatively small sample size means that the LFS can only offer, at best, spatial detail at the region scale, although from March 2005, it has been possible to obtain UA/LAD level data for both previous and usual residence through special license access. With Secure Data Service access, anonymised postcodes are available but for the final quarter of 2009 only. A further limitation was raised by Dennett *et al*. (2007: 90) who noted that, when building an inter-regional flow matrix for counts of individuals derived from the LFS, the *"regional definitions are not constant between the origin (region of residence one year ago) and destination (region of current residence)"*. In one such case, the destination 'Rest of Northern Region' had no corresponding origin, with the closest origin being defined as 'Rest of North East'. As Dennett *et al*. (2007: 90) have asserted, *"[t]his could be dismissed as a labelling error were it not for the unusually high migration to the Northern Region from 'Rest of the North West' … As such it is impossible to tell for certain whether these differences in flows are to be relied upon as accurate differences, or rather the result of boundary change"*.

A survey that is closely related to the LFS is the Annual Population Survey (APS). The APS, published quarterly, is a continuous combined survey of households in GB and has been in

existence since 2004. The fundamental aim of the APS is to achieve a sample large enough to gather a minimum number of economically active respondents (510) in each LAD in England (except London boroughs where the target is 450), so as to produce more accurate attribute estimates at the sub-regional level (Werner, 2006; ONS, 2010b). In relation to the APS' design and its potential value for migration analysis, Cangiano (2010: 7) notes the following: *"The APS sample is obtained by merging waves one and five of four LFS quarters and data from the Annual Local (Area) Labour Force Survey (LLFS) Boosts for England, Scotland [SLFS] and Wales [WLFS]. There are approximately 350,000 individuals per dataset, which makes estimates based on the APS more robust than those obtained from a single LFS quarter"*.

The APS shares all the same characteristics that were highlighted above in relation to the LFS. For instance, the APS suffers from exactly the same issues associated with the omitting of communal establishments and, in terms of its application as a source for internal migration analysis, it is identical to the LFS (Rees *et al*., 2009). However unfortunately, those without special license access do not have the option to generate flow matrices at any geography as no origin identifier is included in the end user access datasets. That said, as Rees *et al.* (2009: 110) assert, *"[d]ata are readily available at GOR level with local authority origin and destination data available through a special licence"*. With many of the variables included in the survey being the same as those in the LFS, with special license access, the APS offers itself as a more robust data source for internal migration analysis, especially at the sub-regional level.

2.3.3 Understanding Society – UK Household Longitudinal Study (UKHLS – incorporating the BHPS)

Understanding Society (UKHLS) is a longitudinal multi-topic household study conducted by the Institute for Social and Economic Research (ISER) at the University of Essex. The overall purpose of the UKHLS, according to McFall (2011: 3), is "*to provide high quality longitudinal data about subjects such as health, work, education, income, family, and social life to help understand the long term effects of social and economic change"*. The study is unprecedented in its size with a UK-wide sample of approximately 40,000 households included in the first wave. Data collection for each wave is conducted over a 24 month period with collection for the first wave having started in January 2009 and ended in January 2011 (McFall, 2011). The overall sample size is made up of a number of smaller components including: the General Population Sample; the Ethnic Minority Boost Sample; the Innovation

Panel; and the BHPS Sample (Burton *et al.*, 2011). As is clear from the sample breakdown, the UKHLS incorporates, and indeed builds upon, the British Household Panel Survey (BHPS) that was phased out in 2011. By offering such a large sample, the UKHLS allows for researchers to gain greater insights into particular population sub-groups, such as teenage parents, older workers or the unemployed, that has not been possible in previous studies with smaller sample sizes (Bryan, 2011). Moreover, its UK-wide 40,000 household sample allows for more detailed geographical analysis across a number of spatial scales, depending on the dataset used (see below). It should also be noted that preparations for administrative data linkages are underway. During the first wave of interviews, each adult participant was asked to provide their consent for the UKHLS to link their survey data to health and education records. Furthermore, the study requested consent of parents to link health data on children aged 0-15 and education data on children aged 4-15. There are plans for further administrative data linkages including records of benefit receipt, participation in government employment schemes, savings and pensions, earnings and National Insurance contributions (Bryan, 2011). Clearly, once completed and made publically available, the data linkage would vastly increase the scope of the study.

The data from the UKHLS are available at varying levels of spatial detail, and can be accessed through ESDS. The basic end user license allows for analysis at GOR level, although analysis at LAD level is possible via the special license dataset. The highly restricted Secure Data Service access allows for any level of geography to be derived due to the availability of national grid references (easting and northing) on each record.

In terms of its value for migration analysis, the UKHLS offers a number interesting possibilities. Beyond its basic advantage as a longitudinal study, in that it follows its members from residence to residence, it includes questions on length of time at current address but also questions that have potential for the analysis of *future* migration propensities and *lifetime* migration propensities:

- If you could choose, would you stay here in your present home or would you prefer to move somewhere else?
- Do you expect you will move in the coming year?
- How many times have you moved to a new address since you were aged 14 (come to the UK to live) either on your own or with family?

These are questions that, aside from the ROP, are quite unique to this study and along with the sample size and potential data linkages, the UKHLS should be taken seriously as a future source of data for those interested in studying internal migration in the UK. Table 5 provides a summary of the characteristics of the major social surveys in the UK with a focus on their application in the analysis of migration/migrants.

2.4 Survey data and 'Beyond 2011'

It is now common knowledge that the ONS is looking for alternative methods to the traditional population census for the collection and provision of small area socio-demographic data. Indeed, it is the intention of the UK Statistics Authority (UKSA) that, given rising costs and an increasingly dynamic population, the 2011 Census will be the last of its kind (ONS, 2011b). As such, the ONS 'Beyond 2011' Programme has been running since April 2011 and, along with close collaboration with the statistics offices of Northern Ireland (NISRA) and Scotland (NRS), is tasked with investigating the alternative options for producing the population and socio-demographic data required by UK users (ONS, 2011b). While potential census based options will be assessed within the programme, the main focus will be on investigating the potential for combining existing administrative datasets with survey datasets, both public and commercial (ONS, 2011b; 2011e). In terms of potential sources of population data, the ONS has highlighted a number of candidates, including: NHS Central Register (NHSCR); Department for Work and Pensions/HM Revenue & Customs Customer Information System (CIS); electoral roll (18 years and over); School Census (5-16 years); HESA (students); birth and death registrations; and the Driving and Vehicle Licensing Agency (DVLA). For sources of socio-demographic data, the ONS will be largely reliant on large scale social surveys, however, further possibilities include: DVLA; utilities; TV licensing; and commercial sources (Calder and Swan, 2011).

Table 5: Overview of social survey data sources

Survey	Reporting period	Coverage	Lowest level of geography	Approx. Sample	Origin	Destin	Broad Themes	Download	Components
Integrated Household Survey	2009	UK	GOR (Special License: LAD)	450,000 people	✓	✓	Socio-economic & demographic characteristics	ESDS (OTHER, SPSS, STATA, TAB)	GLF (GHS), LCF (EFS), EHS, LOS, LFS/APS
Labour Force Survey	1973	UK	GOR (Special License: UA/LA)	100,000 people	✓	✓	Labour market characteristics & conditions	ESDS (SPSS, STATA, TAB)	
Annual Population Survey	Annual since 2004	UK	GOR (special license: UA/LA)	325,000 people	✓	✓	Same as LFS	ESDS (SPSS, STATA, TAB)	LFS, LLFS, WLFS, SLFS
General Lifestyle Survey	2008 - 2012	UK	Special License only: GOR	19,000 people		✓	Lifestyle & family information, socio-economic circumstances	ESDS (SPSS, STATA, SAS, TAB)	Formerly GHS
Family Resources Survey	1992	GB 1992 UK 2002	GOR (Special License: same)	75,000 people		✓	Housing, employment, income, general health	ESDS / DWP (SPSS, STATA, TAB)	Full integration of EU-SILC by 2015
English Housing Survey	2008	England	GOR (Special License: N/A)	41,000 people		✓	Housing circumstances & condition	ESDS (SPSS, STATA, SAS, TAB)	Formerly SEH & EHCS
Life Opportunities Survey	2009	GB	Special License only: LAD	23,000 people		✓	Social barriers to participation for disabled people	ESDS (SPSS, STATA, TAB)	
Living Costs & Food Survey	2008	UK	GOR (Special License: N/A)	12,000 people		✓	Household expenditure & socio-economic data	ESDS (SPSS, STATA, SAS, TAB)	Formerly EFS
Acxiom Lifestyle Survey	Current format 2004	GB	Postcode	750,000 people	✓	✓	Consumer spending, preferences, socio-demographic data	Acxiom	
Non-Census Longitudinal Studies									
British Household Panel Survey	1991 – 2011	UK (from wave 11 onwards)	GOR (Special Licence: CAS Wards) (Secure Data Service Access: easting & northing)	10,000 households		✓	Social & economic change	ESDS (SPSS, STATA, TAB)	
Understanding Society	First wave: 2009	UK	GOR (Special License: LAD & CAS wards) (Secure Data Access: easting & northing)	40,000 Households		✓	Socio-economic circumstances & attitudes	ESDS (SPSS, STATA, SAS, TAB)	BHPS, LSEM

With this in mind, it is clear that the Acxiom Lifestyle Survey may have potential as a data source in a post-census world. As a commercial survey, it offers a sample size that far outweighs anything seen in the official survey source datasets. Moreover, the lifestyle and socio-demographic information contained within it, is equally impressive. Thus, within the specific context of the on-going 'Beyond 2011' programme, it is hoped that the remaining sections of this paper coupled with the research undertakings of the project at large, will not only be useful in broadening the evidence base relating to our knowledge of migration in the UK, but also in benchmarking, validating and integrating the Acxiom ROP data with census, administrative, and other survey statistics.

3 Validation: Benchmarking the ROP against official statistics

This section of the paper is primarily concerned with detailing the validation exercises carried out on the ROP microdata in order to have confidence that the data reflect real characteristics and behaviour. These exercises can be split into three main subsections: aggregate level benchmarking; micro level benchmarking; and spatial benchmarking. However, to begin with, it is appropriate to present details of some of the significant data preparation and cleaning exercises that made the subsequent benchmarking procedures possible.

3.1 Data preparation and cleaning exercise

Excluding the responding household's postcode address, which is cleaned and imputed using the latest Postal Address File (PAF), Acxiom's ROP data are delivered in raw format (Thompson *et al.*, 2010). As such, concerns surrounding missing values and/or 'impossible' values are left for the end user to decide on. To get the most out of the data, such concerns require careful consideration. For instance, in the case of the year and/or month of move to the current address, one would expect some recall bias and therefore missing values may not mean that the individual has never moved, but that they cannot remember the exact year or, to a greater extent, month of move. Moreover, one would expect the likelihood of recall bias to increase with time spent at the current address. At a very basic level, there does appear to be some evidence of simple recall bias associated with this variable when we look at the response rate for the 'year of move' as opposed to the 'month of move' for the January 2005 ROP data. From a total sample of 411,325 records, there are just over 83,000 (20%) 'missing' records with no 'year of move' provided but more than 125,000 (30%) records with no 'month of move' provided.

A further variable that is central to this research project is the previous address. The Acxiom ROP provides a previous address postal in-code (postcode sector) and out-code (postcode district). For the January 2005 ROP data, there are just over 120,000 records (29%) with out-code responses and 103,000 records (25%) with in-code responses. Again, given that the variable is asking for the address of the respondents' previous residence, i.e. not restricted to a specific timeframe, there are concerns associated with recall bias. This is something that could be further exacerbated by the period of time spent at their previous residence, i.e. people who spent years at their previous address may find it easier to recall a full postcode address than those who spent just a short time there. Moreover, because this variable is not a simple 'yes-no' categorical variable, but one that is raw and open to around 2.5 million postcode combinations (Phelps, 2011), the variable requires a great deal of work before it can be deemed suitable for further analysis.

The first task is to check which postal out-codes are valid, which appear to be mis-specified and which are broken/incomplete. In this project, this was done by cross-referencing the response out-codes against a full list of postal areas used by the Royal Mail (Raper *et al.*, 1992) and included in the ONS National Statistics Postcode Directory (ONSPD). A set of codes were produced to indicate those that were usable, those that required cleaning/interpretation and those that were unusable (Table 6). Those that are delivered in a usable format were coded '1', those that only included the postal area (e.g. LS for Leeds) were coded '2', those where no information was recorded were coded '-99' as non-movers, and those that are unusable (e.g. including impossible combinations or characters) were coded '-9'. In addition to these basic codes, 19 additional codes (3-22) were produced, with reference to Raper *et al.* (1992) and the ONSPD, suggesting a requirement for further bespoke cleaning/interpretation.

Table 6: Codes deployed by authors in the cleaning/interpretation of Acxiom ROP previous address 'out-codes'

-99: Non movers	11: Interpreted, 1P (1P – IP for Ipswich)
-9: Moved but broken/wrong code	12: Interpreted, 1V (1V – IV for Inverness)
1: OK	13: Interpreted, L5 (L5 – LS for Leeds)
2: Only Postal Area	14: Interpreted, LV (LV - LU for Luton)
3: Interpreted, Extra 0 (e.g. B02 – B2)	15: Interpreted, P0 (P0 – PO for Portsmouth)
4: Interpreted, S0 (S0 to SO for Southampton)	16: Interpreted, T5 (T5 – TS for Cleveland)
5: Interpreted, 0L (0L – OL for Oldham)	17: Interpreted, W5 (W5 – WS for Walsall)
6: Interpreted, CU (CU – CV for Coventry)	18: Interpreted, Y0 (Y0 – YO for York)
7: Interpreted, 0X (0X – OX for Oxford)	19: Interpreted, CRO (CRO – CR0 for Croydon 0)
8: Interpreted, C0 (C0 – CO for Colchester)	20: Interpreted, B53, … (B53 – BS3 for Bristol)

9: Interpreted, HV (HV – HU for Hull)	21: Interpreted, WU (WU – WV for Wolverhampton)
10: Interpreted, 1G (1G – IG for Ilford)	22: Interpreted, ab1, … (ab1 – AB1 – this only affects Sept 2007 dataset)

The numbers involved in the cleaning and interpretation exercise for each of the three ROP datasets are summarised in Table 7, from which it is clear that the cleaning process has been quite successful in boosting the numbers of usable out-codes. Excluding those coded '2' (postal area only), the exercise has increased the number of usable out-codes by 4.7%, 9.8% and 25.8% for the January 2005, January 2007 and September 2007 datasets respectively. The September 2007 dataset was largely boosted by code '22', and it should be noted that records with this code would have been usable in their raw format. However, for reasons of consistency between the datasets it was thought important to convert these records too. That said, if one was to recode those records with a '22' code to a '1' code, the dataset would still have had an increase in usable out-codes of approximately 7.5%.

Table 7: Counts and percentages for codes used in the cleaning/interpretation of previous address out-code data

	January 2005		January 2007		September 2007	
Code	**Count**	**Percentage**	**Count**	**Percentage**	**Count**	**Percentage**
-9	3,626	3.24	6,809	8.41	7,317	5.74
1	107,019	95.61	73,171	90.38	95,445	74.83
2	337	0.30	123	0.15	786	0.62
3	92	0.08	147	0.18	141	0.11
4	101	0.09	69	0.09	51	0.04
5	12	0.01	15	0.02	17	0.01
6	38	0.03	42	0.05	32	0.03
7	11	0.01	14	0.02	29	0.02
8	102	0.09	55	0.07	65	0.05
9	10	0.01	21	0.03	11	0.01
10	18	0.02	11	0.01	8	0.01
11	30	0.03	21	0.03	20	0.02
12	7	0.01	12	0.01	5	0.01
13	44	0.04	41	0.05	30	0.02
14	20	0.02	18	0.02	18	0.01
15	196	0.18	124	0.15	74	0.06
16	24	0.02	27	0.03	18	0.01
17	27	0.02	52	0.06	9	0.01

18	112	0.1	122	0.15	71	0.06
19	32	0.03	20	0.02	36	0.03
20	55	0.05	26	0.03	30	0.02
21	18	0.02	23	0.03	12	0.01
22	0	0.00	0.00	0.00	23,332	18.29
Total Coded	111,931	100	80,963	100	127,557	100
Total Missing	299,394	-	268,625	-	253,685	-
TOTAL	411,325	-	349,588	-	381,242	-

With the cleaning and interpretation exercise complete, a number of more aggregate geographies were derived from the boosted usable out-codes for use in the benchmarking and validation process. This exercise was performed using the online geography matching and conversion tool GeoConvert. The aggregate geography that is used predominantly in this report is that of the local authority district (LAD). Unfortunately, postal geographies are not compatible with administrative boundaries and a large proportion of postal districts tend to straddle at least two LADs.

Thankfully GeoConvert's output file contains a column highlighting the proportion of the source geography that is contained in the target geography. Table 8 indicates that it would be fair to assume that those from the postal district AB32 are likely to be from LAD 00QB; however, a similar assumption for AB12 is harder to make.

Table 8: Example of GeoConvert postal district to LAD

Postal District	Proportion of source in target	Local Authority District
AB42	1.0000	00QB
AB12	0.6709	00QA
AB12	0.3291	00QB
AB10	1.0000	00QA
AB32	0.9992	00QB
AB32	0.0008	00QA
AB10	1.0000	00QA
KY12	1.0000	00QR

In this project, it was decided that for reasons of simplicity and concerns about available time, the selected LAD would be that with the largest proportion of source in target. However, such a methodology is particularly problematic for the most severe of cases. For

instance, as is shown in Table 9, there are a number of postal districts that straddle more than two LADs and consequently have no single target LAD with a proportion above a 0.5 level.

In the worst case example, the postal district N4 has been converted to LAD 00AP despite it only having a proportion of source in target of 0.38, narrowly larger than the second largest proportion of 0.33. However, on the whole, the approach taken here is deemed to be fair, especially given that only a small proportion of records suffer from the more severe problems noted here. That said, it should be noted that there are alternative methodologies including the use of population-weighted centroids that have the potential to offer more reliable conversions (Norman *et al.*, 2003; Norman and Riva, 2011).

Table 9: Source postal districts split below the 0.5 level for LAD target geography

Postal district	Selected LAD	Selected Proportion	Alternative LAD	Alternative Proportion
CO6	22UG	0.4906	22UC	0.3615
CV7	00CT	0.3769	44UC	0.3453
CW6	13UH	0.4379	13UB	0.3393
DE55	17UB	0.4914	17UC	0.3643
EX18	18UE	0.4663	18UD	0.3198
DY14	39UF	0.4876	47UG	0.2654
L10	00CA	0.4180	00BY	0.3523
LA6	16UG	0.4969	30UH	0.2597
MK17	11UB	0.4391	09UC	0.3042
N4	00AP	0.3755	00AM	0.3269
N11	00AC	0.4914	00AK	0.3232
PE6	32UG	0.4761	00JA	0.4032
PE7	12UD	0.4481	12UE	0.4040
PR4	30UF	0.4863	30UN	0.2542
SN4	46UC	0.4903	00HX	0.4724
SN6	00HX	0.4459	46UC	0.3063

With the Acxiom ROP data now containing codes for a number of aggregate geographies, there is a final requirement to define what constitutes a 'migrant' and what constitutes a 'non-migrant' before we are finally ready to embark on the validation exercise. While there are slight variations from study to study, such definitions are commonly operationalized by defining a migration as a geographical move that crosses a political/administrative boundary and results in a "*change in permanent residence, typically a year or more in duration*" (Frey, 2003: 545). Yet, when Fielding (forthcoming) refers to the term 'migration', he notes that he is referring to: "*[A] change of residence that is at least fairly permanent, and one that*

implies a move over a significant distance – one that is likely to involve a change of residential environment (such as a move to another town) and might also involve a change of employment (such as a job promotion)".

Indeed, as Fielding himself admits, the boundaries to his conceptualisation of 'migration' are not *"nice and sharp"* but instead rather *"fuzzy"* (Fielding, forthcoming). On a similar note, Rees *et al.* (2009: 64) argue that *"it is not useful to define a threshold distance below which migration is labelled residential mobility and above which it is labelled 'proper' migration, because such a threshold is arbitrary"*. However, in reality, when it comes to adopting an operational definition of what constitutes a migrant, the decision is largely influenced by the migration statistics at hand. For instance, the individual records in the PRDS collect information on the NHS patient as well as their home address at the postcode level. These details are updated annually with a migrant being defined as a *"person whose postcode changes between consecutive patient register downloads"* (Jefferies *et al.,* 2003: 5). However, in terms of practical definitions, taking into account of the geographies produced, a migrant is identified only as a person whose change in postcode takes them across either a former HA or LAD boundary (Jefferies *et al.,* 2003).

With this in mind four operational definitions of what are termed 'movers' and 'non-movers' were created with the numbers for each dataset shown in Table 10. The first definition is based on the response to a question asking for the year that the respondent moved to their current address, those that provided an answer were defined as a 'mover' and those who did not were classified as a 'non-mover'. For the second definition, those who provided a usable previous postal out-code, from which spatial aggregations and flows can be derived, were defined as 'movers' and those who did not, 'non-movers'. The third definition is derived from the first, but only includes those who moved in the 12 months prior to the survey date. Finally, the fourth definition is derived from the second and is again restricted to those who moved in the 12 months prior to the survey date. The third definition is used for the micro level benchmarking exercises in section 3.3 while the fourth definition is used in the aggregate level benchmarking (section 3.2) and the spatial benchmarking (section 3.4). Additionally, it should be noted that the actual numbers used in section 3.2 and 3.4 are slightly smaller than those of the fourth definition due to the exclusion of moves to and from Northern Ireland. The removal was based on the fact that the ROP is not representative of the Northern Ireland population (Thompson *et al.*, 2010). To give some idea of the change, the September 2007 ROP, the numbers fall from 12,584 'movers' to 12,417.

Table 10: Four mover non-mover definitions Acxiom ROP, January 2005, January 2007 and September 2007

	1. Ever moved based on year of move	2. Ever moved based on 'usable previous out-code'	3. Moved in last year based on 'year of move'	4. Moved in last year based on 'year of move' & 'usable previous out code'
Jan 2005 ROP				
Cases matching definition	328,158	107,967	17,435	12,232
Cases remaining (defined as non-movers)	83,167	303,358	393,890	399,093
Total Sample size	411,325	411,325	411,325	411,325
Jan 2007 ROP				
Cases matching definition	298,399	74,019	22,013	5,986
Cases remaining (defined as non-movers)	51,189	275,569	327,575	343,602
Total Sample size	349,588	349,588	349,588	349,588
Sept 2007 ROP				
Cases matching definition	269,338	119,225	17,456	12,584
Cases remaining (defined as non-movers)	111,904	262,017	363,786	368,658
Total Sample size	381,242	381,242	381,242	381,242

3.2 Aggregate level benchmarking: Local authority migration counts

While the aggregate level benchmarking has been carried out for all three ROP datasets, only those for the January 2005 data will be included here. The following aggregate level benchmarking exercise involves comparisons of total inflows over a 12 month period at the LAD level in GB. The sources chosen for the aggregate level benchmarking include the Census 2001 SMS, PR-NHSCR 2005 and the Acxiom Aggregate data. While the PR-NHSCR data can be considered as less reliable than the comparable Census 2001 data, it is thought important to include comparisons with this source as it is up-to-date. The Acxiom Aggregate data comparisons are also important because they could provide an indication of how much data manipulation has been undertaken by Acxiom in the production of this apparently representative aggregated data. The comparisons against Census 2001 and Acxiom Aggregate data include intra and inter-LAD flows. The Acxiom ROP intra-LAD flows were removed for comparisons against the PR-NHSCR data as this data source does not record intra-LAD

flows. The decision to focus solely on inflows in this section is based on the fact that the Acxiom Aggregate data only provide district inflow totals, thus making any alternative comparison impossible.

3.2.1 Validation against Census 2001 inflows

Figure 2 provides the basis for a number of interesting comparisons between the Census 2001 and January 2005 Acxiom in-migrant counts. The Acxiom numbers are, of course, far smaller than the census counts, with the distribution centred on roughly the 30 count mark compared with that of the 2001 Census which was closer to 15,000 persons. However, there is clear evidence in both distributions of positive skew and positive kurtosis. We can see that in both histograms there are a substantial numbers of LADs with low counts. The Q-Q plot compares each given value from the sample with the expected value that the score should have if it followed a normal distribution (Field *et al.*, 2012). In a case where data does in fact follow a normal distribution, the Q-Q plot would show a perfect diagonal line. With this in mind, when focusing on the Q-Q plot, the evidence of the non-normal distribution is further exemplified by the substantial clustering/lag towards the lower counts. Indeed, we should expect such characteristics with migration count data, given it is discrete count data and thus follows a discrete (Poisson) probability distribution. In terms of the scatterplot (Figure 3), the results are quite positive with 53.5% of the variation in Census 2001 inflows being explained by the January 2005 Acxiom inflows, furthermore the Pearson correlation coefficient is 0.73.

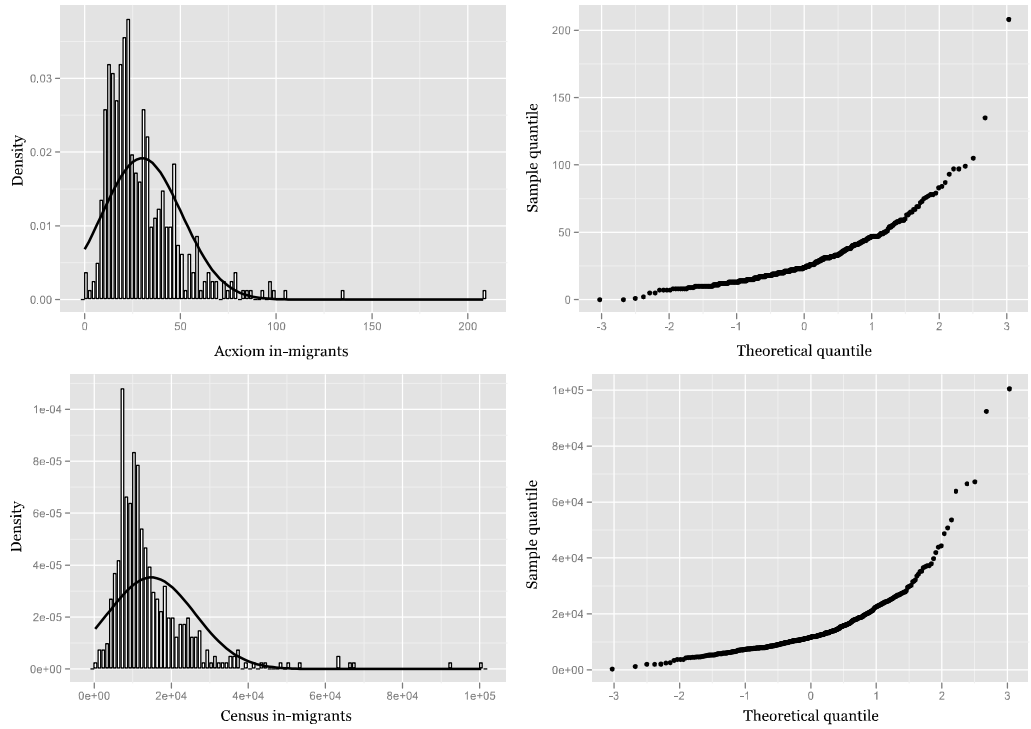


Figure 2: In-migrants to LADs data from Acxiom January 2005 ROP and Census 2001: Tests for normality

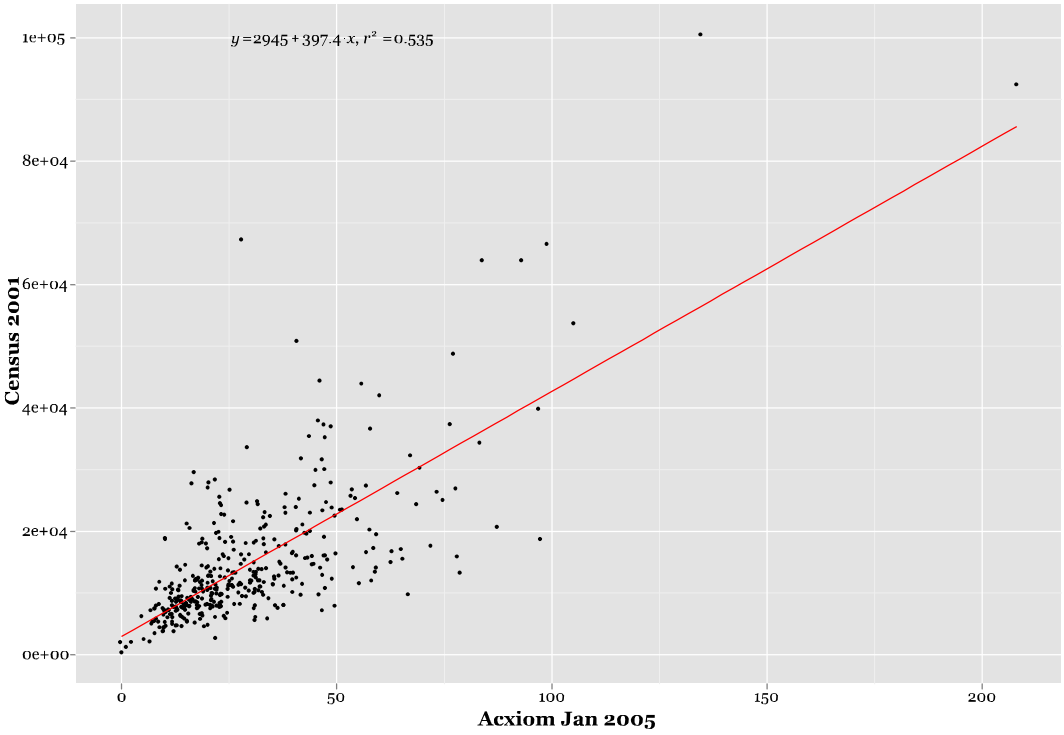


Figure 3: Scatterplot showing the relationship between Acxiom January 2005 ROP and Census 2001 in-migrants to LADs

Given that both 2001 Census and Acxiom ROP data are non-normally distributed, and more specifically suffer from significant positive skew, log10 transformations were performed with the results shown in Figures 4 and 5. It is clear from Figure 4 that the log10 transformations have greatly reduced the positive skew and kurtosis and Figure 5 suggests that we now have a better model fit with 55.9% of the variation explained. Interestingly, a number of extreme outliers, identified here as having a standardised residual of $\geq \pm 3$, can be identified despite the transformation. It is apparent that when compared to Census 2001, Acxiom's January 2005 ROP has a significant under-count for Glasgow City and significant over-counts for the Isles of Scilly and Berwick-upon-Tweed. Given the diverse nature of the three outliers it is doubtful that there is a general pattern to the under/over counts in the Acxiom sample, however, the respective population size of both Glasgow City and the Isles of Scilly (unique also in being an island LAD) will undoubtedly play some role here. Being purely speculative, it is possible that Berwick-upon-Tweed's location on the border between England and Scotland may also have some impact on the Acxiom sample's inflow undercount.

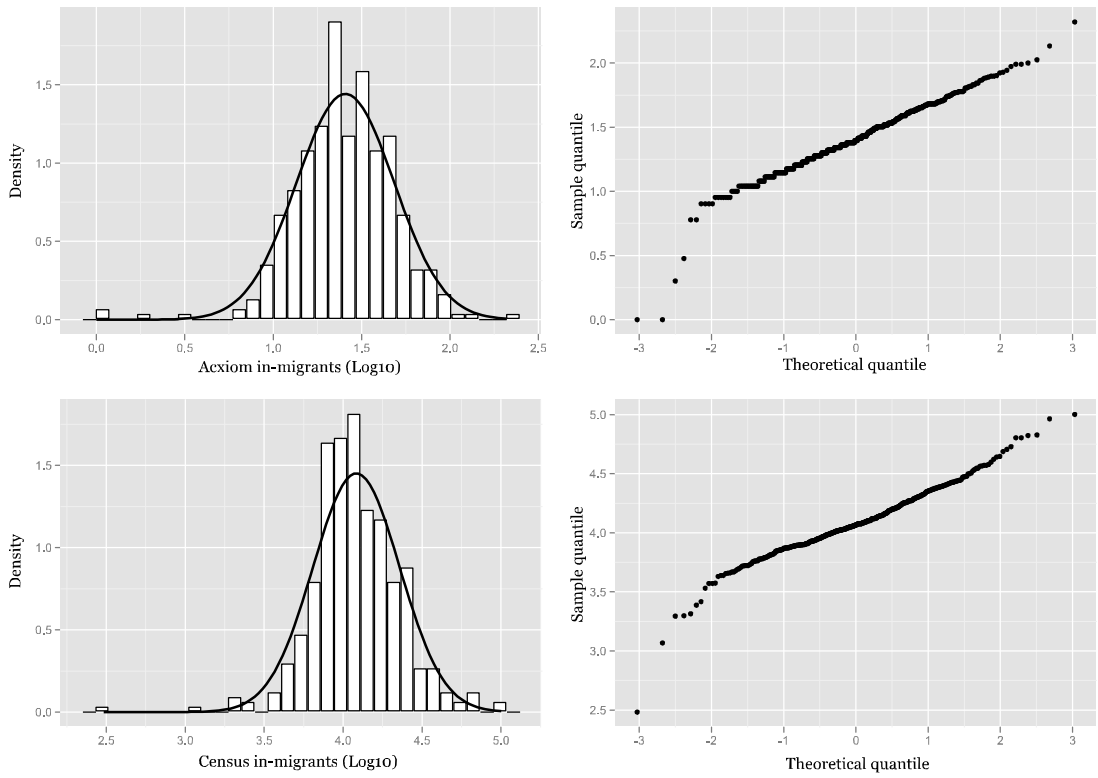


Figure 4: In-migrants to LADs from Acxiom January 2005 ROP (log10) and Census 2001 (log10): Tests for normality

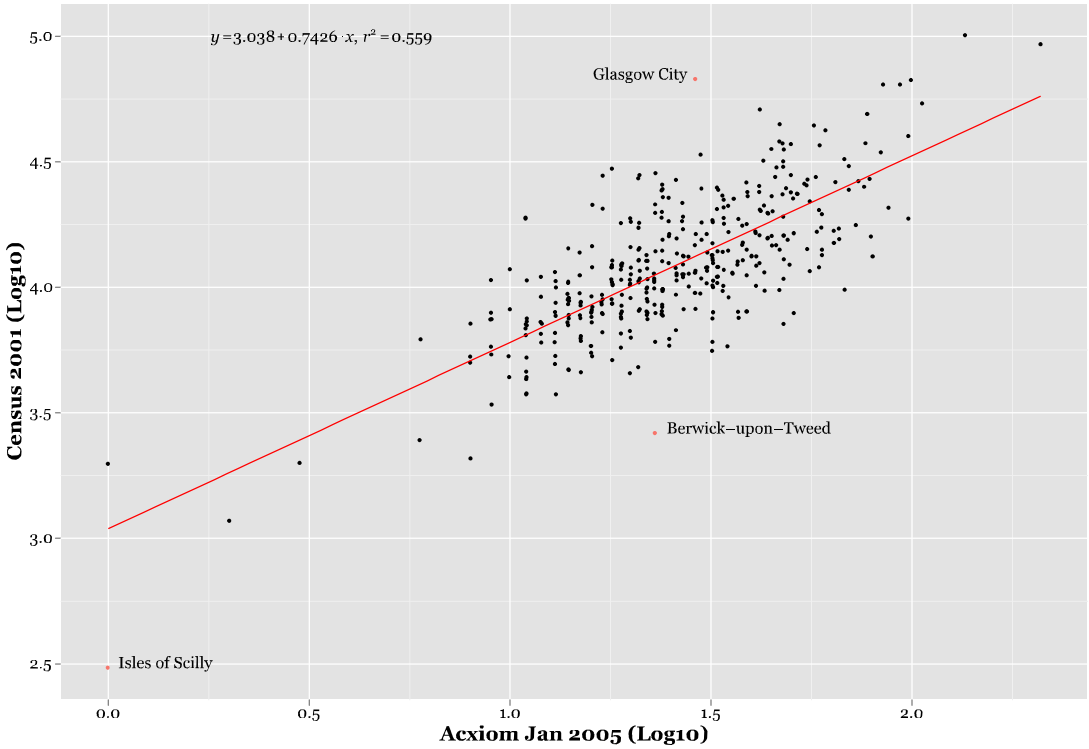


Figure 5: Scatterplot showing the relationship between Acxiom January 2005 ROP (log10) and Census 2001 (log10) in-migrants to LADs

3.2.2 Validation against PR-NHSCR

As was highlighted earlier, this comparison is based simply on in-migrant counts for LADs and therefore excludes those who moved within LADs. This has clearly had an impact on the total counts for each LAD with Figure 6 showing the January 2005 Acxiom data as having a distribution approximately centred on the 11 mark for inflows. Again, the histograms and Q-Q plots in Figure 6 suggest a degree of positive skew and kurtosis although to a lesser degree for Acxiom those seen above. Figure 7 shows an r^2 value of 0.381, suggesting that 38% of the variation in the PR-NHSCR inflows can be explained by their Acxiom ROP equivalents. The explanatory power of Acxiom *vis-à-vis* PR-NHSCR appears to have suffered somewhat from the reduced numbers associated with removal of intra-flows and the redefinition of what constitutes a 'migrant'. While the r^2 together with a Pearson correlation coefficient of 0.62 may be significantly lower than the comparable statistics for the census comparison, there is still evidence of a reasonable fit between the two data sources.

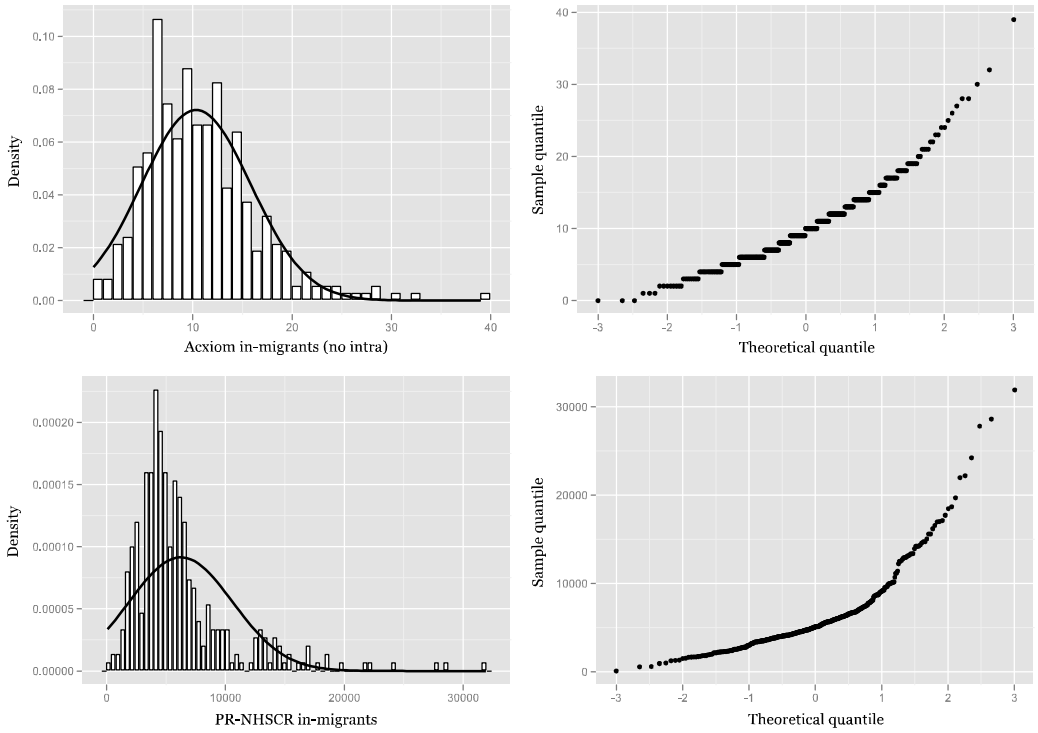


Figure 6: In-migrants to LADs from Acxiom January 2005 ROP and PR-NHSCR 2005: Tests for normality

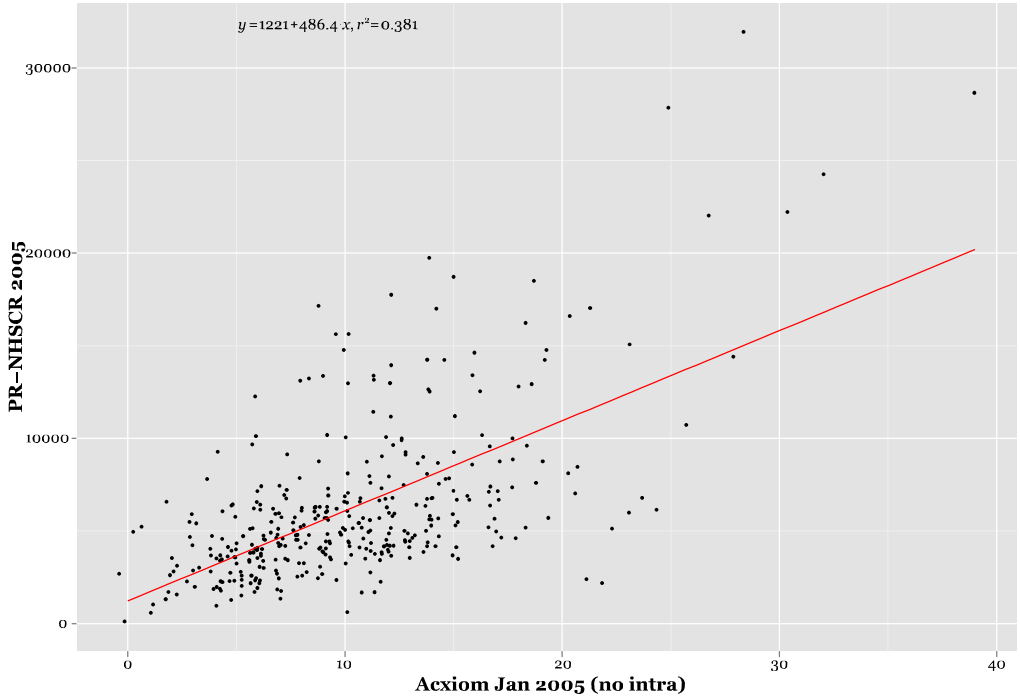


Figure 7: Scatterplot showing the relationship between Acxiom January 2005 ROP and PR-NHSCR 2005 in-migrants to LADs

As the distributions are again non-normal, transformations were applied. However, this time the lighter square root transformation was employed on the Acxiom data. Log10

transformations were applied to the PR-NHSCR data. While the transformations were successful in reducing positive skew and kurtosis (Figure 8), the model fit is marginally worse (Figure 9). However, the transformation does allow us to identify the extreme outliers with two familiar LADs appearing again, Berwick-upon-Tweed and the Isles of Scilly. These two are joined by Blyth valley, with all the outliers appearing to suggest an inflow over-count on the part of Acxiom's ROP sample.

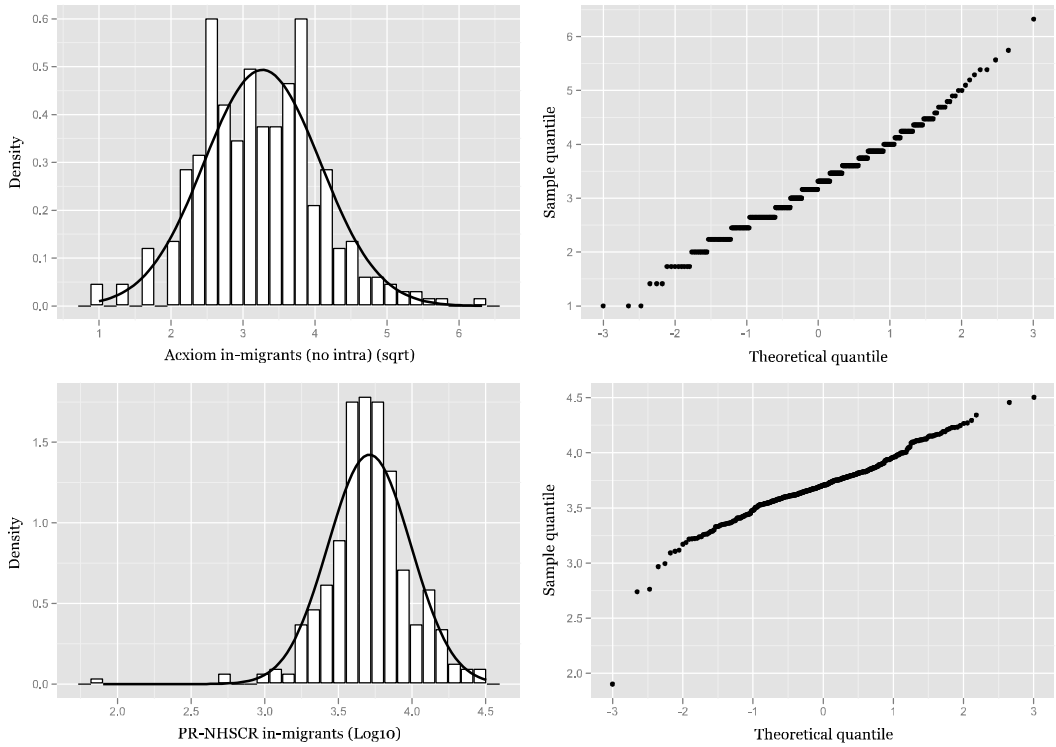


Figure 8: In-migrants to LADs from Acxiom January 2005 ROP (sq.rt.) and PR-NHSCR 2005 (log10): Tests for normality

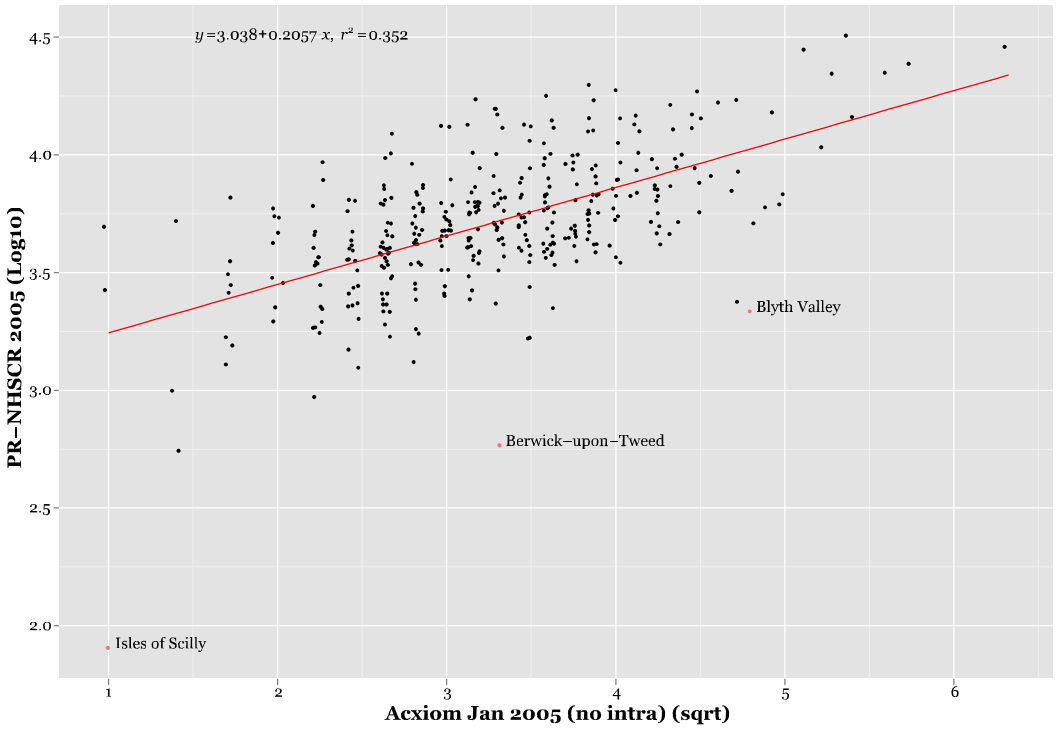


Figure 9: Scatterplot showing the relationship between Acxiom January 2005 ROP (sq.rt.) and PR-NHSCR 2005 (log10) in-migrants to LADs

3.2.3 Validation against Acxiom Aggregate Data

As was discussed in section 2.2, Acxiom produces annual aggregate data that are themselves derived from the Acxiom ROP microdata. The following checks are against Acxiom's Aggregate Data and should give a handle on how much data manipulation has been undertaken by Acxiom in the production of this apparently representative aggregated data. However, precise details of the actual processes of aggregation used by Acxiom are confidential. As Figure 10 shows, there are similarities in the distribution with both datasets showing signs of kurtosis and positive skew. In terms of the fit, Figure 11 suggests that 55.1% of the variation in Acxiom's Aggregate Data can be explained by the ROP microdata. Given that the Acxiom's Aggregate Data is derived from the microdata, it may at first sight be surprising that only 45% of the variation remains to be explained. However, according to Thompson *et al.* (2010: 28) the relatively low r^2 values are arguably an artefact of Acxiom's use of additional data sources *"in an attempt to remove bias in the sample and ensure that the [Aggregate Data] are in line with comparative packages such as the 2001 Census"*. Additionally, the relatively small numbers of 12 month movers with usable out-codes included in the ROP sample may play some part too.

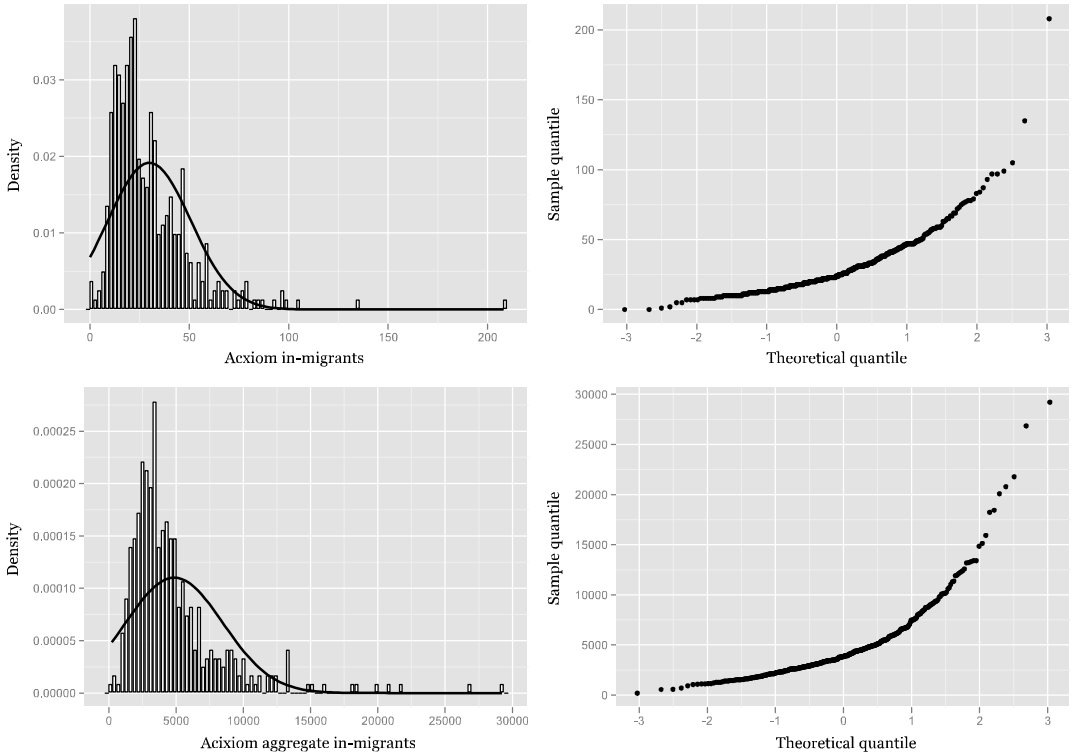


Figure 10: In-migrants to LADs from Acxiom January 2005 ROP and Acxiom Aggregate Data 2005: Tests for normality

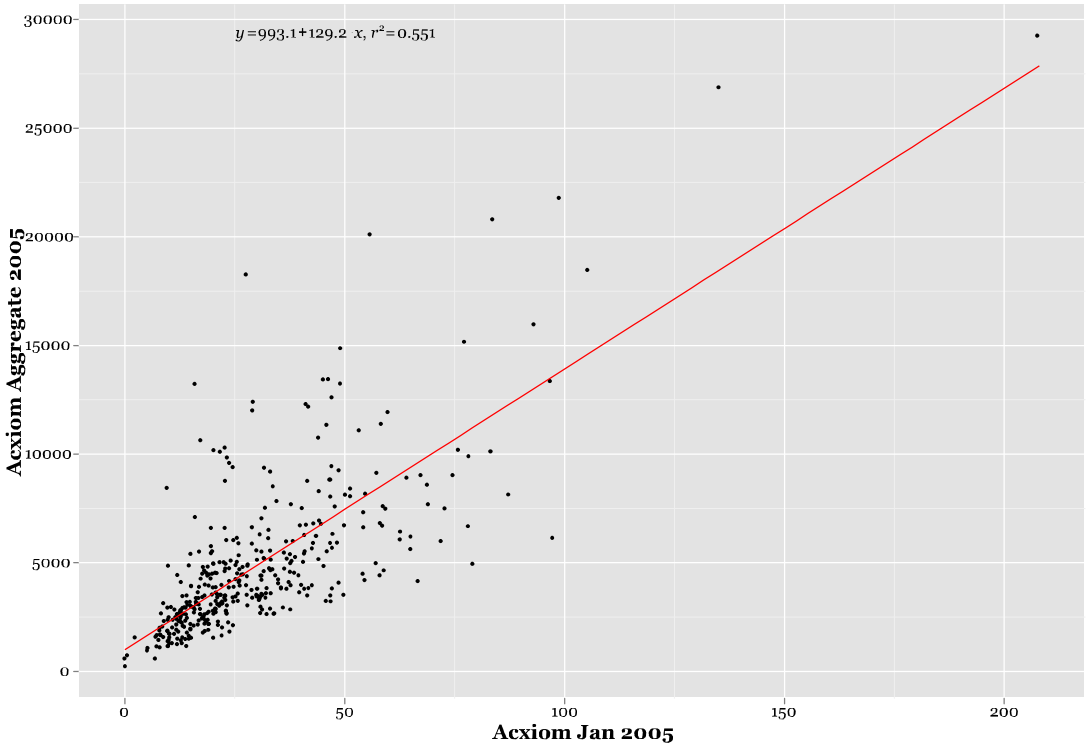


Figure 11: Scatterplot showing the relationship between Acxiom January 2005 ROP and Acxiom Aggregate Data 2005 in-migrants to LADs

As with the other comparisons the data were transformed, using log10 in both cases (Figure 12). Having transformed the data, Figure 13 suggests that the fit has improved with 60.5% of the variation in the aggregate data now explained. The outliers here are potentially important as they give some idea of the manipulation and weighting that Acxiom deemed necessary for its Aggregate Data. The outliers appear to suggest an undercount in the ROP microdata of three central London districts as well as Glasgow City. Aside from the recurrence of Glasgow City, the observed London outliers could be a result of the inherent biases of the ROP, namely those of an underrepresentation of young people and those in the higher income brackets (Thompson *et al.*, 2010).

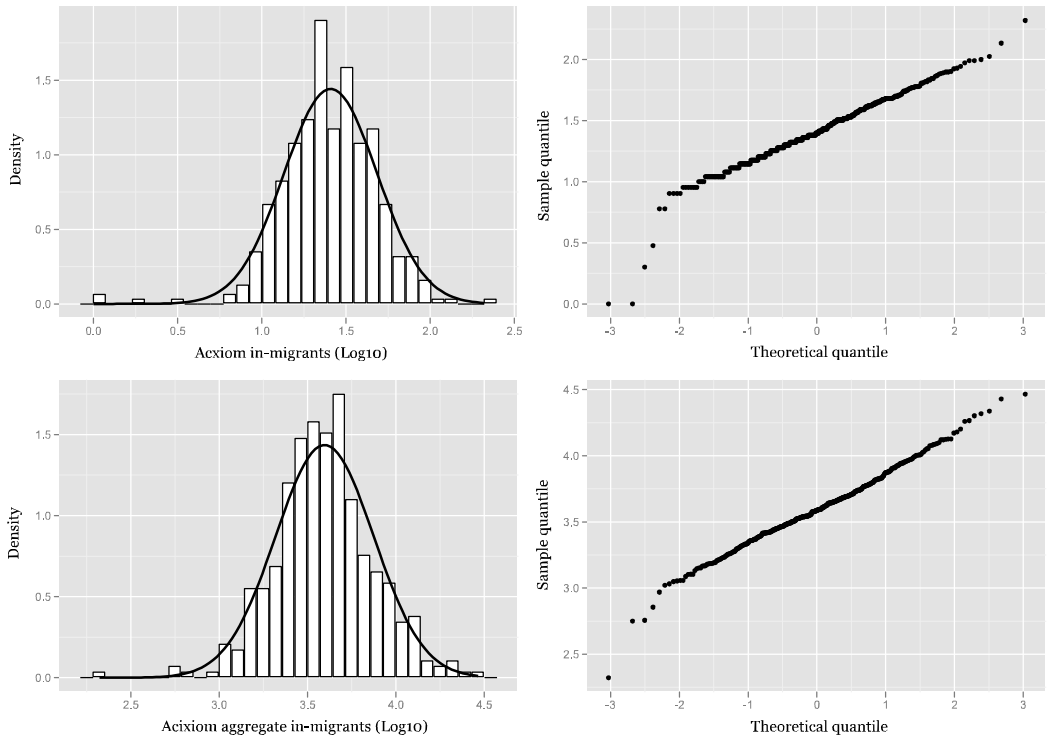


Figure 12: In-migrants to LADs from Acxiom January 2005 ROP (log10) and Acxiom Aggregate Data 2005 (log10): Tests for normality

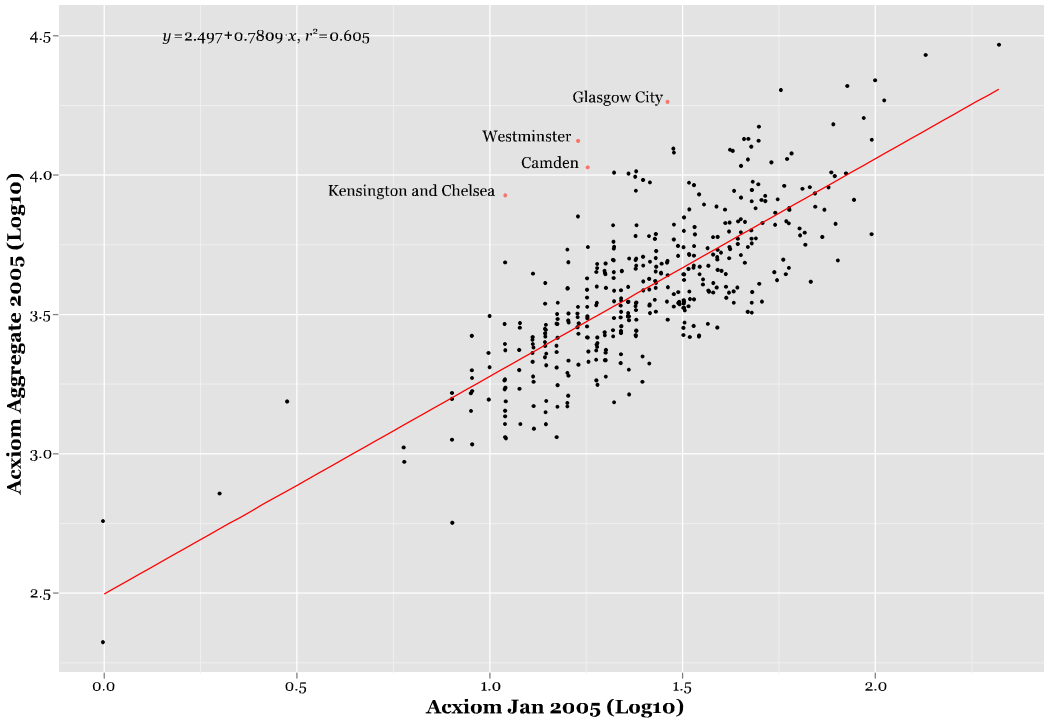


Figure 13: Scatterplot showing the relationship between Acxiom January 2005 ROP (log10) and Acxiom Aggregate Data 2005 (log10) in-migrants to LADs

3.3 Micro level benchmarking: selected variables

In the following sub-sections attention is focussed on three specially selected variables: age, ethnic group and tenure. While these variables were selected for a number of reasons, the main rationale relates to the fact that all three are commonly seen to hold a significant influence over migration patterns, propensities and trends. Furthermore, as has been observed in previous research (Thompson *et al.*, 2010), there are signs that the ROP struggles to capture certain population sub-groups including, for instance, younger people and ethnic minority groups. The alternative sources used in the sub-sections are the Census 2001 and the October to September 2005 APS. As has been mentioned before in this paper, whilst the 2001 Census represents the gold standard when we make our comparisons, the APS is also used as it is timely and offers both the variable detail and sample size required for the micro level benchmarking to be successful.

3.3.1 Age

Biological age alone has no influence over whether one is to migrate or not, but it does act as a rather consistent proxy for certain life course transitions that are known to increase/decrease the likelihood of making a residential move (Rogers and Castro, 1981; Bates and Bracken, 1987; Warnes, 1992; Champion *et al.*, 1998; Champion, 2005a; Stillwell, 2008; Dennett and

Stillwell, 2010). Using 2001 Census data (Figure 14) as a guide, it is clear that migration rates are relatively high for young children but then decline up to the school leaving age before then peaking in the early to mid-20s. What we are witnessing here are a number of commonly cited critical 'stages' within the life course. Firstly, the high rates for the youngest children can be attributed to parents' response to the arrival of a new child (be it the first or subsequent), for example seeking increased space and/or more appropriate environment (Bailey and Livingston, 2005). Following this, the decline is associated with the dependent child years. However, propensities to move increase to their highest point during the subsequent stage of the life course, roughly the years 18-25, wherein it is common for young adults to either move to university, employment or, subsequently, employment following university.

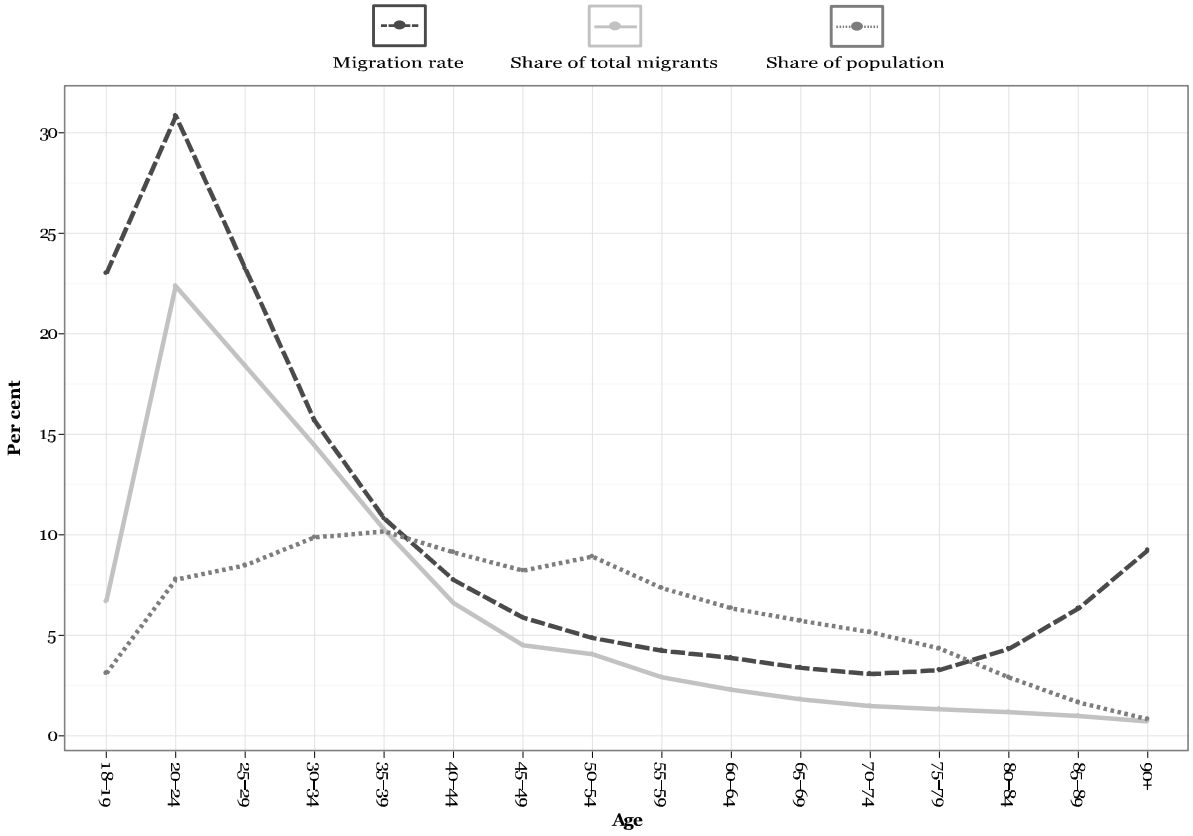


Figure 14: Census 2001 migration rates, migration shares and population distribution by age

The years from the mid-20s to the mid-40s are characterised by a relatively sharp reduction in migration rates and are generally considered the years of family formation and child rearing. The decline then reduces somewhat for the years 45-64, with recent research associating this shrinking in the decline with the transition from parenthood to 'empty nester', prompting the

desire, at least for some, to make a residential move (Wulff *et al*., 2010: 319). The decline finally levels out to a slight increase at 75+, an age commonly associated with a need for closer proximity to family members and services, given the greater requirement of help for the very elderly age groups.

When we compare the same statistics for the January 2005 ROP (Figure 15) and the October-September 2005 APS (Figure 16), we can observe a number of similarities. For both survey samples, we observe the highest propensities to migrate for those aged roughly 18-25; however when compared to Census 2001, the APS does appear to have higher and thus more comparable rates than does the ROP. When we focus on the older age groups, particularly those over 75, there does not appear to be the evidence of migration rates that we observe with 2001 Census. However, the ROP does record some increase for those in the oldest age group 90+.

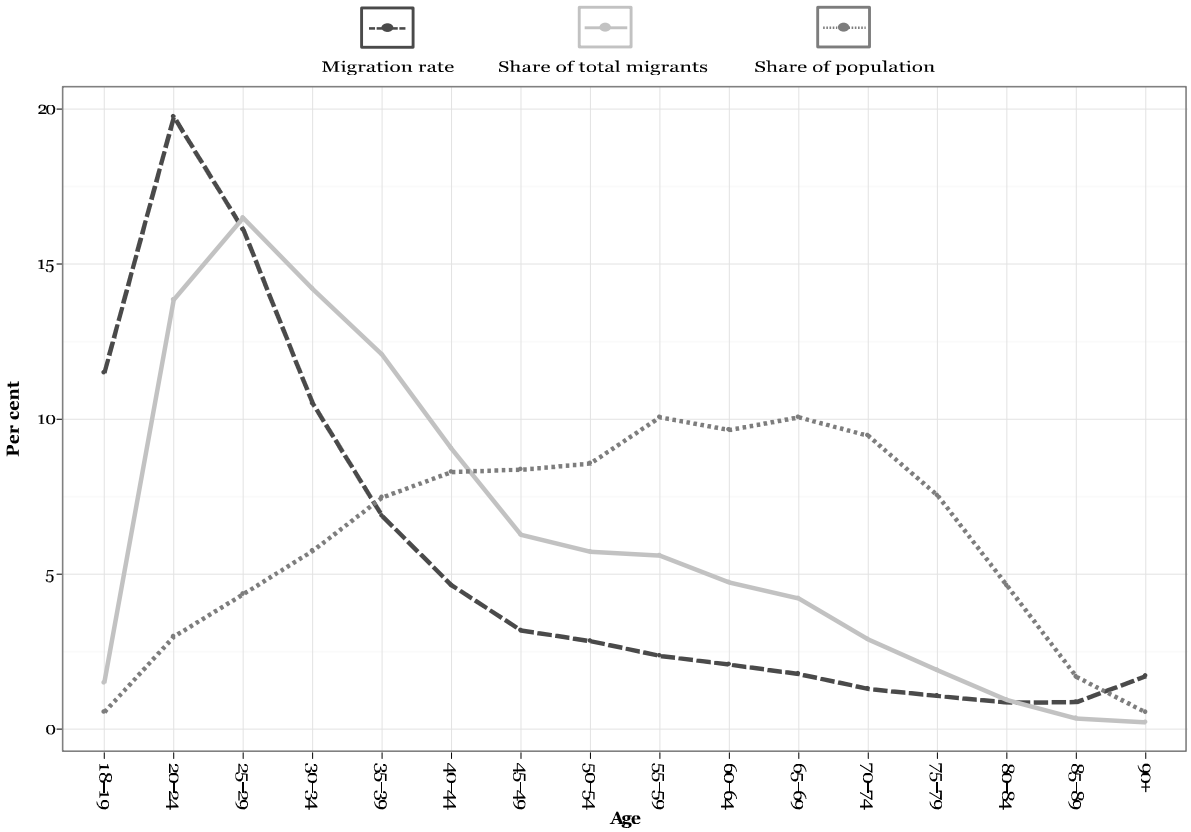


Figure 15: January 2005 ROP migration rates, migration shares and population distribution by age

Beyond the migration rates, the figures do suggest some degree of sample bias with an underrepresentation in the younger age categories, particularly in the ROP sample, and an overrepresentation of those aged 55-80. However, while these biases do exist in both survey

datasets, the general patterns, both in terms of migrant share and rates, remain fairly close to those observed in the 2001 Census.

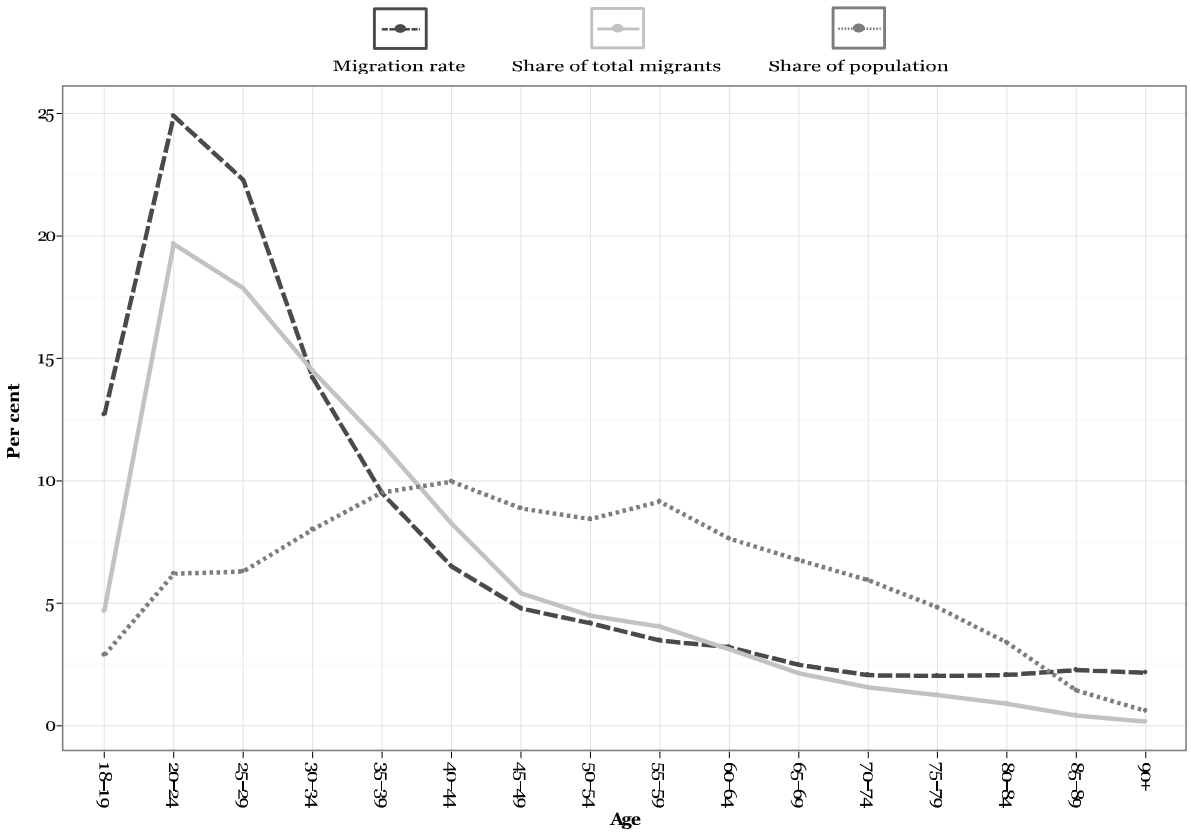


Figure 16: October-September 2005 APS migration rates, migration shares and population distribution by age

3.3.2 Ethnic group

The influence of ethnicity on migration propensities is a further topic of increasing interest (see Champion, 2005a; Stillwell and Duke-Williams, 2005; Large and Ghosh, 2006; Stillwell and Phillips, 2007; Simpson and Finney, 2009; Van Ham and Clark, 2009; Stillwell and Hussain, 2010). In terms of the relationship between ethnicity and migration, one is inclined to agree with Bailey and Livingston (2005) in stating that it is difficult to say what to expect. After all, *"some minority groups have a strong propensity to co-locate, for positive reasons (the importance of extended-family ties or access to particular amenities) as well as negative ones (a defensive reaction to racism or more limited options in housing or labour market terms)"* (Bailey and Livingston, 2005: 17). As research by Stillwell and Hussain (2010) has shown, total migration rates are higher for all non-White ethnic minorities, apart from the Indian population, than they are for the White majority. However, some of this can be explained by the demographic structures of the populations. Indeed, all ethnic minority groups have younger populations than the White majority and thus, given what was discussed

in the sections above, one would expect higher propensities for these populations. (Stillwell and Hussain, 2010: 1386). Beyond this, the variation between the ethnic minority groups is considerable. The Chinese population has been shown to have significantly higher migration rates than the Indian and POSA (Pakistani; Bangladeshi; Other Asian) populations (Stillwell and Hussain, 2010).

These general patterns can be observed in Figure 17 where, in order to make for easier comparison, the ethnic groups included in each dataset were aggregated into broader categorisations. In the case of the January 2005 Acxiom ROP, the original ethnic groups include: White; African; Pakistani; Chinese; Other Asian; Caribbean; Indian; Bangladeshi; and Other. However, the ethnicity question allows the respondent to tick as many boxes as apply to the respondent, thus opening up the potential for more detailed categorisations. Yet this flexibility can also be problematic, for instance it is hard to discern whether or not an individual with reported membership to three or more ethnic groups is genuine, or simply a wrongly specified record. As a result, for the figure below, those who reported three or more ethnicities were grouped, along with the 'Chinese' population, as 'Other'. Table 11 shows some general statistics for the January 2005 ROP, when comparing these against the same statistics for the 2001 Census SMS (Table 12), it is clear that, along with a large proportion of missing values, the ROP sample suffers from an underrepresentation of ethnic minorities. It is also somewhat of a concern that the ROP migration rates are much lower than those in the other sources. While one would expect bias in sample survey data (Crockett *et al.*, 2011), it is apparent that the ROP sample suffers from a greater underrepresentation of ethnic minorities than does the APS (Table 13). Thompson *et al*. (2010) found similar comparative weakness in the ROP ethnic group sample when checked against the EHS and LFS for the Yorkshire and Humber region. Such concerns with the sample could be lessened somewhat through a process of weighting the variables (Crockett *et al.*, 2011), and the potential for employing some form of spatial microsimulation technique (Harland *et al*., 2012) is certainly a possibility with the ROP microdata. Additionally, if internal validation checks between different ROP samples are successful, it would be possible to pool the different datasets together and create one large superset. Such an exercise could potentially boost the numbers of certain underrepresented subgroups and thus allow for more robust analyses of said groups.

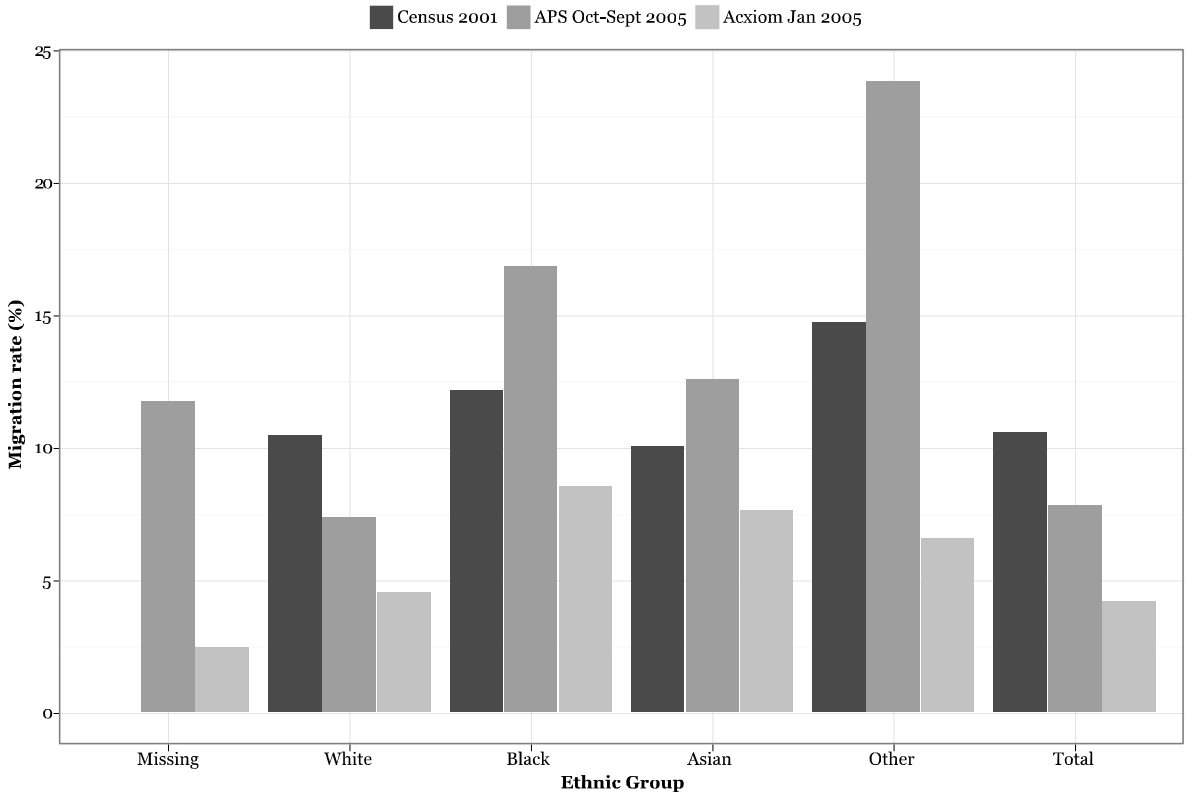


Figure 17: Comparison of migration rates by ethnic group

Table 11: Acxiom January 2005 ROP migration statistics for ethnic groups

	White	Black	Asian	Other	Missing	Total
Population	316,719	2,912	4,695	6,131	80,868	411,325
Share of population	77.00	0.71	1.14	1.49	19.66	100
Non-movers	302,313	2,663	4,335	5,727	78,852	393,890
Movers	14,406	249	360	404	2,016	17,435
Migration rate	4.55	8.55	7.67	6.59	2.49	4.24
Share of movers	82.63	1.43	2.06	2.32	11.56	100

Table 12: Census 2001 migration statistics for ethnic groups

	White	Black	Asian	Other	Missing	Total
Population	52,481,200	1,147,589	2,328,757	1,146,387	0	57,103,933
Share of population	91.90	2.01	4.08	2.01	0	100
Non-movers	46,970,538	1,007,778	2,093,682	977,267	0	51,049,265
Movers	5,510,662	139,811	235,075	169,120	0	6,054,668
Migration rate	10.50	12.18	10.09	14.75	0	10.53
Share of movers	91.02	2.31	3.88	2.79	0	100

Table 13: APS October-September 2005 migration statistics for ethnic groups

	White	Black	Asian	Other	Missing	Total
Population	370,378	5,873	10,359	4,172	212	390,994
Share of population	94.7	1.5	2.65	1.07	0.05	100
Non-movers	342,980	4,882	9,053	3,177	187	360,279
Movers	27,398	991	1,306	995	25	30,715
Migration rate	7.40	16.87	12.61	23.85	11.79	7.86
Share of movers	89.2	3.23	4.25	3.24	0.08	100

N.B those aged under 18 were removed for comparison reasons.

3.3.3 Tenure

Housing tenure has long been considered as having a significant impact on migration propensities (Hughes and McCormick, 1985; Buck, 1994). Traditionally, those living in privately rented accommodation tend to have a greater propensity to move than those in publically rented or privately owned accommodation. Reasons for this include the lower transaction costs associated with moves between privately rented accommodation coupled with the general trend of short-term leases within the private rental sector (Bailey and Livingston, 2005). Beyond this, there are compositional differences associated with those who rent publically, rent privately and own their home. For instance, as one would expect, private renters tend to have a younger age profile than owner occupiers. Additionally, Bailey and Livingston (2005: 11) warn of difficulties in interpreting the direction of causality between tenure and the propensity to migrate, suggesting that *"[c]hoice of tenure may be related to expectations about moving in the near future; if you think you are likely to be moving on, you are less likely to buy and more likely to rent so that renters tend to have higher migration rates"*. Still, while taking heed of such warnings, these longstanding tenure based variations in migration propensity are apparent in the Acxiom ROP data.

Table 14 is a contingency table comparing propensities to move between renters and home owners, the tenure categorisations of 'own home' and 'rent home' are aggregations of the original ROP categories examined in Table 16. Looking at the contingency table, we can see that for the 245,915 individuals who own their own home, just 3.3% moved in the last 12 months and 96.7% did not. However, for the 117,978 renters, 7.6% moved with 92.4% remaining *in situ*. The Pearson's Chi-squared test result is highly significant at the 99% level, indicating that there is a significant association between tenure and residential mobility. In terms of those who did make the move, renters represented 51.4% of the sample with owners representing 47.1%. This is in contrast to those who did not move, who were predominantly

home owners (60.4 %) with renters representing just 27.7% of the sample (Table 14). When we focus on the standardised residuals, we see that significantly more renters moved than we would expect (z = 56.02, p = 0.01) with significantly fewer remaining *in situ* (z = -11.79, p = 0.01). The opposite is the case for home owners where we observe significantly lower numbers of movers (z = -21.61, p = 0.01), and significantly more people remaining *in situ* (z = 4.55, p = 0.01), than we would expect. On the whole, we are seeing a significant difference between renters and home owners *vis-à-vis* the propensity to move.

Table 14: Acxiom January 2005 ROP move/not move by own/rent contingency table

	Not Moved	Moved	Row Total
Own Home			
Count	237,698	8,217	245,915
Row Per cent	96.66%	3.34%	59.79%
Column Per cent	60.35%	47.13%	
Std Residual	4.55	-21.61	
Rent Home			
Count	109,016	8,962	117,978
Row Per cent	92.40%	7.60%	28.68%
Column Per cent	27.68%	51.40%	
Std Residual	-11.79	56.02	
Missing			
Count	47,176	256	47,432
Row Per cent	99.46%	0.54%	11.53%
Column Per cent	11.98%	1.47%	
Std Residual	8.23	-39.13	
Column Total			
Count	393,890	17,435	411,325
Per cent	95.76%	4.24%	100.00%

Pearson's Chi-squared test: $\chi^2(2) = 8291.38$, $p < .001$

Comparably broad categorisations of 'own home' and 'rent home' were produced from the October-September 2005 APS (Table 15). Despite the fact that the APS has a far better response rate on the variable (only 0.1% 'missing'), broadly similar patterns can be observed in the APS and ROP samples. Renters in the APS sample record a significantly higher propensity to move, at 18.1%, than home owners (4.7%). Moreover, as with the ROP, the APS sample suggests that for those who did move, a small majority were renters (53.7%) with home owners representing 46.2%. Furthermore, the standardised residuals suggest the same significant directional patterns to those observed in the ROP data.

Table 15: October - September 2005 APS move/not move by own/rent contingency table

	Not Moved	Moved	Row Total
Own Home			
Count	285,236	14,198	299,434
Row Per cent	95.26%	4.74%	76.58%
Column Per cent	79.17%	46.23%	
Std Residual	17.75	-60.80	
Rent Home			
Count	74,746	16,503	91,249
Row Per cent	81.91%	18.09%	23.34%
Column Per cent	20.75%	53.73%	
Std Residual	-32.19	110.26	
Missing			
Count	297	14	311
Row Per cent	95.50%	4.50%	0.08%
Column Per cent	0.08%	0.05%	
Std Residual	0.62	-2.11	
Column Total			
Count	360,279	30,715	390,994
Per cent	92.14%	7.86%	100.00%

Pearson's Chi-squared test: $\chi^2(2) = 17208.93$, $p < .001$

An advantage of the Acxiom ROP is its detailed breakdown of tenure, and specifically of the renter bracket. Indeed, the data allows us to compare renters' propensities to move based on whether they are private, housing association, or council based. The contingency table below (Table 16), which again has a significant Pearson's Chi-squared test result (p <0.01), provides a more detailed breakdown of the renter bracket. It is apparent that the group with the highest propensity to move is private renters (11.6%), where the standardised residuals suggest that significantly more private renters moved than we would expect (z = 75.89, p = 0.01) with significantly fewer remaining *in situ* (z = -15.97, p = 0.01). Likewise, both council and housing association tenants had higher than expected propensities to move although to a lesser extent than those who rent privately. On the other hand, home owners were represented significantly lower numbers of movers (z = -21.61, p = 0.01) and significantly higher numbers of non-movers (z = 4.55, p = 0.01) than would be expected. Reassuringly, these results appear to support the assertions made above, namely that those living in privately rented accommodation tend to have a greater propensity to move than those in publically rented or privately owned accommodation.

Table 16: Acxiom January 2005 ROP move/not move by various tenure contingency table

	Not Moved	Moved	Row Total
Own Home			
Count	237,698	8,217	245,915
Row Per cent	96.66%	3.34%	59.79%
Column Per cent	60.35%	47.13%	
Std Residual	4.55	-21.61	
Rent (Council)			
Count	48,520	2,451	50,971
Row Per cent	95.19%	4.81%	12.39%
Column Per cent	12.32%	14.06%	
Std Residual	-1.32	6.25	
Rent (Housing Association)			
Count	21,039	1,317	22,356
Row Per cent	94.11%	5.89%	5.44%
Column Per cent	5.34%	7.55%	
Std Residual	-2.53	12.00	
Rent (Private)			
Count	39,457	5,194	44,651
Row Per cent	88.37%	11.63%	10.86%
Column Per cent	10.02%	29.79%	
Std Residual	-15.97	75.89	
Missing			
Count	47,176	256	47,432
Row Per cent	99.46%	0.54%	11.53%
Column Per cent	11.98%	1.47%	
Std Residual	8.23	-39.13	
Column Total			
Count	393,890	17,435	411,325
Per cent	95.76%	4.24%	100.00%

Pearson's Chi-squared test: $\chi^2(4) = 8291.38$, $p < .001$

3.4 Spatial benchmarking

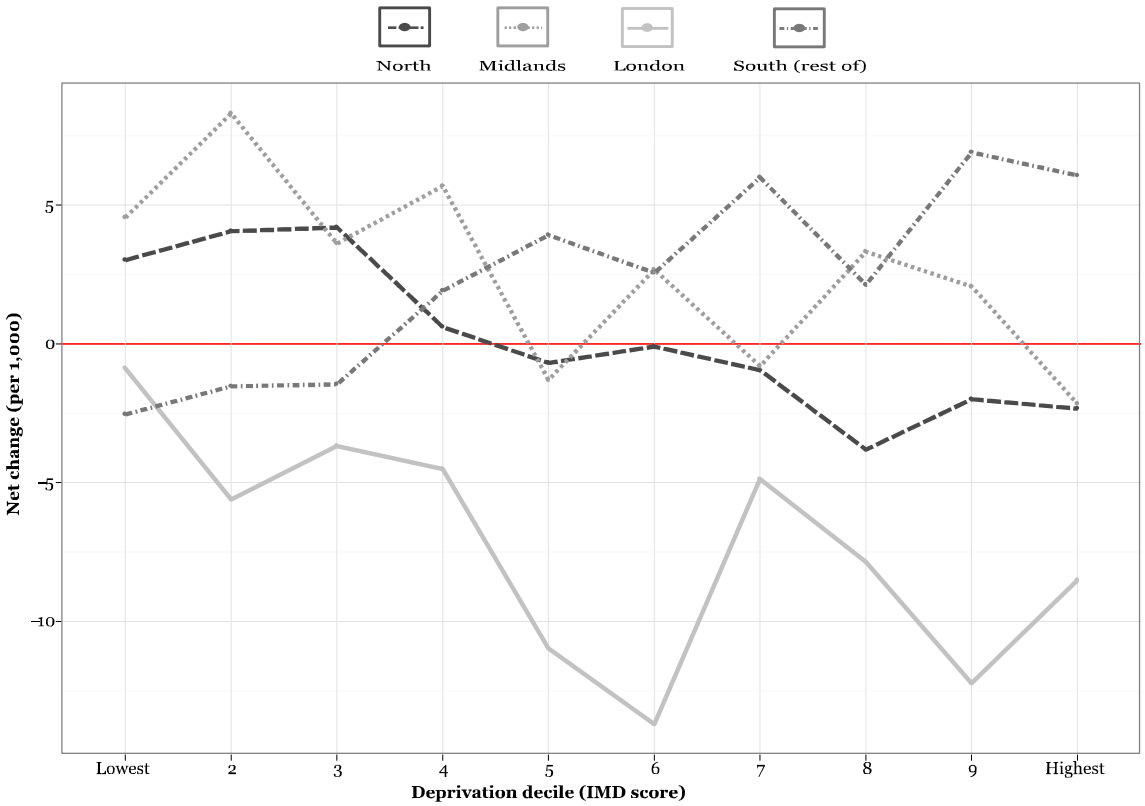
As was mentioned in section 2, when it comes to the spatial detail allowed for, the ROP has a major advantage over its rival surveys. Not only is it possible to generate any geography from the postcode level data, the large sample size makes reasonably reliable analysis possible at the district level (and below). Consequently, in order to benchmark the spatial elements of the ROP, the following section will look at the relationship between area level deprivation and net migration rates (per 1,000) at the district level for England using Census 2001, PR-NHSCR 2005, and January 2005 ROP. The Acxiom ROP sample includes those who moved

within England in the 12 months prior to the survey date (January 2005) and for whom we have usable origin and destination identifiers at the LAD level (n = 10,424). The area level deprivation measure deemed most suitable for use here is the Index of Multiple Deprivation 2004 (IMD2004) with the district scores being the population weighted average of the combined IMD scores for the Super Output Areas (SOAs) in a district (ODPM, 2004). Employing a similar methodology to that used by Bailey and Livingston (2008: 950), the districts were collected into four broad regions in an attempt to maximise the differences in labour and housing market context. These regions include the North (North-East, North-West, and Yorkshire and the Humber), the Midlands (West Midlands and East Midlands), London, and the remainder of the South (East, South-East, and South-West). The districts within each region were ranked into equal deciles (based on the number of LADs) to avoid a concentration of deprivation in the North, the decile averages are presented in Table 17.

Table 17: Region average IMD 2004 score for each deprivation decile

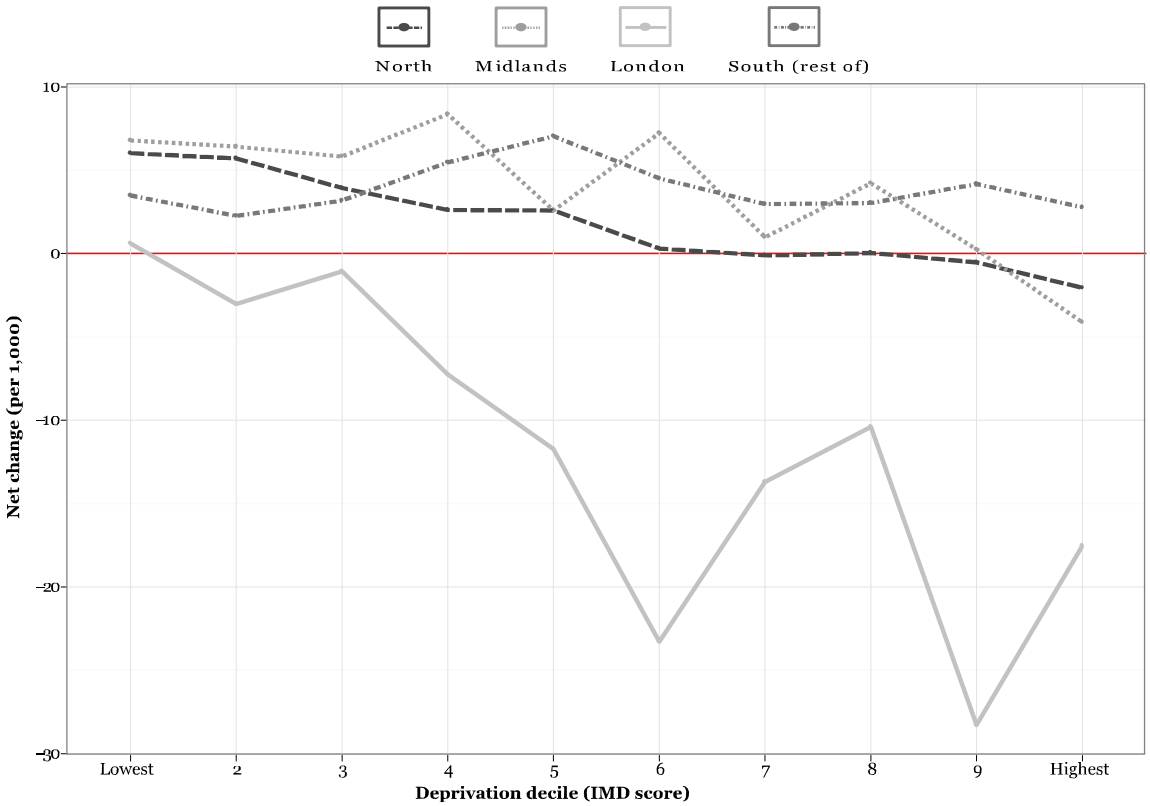
	Lowest	2	3	4	5	6	7	8	9	Highest
North	11.31	14.78	17.61	21.16	24.70	27.46	29.32	31.66	33.65	43.09
Midlands	8.34	11.09	12.46	15.02	16.20	17.72	19.59	23.07	28.26	35.36
London (city region)	11.94	14.09	15.29	19.05	22.60	25.72	30.40	33.52	37.83	44.53
South (rest of)	6.20	8.03	9.19	10.45	12.10	14.14	16.34	19.06	21.65	25.87

According to Figures 18-20, we appear to be observing net migration movements from districts in the most deprived decile to those above them. All three data sources show London has significant net losses for most areas, this is a familiar observation (Duke-Williams and Stillwell, 2010) but the general pattern of greater net losses in the more deprived districts still follows. On the other hand, for the 2001 Census and Acxiom January 2005 ROP, it is clear that the South sees a shift of people in the opposite direction with net losses in the least deprived districts and net gains in the more deprived areas. These findings are consistent with Bailey and Livingston's (2008: 953) analysis, who suggested that such patterns are likely to be attributed to the tight housing markets in this region which have led to demand even for the most deprived neighbourhoods.



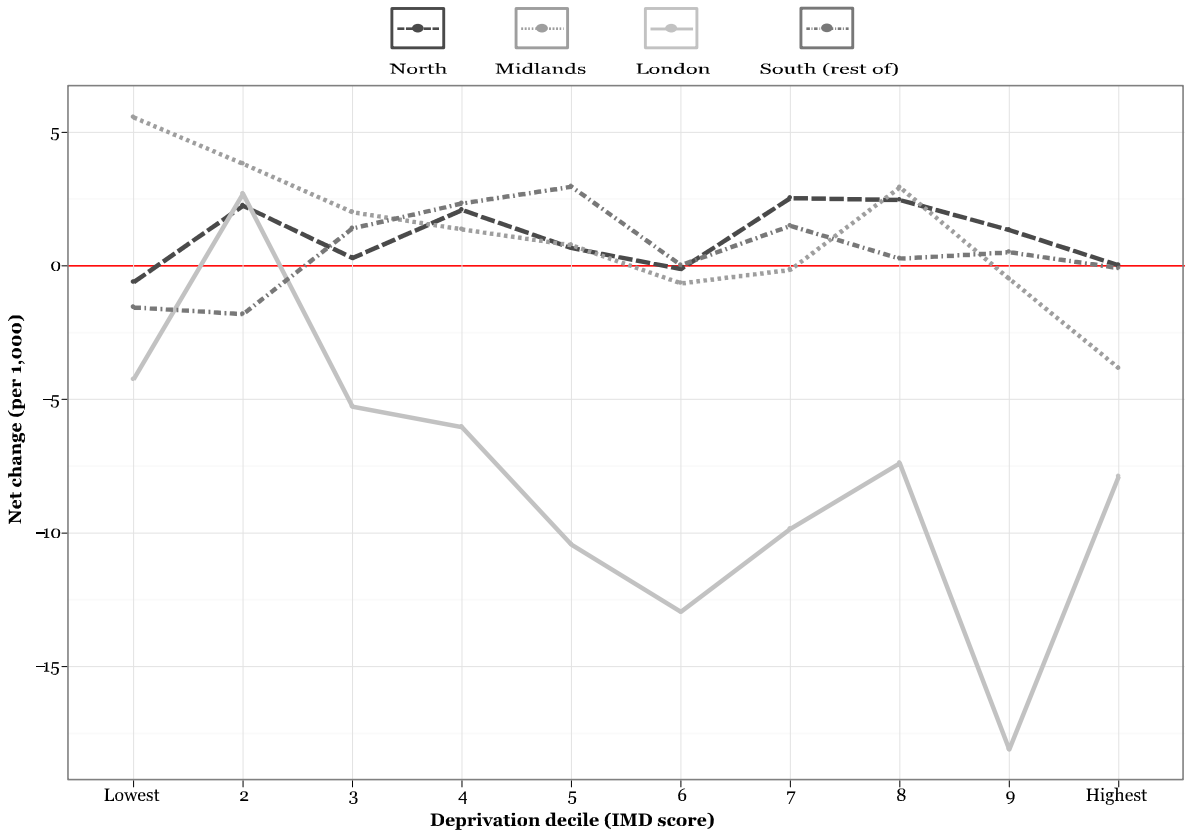
Data: Census 2000-01 SMS inter-district moves.

Figure 18: Census 2001 net migration change by IMD decile at regional level



Data: PR-NHSCR 2004-05 inter-district moves & ONS Mid-year estimates 2004.

Figure 19: PR-NHSCR 2005 net migration change by IMD decile at regional level



Data: Acxiom January 2005 ROP

Figure 20: Acxiom January 2005 ROP net migration change by IMD decile at regional level

Beyond this, Figures 18 and 20 appear to suggest an interesting phenomenon wherein the least deprived decile in the North, Midlands (less so for the ROP data) and South is characterised by either migration rates close to zero or indeed a drop-off of net rates compared to the second least deprived decile. While this may suggest that areas with the least deprivation are more stable, it could also be attributed to the fact that the first decile represents a ceiling in the data wherein there is no possibility for movement beyond. Moreover, across all three data sources, the gross components of the flows suggest that on the whole we see an increase in total migration counts as we move up the deprivation deciles, however, this is not the case for London (Tables 18-20).

In terms of migration effectiveness (Tables 18-20), all data sources suggest that the most deprived deciles have negative ratios suggesting that the majority of migrants have moved out and few have moved in. In the Midlands and the London the values are large suggesting that migration is working to significantly redistribute the population in these regions, thus producing large net effect relative to the volume of migrants. Moreover, the South has relatively low effectiveness values across the deciles suggesting that migration is 'inefficient'

as a mechanism for population redistribution in this region, again possibly being associated with the tight housing markets.

Bringing both the figures and tables in this section together, the results are quite positive considering the Acxiom January 2005 ROP picks up the general patterns observed in the PR-NHSCR 2005 and Census 2001 data. Moreover, if the 2001 Census can be considered as the optimum point of reference for those interested in population statistics and small area analysis (Raymer *et al.*, 2012), it is encouraging to see the ROP sample being more successful in picking up many of the patterns, for instance the reversal of the general shift for the South, than the PR-NHSCR.

Table 18: Census 2001 measures of migration (average) by region and deprivation decile

North	**Lowest**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Highest**
In-migration count	8982.3	11117.7	12182.7	10932.6	11641.6	31905.1	19798.1	17275.1	18296.4	25858.5
Out-migration count	8849.3	10558.9	11891.1	10840.0	11700.1	31696.1	19932.6	17927.1	18621.8	26182.3
Gross-migration count	17831.7	21676.6	24073.8	21772.6	23341.7	63601.2	39730.8	35202.2	36918.2	52040.8
Net-migration count	133.0	558.8	291.6	92.6	-58.6	209.0	-134.5	-652.0	-325.3	-323.8
Migration effectiveness	0.7	2.6	1.2	0.4	-0.3	0.3	-0.3	-1.9	-0.9	-0.6
Midlands	**Lowest**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Highest**
In-migration count	7320.5	8977.1	10821.4	8764.7	9148.7	8061.8	10923.3	12011.8	16996.4	36407.9
Out-migration count	6969.5	8263.9	10360.1	8247.0	9305.6	7863.5	10987.4	11708.0	17028.3	37361.4
Gross-migration count	14290.0	17241.0	21181.5	17011.7	18454.3	15925.3	21910.7	23719.8	34024.7	73769.3
Net-migration count	351.0	713.3	461.3	517.7	-156.9	198.3	-64.1	303.8	-31.9	-953.6
Migration effectiveness	2.5	4.1	2.2	3.0	-0.8	1.2	-0.3	1.3	-0.1	-1.3
London	**Lowest**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Highest**
In-migration count	20421.8	13068.0	21502.7	29868.5	22325.0	27414.0	21634.5	30705.3	27127.3	22877.0
Out-migration count	20655.0	14348.0	22474.7	31116.8	24552.7	30772.0	22732.0	32188.0	29989.3	24471.7
Gross-migration count	41076.8	27416.0	43977.3	60985.3	46877.7	58186.0	44366.5	62893.3	57116.7	47348.7
Net-migration count	-233.3	-1280.0	-972.0	-1248.3	-2227.7	-3358.0	-1097.5	-1482.7	-2862.0	-1594.7
Migration effectiveness	-0.6	-4.7	-2.2	-2.0	-4.8	-5.8	-2.5	-2.4	-5.0	-3.4
South (rest of)	**Lowest**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Highest**
In-migration count	10422.8	11727.3	11180.5	10040.8	10307.5	12817.7	10614.2	13790.1	11766.4	20725.4
Out-migration count	10643.9	11886.2	11317.8	9884.3	9917.4	12459.8	10127.1	13565.9	11118.0	19834.0
Gross-migration count	21066.7	23613.5	22498.3	19925.1	20224.9	25277.5	20741.2	27355.9	22884.4	40559.4
Net-migration count	-221.1	-158.9	-137.3	156.5	390.1	357.9	487.1	224.2	648.4	891.4
Migration effectiveness	-1.0	-0.7	-0.6	0.8	1.9	1.4	2.3	0.8	2.8	2.2

Table 19: PR-NHSCR 2005 measures of migration (average) by region and deprivation decile

North	**Lowest**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Highest**
In-migration count	3562.2	4540.0	4672.2	3793.8	3661.1	9407.8	5228.8	4716.7	5296.7	8516.3
Out-migration count	3078.9	3875.6	4288.9	3505.0	3368.9	9420.0	5176.3	4930.0	5494.4	9273.8
Gross-migration count	6641.1	8415.6	8961.1	7298.8	7030.0	18827.8	10405.0	9646.7	10791.1	17790.0
Net-migration count	483.3	664.4	383.3	288.8	292.2	-12.2	52.5	-213.3	-197.8	-757.5
Migration effectiveness	7.3	7.9	4.3	4.0	4.2	-0.1	0.5	-2.2	-1.8	-4.3
Midlands	**Lowest**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Highest**
In-migration count	4041.3	4358.6	4785.0	3684.3	4448.6	3317.5	4341.4	4258.8	6070.0	13137.1
Out-migration count	3522.5	3757.1	4163.8	2927.1	4231.4	2742.5	4280.0	3838.8	6350.0	15111.4
Gross-migration count	7563.8	8115.7	8948.8	6611.4	8680.0	6060.0	8621.4	8097.5	12420.0	28248.6
Net-migration count	518.8	601.4	621.3	757.1	217.1	575.0	61.4	420.0	-280.0	-1974.3
Migration effectiveness	6.9	7.4	6.9	11.5	2.5	9.5	0.7	5.2	-2.3	-7.0
London	**Lowest**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Highest**
In-migration count	11502.5	8463.3	11956.7	16967.5	12046.7	14330.0	12770.0	18766.7	15336.7	13503.3
Out-migration count	11262.5	9293.3	12463.3	19082.5	14590.0	20186.7	15720.0	21423.3	22206.7	17090.0
Gross-migration count	22765.0	17756.7	24420.0	36050.0	26636.7	34516.7	28490.0	40190.0	37543.3	30593.3
Net-migration count	240.0	-830.0	-506.7	-2115.0	-2543.3	-5856.7	-2950.0	-2656.7	-6870.0	-3586.7
Migration effectiveness	1.1	-4.7	-2.1	-5.9	-9.5	-17.0	-10.4	-6.6	-18.3	-11.7
South (rest of)	**Lowest**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Highest**
In-migration count	6006.9	6184.4	5876.3	4975.6	4991.3	6135.6	4604.1	5881.3	4690.0	8027.5
Out-migration count	5643.1	5925.0	5553.8	4434.4	4360.6	5506.3	3995.3	5670.7	4399.4	7795.6
Gross-migration count	11650.0	12109.4	11430.0	9410.0	9351.9	11641.9	8599.4	11552.0	9089.4	15823.1
Net-migration count	363.8	259.4	322.5	541.3	630.6	629.4	608.8	210.7	290.6	231.9
Migration effectiveness	3.1	2.1	2.8	5.8	6.7	5.4	7.1	1.8	3.2	1.5

Table 20: January 2005 ROP measures of migration (average) by region and deprivation decile

North	**Lowest**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Highest**
In-migration count	16.0	25.3	21.9	23.4	35.6	64.4	46.9	36.3	50.9	38.5
Out-migration count	16.4	22.8	22.4	22.4	34.6	65.0	44.6	35.2	49.6	39.4
Gross-migration count	32.4	48.1	44.3	45.8	70.1	129.4	91.5	71.6	100.4	77.9
Net-migration count	-0.4	2.6	-0.6	1.0	1.0	-0.6	2.3	1.1	1.3	-0.9
Migration effectiveness	-1.4	5.3	-1.3	2.2	1.4	-0.4	2.5	1.6	1.3	-1.1
Midlands	**Lowest**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Highest**
In-migration count	14.1	20.3	27.4	22.0	19.3	16.9	19.4	30.1	37.7	55.9
Out-migration count	12.0	17.9	25.6	21.9	19.0	16.6	19.3	28.0	37.3	67.7
Gross-migration count	26.1	38.1	53.0	43.9	38.3	33.5	38.7	58.1	75.0	123.6
Net-migration count	2.1	2.4	1.8	0.1	0.3	0.3	0.1	2.1	0.4	-11.9
Migration effectiveness	8.1	6.4	3.3	0.3	0.7	0.7	0.4	3.7	0.6	-9.6
London	**Lowest**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Highest**
In-migration count	31.3	19.7	32.7	42.3	23.3	28.3	39.0	27.3	30.0	25.3
Out-migration count	35.5	26.3	40.7	50.8	31.0	37.3	51.5	34.0	47.3	31.0
Gross-migration count	66.8	46.0	73.3	93.0	54.3	65.7	90.5	61.3	77.3	56.3
Net-migration count	-4.3	-6.7	-8.0	-8.5	-7.7	-9.0	-12.5	-6.7	-17.3	-5.7
Migration effectiveness	-6.4	-14.5	-10.9	-9.1	-14.1	-13.7	-13.8	-10.9	-22.4	-10.1
South (rest of)	**Lowest**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Highest**
In-migration count	18.1	22.1	26.2	23.3	22.1	26.3	24.4	34.3	33.3	43.1
Out-migration count	19.1	23.6	25.6	21.6	19.8	25.6	22.9	34.0	31.4	43.4
Gross-migration count	37.2	45.6	51.8	44.9	41.9	51.9	47.4	68.3	64.7	86.5
Net-migration count	-0.9	-1.5	0.6	1.7	2.3	0.6	1.5	0.3	1.8	-0.3
Migration effectiveness	-2.5	-3.3	1.2	3.8	5.5	1.2	3.1	0.4	2.8	-0.3

Moving beyond the focus on deprivation, Figure 21 is a map of inter-district net migration rates (per 1,000) in Great Britain. Using the 2004-05 twelve month movers from the ROP, the map appears to support a clear and persistent pattern observed in many previous analyses on internal migration in Great Britain, namely that of urban/rural shift/counter-urbanisation (Stillwell *et al*., 1992; Champion, 2005b; Dennett and Stillwell, 2008). The vast majority of urban conurbations in GB are characterised by net losses, especially those that represent the major urban centres, while at the same time we can observe net gains for the more rural districts of GB. The Shetland Islands have been highlighted as an extreme outlier, this unusual observation is an example of the relative weakness of commercial data when used to analyse remote and sparsely populated areas, where the sample size is very small.

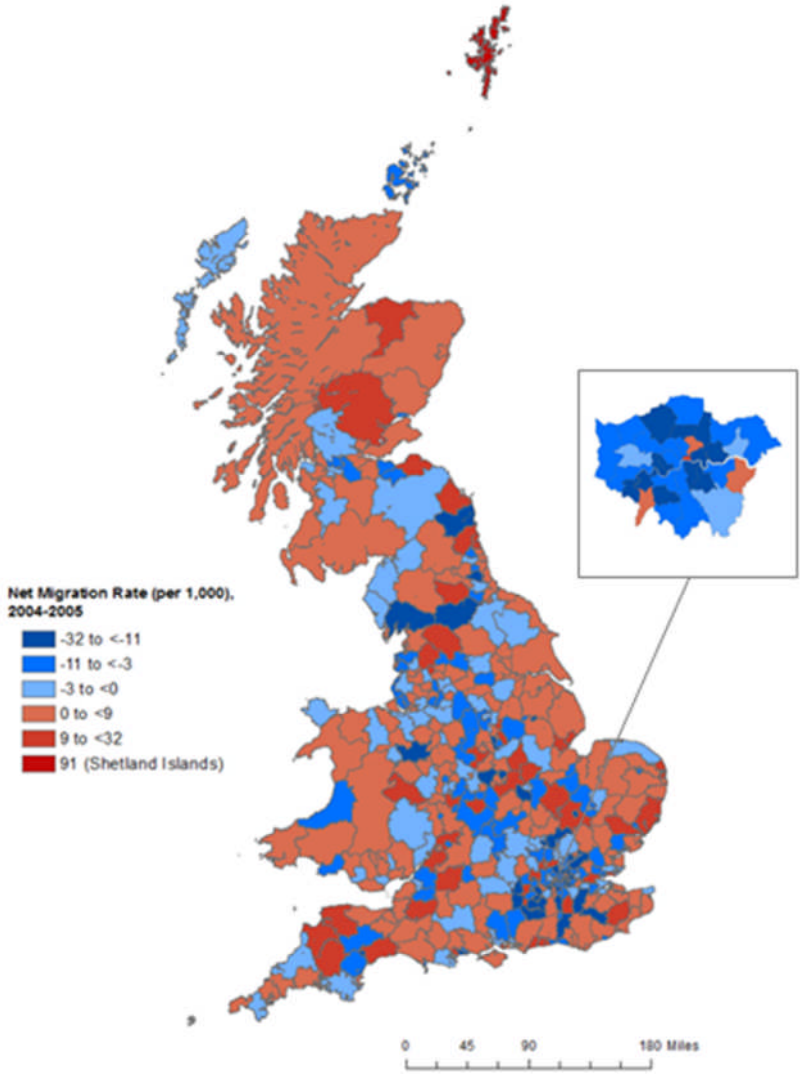


Figure 21: Acxiom January 2005 ROP district net migration rates for GB

4 Acxiom's value added

As has been mentioned a number of times already, the ROP has a number of advantages over many traditional migration sources but a particular strength lies in its variable detail and uniquely for survey sources, its potential for sub-district level spatial analysis.

4.1 Variable detail

One example of the ROP's relative variable potential can be seen in the availability of the variable for gross household income. Indeed, while similar variables are collected in other surveys, the ROP response rate is far superior. For example, the response rate for what is a similar variable in the October to September 2005 APS is well below 5%. However, the ROP sample has attained a response from just over 70% of the total respondents. Figure 22 shows the breakdown of those who moved in the 12 months prior to the January ROP and those who remained *in situ* based on their gross household income.

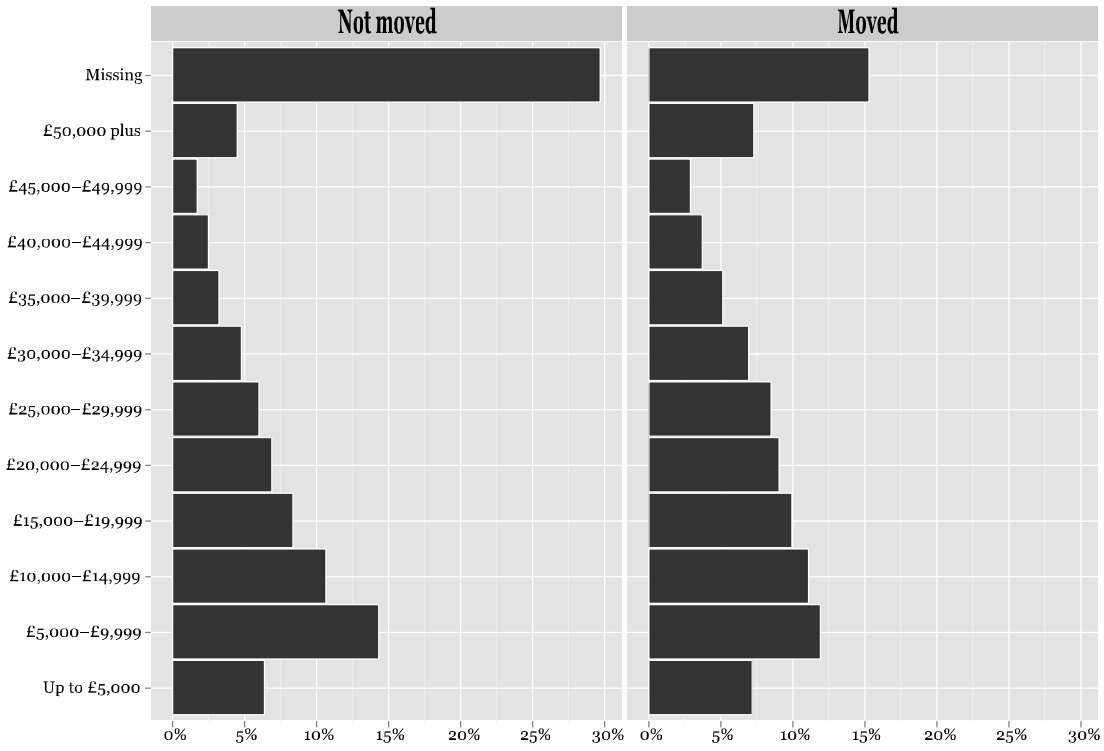


Figure 22: Acxiom January 2005 ROP movers and non-movers by gross household income

Despite a substantial proportion of ROP respondents failing to provide details of their household income (29%) and the fact that the ROP sample under-represents those in the top income brackets (Thompson *et al.*, 2010), it remains apparent that a slightly greater

proportion of movers are in the higher income brackets and a smaller proportion in the lower, when compared to non-movers. Income is a fundamental characteristic upon which migration selection is based (Fielding, 2007; Poston and Bouvier, 2010) and therefore such results are encouraging.

A couple of additional variables that offer real potential for new and interesting insights into internal migration analysis include those that allow us to investigate the potential for future migration propensities as well as those that allow us to measure subjective neighbourhood satisfaction. As such, Figure 23 combines two of these variables in an attempt to discover whether or not, for instance, recent migrants have distinctive neighbourhood satisfaction characteristics when compared to those who have not recently moved or those who are looking to move in the relatively near future.

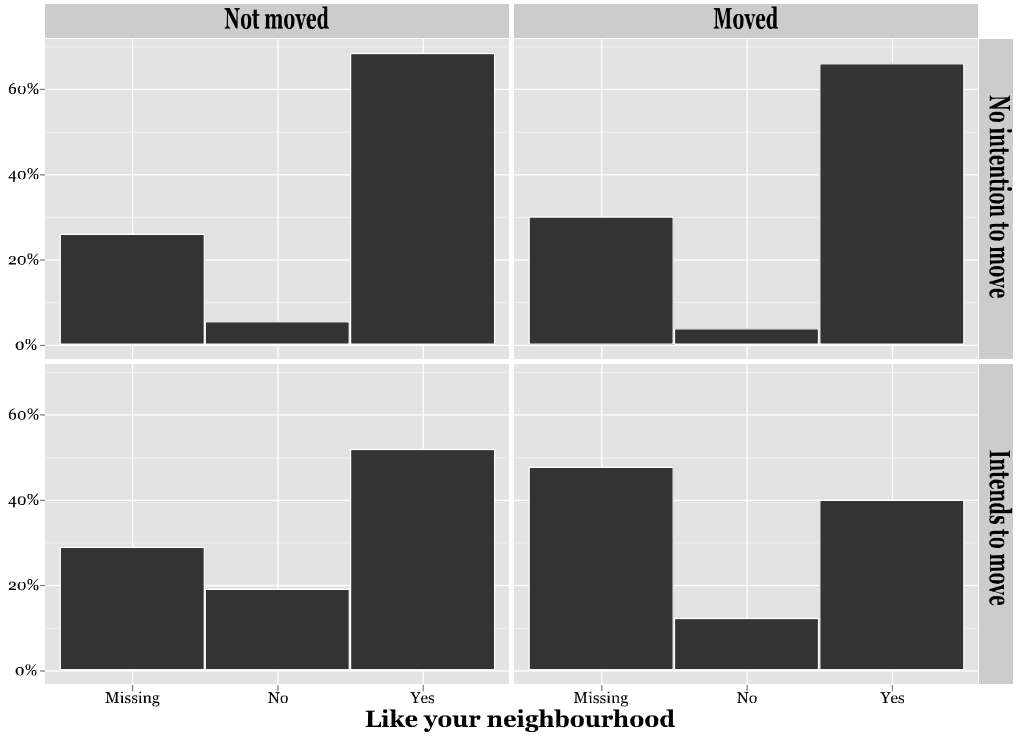


Figure 23: Acxiom January 2005 ROP neighbourhood satisfaction by migrant status and intention to move

When looking at people who have not recently moved and do not intend to move (top left in Figure 23), it is clear that the vast majority (approx. 69%) are generally satisfied with their neighbourhood, something we should perhaps expect. Similarly for those who have recently moved and do not intend to move again, we have a large majority (approx. 66%) who are satisfied with only around 3% of individuals being dissatisfied. For those who have not moved but intend to move in the near future (bottom left), the neighbourhood satisfaction is

lower with around 50% describing themselves as satisfied and just under 20%, the highest for all groups, being dissatisfied. Finally, those who have recently moved and intend to move again in the near future, the number of records with no response to the neighbourhood satisfaction variable is quite large (approx. 48%) and therefore could disrupt the validity of the results here. Despite these concerns, the general picture is still clear, individuals are generally quite satisfied with their neighbourhood, however, the numbers for those who do not like their neighbourhood is still relatively large, but not as large as those who intend to move but have not recently moved. It could be possible that the relatively high 'missing' and relatively low dissatisfaction of those who have recently moved and intend to move again in the near future, is partly to do with the makeup of this subgroup. Indeed, Figure 24 shows us the age breakdown for the respective subgroups suggesting a clear skew towards the more mobile younger age groups in this subgroup. Therefore considering the stage of life course that many of these people will be in, it is most likely that they are short-term residents either moving to university, employment or, subsequently, employment following university, and thus could be expected to feel less attachment and greater indifference when it comes to evaluating their current neighbourhood.

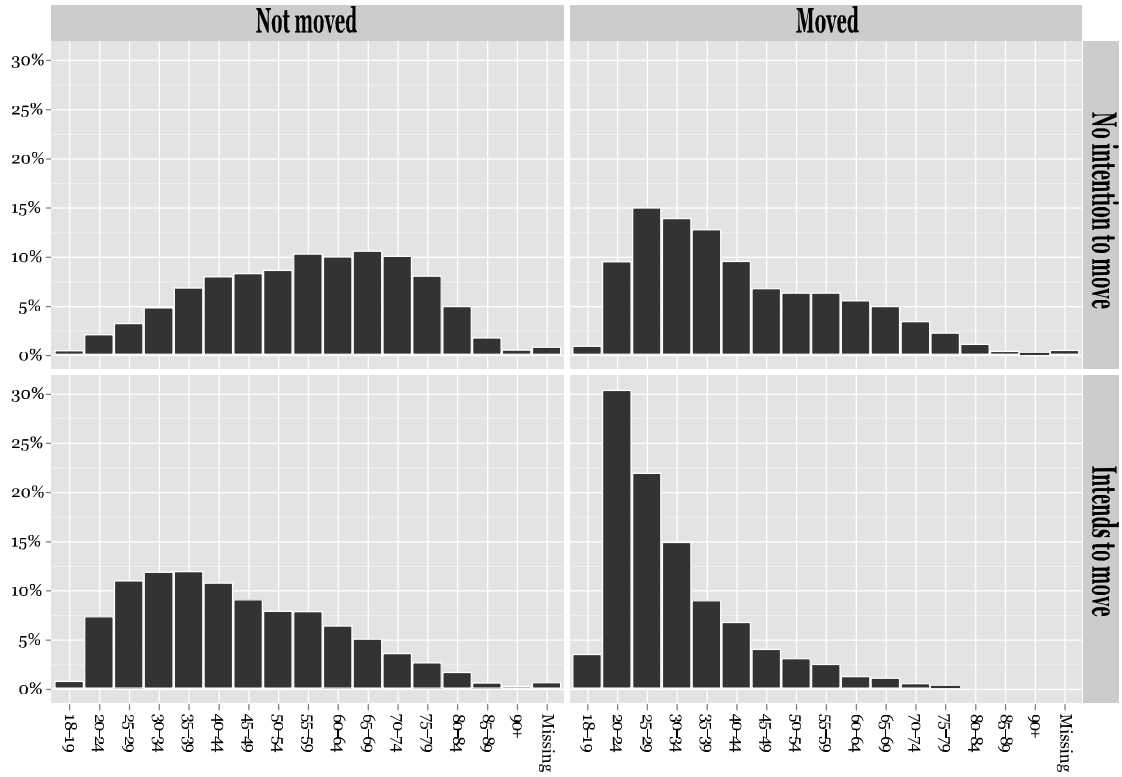


Figure 24: Acxiom January 2005 ROP age groups by migrant status and intention to move

4.2 Spatial detail

With an unrivalled sample size and geographical identifiers at the postcode level, the Acxiom ROP holds real potential for spatial analysis *below* the district level. With this in mind, Figure 25 is a reproduction of the benchmarking exercise in Section 3.4, but this time using migration rates based on ward level moves for the major metropolitan districts of Yorkshire and Humber. Table 19 provides a numerical overview of the deprivation quintiles as well as the relative success of the ROP sample in capturing the district populations. To give some measure of the following analysis, it should be noted that for Yorkshire and Humber, the ROP recorded a total of 939 in-migrants and 954 out-migrants, both of which include moves within the region. This relates to a total sample population of n = 9,272 which includes 12 month movers within GB with usable full postcode address appropriate for aggregation to ward level. The decision to focus specifically on the major metropolitan districts of the region was twofold: firstly limiting the number of comparisons was deemed important for reasons of simplicity and clarity; and secondly, as has been noted by Thompson *et al.* (2010), the ROP sample has a tendency to under-represent populations from rural areas. This will undoubtedly result in small number problems, especially at geographies below the district level, which can manifest themselves in unstable and exaggerated migration values. However, there are possible remedies to consider for future analysis including weighting the variables (Crockett *et al.*, 2011), employing spatial microsimulaiton techniques (Harland *et al.*, 2012), or by pooling the individual ROP datasets to boost the overall sample population.

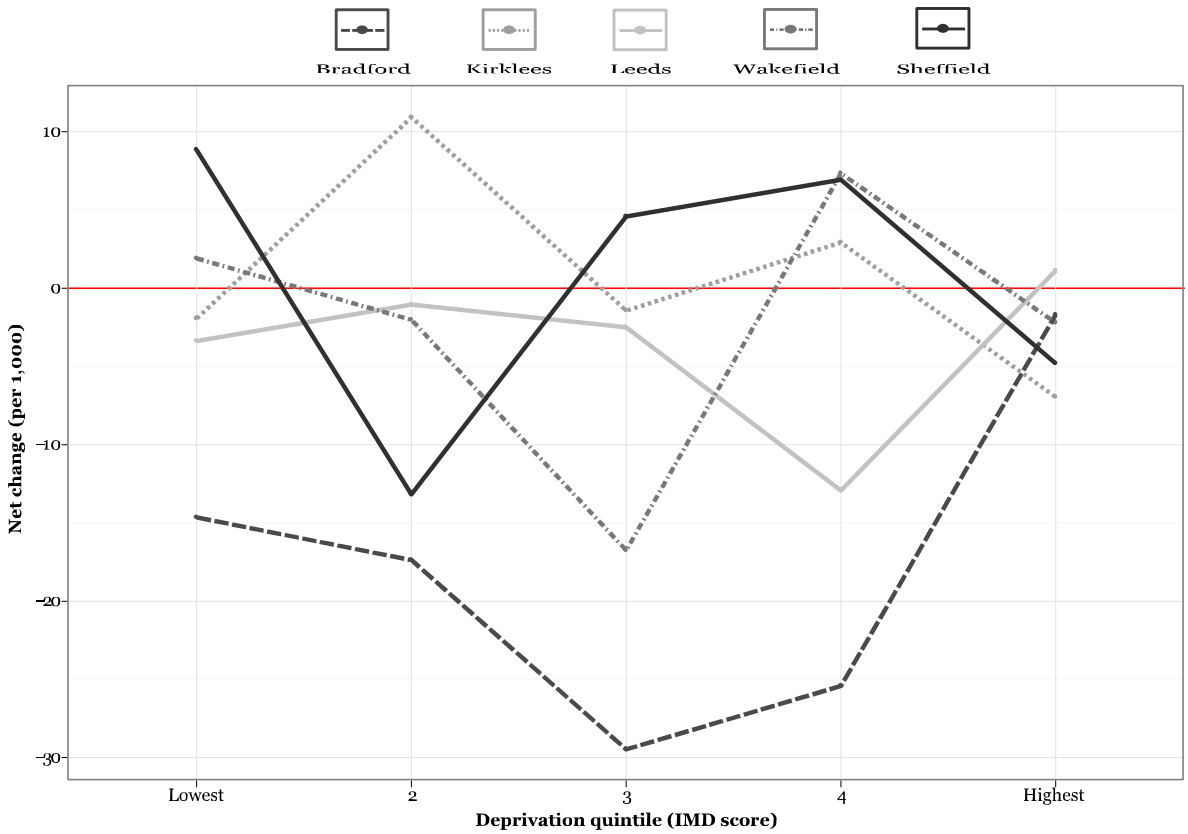


Figure 25: Acxiom January 2005 ROP net migration change by IMD quintile at district level

When we study and compare the bottom quintile and the top quintile for each district, it is apparent that Kirklees, Wakefield and Sheffield see net losses in the most deprived quintile while the least deprived quintiles are seeing either reduced net outflows (Kirklees) or net gains (Wakefield and Sheffield). On the other hand, Leeds appears to have fairly stable net change across the quintiles, when compared to the other districts; however, the least deprived wards see net losses whereas the most deprived wards see net gains.

Table 21: District average IMD 2004 score for each deprivation quintile and sample population capture

	Lowest	2	3	4	Highest	MYE 2005 ('000)	Jan 2005 ROP ('000)	ROP population capture (%)
Bradford	11.9	22.4	32.3	39.4	55.1	480.9	3.0	0.6
Kirklees	13.7	20.9	26.7	33.5	39.0	393.9	3.3	0.8
Leeds	13.0	17.3	25.4	38.7	53.6	734.8	5.7	0.8
Wakefield	16.4	22.8	34.2	36.7	42.1	319.3	2.8	0.9
Sheffield	11.2	18.6	30.1	41.0	55.1	519.8	4.2	0.8

For Bradford, we see the average net change for wards, across all quintiles, suggest net losses, losses that are particularly considerable for quintiles 1-4. While we should be cautious

of these specific results given the ROP captures a relatively low proportion of Bradford's population (Table 21), this does appear to suggest that, on the whole, Bradford is losing relatively large numbers of people, with only the most deprived areas seeing reasonably stable net change. Indeed the 2004-05 PR-NHSCR migration estimates for the Bradford district do confirm that Bradford loses considerably more people to migration than any of the other major metropolitan districts in Yorkshire and the Humber. A further interesting observation that can be made relates to the fact that, aside of the Bradford case, the rates for the least deprived and most deprived quintiles are relatively similar in most cases, this is despite the relatively volatile shifts in net change apparent for those wards that are within the middle 60 per cent for each district (quintiles 2-4).

5 A potential framework for analysis

Much of what has been discussed here has focussed on benchmarking and validating what is a hitherto unused commercial data source, with 'official statistics' using relatively simple descriptive and bi-variate regression based approaches. While the validation process is not complete, the results presented here are certainly promising *vis-à-vis* the reliability of the ROP for future analysis. With this in mind, this section will now provide a brief discussion related to a potential framework for future analysis.

While validation against *external* sources has been a large part of the work presented here, some *internal* validations, that is, validation exercises comparing each ROP sample against one another, represent the next key stage for this research project. One such method of *internal* validation could follow those performed here for the external validations, namely further simple correlation/regression exercises comparing two data cross-sectional sources against one another. Given that research by Duke-Williams and Stillwell (2010) has shown migration rates between districts continues to be surprisingly stable over the last decade or so, it would not appear unreasonable to perform similar correlation exercises between the 2005 to 2007 ROP samples despite their slight shift in the temporal context of the sample data. Beyond this, a more stringent comparison could include an examination of the consistency of specially selected variables. For example, one method would be to observe variations in cohort move counts. For instance, it would be possible to check the number of 'moves' in five year periods (e.g. 1980-85, 1985-90, …) against the same observations in another ROP dataset. If internal validations are successful, it opens up the possibility of combining the

ROP datasets into one 'superset', thus reducing issues associated with sample error, and so forth. It also opens up the possibility of performing analyses at more detailed and finer spatial scales, which itself makes possible exercises to observe/test potential scale/aggregation effects, ecological fallacy and modifiable areal unit problem (Openshaw, 1984; Wrigley, 1995).

However, while the pooling of the ROP datasets is one possibility, there are other methods that could be employed to similar effect. Indeed, while no method represents a panacea, there are options for boosting the sample and thus teasing out certain biases within the ROP. Potential options for this include the use of spatial microsimulaiton techniques for the generation of realistic synthetic small area populations with deterministic reweighting (Ballas *et al.*, 2005; Smith *et al.*, 2011), conditional probability simulation (Monte Carlo simulation) (Birkin and Clarke, 1988; 1989) and simulated annealing (Openshaw, 1995; Voas and Williamson, 2001), all techniques that have been commonly applied in quantitative geography (Harland *et al.*, 2012).

In terms of statistical analysis, the project will employ a multilevel statistical framework to analyse data hierarchically (e.g. individuals within areas within districts) as well as binary and multinominal logit models (Duncan *et al.*, 1998; Snijders and Bosker, 1999; Gould and Fieldhouse, 1997). In doing so it will offer both substantive, methodological and technical benefits to the migration analysis (Jones, 1991). Moreover, thanks to the flexibility of the 'year of move' variable, there is potential in the data for the analysis of repeated cross-sectional migrant cohorts for instance comparing the lifestyle and socio-economic position of different cohorts before and after their move, or whether some cohorts are more immune to the frictional effect of distance on migration than others. Again this repeated cross-sectional design (Gould *et al.*, 1997) can be incorporated within a multilevel model (level 1 = individual, level 2 = time period/cohort, level 3 = spatiality). Beyond this, if the ROP sample datasets can be proved to be suitable, this project may potentially utilise a further multilevel statistical design to analyse cross-classified structures (i.e. individuals nested within both origins and within destinations – both at 'level-2') and multiple membership structures (i.e. where individuals spend differing amounts of time residing in two different residential locations) (Shtttleworth and Gould, 2010; Goldstein, 2003). Multilevel modelling generally has the technical advantages of properly handling heterogeneity and (spatial) autocorrelation through its use of 'shrunken estimation' (using Empirical Bayes and/or Markov Chain Monte

Carlo estimation) (Jones, 1991; Goldstein, 2003). However, there are potential concerns associated with the implementation of the cross-classified research design associated with the precision of model estimates if the areal units in the cross-classification contain very small numbers of sampled individuals (Fielding and Goldstein, 2006); something we know to be the case in the below district analyses. Moreover, it may be found that the model cannot be converged. If this is the case, it further strengthens the case for data pooling and/or the use of spatial microsimulaiton techniques for the creation of realistic synthetic populations.

All in all, the framework for analysis will be designed to consider some of the following multilevel substantive research questions:

- Are there variations in the propensity of moving between different types of people (e.g. demographic, socio-economic, and lifestyle groups)? Are these variations explainable solely by individual characteristics (i.e. composition of areas)?
- Do migration rates of particular groups vary across areas of Britain? If so, is there some remaining contextual area variation (e.g. for geographical areas/districts, and/or geodemographic area type)?
- Are there complex between-area differentials for different types/groups of people?
- Are there variations between individual and area variables/characteristics that together result in different distances travelled?
- Are there complex social-spatial mobility variations between repeated cross-sectional cohorts?

6 Conclusion

This paper has been concerned with outlining some data quality issues and reporting back on some initial validation exercises involving the comparison of ROP data with census, administrative and survey data from alternative sources. Broadly speaking, we can be reasonably satisfied with the reliability of the ROP data. The aggregate level benchmarking showed there to be significant positive correlation between the ROP inflow counts and those of the alternative sources. Similarly, the analyses at the micro level suggests that, despite some sample bias concerns, the overall patterns found in the traditional sources and documented in previous research are picked up by the ROP sample. Furthermore, the spatial benchmarking was again successful in highlighting the reliability of the ROP at the LAD

level. Moving beyond the benchmarking, the paper reveals a number of advantages that the ROP holds over many traditional migration sources, with evidence of its potential highlighted by analysis of the variable detail and uniquely for survey sources, its potential for below district spatial analysis. With this in mind, the potential framework for analysis suggests a number of methods that, once integrated, offer real potential for thoroughly analysing and exploring this unique and impressive data source.

Acknowledgements

The authors wish to thank Acxiom for access to the ROP data. Michael Thomas is funded with a three-year ESRC NCRM TALISMAN PhD Studentship. Any errors in this preliminary draft manuscript remain the authors own.

References

Bailey, N. and Livingston, M. (2005) 'Determinants of individual migration: an analysis of SARs data', *SCRSJ Working Paper No. 3*. Scottish Centre for Research on Social Justice, University of Glasgow.

Bailey, N. and Livingston, M. (2008) 'Selective migration and neighbourhood deprivation: Evidence from 2001 Census migration data for England and Scotland', *Urban Studies*, 45, 4, 943-961.

Ballas D., Clarke, G., Dorling, D., Eyre, H., Thomas, B., and Rossiter D. (2005) 'SimBritain: a spatial microsimulation approach to population dynamics', *Population, Space and Place*, 11, 13-34.

Bates J. J. and Bracken I. (1987) Migration age profiles for local authority areas in England, 1971-81, *Environment and Planning A*, 19, 521-535.

Bell, M., Charles-Edwards, E., Kupiszewski, M., Kupuszewska, D., Stillwell, J. and Zhu, Y. (2012) Comparing internal migration around the world, Presentation at the European Population Conference 2012, Stockholm, 13-16 June.

Birkin, M. and Clarke, M. (1988) 'SYNTHESIS - a synthetic spatial information system for urban and regional analysis: Methods and example', *Environment and Planning A*, 20, 1645-1671.

Birkin, M. and Clarke, M. (1989) 'The generation of individual and household incomes at the small area level using synthesis', *Regional Studies*, 23, 535-548.

Bryan, M. (2011) 'Measuring Work: Prospects for labour market research in understanding Society', in McFall, S.L. and Garrington, C. (Eds.) *Early Findings from the First Wave of the UK's Household Longitudinal Study*. Colchester: Institute for Social and Economic Research, University of Essex.

Burton, J., Laurie, H. and Lynn, P. (2011) 'Appendix: Understanding Society Design Overview', in McFall, S.L. and Garrington, C. (Eds.) *Early Findings from the First Wave of the UK's Household Longitudinal Study*. Colchester: Institute for Social and Economic Research, University of Essex.

Buck, N. (1994) 'Housing and residential mobility in Buck, N., Gershuny, J. Rose, D. and Scott, J. (1994) *Changing Households: the British Household Panel Survey 1990-1992*, Colchester: ESRC Research Centre on Micro-Social Change.

Calder, A. and Swan, J. (2011) 'Beyond 2011: The future for population statistics?', Presentation at the Beyond 2011 User Requirements Workshop, The Carriageworks, Leeds, 03 November.

Cangiano, A. (2010) 'UK data sources on international migration and the migrant population: A review and appraisal', *COMPAS Research Resource Paper*, University of Oxford, Oxford.

Champion, T., Fotheringham, A.S., Rees, P., Boyle, P. and Stillwell, J. (1998) *The Determinants of Migration Flows in England: A Review of Existing Data and Evidence*, University of Newcastle, Commissioned Report for the Department of the Environment, Transport and the Regions.

Champion, T. (2005a) 'Population movement within the UK', in Chappell, R. (ed.) *Focus on People and Migration*, Basingstoke: Palgrave Macmillan.

Champion, T. (2005b) 'The counterurbanization cascade in England and Wales since 1991: the evidence of a new migration dataset', *BELGEO*, 1-2: 85-101.

Crockett, A., Afkhami, R., Rafferty, A., Higgins, V. and Marshall, A. (2011) *Weighting the Social Surveys: ESDS Government*, ESDS, Colchester.

Dennett, A, Duke-Williams, O. and Stillwell, J. (2007) 'Interaction datasets in the UK: An audit', *Working Paper 07/4*, School of Geography, University of Leeds, Leeds.

Dennett, A. and Stillwell, J. (2008) 'Internal migration in Great Britain: A district-level analysis using 2001 Census data', *Working Paper 08/01*, School of Geography, University of Leeds, Leeds.

Dennett, A. and Stillwell, J. (2010) 'Internal migration in Britain, 2000-01, examined through an area classification framework', *Population Space and Place*, 16, 6, 517-538.

Duke-Williams and Stillwell (2010) 'Temporal and spatial consistency' in Stillwell, J., Duke-Williams, O. and Dennett, A. (Eds.) (2010) *Technologies for Migration and Commuting Analysis: Spatial Interaction Data Applications*. IGI Global: Hershey.

Duncan, C., Jones, K. and Moon, G. (1998) 'Context, composition and heterogeneity', *Social Science & Medicine*, 46, 97-117.

Field, A., Miles, J. and Field, Z. (2012) *Discovering Statistics Using R*, London: Sage.

Fielding, A. and Goldstein, H. (2006) 'Cross-classified and multiple membership structures in multilevel models: An introduction and review', *Department for Education and Skills, Research Report No. 791*, University of Birmingham, Birmingham.

Fielding, A. J. (2007) 'Migration and social mobility in urban systems: national and international trends', in Geyer, H.S. (ed.), *International Handbook of Urban Policy*, Edward Elgar: Cheltenham.

Fielding, A. J. (Forthcoming) *Migration in Britain: Paradoxes of the present, prospects for the future*.

Frey, W.H. (2003) 'Internal Migration', in Demeny, P. and McNicoll, G. (eds.), *Encyclopedia of Population*, II: 545-548. Macmillan Reference USA, New York.

Goldstein, H. (2003) *Multilevel Statistical Models*, 3rd edition, London: Arnold.

Gould M.I. and Fieldhouse E. (1997) 'Using the 1991 Census SAR' in a multilevel analysis of male unemployment', *Environment and Planning A*, 29, 611-628.

Gould M.I., Jones, K. and Moon, G. (1997) 'Guest editorial: The scope of multilevel models', *Environment and Planning A*, 29, 4, 581-584.

Harland, K., Heppenstall, A., Smith, D. and Birkin, M. (2012) 'Creating realistic synthetic populations at varying spatial scales: A comparative critique of population synthesis techniques', *Journal of Artificial Societies and Social Simulation*, 15, 1, 1.

Hughes G. and McCormick B. (1985) 'Migration intentions in the UK. Which households want to migrate and which succeed?' *Economic Journal, Conference Supplement*, 95, 113-123.

Jefferies J, Horsfield G, Newman J and Stokes K (2003) *Report on research into revising internal migration estimates*. ONS, London.

Jones, C. (2011) The latest developments on the Integrated Household Survey – Information for LFS users, Presentation at the Labour Force Survey/Annual Population Survey User Meeting, Royal Statistical Society, London, 6 December.

Jones, K. (1991) 'Specifying and estimating multi-level models for geographical research', *Transactions of the Institute of British Geographers, NS*, 16, 148-160.

Large, P. and Ghosh, K. (2006) 'Estimates of the population by ethnic group for areas within England', *Population Trends*, 124, 1, 8-17.

Lomax, N., Stillwell, J., Norman, P. and Rees, P. (2011) 'Data sources and estimation methods for sub-national migration in the United Kingdom', *Working Paper 11/04*, School of Geography, University of Leeds, Leeds.

McFall, S. L. (2011) *Understanding Society – The UK Household Longitudinal Study, Wave 1, 2009-2010, User Manual*, University of Essex, Colchester.

Nam, C.B., Serow, W.J. and Sly, D.F. (eds.) (1990) *International Handbook on Internal Migration*. New York: Greenwood Press.

Norman, P., Rees, P. and Boyle, B. (2003) 'Achieving data comparability over space and time: Creating consistent geographical zones', *International Journal of Population Geography*, 9, 365-386.

Norman, P. and Riva, M. (2012) 'Population health across space and time: the geographical harmonisation of the Office for National Statistics Longitudinal Study for England and Wales', *Population, Space and Place*, DOI: 190.1002/ psp.1705.

ODPM (2004) *The English Indices of Deprivation 2004 (revised)*, ODPM Publications, Wetherby.

ONS (2002) *Review of the Labour Force Survey. National Statistics Quality Review Series: Report No. 12*. Online: www.statistics.gov.uk/about/data/methodology/quality/reviews/labour.asp#nsqr12.

ONS (2010a) *Improved estimation of student migration within England and Wales*. Online: http://www.ons.gov.uk/about-statistics/methodology-and-quality/imps/mig-stats-improve-prog/comm-stakeholders/improvements-2008-pop-est/detailed-methodology-papers/student-adjustment-detailed-methodology.pdf.

ONS (2010b) *User Guide – Volume 1: IHS Background and Methodology 2010*, ONS, London.

ONS (2011a) *Information Paper: Quality and Methodology Information – Internal Migration Estimates*, ONS, London.

ONS (2011b) *Beyond 2011 Consultation on User Requirements: 17 October 2011- 20 January 2012*, ONS, London.

ONS (2011c) *Integrated Household Survey User Guide – Volume 3: Details of IHS variables 2011*, ONS, London.

ONS (2011d) *Performance and Quality Monitoring Report – April to June 2011: Labour Force Survey*, ONS, London.

ONS (2011e) *News Release: ONS takes a fresh look 'Beyond 2011'*, ONS, London.

Openshaw, S. (1984) 'The Modifiable Areal Unit Problem', *Concepts and Techniques in Modern Geography*, 38, Geo Books, Norwich.

Openshaw, S. and Rao, L. (1995) 'Algorithms for reengineering 1991 Census geography', *Environment and Planning A*, 27, 425-446.

Owen, D.W. and Green, A.E. (1992) 'Labour market experience and occupational change amongst ethnic groups in Great Britain', *New Community*, 19, 7-29.

Phelps, M. (2011) *ONS Postcode Directory User Guide: 2011 Edition*, ONS, London.

Poston, D. L. and Bouvier, L. F. (2010) *Population and Society: An Introduction to Demography*, Cambridge: Cambridge University Press.

Raper, J., Rhind, D. and Shepherd, J. (1992) *Postcodes: The New Geography*. Harlow: Longman.

Raymer, J., Rees, P., Blake, A., Brown, J. Disney, G., Lomax, N., Norman, P., Stillwell, J. and Boden, P. (2012) *Conceptual Framework for UK Population and Migration Statistics*, Commissioned Report for the Office for National Statistics, ONS, London.

Rees, P. and Kupiszewski, M. (1999) *Internal Migration and Regional Population Dynamics in Europe,* Strasbourg: Council of Europe Publishing.

Rees, P., Stillwell, J., Boden, P. and Dennett, A. (2009) A review of migration statistics literature, in *Migration Statistics: The Way Ahead?* UKSA, London.

Rogers, A. and Castro, L.J. (1981) *Model Migration Schedules*, Laxenburg: International Institute for Applied Systems Analysis, RR-91-30.

Shutttleworth, I. and Gould, M. (2010) 'The geography of employment recruitment: A multilevel analysis of individual workers, neighbourhoods, and sites in Northern Ireland', *Environment and Planning A*, 42, 1221-1238.

Simpson, L. and Finney, N. (2009) 'Spatial patterns of internal migration: evidence for ethnic groups in Britain', *Population, Space and Place*, 15, 1, 37-56.

Smith, D.M., Pearce, J.R., & Harland, K. (2011) 'Can a deterministic spatial microsimulation model provide reliable small-area estimates of health behaviours? An example of smoking prevalence in New Zealand', *Health & Place*, 17, 618-624.

Smith, P. W. F., Raymer, J., Giuletti, C. (2010) 'Combining available migration data in England to study economic activity flows over time', *Journal of the Royal Statistical Society*, 173, 4, 733-753.

Snijders, T. and Bosker, R. (1999) *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modelling*, London: Sage.

Stillwell, J., Rees, P., Boden, P. (1992) 'Internal migration trends: an overview', in Stillwell, J., Rees, P., Boden, P., (eds.) *Migration Processes & Trends, vol. 2: Population redistribution in the United Kingdom*, London: Belhaven Press.

Stillwell, J. and Duke-Williams, O. (2005) *Ethnic distribution, immigration and internal migration in Britain: what evidence of linkage at the district scale?*, Canterbury: University of Kent, paper presented at the British Society of Population Studies Annual Conference, 12-14 September.

Stillwell, J. and Phillips, D. (2007) 'Diversity and change: understanding the ethnic geographies of Leeds', *Journal of Ethnic and Migration Studies*, 32, 7, 1131-52.

Stillwell, J. (2008) 'Inter-regional migration modelling: a review and assessment', in *Migration and Human Capital: Regional and Global Perspectives*, Poot, J., Waldorf, B. and Van Wissen, L. (eds), Cheltenham: Edward Elgar.

Stillwell, J. and Hussain, S. (2010) 'Exploring the ethnic dimension of internal migration in Great Britain using migration effectiveness and spatial connectivity', *Journal of Ethnic and Migration Studies*, 36, 9, 1381-1403.

Stillwell, J., Boden, P. and Dennett, A. (2011) 'Monitoring Who Moves Where: Information Systems for Internal and International Migration' in Stillwell, J. and Clarke, M. (eds.) *Population Dynamics and Projection Methods, Understanding Population Trends and Processes Volume 4* (pp. 115-164), Berlin: Springer.

Thompson, C., Stillwell, J., Clarke, M. and Bradbrook, C. (2010) Understanding and validating Acxiom's Research Opinion Poll data, *Working Paper 10/06*, School of Geography, Leeds University, Leeds.

Van Ham, M. and Clark, W.A.V. (2009) 'Neighbourhood mobility in context: household moves and changing neighbourhoods in the Netherlands', *Environment and Planning A*, 41, 1442-1459.

Voas, D. and Williamson, P. (2001) 'Evaluating goodness-of-fit measures for synthetic microdata', *Geographical and Environmental Modelling*, 5, 177-200.

Walthery, P. (2011) *Introductory guide to the Integrated Household Survey*, ESDS, London.

Warnes A. (1992) 'Migration and the life course', in A. Champion and A. Fielding (eds.) *Migration Processes and Patterns. Volume 1: Research Progress and Prospects*. London: Belhaven.

Werner, B. (2006) *Reflections on Fifteen Years of Change in Using the Labour Force Survey*. ONS, London.

Wrigley, N. (1995) 'Revisiting the modifiable area unit problem and the ecological fallacy', in Cliff, A., Gould, P., Hoare, A. and Thrift, N. (Eds.) *Diffusing Geography: Essays for Peter Haggett*, Oxford: Basil Blackwell. 49-71.

Wulff, M., Champion, A. and Lobo, M. (2010) 'Household diversity and migration in mid-life: Understanding residential mobility among 45-64 year olds in Melbourne, Australia', *Population, Space and Place*, 16, 307-321.