Working Paper 07/05

INTERACTION DATA SETS IN THE UK: AN AUDIT

Adam Dennett, Oliver Duke-Williams and John Stillwell

Centre for Interaction Data Estimation and Research (CIDER)

School of Geography
University of Leeds
Leeds LS2 9JT
United Kingdom

June 2007

CIDER

UNIVERSITY OF LEEDS

This working paper is an online publication and may be revised.

Full contact details of the authors are:

School of Geography,
University of Leeds,
Leeds,
LS2 9JT

Tel: +44 (0 in UK) 113 34 33300
Fax: +44 (0 in UK) 113 34 33308

Adam Dennett

a.r.dennett@leeds.ac.uk

Oliver Duke-Williams

o.w.duke-williams@leeds.ac.uk

John Stillwell

j.c.h.stillwell@leeds.ac.uk

Contents

Acknowledgement

The Centre for Interaction Data Estimation and Research (CIDER) is a Data Support Unit, funded (RES-348-25-0005) by the Economic and Social Research Council as part of the Census Programme, 2006-2011.

Glossary of Acronyms

AF	Armed Forces
APS	Annual Population Survey
ARIMA	AutoRegressive Integrated Moving Average
ASCII	American Standard Code for Information Interchange
BCS	British Cohort Study
BHPS	British Household Panel Survey
CAMS	Controlled Access Microdata Sample
CAS	Census Area Statistics
CASWEB	Census Area Statistics on the Web
CD	Compact Disk
CeLSIUS	Centre for Longitudinal Study Information and User Support
CHI	Central Health Index (in Northern Ireland)
CHI	Community Health Index (in Scotland)
CHS	Continuous Household Survey
CIDER	Centre of Interaction Data Estimation and Research
CIDS	Census Interaction Data Service
CPS	Continuous Population Survey
CSA	Central Services Agency
CSO	Central Statistical Office
DAS	Destination-Age-Sex
DETR	Department for Environment, Transport and the Regions
DfES	Department for Skills and Education
DOB	Date Of Birth
DOM	Date Of Migration
DVLA	Driver and Vehicle Licensing Agency
DWP	Department of Work and Pensions
EEA	European Economic Area
EFS	Expenditure and Food Survey
ELSA	English Longitudinal Study of Ageing
ESDS	Economic and Social Data Service
ESRC	Economic and Social Research Council
EU	European Union
FE	Further Education
FHSA	Family Health Service Authority
FPC	Family Practitioner Committee
GCSE	General Certificate of Secondary Education
GB	Great Britain
GBLFS	Great Britain Labour Force Survey
GHS	General Household Survey
GOR	Government Office Region

GROS	General Register Office Scotland
HA	Health Area
HBA	Health Board Area
HE	Higher Education
HEI	Higher Education Institution
HES	Hospital Episode Statistics
HESA	Higher Education Statistics Agency
HM	Her Majesty
HO	Home Office
HRP	Household Reference Person
HSE	Health Survey for England
IC	Information Centre
IHS	Integrated Household Survey
ILR	Individual Learner Record
IMPS	Improving Migration and Population Statistics
IOM	Isle Of Man
IPS	International Passenger Survey
IPSS	Integrated Population Statistics System
KS	Key Statistics
LAD	Local Authority District
LEA	Local Education Authority
LGD	Local Government District
LFS	Labour Force Survey
LLSOA	Lower Layer Super Output Area
LS	Longitudinal Study
LSC	Learning and Skills Council
LSCS	Longitudinal Studies Centre - Scotland
LSEM	Longitudinal Study of Ethnic Minorities
LSYPE	Longitudinal Study of Young People in England
MGRP	Moving Group Representative Person
MLSOA	Middle Layer Super Output Area
NCDS	National Child Development Survey
NeSS	Neighbourhood Statistics Service
NHS	National Health Service
NHSCR	National Health Service Central Register
NHSISD	National Health Service Information Statistics Directorate
NI	Northern Ireland
NI CHI	Northern Ireland Central Health Index
NI IC	Northern Ireland Information Centre
NI LFS	Northern Ireland Labour Force Survey
NILS	Northern Ireland Longitudinal Study
NINo	National Insurance Number
NIRS	National Insurance Recording Scheme
NI SAR	Northern Ireland Sample of Anonymised Records
NISRA	Northern Ireland Statistics and Research Agency
NOMIS	National Online Manpower Information System
NPD	National Pupil Database
NSHD	National Survey of Health and Development
NS-SEC	National Statistics SocioEconomic Classification
NTS	National Travel Survey

OA	Output Area
OAS	Origin-Age-Sex
OD	Origin-Destination
ODPM	Office of the Deputy Prime Minister
OMN	OMNibus Survey
ONS	Office of National Statistics
OPCS	Office of Population Censuses and Surveys
OSI	Office of Science and Innovation
PAS	Patient Administration System
PC	Parliamentary Constituency
PCT	Primary Care Trust
PGCE	Postgraduate Certificate in Education
PIAG	Patient Information Advisory Group
PLASC	Pupil Level Annual School Census
PRDS	Patient Register Data System
PUD	Primary Unit Data
QNHS	Quarterly National Household Survey
SAM	Small Area Microdata
SAR	Sample of Anonymised Records
SCAG	Security and Confidentiality Advisory Group
SCAM	Small Cell Adjustment Method
SDC	Statistical Disclosure Control
SEN	Special Education Needs
SHA	Strategic Health Authority
SIC	Standard Industrial Classification
SLS	Scottish Longitudinal Study
SMS	Special Migration Statistics
SNHSCR	Scottish National Health Service Central Register
SOA	Super Output Area
SOC	Standard Occupational Classification
SPSS	Statistical Package for the Social Sciences
ST	Standard Table
STS	Special Travel Statistics
SWS	Special Workplace Statistics
TDOM	True Date Of Migration
TIM	Total International Migration
TIMMIG	TIMe Series MIGration
UA	Unitary Authority
UCAS	University Central Admissions System
UK	United Kingdom
UKBORDERS	United Kingdom Boundary Outline and Reference Database for Education and Research Study
UKLHS	United Kingdom Household Longitudinal Survey
UPN	Unique Pupil Number
UPTAP	Understanding Population Trends And Processes
WICID	Web-based Interface to Census Interaction Data
WRS	Worker Registration Scheme

List of Figures

List of Tables

Preface

Interaction or flow data involves counts of flows between origin and destination areas and can be extracted from a range of sources. The Centre for Interaction Data Estimation and Research (CIDER) maintains a web-based system (WICID) that allows academic researchers to access and extract migration and commuting flow data (the so-called Origin-Destination Statistics) from the last three censuses. However, there are many other sources of interaction data other than the decadal census, including national administrative or registration procedures and large scale social surveys. This paper contains an audit of interaction data sets in the UK, providing detailed description and exemplification in each case and outlining the advantages and shortcomings of the different types of data where appropriate. The Census Origin-Destination Statistics have been described elsewhere in detail and only a short synopsis is provided here together with review of the interaction data that can be derived from other census products.

The primary aims of the audit are to identify those interaction data sets that exist that might complement the census origin-destination statistics currently contained in WICID and to assess their suitability and availability as potential data sets to be held in an expanded version of WICID. Tables or flow data sets are included for exemplification. The paper concludes with a series of recommendations as to which of these data sets should be incorporated into a new information system for interaction flows that complement the census data and also provide opportunities for new research projects.

INTERACTION DATA SETS IN THE UK: AN AUDIT

1 INTRODUCTION

Between August 2001 and July 2006, the Census Interaction Data Service (CIDS) operated as a 'Data Support Unit' under the ESRC/JISC-funded Census Programme for 2001-2006, with the primary function of allowing registered users to access, extract and analyse the 'Origin-Destination Statistics' collected from successive population censuses in 1981, 1991 and 2001. These data sets are referred to as the 'census interaction data sets' and include the Special Migration Statistics (SMS), Special Workplace Statistics (SWS) and Special Travel Statistics (STS), together with estimated flow data derived from them. Essentially, they are data sets that contain flows of migrants or commuters between origins and destinations at various spatial scales. Registered users wishing to access data from these data sets can do so by running queries through the Web-based Interface to Census Interaction Data (WICID) (Stillwell and Duke-Williams, 2003; Stillwell, 2006a, 2006b) and extracting subsets for independent analysis and visualisation. Additional population-at-risk data have been included in the system to enable rates of interaction to be computed within WICID at users' discretion.

Since August 2006, CIDS has become the Centre for Interaction Data Estimation and Research (CIDER) supported by ESRC funding under the Census Programme 2006-2011. CIDER aims to provide access to these detailed data sets through to the next census in 2011 and to redesign the WICID interface so as to further facilitate query building, extraction and data analysis. This has become increasingly important as the number of primary and derived data sets and the number of different 'geographies' held in WICID have expanded. Whilst it is the case that the census interaction data sets have been and remain under-used, it has also become evident that there is a need to ensure that researchers have a good understanding of the data sets contained in WICID and, in particular, an awareness of both the advantages and shortcomings of the various data holdings.

At a time when the results of the 2001 Census are becoming more and more outdated, it also makes sense to complement and enhance our interpretation of the census flow statistics with interaction data from other sources. This is the *raison d'etre* for this paper which reports on an audit that we have conducted of all types of interaction data that are available on a national basis in the UK. Moreover, having developed WICID for the specific function of storing and providing users with access to census interaction data, it is timely to explore the possibilities of adapting the system for other origin-destination flow data sets that will provide new and beneficial insights into human interaction behaviour. It would be extremely useful, for example, to bring data sets into WICID that allow us to identify migration or commuting trends between censuses or to be able to substantiate the magnitude and pattern of flows derived from census sources with flows derived from alternative sources, e.g. for students or immigrants, or even to study less well known population flows not picked up by the standard Census questions, such as those experienced by patients travelling to receive treatment.

The paper identifies a range interaction data sources that have the potential to provide information about the time series context for census data and the trends in interaction behaviour between censuses or for sub-national population estimation and projection. Some of these data are currently available online from internet web sites but may require licensing agreements if they are to be made accessible via WICID; others are much less accessible, have disclosure controls limitations, and require extensive derivation or estimation effort if their potential is to be realised; in all cases, the process of data gathering and assembly requires the establishment of good collaborative working relationships with the providers as well as detailed data provision and licensing agreements.

Whilst British population censuses are accepted as providing the most comprehensive and most reliable interaction data that are available, particularly for small area studies, several of the potential non-census interaction data sets originate from administrative sources and involve the collection of records arising from some transaction, registration or as a record of service delivery. They are collected for administrative rather than purely research purposes and many of these data sets are collected by

Government departments. Jones and Elias (2006) have recently completed a selected audit of administrative data sets, most of which are used to provide stock information. These data sources include variables that provide information about either the migration or the commuting characteristics of NHS patients, school pupils, university students, workers and those attending hospital. In some cases, registration data have much simpler structure than census data and are only available at a relatively aggregate spatial scale but are particularly valuable because they are produced on a regular temporal basis. In other cases, the information on migration or mobility has to be generated from the primary unit data using time-consuming data matching and manipulation algorithms.

Surveys are the third main source of interaction data and, in many cases, surveys such as the Labour Force Survey (LFS) or the International Passenger Survey (IPS) provide reasonably detailed data on migration questions but are of limited value because their sample sizes allow only restricted spatial coverage. The IPS data on immigrants and emigrants are only published at regional scale and even then, users are advised to smooth out irregularities in the data by calculating three year averages. However, in many cases, survey data are particularly valuable because of the cross-classification possibilities that are available with primary unit data, even though the geographical dimension may be limited.

The availability of non-census registration and survey interaction data would increase interest in and use of the WICID system provided by CIDER for academic researchers and students. Furthermore, this development is in line with ESRC emphasis on a more integrated National Data Strategy, a plan to ensure that the data resources required to inform future research issues are in place (Elias, 2006). Under the current ESRC award, CIDER is aiming to assemble non-census-based data that would be collected and incorporated into WICID from the beginning of the second year of the Census Programme (August 2007), commencing with the inclusion of a set of historical NHSCR data and their respective populations at risk that will be explained later in the paper. The integration of this NHSCR data set within WICID will necessitate re-design of the interface and the metadata framework but will provide CIDER with a ‘test case’ in advance of incorporating other time series data sets. CIDER will seek to

establish the feasibility of incorporating interaction data from these and other potential sources during the course of the exercise.

This paper provides a review of the potential sources of interaction data that exist in the UK together with information about their definitional characteristics, estimation methods, attributes, limitations and availability. Sections 3 and 4 deal respectively with different administrative and survey data sources. Section 5 contains our assessment of the various data sources identified and reported in the paper and makes a series of recommendations as to which data sets might be most usefully be incorporated into a future enhanced version of WICID. However, we begin in Section 2 with a short synopsis of Census products including the Origin-Destination Statistics.

2 CENSUS INTERACTION DATA SOURCES

The Census of Population is an extremely important source of two types of interaction data: migration flows taking place in the 12 months prior to the census, and flows of commuters to work (in the UK) and to study (in Scotland). There are various census products from which interaction data can be extracted: the main census tables; the Special Migration, Workplace and Travel Statistics; commissioned tables; the Samples of Anonymised Records; and the Longitudinal Study. These sources are discussed in turn with particular reference to data available from the latest (2001) Census.

2.1 Census Tables

The Office of National Statistics (ONS) produces a range of tables of counts based on the 2001 Census. These include: Key Statistics (KS); Standard Tables (ST); ST Theme Tables; Census Area Statistics (CAS); CAS Theme Tables; CAS Univariate Tables; and Armed Forces Tables. The KS tables provide an overview and summary of the main topics of the 2001 Census in a limited number of simple univariate tables for output areas (OAs) ─ the smallest geographical units of the 2001 Census. The ST data sets provide the most detailed information available from the 2001 Census in a large number of cross-tabulations but only down to ward level in England, Wales and Northern Ireland, and postcode sector level in Scotland. The CAS data are roughly equivalent to those covered in the ST data sets but are available at OA scale like the KS and are less detailed in order to protect the confidentiality of personal information. Armed Forces tables provide information on members of the Armed Forces and data are available down to local authority district (LAD) level for England and Wales only.

Certain tables from amongst these contain interaction data, including:

- KS24: Migration (All people);
- ST008: Resident type by age and sex and migration;
- ST009: Age of household reference person (HRP) and number of dependent children by migration of households;
- ST010: Household composition my migration of households;
- TT033: Migration (People): All people in the area and those who have moved from the area in the past year, within the UK

KS24 provides counts of migrants in various categories including those moving into an 'area' from elsewhere in the UK (in-migrants) and from outside the UK (immigrants), those moving within the area and those moving out of the area (out-migrants) during the previous 12 months. There is also a category for those recorded with no usual address one year ago, some of whom may be in-migrants or immigrants. Table 2.1 presents detail from Table KS24 at district level, showing this classification and its extension to include those 'People in ethnic groups other than White' who move.

Table 2.1. Detail from KS Table 24 at district level

Table KS24 Migration												
All people												
	Who moved into the area from within the UK	Who moved into the area from outside the UK	No usual address one year ago	Who moved within the area	Who moved out of the area	All People in ethnic groups other than 'White'	People in ethnic groups other than 'White' who are migrants	Who moved into the area from within the UK	People in ethnic groups other than 'White' who moved into the area from outside the UK	People in ethnic groups other than 'White' with no usual address one year ago	People in ethnic groups other than 'White' who moved within the area	People in ethnic groups other than 'White' who moved out of the area
The North												
D33700 Accrington Urban Area	2,140	155	633	5,052	2,349	6,595	732	118	74	146	394	169
D33701 Accrington	1,429	89	364	2,112	1,854	5,300	565	99	51	121	294	132
D33703 Church	325	4	39	108	314	100	21	11	-	5	5	13
D33704 Clayton-le-moors	615	25	75	282	527	139	21	8	10	3	-	11
D33705 Great Harwood	644	23	74	515	530	856	104	33	12	16	43	22
D33702 Oswaldwistle	674	14	81	488	670	200	21	10	-	-	9	34
E10200 Ackworth Moor Top	333	22	29	112	351	131	19	4	13	3	-	20
D40400 Addingham	212	10	18	60	238	40	3	3	-	-	-	-
D90300 Adlington	385	26	34	275	399	97	10	3	3	-	4	4
E11400 Adwick le Street	444	18	83	446	551	91	20	9	6	3	3	3
B10100 Alnwick	517	29	60	383	359	37	8	7	-	3	-	6
E82100 Alsager	1,156	34	66	816	894	210	49	36	-	3	10	25
B10200 Amble	284	-	27	363	245	16	6	5	-	-	-	-
C00200 Ambleside	367	18	34	236	263	31	7	-	3	-	3	4
E17200 Anston/Dinnington	826	43	86	868	978	210	52	20	28	-	4	14
C10100 Appleby	151	6	17	101	135	18	-	-	-	-	-	-
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Source: ONS website at http://www.statistics.gov.uk/statbase/

Table ST008, in comparison, contains a spatial breakdown that is similar to that used in KS24 except that 'areas' are distinguished from 'associated areas' within the UK. Counts of migrant flows from individual areas to the aggregate spatial units generated from the following categories:

- Lived at same address
- Lived elsewhere one year ago within same area
- No usual address one year ago
- Inflow
    - Lived outside the area but within 'associated area' one year ago
    - Lived outside the 'associated area' but within the UK one year ago
    - Lived outside UK one year ago
- Outflow

- Moved out of the area but within the ‘associated area’
- Moved outside the ‘associated area’ but within the UK.

Similar categories are used for ST009, ST010 and TT033. The term ‘area’ refers to the particular area level being shown in the table. So, in the case of an ST or CAS table for a ward, this will mean the term ‘area’ translates to the name of the ward. In England and Wales, the ‘associated area’ refers to LAD for tables at ward (electoral division in Wales); parish (community in Wales) or OA level. For all other geographical areas the ‘associated area’ is England and Wales. Data can be extracted online by registered academic users from these tables using the CASWEB service at http://census.ac.uk/casweb/, although KS Table 24 is not contained within CASWEB and must be accessed directly from the ONS web site at http://www.statistics.gov.uk. Data derived from the 2001 Key Statistics and Standard Tables have been used in analyses of patterns of internal migration by Champion (2005) and of international migration by Horsfield (2005).

There are no tables that provide interaction data on commuting equivalent to those above relating to migration other than TT011 which provides flows from each area (OA) to aggregate areas based on the distance traveled to work (<2km, 2-5km, 5-10km, 10-20km, 20-30km, 30-40km, 40-60km, and 60+km).

It is clear from the preceding discussion that the interaction data contained in the main census tables only provide a detailed geographical breakdown of either the origin units or the destination units but not both. In order to obtain flows between local authorities, wards or output areas, it is necessary to use the Origin-Destination Statistics.

2.2 Origin-Destination Statistics

As in 1991, two major migration and commuting interaction data sets are available from the 2001 Census: the Special Migration Statistics (SMS) and the Special Workplace Statistics (SWS). However, in Scotland, the SWS in 2001 were replaced with a new set of Special Travel Statistics (STS) that include journeys to place of study as well as place of work. These data sets are also known collectively as the 2001 Census Origin-Destination Statistics. They are currently accessible to members of the

academic community and data suppliers registered with the Census Registration Service via WICID (http://cider.census.ac.uk). The 2001 data sets have been reviewed by Rees et al. (2002) and Cole et al. (2002). More recently, Stillwell et al. (2005) have explained the structure of the 2001 interaction data sets, documenting the differences in the data sets between 2001 and 1991 and the problems associated with making comparisons between 1991 and 2001. Stillwell and Duke-Williams (2007) have also examined the impact of the small cell adjustment methods (SCAM) used to adjust flows in 2001 to ensure confidentiality and reduce the risk of disclosure.

A summary of the tables and counts from the 2001 and 1991 Censuses (Table 2.2) shows a similar number of tables but considerably more counts in 2001 than in 1991. Data are available in 2001 for three sets of interaction zones: level 1 involves 426 'districts' that include local authorities in Great Britain and parliamentary constituencies in Northern Ireland; level 2 includes 10,608 'interaction wards'; and level 3 contains 223,060 OAs throughout the UK. The STS for Scotland in 2001 contain counts for children aged under 16 and require additional categories in certain tables. The 1991 SWS data identified in Table 2.2 are the 10% sample of journey from home to work flows produced only at ward level and referred to as SWS Set C (Cole et al., 2002).

Table 2.2. Tables and counts in the 2001 and 1991 interaction data sets held in WICID

Data sets	Level 1 (District)	Level 2 (Ward)	Level 3 (OA)
2001 SMS	10 tables (996 counts)	5 tables (96 counts)	1 table (12 counts)
2001 SWS	7 tables (936 counts)	6 tables (354 counts)	1 table (36 counts)
2001 STS	7 tables (1,176 counts)	6 tables (478 counts)	1 table (50 counts)
1991 SMS	11 tables (94 counts)	2 tables (12 counts)	Not available
1991 SWS	Not available	9 tables (274 counts)	Not available

The data in each of these tables are available from WICID, together with data sets of flows adjusted for suppression in 1991, inflated for under-enumeration in 1991, or estimated from 1981 and 1991 data to be consistent with 2001 boundaries. The modelling methodology that underpins the latter estimation is explained in detail in

Boyle and Feng (2003). One new dimension of measurement in 2001 is the introduction of moving groups, as well as migrants and migrant households. Moving groups refer to one or more people moving from the same previous address, for which there is a moving group reference person (MGRP). A detailed report on migration and socioeconomic change in Britain's larger cities, based on data from 2001 SMS Table MG109 on moving group migration by National Statistics Socioeconomic Classification (NS-SEC), is available in Champion et al. (2007).

An ONS consultation document looking forward to the 2011 Census of Population and Housing (ONS, 2005a) indicates that migration and commuting questions similar to those asked in 1991 will be asked again in 2011 and it is likely that separate Origin-Destination Statistics will be produced once again. In order to maximise the success of the 2011 Census, the ONS carried out a test of the procedures to be used in England and Wales on 13 May 2007 on 100,000 households in five local authorities (Bath and North East Somerset, Camden, Carmarthenshire, Liverpool and Stoke-on-Trent). An equivalent test was carried out simultaneously in parts of Northern Ireland by NISRA whilst a test was carried out by GRO(Scotland) in May 2006. There are a number of questions on the ONS test questionnaire from which it will be possible to extract new interaction data. The first of these relates to visitors and simply asks for usual address, thus providing some indication of where visitors come from by age and sex. Secondly, there is the question about country of birth that allows a measure of lifetime migration to be derived but, in addition, there is a question for those born abroad about when they most recently arrived to live in England and Wales. In theory, this should enable the creation of matrices of those born overseas by origin and destination and year of entry. The familiar question relating to place of usual residence one year ago is asked, but there are also questions asking about other addresses at which an individual stays for part of the week or year. The second address is asked for together with information about the reason for staying at the second address. Reasons are categorised as 'to stay with another parent/guardian'; 'to stay when I work away from home'; 'to stay when not at university/boarding school'; 'my holiday/second home'; 'to stay when I'm on duty (armed forces)'; and 'other'. There is also a question about how long the individual stays at the second address: 'less than half the time', 'about half the time' or 'more than half the time'. These

questions have the potential to generate a considerable amount of interaction data relating to temporary mobility.

A major issue relating to 2011 interaction data is what form of adjustment will be used to the data to preserve confidentiality. It is encouraging to see that the Registrars General of Scotland, England and Wales and Northern Ireland have agreed to aim for a common Statistical Disclosure Control (SDC) methodology for 2011 Census outputs (Office for National Statistics, General Register Office for Scotland, Northern Ireland Statistics and Research Agency, 2006) based on the principle of protecting confidentiality set out in the National Statistics Code of Practice. It is considered that 'attribute disclosure' (i.e. learning something about an individual) as opposed to 'identification' of an individual is the key disclosure risk, because identification reveals no new information to the user. *"*In order to meet the agreed interpretation of the Code of Practice, it has thus been agreed that small counts (0s, 1s, and 2s) could be included in publicly disseminated Census tables provided that a) uncertainty as to whether the small cell is a true value has been systematically created; and b) creating that uncertainty does not significantly damage the data*"* (Office for National Statistics, General Register Office for Scotland, Northern Ireland Statistics and Research Agency, 2006, p.1). The exact threshold of uncertainty required has not been decided. Different levels of disclosure control are applied to outputs from the Census according to the mode of access. As we have seen from analysis of the 2001 data (Stillwell and Duke-Williams, 2007), Origin-Destination tables at low geographical levels are amongst those tables likely to be seriously compromised by SDC and it is interesting to note the view of the Census agencies that these tables could be released under access arrangements (e.g. licence or safe setting) where restrictions on access to the data allow less stringent levels of SDC to apply. The recommended SDC methods for all types of 2011 Census outputs will be published in Autumn 2008 for consultation and finalised in Spring 2009.

2.3 Commissioned Tables

Customised output from the 2001 Census may be commissioned from ONS Customer Services when particular cross-tabulations are not available from the standard tables, but commissioned tables incur charges to recover staff and material costs. Once a

table has been delivered and paid for by a customer, it is listed on the ONS website (http://www.statistics.gov.uk/census2001/downloads/com_tab_finder.xls) and is available to all users free of charge on request from the Census Customer Services. All commissioned tables for 2001 data are subject to checks to ensure confidentiality. There is a function on the commissioned tables spreadsheet that allows you to identify tables of interest by entering the topics of interest. By way of example, Table 2.3 is the list of commissioned tables produced when 'origin' and 'destination' appear in the table titles.

A key question is whether it is useful to incorporate some or all of these tables into WICID. Given that many of the tables contain data relating to particular parts of the country (e.g. London, Berkshire, Avon), it seems appropriate to consider including only those commissioned data sets that have national coverage and which cannot be built from existing data held in WICID. Examples might include migration flows between LADs disaggregated by ethnic group and age and sex or international immigrant flows by ward by country of origin.

Table 2.3: Commissioned tables with 'origin' and 'destination' in the table title

Table No.	Table Title	Geography	Table Population	No of XLS Files	Added to Website
C0188a	Origins and Destination of migrants by Age and Sex	Area of Residence at Census: City of Lo	All people	1	24-Feb-05
C0189	Origin and Destination of migrants by Age and Sex	Government Office Regions in England, S	All people	1	07-Dec-04
C0198	Origin and Destination of migrants by Ethnic group, Age and	Area of residence at Census : City of Lo	All people	1	13-Dec-04
C0199	Origin and Destination of migrants by Age, Sex and Ethnic g	Area of residence at census : Metropolit	All people	1	14-Dec-04
C0202	Origin and Destination of migrants aged 16-74 by Areas of w	London as Area of Residence	All migrants into and wi	1	04-Jul-05
C0203	Origin and Destination of migrants aged 16-74 by areas of W	UK (Excluding London) as Areas of Resi	All former London resid	1	04-Jul-05
C0204	Origin and Destination of migrants aged 60 and over in Medi	London Boroughs as Areas of Residence	All migrants aged 60 ar	1	05-Jul-05
C0205	Origin and Destination of migrants aged 60 and over in Medi	South East; Rest of England (excluding	All former London resid	1	05-Jul-05
C0210	Origins and Destination of migrants by Age and Sex	Unitary Authorities in Berkshire	All people	1	25-Feb-05
C0211	Origins and Destination of migrants by Age and Sex	Greater London, Virtual Counties in GOS	All people	1	24-Feb-05
C0216	Origin and Destination of migrants by Age and Sex	Unitary Authorities in Avon	All people	1	06-Sep-04
C0216a	Origin and Destination of migrants by Age and Sex	Area of Residence at Census: Unitary A	All people	1	24-Feb-05
C0217	Origin and Destination of migrants by Age and Sex	Virtual Counties in GOSW, Each other G	All people	1	24-Feb-05
C0281	Origin and Destination of migrants by Age and Sex	Central, Inner and Outer London	All people	1	06-Sep-04
C0330	Origins and Destination of migrants by Age and Sex	East England(GOR), South East Englan	All people	1	26-Nov-04
C0342	Origins and Destination of migrants by Age and Sex	Area of Residence at Census: City of Lo	All People	1	24-Feb-05
C0343	Origins and Destination of migrants by Age and Sex	Area of Residence at Census: London B	All People	1	24-Feb-05
C0497	Workplace table - Origin Destination - Sex by Highest Level	Area of residence : CAS wards in Londo	All people aged 16 to 74	7	02-Nov-06
C0498	Workplace table - Origin Destination - Age by Highest Level	Area of residence : CAS ward in Londor	All people aged 16 to 74	1	12-Feb-07
C0528	Origin and Destination of migrants by Ethnic group	Local Area Districts (LADs) in England	All migrants resident in	2	13-Sep-05
C0609	Age and Country of birth by Origin and destination of migrant	England and Wales; England; West Mid	All People	2	26-May-06
C0648	Origin destination of migrants by religion	Local authorities in England and Wales	All migrants	1	31-Aug-06
C0649	Origin and destination of workers by religion	Local authorities in England and Wales	All people aged 16 to 74	1	31-Aug-06

Source: http://www.statistics.gov.uk/census2001/downloads/com_tab_finder.xls

One specific set of interaction data that ONS have considered producing at a national level are origin-destination statistics for super output areas (SOAs), both at middle and lower order levels. However, Keith Spicer, an ONS statistician, has indicated that there are problems with producing these data because of the differencing between

wards and the SOA hierarchy. ONS have looked that this problem specifically for Wales and found that there were many instances where it would be possible to difference and effectively generate OA to area x counts or area x to OA counts and even OA to OA counts with considerable detail on individual Census respondents. The position is similar in England and hence it is not possible to supply any of the origin-destination tables at SOA level where they have previously been released for wards. Whilst it is possible for CIDER to produce SOA-SOA interactions flows by aggregation of OA-OA matrices using look-up tables, these will inherit the errors associated with the application of SCAM at the OA level.

2.4 Samples of Anonymised Records

The Samples of Anonymised Records (SARs) were introduced as a new innovation in the UK as one of the outputs of the 1991 Census, and offer a considerable degree of flexibility for multivariate analysis of individuals (Dale, 1998). The SARs comprise a set of records relating to individuals and (where appropriate) households, with personal data such as names and addresses removed. Some fields in the records retain the full original coding, whereas others are recoded in order to prevent disclosure of information for identifiable individuals. There are a variety of spatial variables in the SARs, including residential location at the time of the Census, location of address one year ago for migrants, and country of birth. Some locational variables are simple aggregates of more detailed base values, for example ‘county’ in the 1991 Individual SAR, which is aggregated from the more detailed SAR area variable.

In order to generate 'interaction data', any two spatial references can be cross-tabulated, with possible disaggregation by any other chosen variable(s). In keeping with the focus of this audit, the commentary on the SARs will concentrate mainly on migration and commuting aspects of the data, with additional reference to other potential interaction matrices.

2.4.1 1991 SARs

A total of four SAR files were generated from the 1991 Census; two relating to Great Britain, and two relating to Northern Ireland. It is possible to coalesce these into a single UK pair of files by limiting the variables to those that are consistent between the two sources. In both cases, there was an Individual SAR (a 2% sample of

individuals in the Census), and a Household SAR, a 1% sample of households. The Household SAR contains a household record for each sampled household, followed by a set of individual records for each individual within the sampled household. The Individual SAR includes limited household level information for each sampled individual, but does not include detailed information about other persons in the sampled person's household.

In the case of Great Britain, the SAR records were extracted from the 10% set of Census records for which all entries on the original Census form had been coded. The Household SAR was extracted first and the relevant records removed; the records for the Individual SAR were extracted from the remaining set. Thus, there was no overlap between the individuals or households contained within either SAR. The Northern Ireland (NI) SAR files were produced in a similar manner, however there had been 100% coding of all forms in Northern Ireland, and thus the SARs were drawn from the entire population. In both cases, the sample excludes fully imputed households, and more obviously, households missed by the Census.

Separate geographies were used for the primary reporting areas (i.e. the location at the time of the Census) of the Individual and Household SARs. The individual SAR was based on LADs, and includes all districts with a population of at least 120,000 persons. LADs with a smaller population than this were amalgamated with neighbouring LADs. These rules resulted in 278 areas being recognised in Great Britain, and 10 areas in Northern Ireland. For the household SAR (and for some broad coded variables in the Individual SAR), the geography consists of 12 zones based on the statistical Standard Regions of Britain, with the South East being subdivided into Inner London, Outer London and Rest of South East. Northern Ireland was recognised as an additional region where appropriate.

Table 2.4 below summarises the type of geography and number of areas (in square brackets) for available spatial variables in the SAR data sets. The term ‘Categorised’ in Table 2.4 is used to refer to a potential spatial variable (i.e. a variable for which a detailed spatial reference was collected on the original Census form) that is not included in a spatially disaggregate form in the SAR. For example, workplace

location was collected using the postcode of the place of employment, allowing detailed data sets such as the Special Workplace Statistics to be generated. However, in order to reduce the risk of disclosure, this detail is not included in the SAR. Instead, workplace information is presented as a categorical variable, with classes including 'At home/No Fixed Place', 'In SAR area', 'Outside SAR area, inside GB', et cetera. Similarly, term-time address for students has categories such as 'This address' and 'Elsewhere in same region'.

Table 2.4: Type of geography and numbers of areas for spatial variables in the 1991 SAR data sets

	1991 GB Individual SAR	1991 GB Household SAR	1991 NI Individual SAR	1991 NI Household SAR
Sample size	2%	1%	2%	1%
		Geographic variables		
Location at time of Census	SAR area [278]	SAR region [12]	NI SAR area [10]	Not further coded
Migrant origin	SAR region [12] + outside GB + n/a et cetera	SAR region [12] + outside GB + n/a et cetera	Collapsed SAR region [11] + outside GB	Collapsed SAR region [11] + outside GB
Country of birth	Grouped country of birth [42]	Grouped country of birth [42]	Grouped country of birth [42]	Grouped country of birth [42]
Visitor area	SAR region [12] + outside GB	SAR region [12] + outside GB	SAR region [12] + outside GB	SAR region [12] + outside GB
Term time address	Categorised	Categorised	Categorised	Categorised
Workplace	Categorised	Categorised	Categorised	Categorised

Whilst the SAR area geography of the Individual SAR is reasonably detailed, there is little spatial detail available in other variables. Migrant origin is limited to the standard region based geography, and workplace location is only available as a broadly coded categorical variable, or as a broadly coded 'distance to workplace' observation. This limits the potential for use of the SAR as interaction data. Nevertheless, the multivariate flexibility of the SAR means that there is considerable scope for analysis of the characteristics of migrants and commuters at an aggregate level. It is possible to generate cross-tabular frequency counts of total migrants by SAR area destination or total commuters travelling into a SAR area workplace location that are based on any other Census variables, thus including ones that are not used in the SMS or SWS.

Because of their large sample size and the ability to cross-tabulate variables not available from the main census tables, the 1991 SARs have been used to identify the characteristics of migrants. One example of this is the migration of the elderly to join existing households by Al-Hamad et al. (1997). The 1991 SARs have also been used to assess the impacts of tenure on long-distance migration compared with short-distance migration by Boyle (1995), suggesting that long-distance migrants are less likely to move into council housing than other tenures. Unfortunately, the potential for use of SAR data as interaction data is particularly limited in the Northern Ireland SAR – the migrant origin variable includes an 'outside GB' origin class, but there is no distinction between migrants from Northern Ireland and those from outside the UK.

2.4.2 2001 SARs

The range of SAR files produced was expanded with the 2001 Census, with five files being produced. The 2001 SAR outputs comprise: the Individual SAR (Licensed), the Individual Controlled Access Microdata Sample (Individual CAMS), the Special License Household SAR, the Household Controlled Access Microdata Sample (Household CAMS) and the Small Area Microdata (SAM). These files have been generated on a UK basis, with the exception of the Special License Household SAR, which was produced for England and Wales only. The Household CAMS includes data for the whole of the UK, in contrast to the Licensed Household SAR. As with the 1991 SARs, there is no overlap between any of the 2001 SAR files.

The two CAMS files offer more detailed versions of the respective licensed files, and they are made available under more restrictive conditions. These files are only made available by application and in a safe setting; that is, researchers have to travel to a secure location (currently, ONS offices at Titchfield, London, Newport and Southport), and use the data on secure machines in those locations.

As with the 1991 SAR files, there are several potential locational variables which could be used to generate interaction data. For the Individual CAMS, a much more detailed residential geography is available than was the case with the 1991 SARs. As before, it is based on the LAD but with much lower thresholds used for amalgamation. There are only three cases in which LADs with small populations have been merged with neighbours. For other variables, the spatial geography is similar to

that of the 1991 SARs, being based on an expanded Government Office Region (GOR) geography. The formal GOR geography refers only to England, and contains 10 regions. The expanded GOR geography used in the 2001 SAR improves on this by adding Wales, Scotland and Northern Ireland as additional regions, and by splitting London into two regions, Inner and Outer London. This geography is extended where appropriate for some variables; for example to include 'Outside UK' et cetera for migrant origins. A summary of the resolution used to report various spatial variables is shown in Table 2.5.

Table 2.5: Type of geography and numbers of areas for spatial variables in the 2001 SAR and CAMS data sets

	2001 Individual SAR (Licensed)	2001 Individual CAMS	2001 Special Licensed Household SAR	2001 Household CAMS	2001 Small Area Microdata
Sample universe	UK	UK	E&W	UK	UK
Sample size	3%	3%	1%	1%	5%
			Geographical variables		
Location at time of Census	Expanded GOR [13]	Modified LAD [423]	n/a	Modified LAD [423]	Modified LAD [423]
Migrant origin	Expanded GOR [13] + outside UK	Modified LAD [423]	Categorised	Categorised	Expanded GOR [13] + outside UK
Country of birth	Broad grouping [16]	Detailed country of birth [329]	Broad grouping [16]	Detailed country of birth [329]	UK components [4] + outside UK
Term time address	Categorised	Categorised	Categorised	n/a	Categorised
Workplace	Categorised	Categorised	Categorised	Categorised	Categorised

Whilst the sample size for the Individual SAR has increased from 2% in 1991 to 3% in 2001, the value for use as interaction data is diminished, due to the reduction in resolution of the primary geography from 278 regions (in GB) to 13 regions in 2001. This is disappointing, although the potential for use in aggregate analysis of characteristics of migrants or commuters remains significant.

The Licensed Household SAR has the same sample size as used for the 1991 Household SAR, but has little or no potential for use as interaction data, due to the removal of the primary geography (i.e. area of residence at the time of the Census) in order to reduce the risk of disclosure. At the same time, other spatial variables including migrant origin have been included only as a categorical variable indicating whether or not the migrant origin was in the same district et cetera.

The two CAMS files have more potential for use in interaction data analysis. The Individual CAMS has a LAD based geography for both migrant origins and destinations, thus offering similar spatial detail to 2001 SMS Level 1. The data can be disaggregated by any chosen variable, although the sample size, coupled with the generally low incidence of one-year migrants in all Census data (around 12% of individuals were identified as migrants) will tend to restrict the ability to carry out multivariate analysis. The household CAMS file has a detailed primary geography, but only a categorical version of migrant origin. Both CAMS files feature very detailed versions of the country of birth variable, allowing spatially detailed analysis of life-time mobility.

The 2001 outputs also saw a new flavour of microdata: the Small Area Microdata (SAM). This is an individual sample (with the same sample size as the 2001 Individual SAR) which sacrifices attribute detail in order to permit greater spatial detail. At the same time, the sample size is increased to 5%. For migration analysis, the SAM has the advantage of a detailed destination geography, although the origins remain as the expanded GOR geography. Thus, for the study of in-migrants, considerable detail can be discerned. Crucially, this can be done with multiple variables; the researcher is not limited to the tables produced as part of the SMS. However ─ as with all other 2001 SARs and the 1991 SARs ─ workplace address is provided solely as a 'movetype' classification, meaning that the data are not suitable for use as interaction data.

2.4.3 Summary

The 1991 SARs data, and the licensed versions of the 2001 SARs data are of limited use for spatially detailed analysis of the data as interaction data. However, they retain the general advantage of microdata as an opportunity for flexible multivariate analysis, and thus have potential use for the aggregate study of characteristics of those involved in spatial interactions (i.e. migrants and commuters). In general, the SARs are more useful for interaction data use with respect to migrants than to commuters, as there is no spatial coding of workplace location. The focus on local aspects of commuting (e.g. whether or not the individual works and lives in the same district) limits usefulness for some applications, such as destination specific analysis of the

characteristics of long-distance commuters. There are other examples of spatial data collected in the Census but not included in the SAR. For migrants from overseas, the country of origin was collected, but this is not used in the SAR data.

The 2001 SAM improves on the spatial resolution used in the licensed versions of the 2001 SAR data. This increases the potential for use as interaction data, although the coding of migrant origin remains coarse, and there is no spatial disaggregation of workplace location. The 2001 CAMS files also offer improvements over their licensed counterparts. In particular, the 2001 Individual CAMS offers considerable scope for interaction based analysis of migrants. Both origin and destination are coded at LAD level (with the exception of a very small number of amalgamations), and country of birth is also available with a detailed coding. The file is available on a UK basis, and has an increased sample size compared to that of the 1991 Individual SAR. The main drawback (aside from the access arrangements) is that the sample size is still too small to allow analysis of migrants using more than a few variables.

2.5 Longitudinal Studies

2.5.1 What are they?

Longitudinal studies are data sources that contain multiple observations of a population of interest over a period of time. This includes both surveys which are repeated at intervals for a known set of respondents, and more general instruments from which a sample is extracted, and externally linked to records for the same persons from earlier collection rounds. Examples of the former type of longitudinal study include the UK Cohort Studies, in which a selected sample are surveyed in multiple sweeps over the course of their life. Examples of the latter type include the census based Longitudinal Study, which is derived from samples extracted from each decennial census. Data of these kind provide a valuable resource for many researchers in a variety of fields. This review concentrates on their potential for use specifically as a form of interaction data.

2.5.2 Longitudinal Studies in the UK

There are three major longitudinal studies in the UK that are based on census data, with linked administrative records from other sources including vital events and

registration data. These are: the ONS Longitudinal Study of England and Wales the Scottish Longitudinal Study, and the Northern Ireland Longitudinal Study. These differ in a variety of ways including the length of the time period covered, the sampling fraction used, and the types of other data linked into the study.

The ONS Longitudinal Study of England and Wales (originally known as the OPCS Longitudinal Study) is the longest established of the three studies. It contains linked data from 1971, 1981, 1991 and 2001 Censuses. The sample is selected on the basis of four (undisclosed) birth dates, giving a sample fraction of around 1%. Persons born on one of these days are extracted from each Census and attempts are made to link them to established records from earlier Censuses or to administrative records. In addition to the core sample members, records are also extracted and added to the LS for other persons in the sample member's household, although these additional persons are not (unless they also happen to be a sample member) tracked in later censuses, unless they are still living in a sample member's household. This study is generally referred to as 'the LS' (that is, the short acronym is used to refer to the LS of England and Wales, rather than one of the other two). It also contains linked vital events data and cancer registrations. Access to the LS is facilitated by the Centre for Longitudinal Study Information and User Support (CeLSIUS) (http://www.celsius.lshtm.ac.uk/).

The linked LS has enabled researchers to examine changing patterns of settlement and local geography as well as factors affecting long-term migration. The link between inter-regional migration and social mobility has been explored by Fielding (1992) to identify the South East region as an 'escalator'. The relationships between counterurbanisation and social mobility have been investigated by Fielding (1998) and various studies have tracked the spatial distribution of the population in different parts of the country (Williams, 2000; Davies et al., 2006), migration relating to health and deprivation (Norman et al., 2005) and the geographical and social dynamics of ethnic groups (Platt et al., 2005).

The Scottish Longitudinal Study (SLS) has been established by the Longitudinal Studies Centre – Scotland (LSCS) (http://www.lscs.ac.uk/sls/), currently based in the University of St Andrews and the offices of the General Register Office for Scotland

(GROS) in Edinburgh. The SLS is similar in general terms to the England and Wales LS, although there are some significant differences. It is a continuous study, incorporating data from the 1991 and 2001 censuses. It is a 5.5% sample, based on 20 birth dates, rather than a 1% sample, based on four birth dates, as in England and Wales. The SLS will provide linked information for approximately 274,000 individuals in Scotland. For each individual, the SLS has all the variables that can be extracted from the complete 1991 and 2001 census forms. Census data includes a range of information including occupation, economic activity, housing, ethnicity, age, sex, marital status, health, education and religion as well as place of usual residence 12 months before the census and details relating to the journey to work or study.

Data are also collected for the sample members on vital events. Events currently being collected include births, stillbirths and infant mortality occurring to sample members (where the mother and/or the father is an SLS member), widow(er)hoods (where the SLS member is the surviving spouse), deaths, re-entries after earlier emigrations, and marriages (where the bride and/or groom is the sample member). In Scotland, health data are highly confidential and sensitive, so linkage to cancer and hospital admissions data are not permitted. However, a method of flagging the NHS Information Statistics Directorate (ISD) data sets with encrypted SLS identifiers has been devised which will allow the linkage of particular variables for SLS members as required and Scottish cancer registration data will be linked into the SLS data as requested for specific approved research studies. In addition, however, it is possible in Scotland to make the same linkage into hospital admissions information.

The SLS has been designed to be compatible with the England and Wales LS so that an integrated British LS is a realistic possibility in the future (four of the 20 birth days match those used in the England and Wales LS). It should be possible for researchers to select similar datasets from the LS and SLS to allow comparisons to be made throughout Britain, although care will need to be taken about how migrants between England and Wales and Scotland are handled.

The Northern Ireland Longitudinal Study (NILS) is the most recently started study, and contains data from the 2001 Census only. Unlike the other two studies which are

supported by ESRC Census Programme units, the NILS is supported directly by the Northern Ireland Statistics and Research Agency (NISRA) (http://www.nisra.gov.uk/nils/). NILS members are selected on a total of 104 birth dates, giving a much larger sample size than the other two studies of around 28%. The linked administrative data include birth and death registrations, health service related migration data, and information about members' households from the Valuations and Land Agency.

Finally, a UK Household Longitudinal Study (UKLHS) has been funded by the ESRC and the Office for Science and Innovation (OSI) Large Scale Facilities Programme and is due to start its first wave of data collection in 2008. The UKLHS will consist of three components: (i) a wholly new sample of households; (ii) an ethnic minority boost sample, and (iii) a sample (up to 100%) drawn from the existing BHPS. It will yield a sample of at least 40,000 households, making it the largest type of study of its kind in the world. The BHPS will be integrated into the UKLHS at wave 2 (2009-2010) and the Longitudinal Study of Ethnic Minorities (LSEM) will also be incorporated into the UKHLS.

2.5.3 Cohort Studies in the UK

There are four significant birth cohort studies in the UK as follows:

- the MRC National Survey of Health and Development (NSHD) (the British 1946 birth cohort study);
- the National Child Development Study (NCDS) (the 1958 birth cohort study, originally known as the Perinatal Mortality Survey);
- the 1970 British Cohort Study (BCS1970); and
- the Millennium Cohort Study (cohort born in 2000/2001).

The NSHD is currently maintained in the Department of Epidemiology and Public Health at University College London, the later three studies are maintained by the Centre for Longitudinal Studies at the Institute of Education. Data for the latter three studies ares available via the Economic and Social Data Service (ESDS) (www.esds.ac.uk).

Whereas the longitudinal studies are based around linked census data, which contain – in general – the same questions each time; the birth cohort surveys use different questionnaires in each sweep. These tend to contain core questions that are asked at each sweep, plus additional questions that reflect changing interests and research priorities. Clearly, the questions that are asked to (the parents of) young children in the earliest waves of any birth cohort study will be very different to those asked as the survey members grow to adulthood and subsequently into retirement. Whilst adult members of the earlier cohort studies have been asked numerous questions in each wave about employment and occupation related issues, it would appear from examination of the available data that specific questions about the location of members' workplaces have not been regularly asked. Thus, the potential for use as journey-to-work interaction data is very limited. In contrast, the very nature of the studies, which track individuals over time, means that a near complete record of residential history is maintained, giving rise to very rich migration based interaction data. However, as discussed below, the potential for such use in practice is limited.

Other significant cohort studies, not based directly on birth cohorts include the British Household Panel Survey (BHPS); the English Longitudinal Study of Ageing (ELSA); and the Longitudinal Study of Young People in England (LSYPE).

The BHPS is a sample of all adult members of households in the UK. In the first survey in 1991, just over 5,000 households were selected. Membership of the survey is a characteristic of individuals in households, and thus when survey members move into a new household, all adult members of the new household are added to the sample; the survey now covers some 10,000 households. Surveys are conducted annually, with core questions including those relating to household composition, residential mobility, education, health and occupation. In addition, a varying set of additional questions are asked, based on emerging policy and research priorities. Access to the BHPS is supported by ESDS and the ESRC United Kingdom Longitudinal Studies Centre (http://www.iser.essex.ac.uk/ulsc/bhps/).

The ELSA is a sample drawn from existing respondents of the Health Survey for

England (HSE). The initial sample (surveyed in 2002) was of around 12,000 persons aged 50 or over, together with their partners. Being based on the HSE, the survey contains health data (including basic physiological measurements and blood samples analyses) as well as the basic demographic data common to the other surveys discussed in this section. There have been two additional surveys since the first one.

The LSYPE is the newest of these cohort studies. Sample members are young people who were in Year 9 (or equivalent) in schools in England in February 2004 (born between 1 September 1989 and 31 August 1990) and their parents or carers. Interviews are conducted annually. The initial sample featured young people and adults in over 15,000 households. The data have been linked to a number of administrative data sets including the Pupil Level Annual Schools Census (PLASC), although these linked versions are not available for direct use. The questions asked include a full history of schools attended, potentially allowing interaction data of the journey-to-school type to be derived.

2.5.4 Potential for use as interaction data

Both the cohort studies and the longitudinal studies have considerable potential scope for research into spatial mobility. As indicated above, the LS has been used already to study migration. Indeed, given that migration is an event that takes place (for most people) on a relatively infrequent basis, these types of study are far better at capturing an effective picture of mobility during the lifecourse than a series of discrete snapshots (i.e. the censuses considered in isolation). However, this high potential is also paradoxically something of a constraint on the data sources. The amount of data held means that these sources have the potential to be highly disclosive of the sample members. For this reason, access to the data is generally strongly controlled. Where relatively free access is allowed (data from the many of the cohort studies can be downloaded by academics having agreed to usage conditions), the level of detail that is used for 'spatial' fields tends to be very limited. Where any information is given about place of residence, it tends to be no more detailed that Government Office region. Furthermore, little or no data is released about the previous residential address, except whether or not it was in the UK. In contrast, data are included in a number of these studies about reasons for moving from a previous address.

The longitudinal studies, being census based contain – potentially – the full census data on previous residences and on workplace location. However, due to disclosure control reasons, it is unlikely that data will be released that contains a detailed spatial coding. Access to the LS data sets is mediated through experts who advise on requests and who are also responsible for vetting output. For any output to be released it must satisfy disclosure control regulations.

3 ADMINISTRATIVE INTERACTION DATA SOURCES

3.1 NHSCR Movement Data for England and Wales

Measuring internal migration between censuses is not straightforward, as there is currently no compulsory system within the UK to record the movement of the population. It is necessary to use 'proxy data' to make estimates of migration. One source of migration estimates is the registration system that records National Health Service (NHS) patients who migrate and change their doctor. The NHS Central Register (NHSCR) at Southport records movements of patients between Health Authority (HAs) areas in England and Wales and the Census Office has developed systems for capturing the reporting of re-registrations of patients between areas used to administer the general practitioner services of the NHS.

Since the early 1980s, individual anonymised records from the NHSCR known as Primary Unit Data (PUD) have been available from the Office of Population Censuses and Surveys (OPCS), now the Office of National Statistics (ONS), in quarterly data files. Whilst researchers have highlighted many of the conceptual and definitional characteristics of the data and commentators have welcomed its availability as a source from which to gain some understanding of migration trends between censuses of population, the methods of creating and processing the re-registration information centrally have remained something of a mystery.

A time series of NHSCR data for a set of 100 zones in the UK was assembled in the School of Geography at the University of Leeds from 1975 to 1998 (Duke-Williams, 2004) and it is these data that we seek to explain in the following sections. We begin with a short review of the major characteristics, benefits and limitations of the NHSCR data since these are explained in detail elsewhere (Stillwell et al., 1992). The variables, formats, coding systems, and derivation and imputation methods used in processing the registration data are described and then the data outputs are exemplified and a number of quality issues are explored, such as the extent of imputation, the existence of matching records, the timing of registrations and the implications of imputing the date of migration. In order to avoid the problem of seasonality in migration behaviour, the data are usually aggregated from quarterly

periods into twelve month periods. In the final subsection, we illustrate annual fluctuations in the annual time series of NHSCR data at different spatial scales over the late 1970s, 1980s and 1990s.

3.1.1 NHSCR data for 1975-1998

NHSCR data have become an increasingly important resource for the continuous monitoring of population redistribution patterns and trends (Stillwell, 1994) and were used during the 1990s in the preparation of sub-national population estimates and projections, as reviewed by Stillwell and Debenham (2001). The spatial system for recording patient re-registrations changed in the late 1996 from a system of Family Health Service Authorities (FHSAs) in England and Wales and Health Board Areas (HBAs) in Scotland to one of Health Authorities (HAs). Up to this point, the NHS Central Register for England and Wales at Southport, together with the equivalent register in Scotland, provided ONS with details of all NHS patient re-registrations which involved the transfer of records between doctors in two different FHSAs. These were formerly (before 1990) known as Family Practitioner Committee (FPC) areas and were equivalent, in England and Wales, to groups of London boroughs, metropolitan districts and shire counties that existed prior to local government reorganisation in the mid-1990s. Unfortunately, the conversion of records from old FHSAs to new HAs was only undertaken during 1999, as part of the New Health Authorities Programme, and consequently NHSCR data for 1996-98 still related to the old FHSA areas in England and Wales.

The re-registration system worked as follows: the 'new' or destination FHSA with which the patient registered sent the details of the patient (name, date of birth, NHS number) to the NHSCR where the information was processed centrally before being sent to the 'old' or origin FHSA where the individual had previously been registered, informing them that the patient could be removed from their files. Entries in the NHSCR include the date of birth mentioned above together with the sex, the codes of the FHSA that the patient has been registered with in the past as well as the new FHSA code. The operation of the patient registration system has been explained in detail by Bulusu (1991).

The registration data available from the NHSCR are defined as 'movement' data and their measurement is conceptually different from that of 'transition' data available from the census (Rees, 1977). The movement approach involves a count of the number of moves that occur in a given time interval with the age of the migrant at the time of the move recorded or imputed. This differs from the census transition count of individuals who migrate at least once during a period whose age is recorded at the end of the period. Consequently, the NHSCR flows tend to be larger in volume than census counts because they should capture multiple and return moves.

The advantages and shortcomings of the data have been identified by several authors (see, for example, Stillwell et al., 1992; Champion et al., 1998). In comparison with the census, the NHSCR data have the following advantages:

- they provide a measure of all the moves that take place during a period that involve registration with a new doctor in a different FHSA unlike census data which misses multiple migrations, migrants who are born in the period and migrants who die during the period;
- they include students who register with a doctor or health service facility in the area of their further or higher education institution;
- their quarterly availability means that they provide a continuous record of migration over time rather than a once a decade snapshot like the population census;
- they should be quite comprehensive because individuals are given an NHS record at birth and those in private health schemes usually maintain their NHS registration during their lives even if they do not use the facilities provided; and
- they have the potential for monitoring migration within FHSAs and well as international migration, although the administrative mechanisms for registering these types of moves would need further consideration.

However, NHSCR data during this period also have a number of limitations in comparison with census data, including the following:

- they only provide data about the age and sex of individual migrants and no other attribute whilst the population census provides a richer compositional breakdown as well as some information about household migration;
- the propensity to register with a new doctor will vary between different migrant subgroups. Some individuals, notably young adult males, may not register with a doctor until they need to, which may be several months after they have moved, whilst others such as the elderly or pregnant women are likely to register immediately if not before they have made a move;
- they only provide data on flows between relatively coarse spatial units whereas output data from the population census can be obtained at ward and enumeration district level;
- some short distance moves that happen to cross an FHSA boundary will not be recorded because the individuals will not bother to register with a new doctor;
- individual members of HM armed forces are only recorded as entries to and exits from the services so that only the origin FHSAs of recruits and the destination FHSAs of those discharged are recorded; and
- long-stay psychiatric patients and prisoners are excluded altogether.

Despite these shortcomings, the NHSCR data has been used in a number of studies of time series trends (e.g. Bulusu, 1989; 1990; Devis, 1984; Rosenbaum and Bailey, 1991; Stillwell et al., 1992; Stillwell, 1994; Stillwell et al., 1996) and have been compared against 1971 Census data in 1971 by Ogilvy (1980), 1981 Census data by Devis and Mills (1986) and Boden (1989) and 1991 Census data by Stillwell et al. (1995). More recently, the data have been used for a major migration modelling study commissioned by the DETR (ODPM, 2002; Champion et al., 2003; Fotheringham et al., 2004) and for examining trends in internal migration by Kalogirou (2005).

3.1.2 Availability of NHSCR data

Pre-computerisation

From mid-1975, the NHSCR data were produced by OPCS on a quarterly basis. Prior to computerisation in 1990, it was assumed that there was a time lag of three months between the actual date of migration and the date at which the individual concerned

registered with a new doctor. As a result the magnetic tapes containing the data for a given quarter were assumed to refer to those events that occurred in the preceding quarter. The registrations were for 97 FPC areas in Britain, with Scotland being treated as a single area. From mid-1975 to mid-1983, OPCS supplied the data in files output from their own processing software in aggregate form which included:

- an array of aggregate origin-destination flows;
- an array of out-migration totals for each area by quinary age group and sex;
- an array of in-migration totals for each area by quinary age group and sex.

After mid-1983, when OPCS decided to release individual records (PUD), the volume of data rose significantly (to over 4 million records per year) as each out-migration record was duplicated as an in-migration record. As indicated in Table 3.1, which summarises the main changes in the record structure, data for 15 sub-divisions of Scotland became available from the third quarter of 1983 and Middlesex was split into five separate parts in 1985, creating the system of 115 zones that have remained unchanged up until 1999.

Table 3.1: Main changes in the NHSCR record structure

Period	Form of record	Zones	Description
1975Q3 – 1983Q2	Table cell	97	Data taken from a series of output tables produced by OPCS. GB divided into 97 zones with Scotland as a single zone
1983Q3 – 1985Q2	Individual record	111	PUD made available as 100% sample. Scotland subdivided into 15 HBAs
1985Q3 – 1990Q3	Individual record	115	'Middlesex' subdivided into five areas plus an unstated area
1990Q4 – 1995Q4	Individual record	115	Register computerised. Time lag reduced from 3 to 1 month for flows England and Wales. Record length increased to Include categories such as date of birth and date of migration
1996Q1– 1998Q4	Individual record	115	PUD data made available on CD
1996Q1– 1998Q4	Individual record	115	PUD data made available on CD

Computerisation

The NHSCR stopped updating their manual records at the end of 1990. The computerised system was set up during January to March 1991 and went live in April 1991. At that time, moves were recorded not by the date that they were processed at the NHSCR but by the FHSA acceptance date. The system changed in the fourth quarter of 1990, with data for the first three quarters of 1990 being produced from the old NHSCR system. The data for the fourth quarter of 1990 were produced for both the old and the new systems and thereafter, all re-registrations were entered onto the computer and the time lag between an individual moving and acceptance by the NHSCR system was reduced to one month.

One of the problems associated with computerisation was that a backlog of processing built up and this meant that, for a period of time after implementation of the computerised system, it was necessary to complete the destination FHSA registration process quickly without concern for the origin FHSA. This led to an undercount of the migration taking place during the final two quarters of 1990 which was confirmed by a comparison of the NHSCR migration from Scotland to England and Wales over this period with data on cross-border migration kept by the General Register Office for Scotland (Hornsey, 1993). Any problems caused by computerisation had disappeared by the end of 1991, and from 1991 onwards, Hornsey indicates that the migration data were of better quality.

Although there were changes in the format of the data, the essential information did not change until 1999. The only development of practical significance has been the transition to supplying the data on compact disk rather than magnetic tape in 1996, a change which has facilitated the transfer of data onto user's machines (earlier PUD files are available from ONS on CD). The NHSCR data were supplied by ONS to the National Online Manpower Information System (NOMIS) at the University of Durham and NOMIS users could access the data commencing from the quarter ending March 1984.

3.1.3 Changing administrative areas, 1996-1999

In 1996, the 98 FHSAs in England and Wales became 89 HAs and, three years later (during 1999), the work to convert all the NHS records currently held by FHSAs to the new HAs was undertaken. Scotland, Northern Ireland and the Isle of Man were unaffected by the changeover. In most cases, the old FHSAs in England and Wales simply became new HAs with no change in their boundaries; in some cases, two old FHSAs were merged in their entirety into one new HA; and in other cases, old FHSA boundaries were changed to form new HAs. The conversion took some time (12 months) since not all FHSAs were converted simultaneously. One of the problems was that, during the transition period, the re-registration data held by the NHSCR were for a mixture of old FHSAs and new HAs. ONS provided tabular information to their customers during this interim period for :

- FHSAs that had not changed their boundaries;
- HAs that had been created by merging two former FHSAs; and
- groups of FHSAs where FHSA boundaries had been changed.

During 1998, a small number of HAs had already moved some patients between FHSAs to realign boundaries in advance of the conversion programme. This meant that the NHSCR contained records of 'moves' that were the result of boundary changes rather than actual patient migrations! For example, ONS provided estimates of the so-called 'non-migrant moves' recorded for the quarter ending mid-1998 so that the 'true flows' into and out of each FHSA to the rest of England and Wales could be determined. The largest estimated inflow of non-migrants during this quarter was 2,113 to the London FHSA of Merton, Sutton and Wandsworth. There were six other FHSAs with non-migrant estimates over 1,000 and all but 29 FHSAs had some inflow adjustment. The adjustment of outflows for non-migrants affected fewer FHSAs but City and East London, Merton, Sutton and Wandsworth and Manchester all had estimated non-migrant outflows over 2,000 and the estimates for Lambeth, Southwark and Lewisham were over 5,000.

3.1.4 Quarterly data files

What information do the quarterly files contain? The information available on each movement in the NHSCR system during 1975-1998 is minimal and includes the migrant' name, sex, date of birth, and origin and destination FHSAs. The name is not contained on the record because the record must remain anonymous and, unfortunately, there is no unique identifier attached to the record. The information on input may be either correct, incorrect or partial or missing. It therefore becomes necessary to verify the data where possible and to derive or impute further variables there missing values occur. It is also necessary to indicate on each record where imputation has been undertaken. Consequently, the file of NHSCR data for any one quarter contained 15 'data items' for each non-identifiable migrant 'entity'. These items are summarised in Table 3.2.

Table 3.2: Current entity descriptions

Data item	Format	Size	Values
Migrant number	Numeric	5	00001-99999
Processing period	Alpha-numeric	2	Denotes week number for England and Wales data (01-53) and quarter and format for Scotland data (A1-D1 Scotland format 1 data) (A2-D2 Scotland format 2 data)
Processing year	Numeric	2	yy
FHSA code (New)	Numeric	3	3 digit code
FHSA code (Old)	Numeric	3	3 digit code
Sex	Numeric	1	1 = male; 2 = female
Type of move	Numeric	2	01-12
Date of birth (DOB)	Numeric	8	ddmmccyy
Acceptance date for new posting	Numeric	8	ddmmccyy
Age at migration	Numeric	3	000-125
Date at migration	Numeric	8	ddmmccyy
Date of birth imput-ation indicator	Numeric	2	10-12
Sex and acceptance date imputation indicator	Numeric	1	0 = no imputation; 1 = sex imputed only; 3 = acceptance date imputed only; 4 = sex and acceptance date imputed

True date of migration	Numeric	8	ddmmccyy
Age at mid-year	Numeric	3	000-126

Source: ONS Internal Migration Redevelopment Functional Specification

The first three items are either assigned automatically by the system or identified by the user at the start of processing the data.

- Migrant number: this number is automatically assigned to the record by the system and is unique only within the processing period and the processing year. The migrant number resets to 1 when the processing period changes. Alphabetic characters are occasionally found in addition to numeric values from 1 to 99999 and may be used when codes beyond 99999 are required.
- Processing period: for England and Wales, this refers to the processing week number input at the start of processing by the system user and therefore takes a value from 1 to 53; for Scotland, it refers to the processing quarter (A, B, C or D) at the start of processing. The NHSCR for England and Wales does not contain data on flows between HBAs in Scotland; nor does it include flows from Scotland to Northern Ireland. Flows from Scottish HBAs to the FHSAs in England and Wales are processed using the same format as the main data; flows from England and Wales FHSAs to Scotland and from Northern Ireland to Scotland have two separate formats.
- Processing year: the year in which the data are being processed which is set by the user at the start of processing. Currently the field width for this variable is two characters.

The next three variables are defined by the user.

- FHSA code (new) and (old): the codes for these two variables, the origin and destination FHSAs, are derived from short coded values (2 or 3 letters) input by the user and converted by a set of control data. Both the origin and destination FHSA must exist in the control data and the old FHSA must not equal the new FHSA. Five digit codes are provided in the control data for the 117 geographical areas in the UK (Table 3.3) where the first digit indicates the region in which the FHSA is located (0-9), the second indicates the metropolitan county or region remainder in which it is located (1-3), and the last three digits are unique FHSA codes used in the data file. Table 3.3 also shows how the FHSAs relate to the new

HAs in England and Wales, although there are boundary changes which further complicate the changeover.

Table 3.3: Codes for UK areas used in NHSCR

Region	Code and area (FHSA in England and Wales)	New area (HA in England and Wales)
Northern Ireland (06)	026 Northern Ireland	Northern Ireland
Scotland (07)	027 Ayr and Arran	Ayr and Arran
	028 Border	Border
	029 Argyll and Clyde	Argyll and Clyde
	030 Fife	Fife
	031 Greater Glasgow	Greater Glasgow
	032 Highland	Highland
	033 Lanark	Lanark
	034 Grampian	Grampian
	035 Orkney	Orkney
	036 Lothian	Lothian
	037 Tayside	Tayside
	038 Forth Valley	Forth Valley
	039 Western Isles	Western Isles
	040 Dumfries and Galloway	Dumfries and Galloway
	041 Shetland	Shetland
	042 Scotland pre-1974	Scotland pre-1974
Isle of Man (08)	043 Isle of Man	Isle of Man
Tyne and Wear	045 Gateshead	Gateshead and South Tyneside
(11)	050 Newcastle	Newcastle and North Tyneside
	055 North Tyneside	Newcastle and North Tyneside
	060 South Tyneside	Gateshead and South Tyneside
	065 Sunderland	Sunderland
North Remainder	070 Cleveland	Cleveland
(13)	075 Cumbria	Cumbria and Lancashire
	080 Durham	Durham
	085 Northumberland	Northumberland
South Yorkshire	090 Barnsley	Barnsley
(21)	095 Doncaster	Doncaster
	100 Rotherham	Rotherham
	105 Sheffield	Sheffield
West Yorkshire	110 Bradford	Bradford
(22)	115 Calderdale	Calderdale and Kirklees
	120 Kirklees	Calderdale and Kirklees
	125 Leeds	Leeds
	130 Wakefield	Wakefield
Yorks and Humbs	135 Humberside	Humberside
Remainder (23)	140 North Yorkshire	North Yorkshire
East Midlands	145 Derbyshire	Derbyshire

(33)	150 Leicestershire	Leicestershire
	155 Lincolnshire	Lincolnshire
	160 Northamptonshire	Northamptonshire
	165 Nottinghamshire	Nottinghamshire
East Anglia (43)	170 Cambridgeshire	Cambridgeshire and Norfolk
	175 Norfolk	Cambridgeshire and Norfolk
	180 Suffolk	Suffolk
Greater London	185 City, Hackney, Newham &Tower H	City and East London
(51)	190 Redbridge & Waltham Forest	Redbridge and Waltham Forest
	195 Barking & Havering	Barking and Havering
	200 Camden & Islington	Camden and Islington
	205 Kensington, Chelsea & Westminster	Kensington, Chelsea and Westminster
	210 Richmond & Kingston	Richmond and Kingston
	215 Merton, Sutton & Wandsworth	Merton, Sutton and Wandsworth
	220 Croydon	Croydon
	225 Lambeth, Southwark & Lewisham	Lambeth, Southwark and Lewisham
	230 Bromley	Bromley
	235 Bexley & Greenwich	Bexley and Greenwich
	236 Enfield & Haringey	Enfield and Haringey
	237 Barnet	Barnet
	238 Hillingdon	Hillingdon
	239 Brent & Harrow	Brent and Harrow
	240 Middlesex - not stated	
	241 Ealing, Hammersmith & Hounslow	Ealing, Hammersmith and Hounslow
South East	245 Bedfordshire	Bedfordshire
Remainder (53)	250 Buckinghamshire	Buckinghamshire
	255 Essex	Essex
	260 Hertfordshire	Hertfordshire
	265 Berkshire	Berkshire
	270 East Sussex	East Sussex
	275 Hampshire	Hampshire
	280 Isle of Wight	Isle of Wight
	285 Kent	Kent
	290 Oxfordshire	Oxfordshire
	295 Surrey	Surrey
	300 West Sussex	West Sussex
South West (63)	305 Avon	Avon
	310 Cornwall	Cornwall and Isles of Scilly
	315 Devon	Devon
	320 Dorset	Dorset
	325 Gloucestershire	Gloucestershire
	330 Somerset	Somerset
	335 Wiltshire	Wiltshire
West Midlands	340 Birmingham	Birmingham
(71)	345 Coventry	Coventry
	350 Dudley	Dudley

	355 Sandwell	Sandwell
	360 Solihull	Solihull
	365 Walsall	Walsall
	370 Wolverhampton	Wolverhampton
West Midlands	375 Hereford & Worcester	Hereford and Worcester
Remainder (73)	380 Shropshire	Shropshire
	385 Staffordshire	Staffordshire
	390 Warwickshire	Warwickshire
Greater Manchester	395 Bolton	Wigan and Bolton
(81)	400 Bury	Bury and Rochdale
	405 Manchester	Manchester
	410 Oldham	Tameside and Oldham
	415 Rochdale	Rochdale and Bury
	420 Salford	Salford and Trafford
	425 Stockport	Stockport
	430 Tameside	Tameside and Oldham
	435 Trafford	Salford and Trafford
	440 Wigan	Wigan and Bolton
Merseyside (82)	445 Liverpool	Liverpool
	450 St Helens & Knowsley	St Helens and Knowsley
	455 Sefton	Sefton
	460 Wirral	Wirral
North West	465 Cheshire	Cheshire
Remainder (83)	470 Lancashire	Cumbria and Lancashire
Wales (93)	475 Clwyd	Clywd
	480 Dyfed	Dyfed
	485 Gwent	Gwent
	490 Gwynedd	Gwynedd
	495 Mid Glamorgan	Mid Glamorgan
	500 Powys	Powys
	505 South Glamorgan	South Glamorgan
	510 West Glamorgan	West Glamorgan

Four of the coded spatial units are not strictly FHSAs or HBAs: Northern Ireland and the Isle of Man are included as single regions; there is a code for Scotland pre-1974; and there is a code for Middlesex-unstated. In 1985, Middlesex was subdivided into Enfield and Haringey, Barnet, Hillingdon, Brent & Harrow, and Ealing, Hammersmith and Hounslow. As a destination, Middlesex-unstated has not been used since then, but it has been used as an origin, albeit with very small flows as indicated in Table 3.4 which shows flows from Middlesex and its constituent parts to Leeds between 1975-76 and 1991-92.

In addition to these geographical areas within the UK, there are also codes for other 'Outside UK' areas as follows:

0 0 000 Not stated
0 1 001 Psychiatric or prisoner
0 4 005 Abroad 1 – no birthplace stated
0 4 010 Abroad 2 – country of birth Ireland
0 4 015 Abroad 3 – country of birth other than Ireland
0 4 016 Abroad – not first acceptance (returned migrant)
0 5 020 Armed forces
0 5 025 Armed forces dependents

Table 3.4: NHSCR flows from Middlesex to Leeds, 1975-76 to 1991-92

Period	Enfield & Haringey	Barnet	Hillingdon	Brent & Harrow	Ealing, Hammer & Hounslow	Middlesex Un-stated
75q3-76q2		0	0	0	0	530
76q3-77q2		0	0	0	0	500
77q3-78q2		0	0	0	0	510
78q3-79q2		0	0	0	0	430
79q3-80q2		0	0	0	0	340
80q3-81q2		0	0	0	0	480
81q3-82q2		0	0	0	0	330
82q3-83q2		0	0	0	0	411
84q3-85q2		0	0	0	0	484
85q3-86q2		26	6	27	47	312
86q3-87q2		56	38	89	130	25
87q3-88q2		69	29	99	272	4
88q3-89q2		71	60	142	192	4
89q3-90q2		109	40	164	153	11
90q3-91q2		130	48	154	158	3
91q3-92q2		125	46	141	178	0

Source: NHSCR quarterly data files

- Sex: the code for sex is input as either 1 for male, 2 for female or 3 for unstated. In the latter case, sex is imputed as 1 for male or 2 for female alternately and the 'sex and acceptance date imputation indicator' is assigned the value of 1.

The next variable is derived from the codes indicating the origin and destination FHSAs and is used when aggregating the data.

- Type of move: this variable identifies the move in one of eleven aggregate categories as follows:

- Within England - different FHSA, same region
- Within England - different FHSA, different region
- Within Wales
- Between England and Wales
- Between Scotland and England and Scotland and Wales
- Between Northern Ireland and the Isle of Man
- Between England and Northern Ireland, Wales and Northern Ireland
- and from Northern Ireland to Scotland
- Between Outside UK (abroad) and England or Wales,
- and from Outside UK (abroad) to Scotland or Northern Ireland
- Between Armed Forces and England, Wales, Scotland and Northern Ireland
- Between Outside UK (not stated, psychiatric or prisoner) and England or Wales
- Between England and Isle of Man or Wales and the Isle of Man.

The next two variables are both dates and are important because they are used as the basis for imputing key information, the age at migration and date of migration.

- Date of birth (DOB): the field has eight characters, the first two relate to the day (dd), the second two to the month (mm), and last four to the century and year (ccyy). If any of this DOB information is missing, then the following imputation occurs:
    - if the day number is missing, the day element is imputed as 15 (roughly halfway through the month) and the value of the DOB imputation indicator is incremented by 1;
    - if the month element is missing, the day element is checked. If it is 31, then it cannot be June and thus is imputed as 05 (May) or 07 (July) alternately; otherwise the month is imputed as 06 (June). The value of the DOB imputation indicator is set to 2;
    - if the year is missing, century (cc) and year (yy) are imputed as 9999 and value of DOB imputation indicator is set to 9;
    - if the century is missing and the processing year is less than DOB year, then the century is imputed as 18 and the DOB is set to 4;
    - if the century is missing and the processing year is greater than the DOB year, then the century is imputed as 19 and DOB imputation indicator is set to 4;
    - if the century is missing and the processing year is equal to the DOB year and the month has been imputed, then the century is imputed as 19 and the DOB imputation indicator is set to 4;

    - if the century is missing, the processing year is equal to DOB year and processing month is greater than DOB month, then the century is imputed as 18 and the DOB imputation indicator is set to 4; and
    - if the century is missing, the processing year is equal to DOB year and the processing month is not greater than the DOB month, the century is imputed as 18 and the DOB imputation indicator is set to 4.
- Acceptance date for new posting: This date is when the individual registration with the new FHSA is accepted. It must be a valid date and if the day is missing, it is imputed as 15 (middle of the month) and the value of the sex and acceptance indicator is set to 2; if the year is greater than 85, then the acceptance century is imputed as 19; if the acceptance year is not greater than 85, then the acceptance century is imputed as 20. This seems to be a mechanism that will take effect from the millennium onwards.

The next two variables, the age at migration and the date of migration, are calculated using existing information on date of birth or acceptance date.

- Age at migration: this is calculated as the difference between the full DOB and the 'true date of migration' or the date of migration if no true date of migration is given; if the year in DOB field is set to 9999, the age of migration is set to 999.
- Date of migration (DOM) and True date of migration (TDOM): if the latter is not given, then the DOM is imputed as the acceptance date minus 30 days (for England and Wales), the assumption being that the actual migration will take place one month before registration on average. For Scotland, the assumption is 90 days or three months.
    - It is unclear how the TDOM (where present) is determined. It might be that this is the date that is written on the re-registration form by the migrant. According to the documentation:
    - If the DOM is earlier than the 'earliest migration date', then DOM is set to the earliest migration date and a TDOM field set to the imputed date is held in the database. The DOB is checked against the TDOM or DOM if the latter is unavailable.

- If the DOB is greater than the TDOM/DOM and has been imputed, then DOB is set to TDOM and DOB imputation indicator set to 08. Age at migration will be set to zero.
- If the DOB is greater than the TDOM/DOM, has not been imputed and is greater than the acceptance date, then DOB is invalid.
- If the DOB is greater than the TDOM/DOM, has not been imputed and is less than the acceptance date, then age at migration is set to zero.

There is no full documentation for these procedures so the definition of 'Earliest Migration Date' is not known. However, the implication is that TDOM is only ever used as a way in which to store an alternatively imputed migration date. Where TDOM exists it is always a 30 day offset from acceptance date.

There are two fields on the record that are assigned to contain information about the data imputed for the individual.

- DOB imputation indicator: this indicator is set according to the imputations that have been performed on the date of birth field. The values are as follows:

  00 No imputation
  01 Day
  02 Month
  03 Day and month
  04 Century
  05 Day and century
  06 Month and century
  07 Day, month and century
  08 DOB later than True DOM
  09 Century and year
  10 Day, century and year
  11 Month, century and year
  12 Day, month, century and year

- Sex and acceptance imputation indicator: This indicator is set according to the imputations performed on the sex and acceptance date values as follows:

  0 No imputation
  1 Only sex imputed
  2 Only acceptance date imputed
  3 Sex and acceptance date imputed

The final variable is another imputed age variable.

- Age at mid-year: this is imputed as the difference between the full DOB and the 30 June in the processing year. If the century and DOB are set to 999, then the age at mid-year is set to 999. If the age at mid-year is greater than 126, then it is assumed that the DOB is incorrect.

Thus, in the example data shown in Figure 3.1 for 1998Q2, the first record (00001) was processed in the 19th week of 1998 and relates to a male (1) moving from Brent and Harrow FHSA (239) to Birmingham (340). The type of move was between FHSAs in different regions within England (02). The migrant was born on 6 June 1970 and his registration was accepted on 30 May 1998. He was aged 27 on the day (3 April 1998) when he moved. There was no imputation of his date of birth, sex or acceptance date and he is estimated to be 28 at mid-year 1998.

(i) Raw data

```
000011998239340102060619700305199802703041998000      028
000021998500485203170819820605199801506041998000      015
000031998239215201051119740605199802306041998000      023
000041998500485103171019480605199804906041998000      049
000051998325016209051219570605199804006041998000      040
```

(ii) Raw data separated into constituent fields

```
00001|19|98|239|340|1|02|06061970|03051998|027|03041998|00|0|     |028
00002|19|98|500|485|2|03|17081982|06051998|015|06041998|00|0|     |015
00003|19|98|239|215|2|01|05111974|06051998|023|06041998|00|0|     |023
00004|19|98|500|485|1|03|17101948|06051998|049|06041998|00|0|     |049
00005|19|98|325|016|2|09|05121957|06051998|040|06041998|00|0|     |040
00006|19|98|500|470|1|04|23111976|06051998|021|06041998|00|0|     |021
```

Figure 3.1: Five records of primary unit data for 1998Q2

3.1.5 Exemplification of the level of imputation

How much imputation is undertaken on the basic NHSCR data? Table 3.5 indicates the level date of birth imputation associated with two selected quarters in the 1990s, the first quarter of 1998 and the same quarter 5 years previously. There are 499,425 records in the 1993Q1 file and 563,049 records in the 1998Q1 file. The evidence from these two quarters is that 99.93% of the records required no date of birth imputation, with only a very small fraction of cases (0.07% in both quarters) requiring day,

month, day and month, date of birth later than true date of migration, or century and year imputations.

The statistics for the second imputation indicator, the sex and acceptance date indicator, suggest that the records are slightly less reliable in this respect (Table 3.6). In 1998Q1, nearly 13,000 records (2.3%) required acceptance date imputation and consequently imply that the date of migration, and therefore the age at migration, may be inaccurate.

Table 3.5: Records with date of birth imputation, 1993Q1 and 1998Q1

Imputation indicator	1993Q1 Number	Records %	1998Q1 Number	Records %
00 None	499,092	99.93	562,631	99.93
01 Day	23	0.00	0	0.00
02 Month	8	0.00	0	0.00
03 Day and month	203	0.04	411	0.07
04 Century	0	0.00	0	0.00
05 Day and century	0	0.00	0	0.00
06 Month and century	0	0.00	0	0.00
07 Day, month and century	0	0.00	0	0.00
08 DOB later than TDOM	43	0.01	7	0.00
09 Century and year	56	0.01	0	0.00
10 Day, century and year	0	0.00	0	0.00
11 Month, century and year	0	0.00	0	0.00
12 Day, month, century and year	0	0.00	0	0.00

Source: NHSCR quarterly data files

The extent of imputation illustrated for 1993Q1 is typical of the levels of imputation for other quarters in the time series and indicates that a very high proportion of the data has not required any adjustment.

Table 3.6 Records with sex and acceptance date imputed, 1993Q1 and 1998Q1

Imputation indicator	1993Q1 Number	Records %	1998Q1 Number	Records %
0 No imputation	497,206	99.56	550,163	97.71
1 Only sex imputed	43	0.01	16	0.00
2 Only acceptance date imputed	2,155	0.43	12,870	2.29
3 Sex and acceptance date imputed	21	0.00	0	0.00

Source: NHSCR quarterly data files

3.1.6 Matching records problem

It is presumed that the probability of there being two or more records with identical values for the fields 'New FHSA', 'Old FHSA', 'Date of birth' and 'Date of migration' is low, and is most likely to occur in the case of twin children (or triplets, etc.) moving with their parents. It is conceivable that similarly identical records may occur for unrelated individuals, but it is presumed that the probability of this occurring would be extremely low. All other fields in matching records would be expected to be identical with the exception of migrant number and sex (as twins do not need to be the same sex). Table 3.7 shows the number of records in recent NHSCR files (from 1996Q1), together with the number that match one or more other records on the selected fields. The larger absolute and proportional values for quarter 3 are expected since there is likely to be a higher propensity for families with school age children to migrate during the summer vacation.

Table 3.7: Incidence of matching records, 1996Q1-1998Q2

Period	Number of records	Number of matching records	% Matching Records
1996Q1	540,349	2,684	0.50
1996Q2	535,081	2,485	0.46
1996Q3	801,125	6,814	0.85
1996Q4	616,023	3,364	0.55
1997Q1	543,241	2,797	0.51
1997Q2	554,951	2,705	0.49
1997Q3	828,098	6,109	0.74
1997Q4	620,195	3,991	0.64
1998Q1	563,049	2,958	0.52
1998Q2	588,455	2,840	0.48

Source: NHSCR quarterly data files

Table 3.8 shows the distribution of matching records (whether true, coincidental or due to error) by the number of matches. The entries in the table show the number of separate sets of identical records which were found in each quarter's data - for example, in 1996Q1 there were 1,286 sets of two identical records, 28 sets of 3, one set of four and two sets of five identical records.

Table 3.8: Distribution of two to five matching records, 1996Q1-1998Q2

Matching records	1996 Q1	1996 Q2	1996 Q3	1996 Q4	1997 Q1	1997 Q2	1997 Q3	1997 Q4	1998 Q1	1998 Q2
2	1,286	1,190	2,851	1,464	1,313	1,283	2,739	1,419	1,395	1,385
3	28	31	81	36	28	34	68	32	24	18
4	1		13	2	4	2	5	13	3	2
5	2	5	4	7	3	3	4	10	2	

Source: NHSCR quarterly data files

If the data are correct, and the propensity to migrate by families with multiple-birth children is not significantly different to anybody else, then one would expect a similar distribution of matches to the pattern of multiple births in the general population. Table 3.9 shows the breakdown of multiple births for England and Wales in 1997. It is apparent that the chance of coincidental matching in the NHSCR data is much higher than expected and this implies that there may be some systematic error causing duplication of records in all data sets.

Table 3.9: Distribution of multiple births in 1997

	Maternities	Number of children	% of all children
All maternities	636,015	645,525	100.00
Singletons	626,814	626,814	97.10
Twins	8,899	17,798	2.76
Triplets	295	885	0.14
Quads +	7	28	<0.01

Source: Table 6.4, Birth Statistics for England and Wales (FM1(26))

Matching records in the NHSCR quarterly files are not confined to sets of 2, 3, 4 or 5 identical sets. It is clear from Table 3.10 that there also a large number of cases with many records matching on the fields origin, destination, date of birth and date of migration.

Table 3.10: Distribution of six and over matching records, 1996Q1-1998Q2

Matching records	1996 Q1	1996 Q2	1996 Q3	1996 Q4	1997 Q1	1997 Q2	1997 Q3	1997 Q4	1998 Q1	1998 Q2
6			4	4	2			3	2	
7	1		4	1	1	1	1	4	2	
8			3				1	4	3	
9			3					4		
10			1	1			1	2		
11				1	2		1	2		
12			1				1			
13								1		
14					1			1		
15				1			1			
17			1							
18				1			2	1		
19				1				1		
20			1							
21			1				1	1		
22			1	1			1	1		
23			1					1		
24			1							
30								1		
32			2							
34								1		
35							1			
37			1							
38							1			
40			1							
42								1		
43							2			
57				1						
61				1						
66			1							
67			2							
68								1		
69								1		
92								1		
102			1							
108								1		
113								1		

Source: NHSCR

1996Q3 is the first quarter where there are a number of multiple matching cases, the worst case being that of 102 matching records. These have the structure:

From: 26 To: 36 DOB:15/12/1977 DOM:16/ 9/1996
where the origin and destination codes represent Northern Ireland (26) and Lothian (36) respectively.

Analysis of frequency tables used to create Tables 3.8 and 3.10 show that many incidences of records with high match counts have Northern Ireland as their old FHSA. Table 3.11 shows the same information but with all multiple counts originating in Northern Ireland omitted. These counts have a much more realistic distribution; however it is clear that some multiple birth migrants must originate from Northern Ireland, so clearly excluding all multiple record flows originating in Northern Ireland is too heavy handed and would lead to some correct data being missed.

Table 3.11 Distribution of matching records in NHSCR data excluding matching records with Northern Ireland origins

Matching records	1996 Q1	1996 Q2	1996 Q3	1996 Q4	1997 Q1	1997 Q2	1997 Q3	1997 Q4	1998 Q1	1998 Q2
2	1,264	1,168	2,807	1,402	1,259	1,234	2,684			
3	25	24	60	25	13	25	56			
4			2				1			
5										
6										

Source: NHSCR quarterly data files

Table 3.12 repeats the information in Table 3.11, but shows it as percentages of all records (excluding multiple records originating in Northern Ireland, but including singleton records originating there). The values seem relatively stable, with the higher values in the Q3 datasets. Table 3.9 shows that 2.9% of children born in 1997 were as part of twins and other multiple groups. Assuming that this rate is constant, and a life expectancy of 78 years, we expect around 0.56% of the population to be both under 16 and born as part of a multiple birth; this value lies within the range of percentages shown in Table 3.12. The values in Table 3.11 would therefore seem to be a

reasonable observation of matching records within the NHSCR as being school-age children.

Table 3.12: Identical migrants as a percentage of all migrants excluding matching records with Northern Ireland origins

Matching records	1996 Q1	1996 Q2	1996 Q3	1996 Q4	1997 Q1	1997 Q2	1997 Q3	1997 Q4	1998 Q1	1998 Q2
2	0.46	0.44	0.70	0.46	0.46	0.44	0.64			
3	0.01	0.01	0.02	0.01	<0.01	0.01	0.01			
4			0.001				<0.001			
5										
6										

Source: NHSCR quarterly data files

Whilst the processing period is assigned as a value from 1 to 53 in the NHSCR data for England and Wales, the processing period for Scottish data is identified by an alphabetic character referring to a particular quarter. Table 3.13 shows the distribution of 'identical' migrants if the Scottish records are excluded and with no other filtering - such as for Northern Ireland. These figures look 'reasonable'. However, the counts for 1996Q1 are too low with this type of filtering. In the case of this file, unexpected letters appear in the 'migrant number' field. This is the only quarter in which this happens. Filtering 1996Q1 for letters in the processing period only returns results that are comparable with all other periods (1,025 twins, 25 triplets). Table 3.13 suggests that the problems are neither exclusively limited to Scottish data or to flows originating from Northern Ireland.

Table 3.13: Identical migrants excluding migrants with alpha characters in any field

Matching records	1996 Q1	1996 Q2	1996 Q3	1996 Q4	1997 Q1	1997 Q2	1997 Q3	1997 Q4	1998 Q1	1998 Q2
2	832	1,127	2,772	1,352	1,207	1,193	2,598			
3	21	24	58	25	13	25	55			
4			2				1			

Source: NHSCR quarterly data files

3.1.7 Age and sex

The demographic composition of the population is clearly of great relevance to planners and policy-makers charged with the provision of services, houses, schools, public transport and so on. Since migration is the key component of population change in most sub-national areas, the demographic structure of out-migration and in-migration flows is very relevant. In Figure 3.2, we illustrate the age and sex breakdown of the moves occurring in 1998Q1 in the form of a migration pyramid. The pattern is a familiar one of absolute levels of migration of both males and females declining in the childhood ages, increasing rapidly following school leaving age, peaking in the mid 20s and then declining almost monotonically until very old age. It is not surprising to observe the higher levels of movement for females in the 20-30 age range given the sex-specific and age-specific differences in registration propensities, whereas the excess of female over male movers aged over 75 is a function of higher life expectancies for women and the fact that there are more of them to move. In the remaining adult ages, more male than female re-registrations took place.

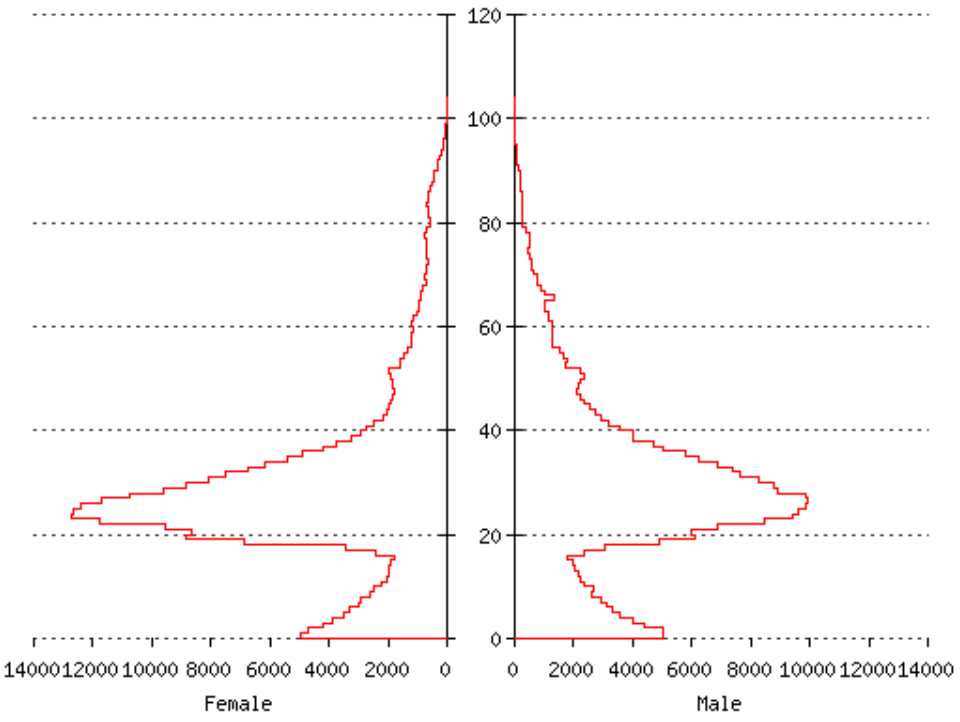

Figure 3.2: Age and sex structure of moves in 1998Q1

3.1.8 Geographical patterns

The geographical pattern of migration can be derived by the user from analysing the records at selected geographical scales. However, the data processing system creates a type of move variable that allows for the rapid calculation of moves according to the spatial categories illustrated in Table 3.14. In the first quarter of 1998, for example, 41% of moves took place between FHSAs in the same region of England, whereas 28% were inter-regional. Within Britain, more moves took place between Scotland and England and Wales than between England and Wales. Over 100,000 moves (18%) occurred between the UK and foreign countries during this quarter and nearly 16,000 moves of armed forces personnel and their dependents were registered.

Table 3.14: Types of move, 1998Q1

Type of move	Number	%
01 Within England - same region	231,003	41.03
02 Within England - different region	156,922	27.87
03 Within Wales	6,158	1.09
04 Between England and Wales	22,113	3.93
06 Between S and E and S and W	23,374	4.15
07 Between N Ireland and the Isle of Man	48	0.01
08 Between E and NI, W and NI and from NI to S	4,531	0.80
09 Between Outside UK (abroad) and E or W, and from OUK (abroad) to S or NI	101,806	18.08
10 Between AF and E, Wales, S and NI	15,980	2.84
11 Between OUK (ns, ps or pr) and E or W	211	0.04
12 Between E and IoM or W and the IoM	873	0.16

Source: NHSCR quarterly data file

Further details of the overseas and armed forces movements are presented in Table 3.15. The majority of international movements (over 77,000) are immigrants from other countries of the world, with a small proportion of movers with no birthplace stated and around 5,000 return migrants. The register also contains 18,000 emigrants in this quarter whose birthplace was not stated but does not contain any further

information about emigration. Armed forces recruitments and discharges are recorded by their origin or destination FHSAs respectively in roughly equal proportions and there are only a very small number of psychiatric patients, prisoners and those with origins unstated in this particular quarter.

Table 3.15: Other categories of move, 1998Q1

	In	Out	Net
Not stated	0	119	-119
Psychiatric or prisoner	0	92	-92
Abroad 1 – no birthplace stated	18,083	2,465	-2,465
Abroad 3 – other than Ireland	0	75,338	-75,338
Abroad – return migrant	0	5,013	5,013
Armed forces	3,479	3,778	-299
Armed forces dependents	4,130	4,593	-463
Total	25,692	92,305	-66,613

Source: NHSCR quarterly data file

3.1.9 Time of migration

The 'date of migration' is derived as being the date 30 days prior to the date of acceptance for England and Wales data and 90 days for Scotland data. It remains unclear whether this time lag difference is allowed for in the data. Figure 3.3 exemplifies how the count of registrations by date of migration varies over one quarter (1998Q1). The first daily count in the sequence is for Thursday 1 January and a weekly pattern is evident from the registrations: the norm is for the lowest count to occur on a Friday (only 450 registrations on 2 January, for example), with Thursdays having the second lowest count (3,843 on 1 January), and all other days having counts over 8,000, although there are occasional exceptions such the low counts on Tuesday 3 March and Sunday 14 March. This daily pattern is difficult to explain since one might expect the lowest migrations to occur on Sundays. The pattern may result from the derivation of the migration date from the acceptance date. It would possible to adjust the derivation assumption so that the counts fit a more logical daily pattern.

1	3843	1	1998	3843				
2	450	1	1998	450				
3	8492	1	1998	8492				
4	9773	1	1998	9773				
5	9306	1	1998	9306				
6	8717	1	1998	8717				
7	8277	1	1998	8277				
8	1759	1	1998	1759				
9	250	1	1998	250				
10	8669	1	1998	8669				
11	9425	1	1998	9425				
12	9814	1	1998	9814				
13	8518	1	1998	8518				
14	7902	1	1998	7902				
15	6134	1	1998	6134				

Figure 3.3: Daily registrations counts, 1998Q1

In addition to the date of migration, there is also a 'true date of migration' field. As indicated by Table 3.16 (which shows the number and percentage of true migration dates in quarters between 19961Q1 and 1998Q2), only a very small percentage of records have a TDOM value. For 1998Q1 for example, 7,444 or 1.32% of records have a completed 'true date' field, but where it does exist, there appears to be no systematic relationship between 'date of migration' and 'true date of migration'. The 'true date' may often be several years earlier than the quarterly period that data file describes. However, all such records appear to have an 'acceptance date' contemporaneous with 'true date of migration. Table 3.17 shows the number of migrants in the 1998Q1 file who have a true migration date, grouped by the year and quarter of the claimed true date of migration.

Table 3.16: Records with true migration dates, 1996Q1 to 1998Q2

Quarter	Number	%
1996Q1	26,137	4.84
1996Q2	12,647	2.36
1996Q3	11,450	1.43
1996Q4	12,506	2.03
1997Q1	10,088	1.86
1997Q2	10,747	1.94
1997Q3	8,973	1.08
1997Q4	13,512	2.18
1998Q1	7,444	1.32
1998Q2	5,914	1.01

Table 3.17: True quarters of migration for 1998Q1 migrants

Quarter	Migrants	Quarter	Migrants	Quarter	Migrants	Quarter	Migrants
				1967Q3	1		
						1971Q4	1
						1974Q4	1
		1976Q2	1	1976Q3	1		
1977Q1	1			1977Q3	1	1977Q4	1
			2	1979Q3	1		
		1979Q2					
		1980Q2	3	1980Q3	1	1980Q4	1
					1	1981Q4	1
				1981Q3			
1982Q1	1			1982Q3	1	1982Q4	1
1983Q1	1	1983Q2	1	1983Q3	1	1983Q4	4
		1984Q2	4	1984Q3	3	1984Q4	1
		1985Q2	3	1985Q3	1	1985Q4	3
1986Q1	4	1986Q2	2	1986Q3	5	1986Q4	5
1987Q1	5	1987Q2	3	1987Q3	10	1987Q4	8
1988Q1	7	1988Q2	12	1988Q3	13	1988Q4	19
1989Q1	12	1989Q2	23	1989Q3	19	1989Q4	18
1990Q1	18	1990Q2	22	1990Q3	23	1990Q4	14
1991Q1	6	1991Q2	13	1991Q3	13	1991Q4	10
1992Q1	13	1992Q2	11	1992Q3	15	1992Q4	33
1993Q1	37	1993Q2	30	1993Q3	13	1993Q4	8
1994Q1	18	1994Q2	14	1994Q3	23	1994Q4	21
1995Q1	37	1995Q2	52	1995Q3	74	1995Q4	126
1996Q1	124	1996Q2	154	1996Q3	241	1996Q4	262
1997Q1	324	1997Q2	501	1997Q3	1,080	1997Q4	2,289
1998Q1	1,617						

Source: NHSCR quarterly data file

The same problem applies to acceptance dates. Figure 3.4 shows all acceptance dates listed for the four quarters that are labelled as '1997'. It is apparent that there are a number of acceptance dates before 1 January 1997 (or rather, before 1 February 1997, which is the earliest acceptance we would expect to see given the assumed lag of 1 month). The number of records with dates prior to 1997 is not huge, but they are highlighted in the plot because a logarithmic scale has been used on the y axis to highlight the smaller counts.

Figure 3.5 shows an expanded section of Figure 3.4, showing only those acceptance dates after 1 January 1997. A seasonal cycle is apparent with the number of acceptances reaching a peak between 1 September and 1 November. This suggests

that August and September are the peak months of the year for migration. This is more evident from Figure 3.6 which shows both the acceptance and migration dates for moves where either or both dates are later than 1 January 1997. The movements are graphed as weekly totals. Week 1 is the week starting on the first Sunday in January. Week 0 is the week prior to this and weeks beyond the end of 1997 are numbered beyond 52.

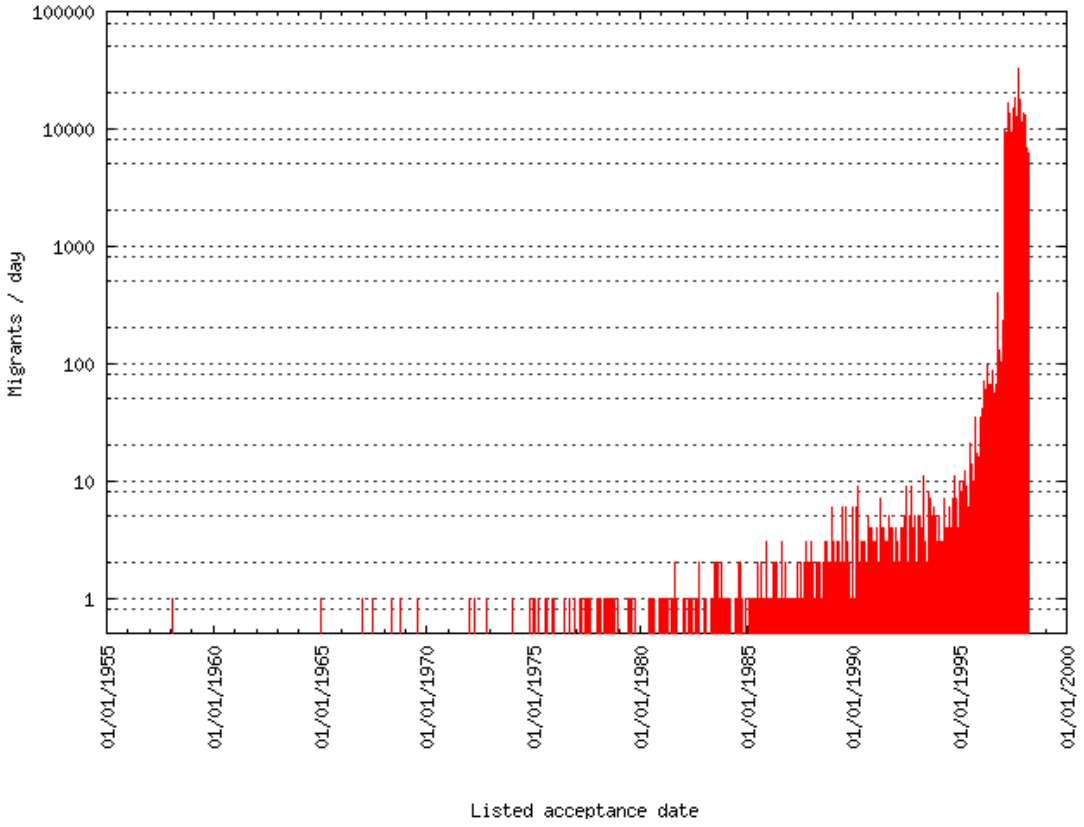

Figure 3.4: Acceptance dates for migrants contained in 1997 quarterly files

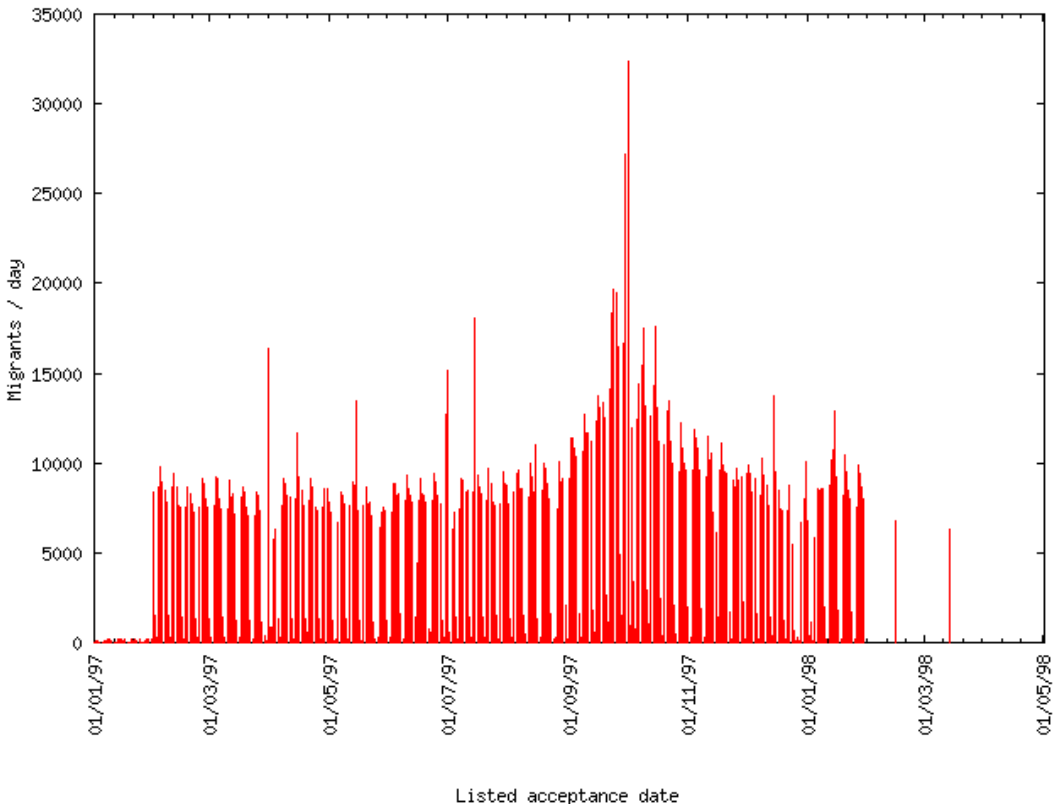

Figure 3.5: Acceptance dates in 1997 for migrants contained in 1997 quarterly files

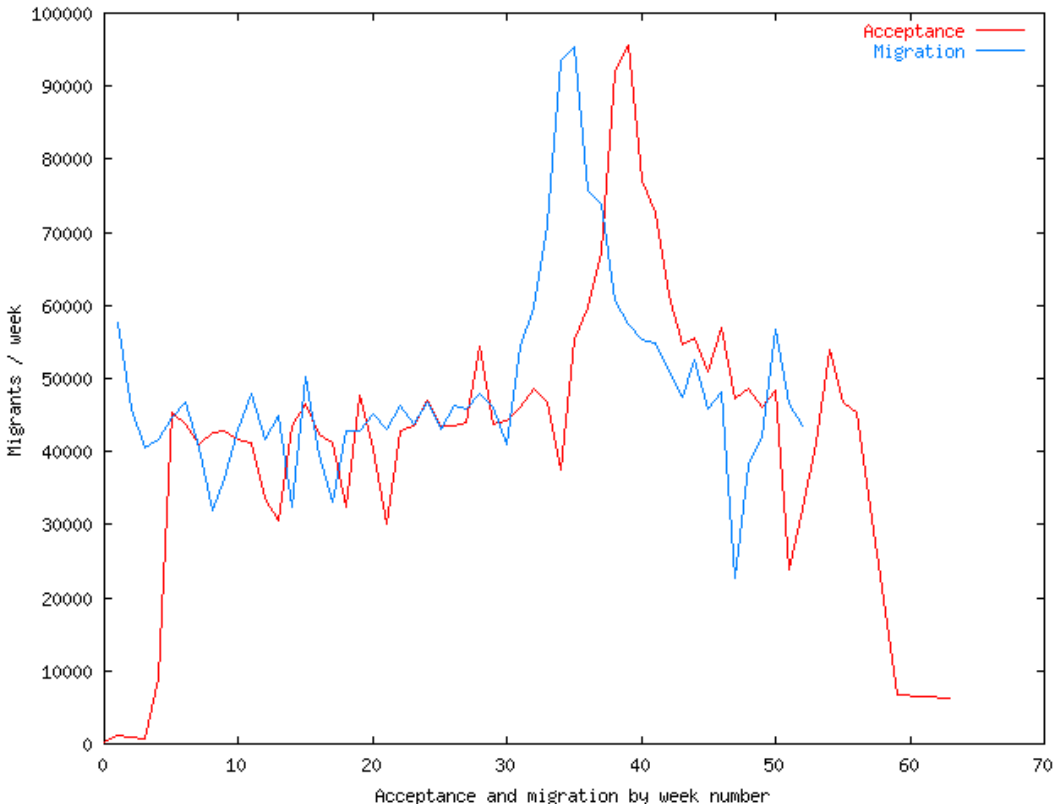

Figure 3.6: Acceptance and migration dates for migrants in 1997 quarterly files

The imputation assumption of one month prior to acceptance explains the difference between the two schedules although the two lines do not always move in parallel. The mean weekly movement total is around 40,000, but this begins to rise in August and in 1997, peaked in week 35 (late August-early September) at around 95,000 moves, before falling back to the norm by mid-October.

3.1.10 TIMMIG system

In order to avoid weekly and seasonal fluctuations in re-registration propensities, it is common practice to merge quarters together to study migration over a year. In order to manipulate the data into a form in which it can be used for analysis , it is necessary to create a framework for storing the large amounts of primary and quarterly data involved and to design a system to allow users to extract data that they require in a relatively user friendly way. A system was constructed at Leeds to do this called TIMMIG (TIMe series MIGration) (Duke-Williams and Rees, 1993) and allowed user access to mid-year to mid-year data from 1975-76 to mid-year 1997-98. Data for a consistent set of 115 geographical areas were available from TIMMIG from 1986 as follows:

- an aggregate origin-destination (OD) array;
- an origin-age-sex (OAS) array of out-migration totals for five year age groups;
- a destination-age-sex (DAS) array of in-migration totals for five year age groups.

The propensity to migrate or the level of migration taking place in a country varies over time. Long-term declines in migratory movements have been detected in France (Courgeau and Pumain, 1993), Netherlands (Kuijper and Noordam, 1992) and Belgium (Poulain, 1996), for example, although frequently there are shorter-term fluctuations against the general trend. NHSCR data from TIMMIG can be used to indicate how the level of internal migration in Britain has changed over the period from mid-1976 to mid-1998 (Figure 3.7) at three alternative spatial scales: FHSAs, NUTS 2 regions and standard regions.

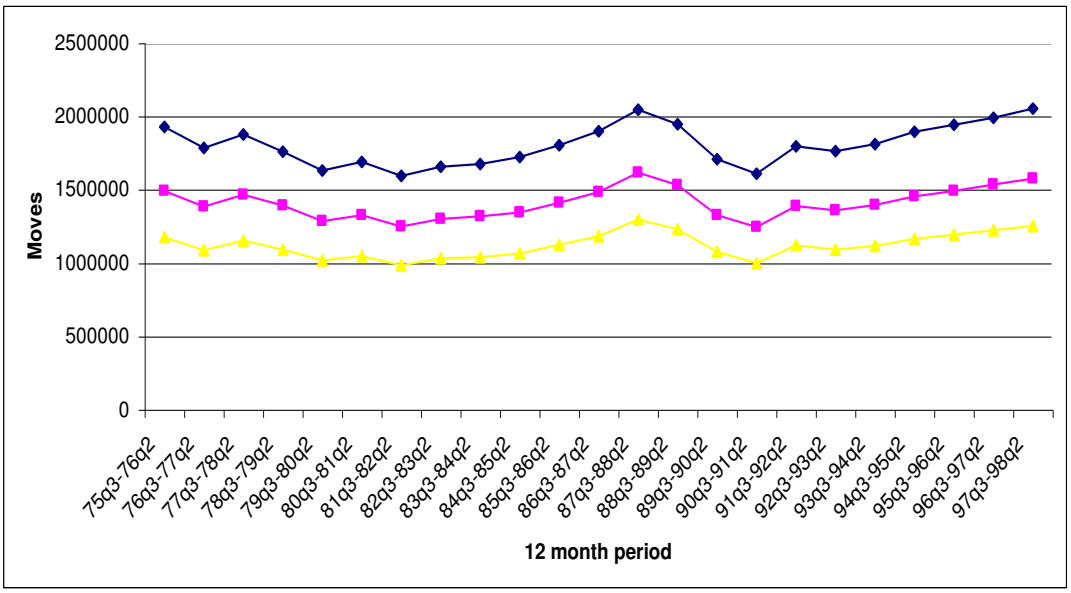

Figure 3.7: Levels of movement, at three spatial scales, 1975-98

3.1.11 Summary

Six points are identified here that summarise the NHSCR quarterly data from 1975-1998. Firstly, levels of imputation in the NHSCR data over the 1975-98 period are relatively low, with date of acceptance being the variable that is most problematic. Secondly, there is a problem with the number of matching records but this appears to be confined primarily to flows originating from Northern Ireland. Thirdly, there are gender differences in the age structure of movements that may reflect differing registration propensities as well as migration propensities. Fourthly, the data are useful for the analysis of internal net migration patterns and does provide some information on immigration, but the lack of emigration information does not allow the net consequences of international flows to be identified. Fifthly, the pattern of daily counts of moves is difficult to explain but this problem is removed when the data are aggregated. Finally, the files do contain a non-trivial number of migrants those imputed true date of migration is before the quarter in which they registered.

CIDER has requested permission from ONS to make the mid-year to mid-year movement flows from the NHSCR by quinary age group and sex from 1975-76 to 1997-98 available to users through the WICID system.

3.2 NHSCR and Patient Register data since 1998 for England and Wales

Since 1998, NHSCR data sets have continued to be produced for a national geography of health-related administrative areas, but the boundaries of thee areas have changed and ONS have begun to exploit the information that is available from individual records of patients contained within the registers held by the health authorities.

3.2.1 NHSCR data

Since 1998, the NHSCR data set is no longer based at the FHSA level. As indicated previously, between 1999 and 2000 the data were based on interim codes which were a mixture of old FHSA codes and new HA codes. From 2001, the data are based entirely on HA codes. Thus, the NHSCR currently supplies ONS with weekly data on moves between HAs in England and Wales in the previous week. These data are processed and tabulated for quarterly and 12-monthly periods ending March, June, September and December. In terms of origin-destination statistics, data are produced by ONS for flows between GORs that are available on the web site (Table 3.18 is an example) and published in Population Trends, Key Population and Vital Statistics and other ONS publications such as Social Trends and Regional Trends.

Table 3.18: NHSCR inter-regional migration matrix in year ending December, 2005

NHSCR Inter-regional migration movements within the UK[1] in the year ending DECEMBER 2005

Thousands

	Region of origin													
	United Kingdom	England	North East	North West	Yorkshire and The Humber	East Midlands	West Midlands	East	London	South East	South West	Wales	Scotland	Northern Ireland
Region of destination														
United Kingdom	.	118.2	39.3	103.1	92.6	96.7	98.6	124	242.8	201.0	107	50.0	44.7	12.7
England	98.3	.	33.3	83.1	83.0	88.9	84.9	114	227.0	181	91.1	47.5	40.8	10.1
North East	39.9	34.3	.	6.0	8.9	3.2	2.3	3.0	4.2	4.3	2.4	1.0	3.7	0.9
North West	102.1	84.8	5.9	.	17.4	9.3	12.3	7.4	13.1	12.0	7.3	8.4	6.5	2.5
Yorkshire and The	94.1	86.4	8.9	17.8	.	15.9	7.8	8.8	10.9	10.9	5.4	2.6	4.3	0.8
East Midlands	105.8	99.2	3.0	9.1	16.5	.	15.6	17.1	13.3	17.7	7.0	2.9	3.1	0.6
West Midlands	94.0	82.6	2.2	12.0	7.3	14.1	.	7.8	13.0	14.1	12.1	7.8	2.7	0.8
East	138.7	131.4	2.6	6.8	7.0	13.3	7.0	.	60.0	25.7	9.0	2.9	3.6	0.9
London	161.2	148.1	4.7	11.9	10.4	10.8	11.8	29.4	.	53.5	15.6	5.0	6.7	1.4
South East	216.5	201.7	3.9	11.2	9.5	13.9	13.0	27.6	90.3	.	32.3	7.1	6.3	1.3
South West	132.3	117.7	2.1	8.2	6.0	8.4	15.1	12.7	22.2	43.0	.	9.9	3.8	0.9
Wales	55.9	53.9	0.9	9.7	2.7	3.0	9.0	3.6	5.7	8.9	10.2	.	1.6	0.4
Scotland	59.2	55.0	4.6	8.5	6.0	4.1	3.8	5.5	8.3	9.2	4.8	2.0	.	2.2
Northern Ireland	12.2	9.4	0.4	1.8	0.8	0.6	0.9	0.9	1.7	1.5	0.8	0.5	2.3	.

1 Based on patients re-registering with NHS doctors in other parts of the United Kingdom.

Source: NHSCR, GROS and NISRA

Source: http://www.statistics.gov.uk/STATBASE/ssdataset.asp?vlnk=9440

The problem of boundary change since the transition to HAs has been persistent. In April 1996, HAs were introduced to supersede FHSAs, a change that resulted in some boundary changes as well as the merging and splitting of some FHSAs to create new HA areas. Table 3.19 attempts to present some idea of the splits and mergers of FHSAs that took place to create the new HAs. The left hand column shows the FHSAs used in the 100 zone system developed by Duke-Williams (2004) which include the 98 FHSAs for England and Wales, with Scotland and Northern Ireland occupying a zone each. Note that where one FHSA or HA appears in a column, it is made up of all or part of the other FHSAs or HAs listed in the adjacent column.

Table 3.19: Changes between FHSAs (100 zones) and HAs (2001 pre-2002 change) in England and Wales

Change	FHSA	HA
1	Devon	South and West Devon North and East Devon
2	Somerset	Somerset Avon
3	Avon Somerset Gloucestershire	Avon
4	Gloucestershire	Gloucestershire Avon
5	Isle of Wight	Isle of Wight, Portsmouth and South Hampshire
6	Hampshire Isle of Wight	Isle of Wight, Portsmouth and South Hampshire
7	Hampshire	North and Mid Hampshire Southampton and Southwest Hampshire Isle of Wight, Portsmouth and South Hampshire
8	Surrey	West Surrey East Surrey
9	Kent	East Kent West Kent
10	Essex	South Essex North Essex
11	Hereford and Worcester	Herefordshire Worcestershire
12	Shropshire	North Staffordshire
13	Staffordshire	North Staffordshire South Staffordshire
14	Derbyshire	South Derbyshire North Derbyshire West Pennine
15	Nottinghamshire	Nottingham North Nottinghamshire
16	Cheshire	South Cheshire

		North Cheshire
17	Salford Trafford	Salford and Trafford
18	Tameside	West Pennine
19	Oldham	West Pennine
20	Oldham Tameside Derbyshire	West Pennine
21	Bury Rochdale	Bury and Rochdale
22	Wigan Bolton	Wigan and Bolton
23	Lancashire	South Lancashire East Lancashire North-West Lancashire Morcambe Bay
24	Cumbria Lancashire	Morecambe Bay
25	Cumbria	North Cumbria Morcambe Bay
26	Humberside	South Humber East Riding and Hull
27	Gateshead South Tyneside	Gateshead and South Tyneside
28	Newcastle Upon Tyne North Tyneside	Newcastle and North Tyneside
29	Mid Glamorgan	Gwent Bro Taf
30	Gwent	Gwent Mid Glamorgan
31	Mid Glamorgan South Glamorgan	Bro Taf
32	Dyfed Powys	Dyfed Powys
33	Clywd Gwynedd	North Wales
34	Enfield and Haringey Barnet	Barnet, Enfield and Haringey

The differences between the two boundary sets are further exemplified in Figure 3.8. It is evident that the change from FHSAs to HAs was not straightforward. In many cases, boundaries have remained the same, such as those for the areas to the north of London such as Cambridgeshire, Bedfordshire and Buckinghamshire. In some cases, old FHSAs have been split into different HAs – for example, Hampshire became North and Mid Hampshire, Southampton and South West Hampshire, and part of Isle of Wight, Portsmouth and South East Hampshire. In other cases, new HAs have been

made up of more than one FHSA. For example, the new HA of Barnet, Enfield and Haringey is made up of the old FHSAs of Barnet and Enfield and Haringey.

Where NHSCR data were used by the ONS to produce migration estimates, processing of the data continued using FHSAs until 1998. Between 1999 and 2000, the NHSCR data were produced for a mixture of FHSAs and HAs, with data only being produced entirely for HAs from the middle of 2001 onwards.

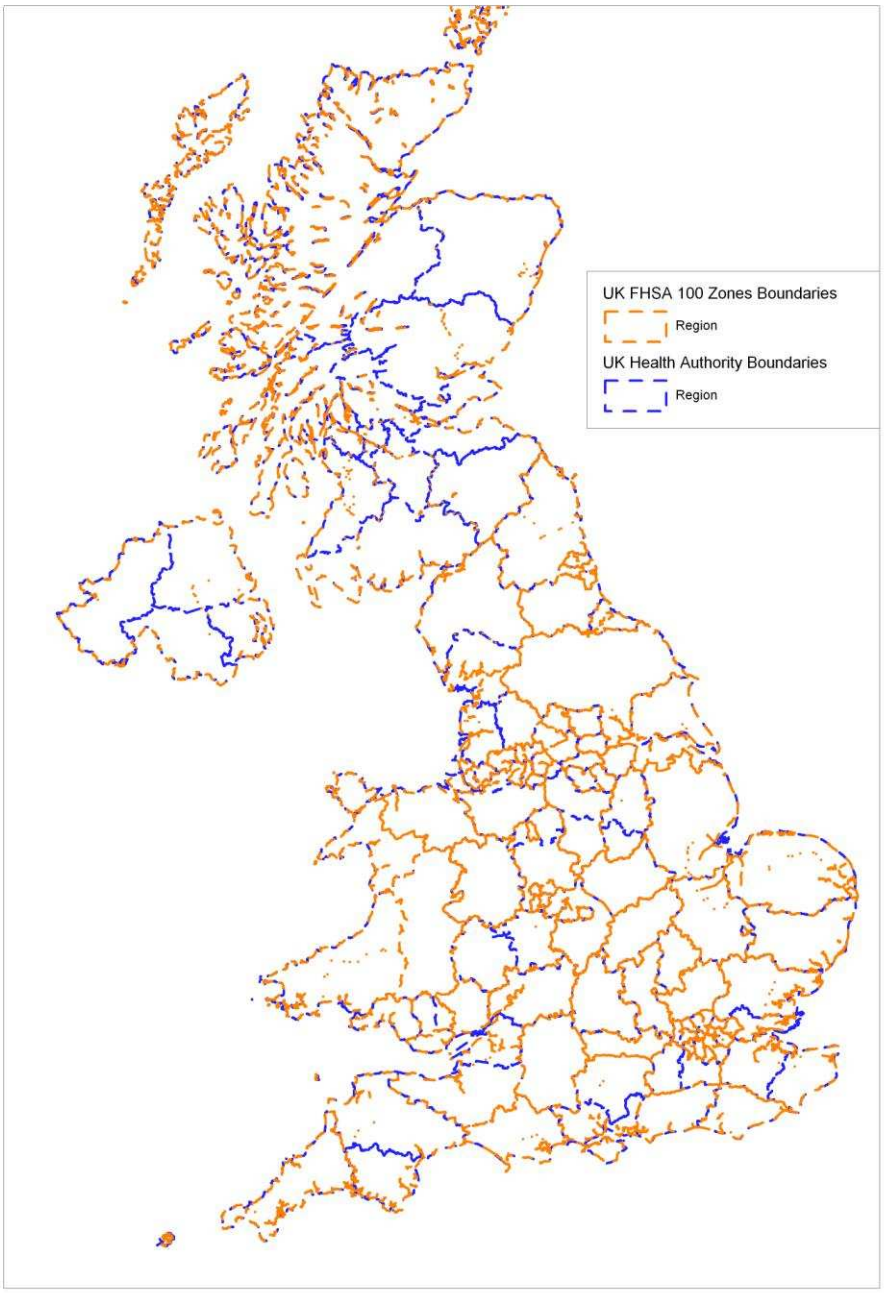

Figure 3.8: Compatibility of FHSA and HA boundaries

Analysis of published internal migration statistics from the ONS (http://www.statistics.gov.uk/statbase/explorer.asp?CTG=3&SL=4218&D=4219&DCT=32&DT=32#4219) reveals the boundary changes that took place in the areas for which statistics were published, leading up to the final mid-2001 HA definitions which are still presently being used by ONS to disseminate migration information. From previous assertions by the ONS that the changes in geography that took place were from FHSAs to HAs, it might be assumed that any changes observed are of this nature. There is some confusion however, as some of the changes do not appear to be FHSA to HA. Indeed some statistics are produced for areas that are neither HAs nor FHSAs.

In the year before 2001, four of the HAs did not exist: The HA of Hertfordshire was made up of the area (constructed from LADs) of East and North Hertfordshire and the area of West Hertfordshire. The HA of Bexley, Bromley and Greenwich was made up of the area of Bromley and the area of Bexley and Greenwich. The HA of Barnet, Enfield and Haringey was made up of the FHSA of Barnet and the FHSA of Enfield and Haringey. The HA of Isle of Wight, Portsmouth and South East Hampshire was made up of the FHSA of Isle of Wight and the area of Portsmouth and South East Hampshire. In the year preceding 1999-2000, there were no changes to the health geographies used by the ONS to disseminate migration statistics. In the year preceding 1998-1999, the only change was that the HA/FHSA of Norfolk did not exist, but was made up of the areas of East Norfolk and North West Anglia.

It should be noted that during this transitional period, the boundaries of the HAs used by the ONS did not change. HAs in 1998 were exactly the same as those used in 2001 when the ONS began producing migration estimates for HAs only. A map of these boundaries is shown in Figure 3.9 which includes health units for Scotland and Northern Ireland. Since the boundaries of HAs were drawn up in April 2001, there have been a number of revisions of the UK's health geography. These revisions happened on 1 April 2002, 1 April 2003, 1 July 2003 and 9 February 2004 (see http://www.statistics.gov.uk/geography/health_geog.asp for full details). The first of these changes was probably the most significant with the 93 HAs in England being replaced by 28 'Strategic Health Authorities.' (SHAs) The boundaries in Wales, Northern Ireland and Scotland remained the same at this time. These new SHAs were

essentially direct aggregations of HAs. Subsequent revisions changed firstly the boundaries in Wales, and then, less significantly for geographers, removed the directorates of health and social care (a higher level division) of which all SHAs were part. The last revision of UK health geography to date happened on 1 July 2006, when the number of SHAs was reduced from 28 to 10 – the boundaries of all but two of these new SHAs (South Central and South East Coast, which together make up the South East GOR) being coterminous with the boundaries of Government Office Regions (GORs).

Figure 3.9: 2001 Health Authority boundaries in the UK

Whilst the number of boundary changes since 2001 could present problems, as previously mentioned, migration estimates produced by the ONS from NHSCR data since 2001 have been for the 2001 HAs. As such, any attempt to create a time series of data will only need to deal with the boundary changes from FHSA to HA. Both the FHSAs used in the 100 zone system and the 2001 HAs are aggregations of the 2001 UK district boundaries. With this being the case, it is entirely possible to compare NHSCR data for HAs directly with 2001 census data. The look-up table has been built to facilitate a geographical conversion from district to HA level data.

Whilst NHSCR data aggregated into tables of gross inflows and outflows by broad age group (15 and under, 16-59, 16-64, 60+, 65+) are available for HAs (and LADs) from the ONS web site, no origin-destination flows between HAs are available directly. ONS have agreed to provide HA-HA flows for inclusion in WICID that are based on these registration data for annual periods from 1999 onwards, disaggregated by broad age group and sex and rounded to the nearest five. The data would be also be subject to the conditions of the crown copyright.

In terms of long-term availability of the NHSCR data, ONS have indicated (Catherine O'Hara, Manager, Internal Migration, Migration Statistics Unit, ONS, private correspondence) that the NHS is re-engineering its patient register database which will ultimately result in the loss of the current NHSCR. ONS are currently in the process of evaluating this new system as a source of internal migration data. If the evaluation exercise is successful, ONS expect a complete transition by 2009/2010. From that date they anticipate still producing migration estimates at the HA level (as well as for other geographies), but the estimates may not be directly comparable as the building blocks will be at postcode level rather than HA level.

One final point of note is that whilst HA boundaries are currently being used by ONS to produce internal migration estimates, as an administrative geography the HA has long since been retired, replaced subsequently by the revisions described above.

3.2.2 Patient register data

The NHSCR system in England and Wales only records movements between HAs and, in the past, the Migration Statistics Unit at ONS has used information from

electoral registers and the most recent census to apportion NHSCR inflows and outflows between constituent local authorities (LAs). The inadequacy of the electoral registers in the estimation of sub-HA flows led ONS to investigate the patient registers held by every HA in England and Wales (Scott and Kilbey, 1999; Chappell et al., 2000). These registers contain the NHS number, gender, date of birth, date of acceptance at the HA and, importantly, the postcode of address, for each patient. With postcode unit information being available, it is possible theoretically to create aggregate migration matrices for any level of geography. Furthermore, it is also possible to produce migration estimates for a range of different current and historical geographical boundaries. This is a significant advantage over the FHSA/HA boundaries (and aggregations thereof) that previous NHSCR migration estimates were restricted to**.**

NHSCR and patient register data differ in their composition, in that the NHSCR data are counts of moves from one area to another in a particular period of time. That is, if a patient moves from one area to another more than once during the period, a record is made of each movement. Patient register data, on the other hand, are counts of persons migrating. That is, the migrant is defined if their address at the beginning of the period is different to their address the end of the period. The migration flow in this case, therefore, is the flow from the address at the beginning of the period to the address at the end of the period, and does not include any other changes of address that may have happened over that time. Conceptually, patient register data is transition data and similar to census migration data.

With patient registration data only recording migrations if the address at the beginning of the period is different from the address at the end of the period (and if two addresses are present on the register), some categories of migrant are unaccounted for who would be identified by the NHSCR. Firstly, newborn infants born during the period are not included as they will not have had an address at the beginning. Secondly, for the same reason, international immigrants will not be included, along with persons who have been discharged from the armed forces. Other migrants who would not feature on the patient registers are those that move during the period, but then leave the register before the end. The migrants may leave the register either

through death, emigration or enlistment into the armed forces (Scott and Kilbey, 1999).

Whilst it should be noted that there are differences between the migrations recorded by the NHSCR and patient registers, close examination of the figures reveals that migrants missed by the patient registers only constitute a very small proportion of total migrants over a 12 month period. As is noted by Scott and Kilbey (1999), in the period mid-1996 to mid-1997, the change in population that would result from using the patient register data instead of NHSCR data would be less than 0.05% in 81 of the 98 FHSAs. Even so, in an effort to produce the most accurate estimates of migration, it is suggested that migration estimates made from patient registers be constrained by figures from the NHSCR. That is, where direct comparisons can be made at the HA scale with both data sources, migration estimates from the patient registers will be adjusted to include those migrants who may be omitted. Unfortunately, no explanation is provided on how these migrants may be redistributed by local authority or indeed lower geographies once identified.

Acting on the advice of extensive consultations about patient register data, ONS have used patient register and NHSCR data in combination to produce migration estimates for England and Wales since 1999. By obtaining a download from each patient register on an annual basis and by combining all the HA patient register extracts together, a total register for the whole of England and Wales has been created, the Patient Register Data System (PRDS). Comparing records in one year with those of the previous year by linking on NHS number enables identification of each person who changes their postcode. The download is taken at 31 July each year to enable migration estimates to be made for the year ending 30 June. This is consistent with the assumption that people delay registering with a new GP for a month after they move.

Tables are created from the PRDS, combined with NHSCR data and used as an indicator of population movement. A range of tables are available on the National Statistics website (http://www.statistics.gov.uk/statbase/explorer.asp?CTG=3&SL=4218,4819&C=4819#481). These tables are available from 1997 to 2005 and cover inflows and outflows rounded to the nearest 100, disaggregated (in some cases) by

age and gender. Flows are either between local authority or health authority, but are not tabulated in matrix form, just as inflows and outflows from each area.

Non-standard tables that are not published on the website are available on request from the ONS Migration Statistics Unit. These tables include origin-destination matrices of flows between UA/LADs. Three examples of such tables are:

- GOR, unitary and local authority flows by summary age and sex;
- GOR, unitary and local authority flows by broad age group and sex; and
- GOR, unitary and local authority flows by quinary age group and sex.

One of the products is a table (ONS Table 2a) of aggregate movements between unitary and local authorities in England and Wales, with no within area moves included. Data are rounded to the nearest 10 people, row and column totals are rounded to the nearest 100 and there is no age-sex breakdown. These data are available free of charge for years ending June 1999 onwards and the tables are re-issued each year in the autumn of the year after the data year. Table 3.19 presents an example of ONS Table 2a.

Table 3.19: Example of ONS Table 2a data

Table 2a *Movements between LAs in England and Wales during the year ending June 2003.

Origin / Destination		Darlington UA	Hartlepool UA	Middlesbrough UA	Redcar and Cleveland UA	Stockton-on-Tees UA	Chester-le-Street	Derwentside
		00EH	00EB	00EC	00EE	00EF	20UB	20UD
Darlington UA	00EH	-	10	80	80	290	0	40
Hartlepool UA	00EB	40	-	110	40	280	10	20
Middlesbrough UA	00EC	50	70	-	1060	700	0	20
Redcar and Cleveland UA	00EE	70	10	1110	-	290	10	0
Stockton-on-Tees UA	00EF	250	150	1430	460	-	20	0
Chester-le-Street	20UB	10	0	0	10	10	-	210
Derwentside	20UD	10	10	0	10	0	240	-

example only

Source: ONS

Another table (ONS Table 2b) shows movements between a chosen unitary or local authority and all other unitary and local authorities in England and Wales. It shows the inflow and outflow to the chosen authority from all other areas. Moves within areas have been excluded. Data are rounded to the nearest 10 people, totals for England and Wales are rounded to the nearest 100. Age is broken down into broad age groups. There is no sex breakdown shown in this table as indicated in Table 3.20.

Table 3.20: Example of ONS Table 2b data

Moves into and out of Bridgnorth
During the year ending June 2003

Area	Code	Inflow						Outflow	
		0-15	16-24	25-44	45-64	65+	All ages	0-15	16-24
ENGLAND AND WALES		**500**	**400**	**1,000**	**600**	**300**	**2,800**	**400**	**500**
ENGLAND		**500**	**300**	**1,000**	**600**	**300**	**2,700**	**400**	**500**
Darlington UA	00EH	20	-	0	-	-	10	-	-
Hartlepool UA	00EB	-	-	-	-	-	-	30	-
Middlesbrough UA	00EC	-	-	-	-	-	-	0	-
Redcar and Cleveland UA	00EE	-	10	-	-	-	30	10	-

example only

Source: ONS

A third table (ONS Table 3) includes movements by broad age groups in ten regional square matrix tables of origin and destination. Broad age groups shown are 0-15, 16-24, 25-44, 45-64, 65+ and 'All ages'. Each regional table shows movements between unitary and local authorities within the chosen GOR and each other GOR in England, plus Wales. Data are rounded to the nearest 10, row and column totals are rounded to the nearest 100. Destination figures are grouped into broad age groups. There is no sex breakdown shown in these tables as indicated in Table 3.21.

Finally, data are also available for quinary age groups by sex for inflows to, outflows from and net migration balances for each GOR and each unitary authority of local authority district in England and Wales, as exemplified in Table 3.21.

Table 3.21: Example of ONS Table 3 data

				A	B	D	E
Totals				**63,500**	**185,400**	**151,200**	**198,700**
NORTH EAST	**A**	**All ages**	**65,300**	**40,500**	**6,050**	**7,010**	**4,260**
		0-15	**13,200**	8,500	800	1,460	650
		16-24	**21,200**	7,050	1,490	1,640	1,660
		25-44	**20,400**	14,750	2,150	1,680	740
		45-64	**6,700**	5,350	1,410	930	730
		65+	**3,800**	4,850	200	1,300	480

example only

Source: ONS

Table 3.21: Example of quinary age group data

INTERNAL MIGRATION WITHIN THE UNITED KINGDOM
GOVERNMENT OFFICE REGIONS AND LOCAL AUTHORITIES IN ENGLAND, AND WALES
FLOWS BY QUINARY/BROAD AGE GROUP AND SEX
DURING THE YEAR ENDING JUNE, 2003

		Persons			Male		
AREA	**Age**	**Inflow**	**Outflow**	**Balance**	**Inflow**	**Outflow**	**Balance**
NORTH EAST	**All ages**	42,450	40,550	1,900	20,530	20,190	340
	0-4	2,650	2,100	550	1,350	1,130	230
	5-9	1,910	1,690	220	940	840	100

example only

Source: ONS

As yet, despite the potential for estimating migration between areas from the postcode unit level upwards, no attempt has been made by ONS or anyone else to estimate migration for anything lower than local /unitary authority level. Although LAD-LAD patient registration change data are produced by ONS, rounded to the nearest 10, it will not be possible to make these available via WICID to users in academic institutions (Catherine O'Hara, private correspondence).

3.2.2 NHSCR data for Scotland and Northern Ireland

Preliminary research suggests that there has been little if any attempt to produce a consistent set of migration estimates across the UK based on patient re-registration data for HAs. For example, ONS does not collect NHSCR data on flows between the 15 HAs within Scotland or the four Health and Social Services Boards within Northern Ireland. NHSCR counts from ONS are available for flows between HAs in

England and Wales and HAs in Scotland together with Northern Ireland as a whole, but these data may not be consistent with the counts of the cross-border flows captured in the NHSCRs in Scotland and Northern Ireland. In the first instance, it would be necessary to gather data from ONS, GRO Scotland and the Central Services Agency (CSA) in Northern Ireland to assess the extent of inconsistency.

Contact has been made with Cecilia Macintyre (GRO Scotland); and Sandy Fitzpatrick (Head of Information and Research Unit, Family Practitioner Services Directorate, CSA), both of whom have expressed a willingness to discuss data provision and its dissemination through CIDER. The probable difficulties associated with reconciling NHSCR data may justify an independent request for collaboration between CIDER and GRO Scotland (together with ONS and CSA if possible) in order to develop a time series of migration data for use by researchers more widely.

3.3 NHSCR/Patient Register data since 1998 and for Scotland and Northern Ireland

Since 1999 in England and Wales, a combination of patient register and NHSCR data has been used to provide more reliable estimates of inter-censal migration between local authority areas. It appears that a similar combination of locally held patient register data and NHSCR data is now in use in Scotland to estimate internal migration. In Scotland, patient register data is known as the Community Health Index (CHI). The CHI holds almost identical information to England and Wales patient registers – information including postcode, date of birth, gender, details or registered GP and the date joined GP lists.

Whilst patient register data were used in migration estimates from the mid-1999 population estimates in England and Wales, it was not until 2002 that these data were used in Scotland. (GRO Scotland, 2007). As the same method of identifying migrants from patient records in England and Wales is used in the CHI in Scotland, issues with the data very similar to those experienced in England and Wales have been identified. For example, there is the omission of newborn and recently deceased migrants, immigrants, emigrants and newly signed up or released armed forces personnel that needs to be dealt with. Exactly the same method of constraining patient register

estimates to NHSCR estimates is used to deal with this problem. (GRO Scotland, 2006).

Despite the use of CHI data in Scotland which should provide details of estimated migration flows between Council Areas, flow matrices between these and other comparable areas in Britain are not made readily available. Table 3.22 shows an extract from the mid 2005 population estimates for Scottish Council Areas. Net migration statistics are provided down to the single person, but no other information is provided. There are certainly no flow matrices which could be used to determine flow rates between individual council areas in Scotland, and indeed other comparable areas in the UK.

Table 3.22: Sample of migration statistics in Scotland by Council Area, 2005

	Estimated population 30 June 2004	Births	Deaths	Natural change	Estimated [1] net civilian migration	Other [2] changes	Estimated population 30 June 2005	Population change	
								Number	%
SCOTLAND	5,078,400	54,259	56,538	-2,279	19,296	-617	5,094,800	16,400	*0.3*
Council areas									
Aberdeen City	203,450	2,117	2,175	-58	-1,039	17	202,370	-1,080	*-0.5*
Aberdeenshire	232,850	2,442	2,134	308	2,525	-243	235,440	2,590	*1.1*
Angus	108,560	1,081	1,338	-257	871	-4	109,170	610	*0.6*
Argyll & Bute	91,190	759	1,143	-384	264	-200	90,870	-320	*-0.4*
Clackmannanshire	48,240	519	516	3	502	-115	48,630	390	*0.8*
Dumfries & Galloway	147,930	1,418	1,868	-450	864	-4	148,340	410	*0.3*

Source: GRO Scotland, 2006

Patient register data in Northern Ireland is known as the Central Health Index (NI-CHI). In the late 1990s, CHI data in Northern Ireland underwent enhancement through more complete postcoding of the data, and through aggregate statistics being made available at local government district (LGD) and parliamentary constituency (PC) levels. A comparison has been carried out comparing the migration estimates obtained from the 2001 Census and NI-CHI data (NISRA, 2005). It was concluded that estimated migration flows across LGDs in Northern Ireland were sufficiently similar to justify the use of NI-CHI in estimating inter-censal migration flows within the country.

Where migration is happening between Northern Ireland and Great Britain, flow estimates are based upon NHSCR and Scottish NHSCR (SNHSCR) data, as well as the CHI. Pre-2000, inflows from Great Britain to Northern Ireland were distributed

across Northern Ireland using the electoral register as a proxy (NISRA, 2005). However, with the improvement of NI-CHI data in the late 1990s, it is these data that are now used to estimate and distribute in-migration flows. A similar exercise to the inter-LGD comparison using the 2001 Census was carried out to assess the suitability of NI-CHI data for the purpose of estimating inflow and outflow between Northern Ireland and Great Britain (NISRA, 2005). The result of this was that broadly similar flows were recorded from both sources, leading NISRA to use NI-CHI to create mid-year population and migration estimates from 2001 onwards.

Selected migration statistics are available from the NISRA website by LGD from 2001 to 2005 (http://www.nisra.gov.uk/demography/default.asp?cmsid=20_21&cms=demography_Population+statistics&release=). Table 3.23 shows however, that whilst it is possible to identify the net impacts of internal and external migration on LGD populations down to the single person, it is impossible to ascertain from where or to where migrants are coming from or where they are going to.

Table 3.23: Sample of migration statistics in NI by LGD, 2004-05

	Estimated population 30 June 2004	Births	Deaths	Natural change	Migration	
					Internal	External
NORTHERN IRELAND	1,710,300	22,436	14,390	8,046	-	6,671
Eastern Board	664,100	8,086	6,287	1,799	-796	1,386
Ards	74,600	895	700	195	352	84
Belfast	269,000	3,337	2,849	488	-2,076	612
Castlereagh	65,800	698	612	86	-177	-39
Down	66,800	891	543	348	195	169
Lisburn	110,200	1,441	828	613	628	252
North Down	77,600	824	755	69	282	308

Source: NISRA (2007)

3.4 Pupil Level Annual School Census (PLASC)

Whilst the census in Scotland provides details of the daily travel to study for students and children, similar data is not produced for England and Wales or Northern Ireland, However, the annual Pupil Level Annual School Census (PLASC) does collect data from each education authority in England and Wales on the location of pupils and the

schools that they attend, potentially providing an extremely useful data set on the journey to school, if confidentiality issues can be addressed.

Various data sets are collected and held by the Department for Education and Skills (DfES) within a centralised 'data warehouse'. These include the National Pupil Database (NPD), local authority data, school level data, school workforce data and geographical data (Ewens, 2005a; Jones and Elias, 2006). The NPD was established in 2002 and contains linked individual pupil records for all children in the state school system which is updated annually. Each pupil is given a unique pupil number (UPN) and has an associated set of attributes: age, gender, ethnicity, special educational needs, free school meal entitlement, key stage assessments, public exam results, home postcode and school attended. It is the availability of the last two attributes which gives the possibility of identifying the journey from home to school for each pupil.

The NPD combines information from the PLASC with information on pupil attainment, reference data on schools and LEAs. PLASC is the foundation of the NPD, including variables such as ethnicity, a low-income marker, and information on Special Education Needs (SENs). It also includes school attributes such as data on teaching and support staff, classes as taught, and admissions appeals. Separate information on pupil attainment is linked with the PLASC information using the data warehouse. The data includes those students in post-compulsory education at school (e.g. A level students in state schools) but not those studying outside the school system post-16 in further education (FE) establishments. Information about this group of students in England is held by the Learning and Skills Council (LSC) in the form of Individual learner Records (ILRs). In a recent project between the DfES and the LSC, ILR data have been matched with PLASC records, creating a micro data set that allows pupils to be tracked beyond age 16 and into FE.

The linking of pupils from one year to the next using the UPN means that a longitudinal profile of each pupil is available whose extent depends on how long the pupil has been in the education system. Potentially, this means that pupils can be tracked over time and their transitions through the education system can be identified, including their movements between schools and between different home addresses

(Harland and Stillwell, 2007a). PLASC data are therefore a potential source of data on commuting to school, on pupil mobility between schools and of child migration from one usual residence to another. At the moment, there is no indication of any system is in place to process the data to generate this type of information. However, recent research based on PLASC includes studies of the mobility of English school children (Machin et al., 2006) and moving home and changing school in Greater London (Ewens, 2005b), and work is currently underway (Harland and Stillwell, 2007b), based on PLASC data for Leeds supplied by the LEA (Education Leeds) that allows residential migration and movement between schools to be quantified as well as the movements from home to school.

3.5 Higher Education Statistics Agency (HESA) data

One of the problems apparent when comparing census data between 2001 and 1991 is the inclusion of students in the former but not in the latter. In many cases, the movement of students from parental domicile to college or university is only a temporary phenomenon and students return home after their period in higher education (HE) or move to more permanent residence when they secure employment. Nevertheless, the movement of students to study at a higher education institution (HEI) is a very substantial in volume terms and has huge implications for local authorities and service providers.

The Higher Education Statistics Agency (HESA) is the central source for HE statistics, collecting and providing data on students and staff in HEIs as well destinations for HE graduates. There is a 'student dataset' that includes variables such as: A/AS level/Highers points scores, age, disability, gender, source of tuition fees, subject area of study together with domicile and location of HE institution. There is also a 'first destination dataset' that includes information on activity, qualification required for job, employer size, SIC, SOC and location of employment or institution of further study. Initial investigations indicate that the HESA holds postcodes for the permanent residence of students prior to entry to an HEI and the postcode of the administrative headquarters of the HEI attended. In principle, data are therefore available on the flows of students by postcode unit of pre-entry domicile and HEI of study although there are data protection issues since flows of students from a

particular LAD attending a particular university, disaggregated by age and gender, in many cases will involve only relatively small counts of individuals. HESA does not allow general access to microdata and it is unlikely that flows between postcode units would be made available for confidentiality reasons.

Efforts are currently underway to match HESA data to pupil data from the NPD, thus creating a longitudinal data set which would allow pupils to be traced out of school and into university and to match pupils GCSE results, their Year 13 results and also their application to university through UCAS (see project by Vignoles et al., 2006, funded under ESRC Teaching and Learning Research Programme). One of the deficiencies of these data will be that the destination HEI of study is not the place of usual residence. Consequently, research will be required to define a method of allocating inward flows to districts smaller area in order to produce an area to area flow matrix. This method might make use of data from the 2001 Census on the usual residence of the student population.

Discussions between CIDER and HESA (Jonathan Waller, Head of Information Services at HESA) have been positive, although HESA do make a charge for all data they supply and this charge will need to be recovered. HESA is willing to provide CIDER with aggregate data at middle layer SOA (MLSOA) scale from the student dataset (of which there are just over 7,000 areas in England and Wales). The aggregation of data to this geography would be very useful, not least because it appears that ONS are keen to keep SOAs (if not OAs) as a consistent geography beyond the 2011 Census (http://www.onsgeography.net/). If this is the case, then time series analysis of these data will be made a much simpler task. In addition, ONS currently publishes other useful population data at the SOA scale, and so data linkage will be more easily achieved. For example, mid-year population estimates are now being produced by ONS for lower layer SOAs and MLSOAs, as well as for census wards. The greater stability of SOA boundaries is likely to ensure their usefulness beyond the time-frame offered by wards.

Furthermore, the use of SOAs in preference to wards is desirable as whilst many would feel that they have a better understanding of ward geography (principally

because the concept of wards has been around much longer than that of OAs or SOAs), this may in fact not be the case. Ward definitions and boundary changes can be overlooked, and this erodes the reliability of any comparisons. As can been seen from Figure 3.10, there are in fact four main conceptions of wards within the system of geographies used in the UK: CAS wards, ST wards, statistical wards and electoral wards. The where an arrow links geographies, these can be aggregated in the direction of the arrow. For example a unit postcode aggregates into a postcode sector, CAS postcode sector and an Output Area, or a London borough will aggregate into Greater London. A dashed double-ended arrow shows that geographies have boundaries that are essentially coterminous and are different only in name (such as a local education authority and county). Where a dashed line links geographies, this shows that whilst they may be related or very similar there are some key differences, such as the small variations in the number of zones or zone boundaries. Whilst each type of ward is broadly similar to each other type, there are subtle differences between the definitions and boundaries of each type, and in some cases (notably electoral wards) these differences have changed and are likely to continue to change over time. These issues with the definition of ward geographies point to the SOA as being a potentially more stable and accurate unit for geographical comparisons over time.

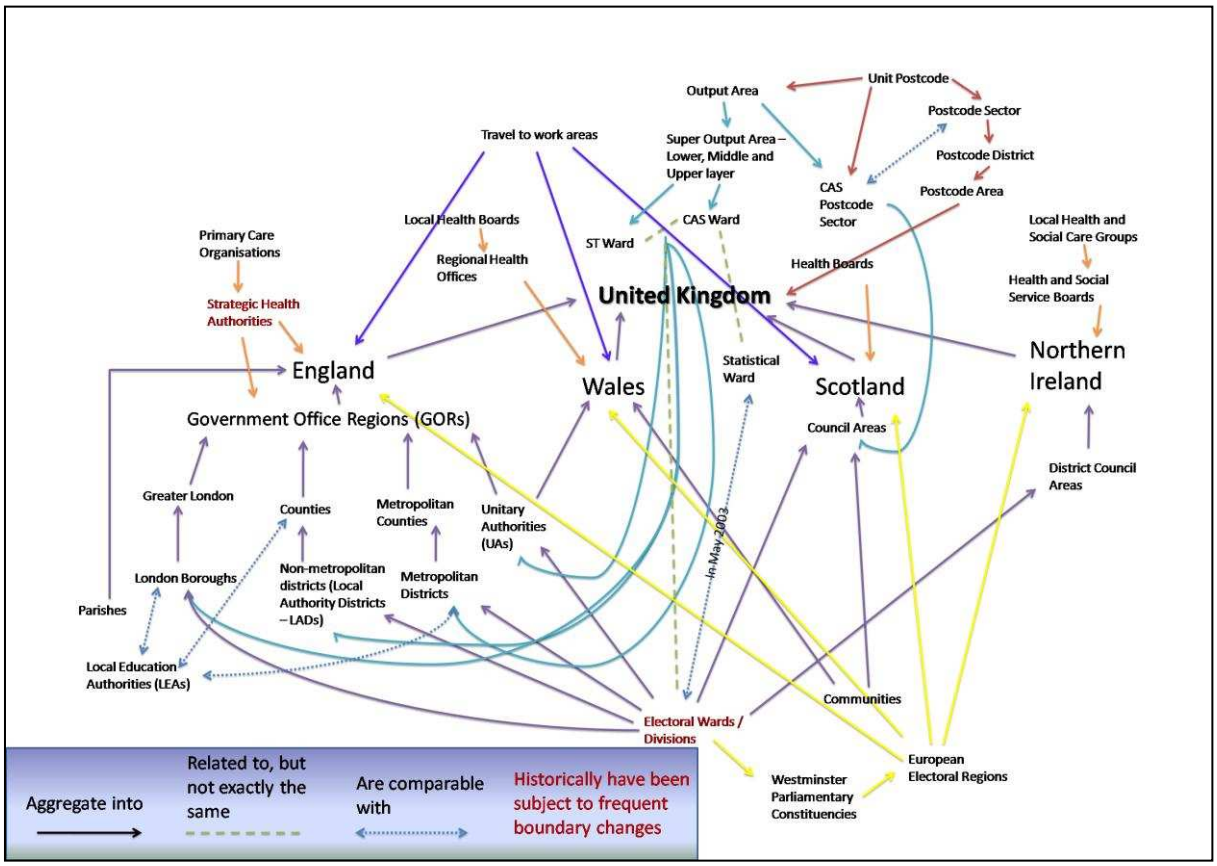

Figure 3.10: A schematic representation of the hierarchy and connectivity of UK geographies, 2007

In summary, HESA will supply CIDER with disaggregated flow data at the MLSOA level. Disaggregation will be by ethnicity, age, gender, qualifications, disability, term-time accommodation, source of tuition fees, subject area of study, and level of study. Ethnicity would be broken down along the lines of white/non-white. Further disaggregation of the non-white category unfortunately will not be possible. HESA are not prepared to do this as their data on ethnicity are only collected for the purpose of monitoring and promoting equal opportunities. When collecting these data, individuals and institutions are assured that this is the only purpose for which the data are collected. Supplying CIDER with detailed data on ethnicity would breach this assurance. Age will be broken down into the quinary age groups comparable with those available in the ONS mid-year population estimates. For example, some of the quinary age groups featured include 15-19, 20-24, 25-29. HESA are prepared to supply CIDER with data in age categories of this type, and could supply single year age groups for coarser geographies if required. Disability data will only be disaggregated into the categories of disabled/not known to be disabled/not known. HESA have also specified that disability should not be cross-tabulated with ethnicity due to issues of disclosure. Cross-tabulation of other variables such as age and gender

with disability and ethnicity, however, will not be an issue, provided appropriate non-disclosure thresholds agreed by CIDER and HESA are maintained.

Data relating to the qualifications obtained by migrant students on the HESA database will necessitate linkage to another dataset held by HESA. This may result in additional costs, and so inclusion of a qualifications variable will only be possible if costs are relatively low. Level of study, however, will not require the use of additional datasets, and so flows will be able to be disaggregated by this variable. Information on postgraduates (studying for doctoral, masters, postgraduate bachelor's degrees and postgraduate diplomas and certificates – including PGCEs) is supplied along with undergraduates, despite postgraduates not entering the HE system through the undergraduate UCAS route. This is because postgraduate information is supplied to HESA directly from HEIs.

3.6 Hospital Episode Statistics (HES) data

The Hospital Episode Statistics (HES) are held in a central data store and include details of all patient admissions to NHS hospitals in England. Data for NHS hospitals in Northern Ireland, Scotland and Wales are collected separately by respective national offices. HES data contain information on each hospital episode (Liffen et al., 1988). Data are available for every financial year from 1989-90 onwards, and each record holds around 100 personal, medical and administrative details of each patient admitted to hospital in England. This includes geographical information about the location of treatment and where the patient lived. Around 12 million new records are added to the dataset each year, with most of the variables collected at point of contact from the Patient Administration System (PAS). However, there are also 'derived' variables which are imputed from other information contained within the HES.

Access to HES data can be achieved through the HESonline website (http://www.hesonline.nhs.uk/), however the use of HES microdata (episode level) is strictly controlled due to high risk of disclosure. Requests for this data in the form of database extracts or custom tabulations are currently made to the NHS Information Centre through their external data custodians, Northgate Information Solutions. If requests are made for data items of increased sensitivity (such as OA code of patient

residence), approval is needed from the Security and Confidentiality Advisory Group (SCAG) and/or the Patient Information Advisory Group (PIAG).

If permission is granted by the advisory groups, it should be possible to obtain data for patients relating to their residential location (which could be as detailed as postcode unit or OA) and their location of treatment (which in theory could be as detailed as hospital postcode – this information, however, would need to be derived externally by users from the 'site' or 'provider' code of treatment featured in the database). These interaction data can be further disaggregated by variables including gender, age, ethnicity, admission/discharge date, length of treatment spell and illness/diagnoses/operation type. Disaggregation could also be by maternity or psychiatric identifier variables which may be seen as different from the standard variables due to the nature of admittance (i.e. non-standard medical admittance).

As stated previously, for the purposes of studying interaction flows it is necessary to access either database extracts or custom tabulations. Aggregated HES data are available to download 'off-the-shelf' from the HES website, although these tables are of very limited use for studying interaction. The only geographical data available from this source identifies either the strategic health authority of residence for patient episodes, or the hospital care provider/HA location for patient episodes. These geographical data do represent aggregated origin and destination statistics, but it is not possible to link the tables (for example through a common identifier) to obtain flow information.

In addition to the main HES dataset (which is concerned with in-patient episodes), outpatient data has been recorded on the IC (HES) Outpatients dataset since 2003. This dataset has more than 40 million new records added to it each year and features records for all outpatient attendances in England. At present the dataset is labelled 'experimental' as there are a number of known weaknesses in the data. These weaknesses principally relate to significant local variations in data completeness (The Information Centre, 2006), with some local administration centres providing far more complete patient records to the central dataset than others. As with the main HES dataset, the HES outpatients dataset also contains information about the location of

treatment and residence for each patient. Again, it is possible to directly identify the postcode or OA of residence for patients, while the location of treatment (if required at a level lower than HA or PCT) would need to be derived for the site of treatment or provider code – an external exercise, but not one that would prove too difficult as provider codes are included in downloadable information from the HESonline website.

A certain amount of care needs to be exercised when using HES data, especially when comparing years. There are fluctuations in the data which, as stated by HESonline (2007), may have occurred as a result of 'organisational changes, reviews of best practice within the medical community, the adoption of new coding schemes and data quality problems that are often year specific.' With this being the case, any observations of changes over time need to be carefully checked.

Little, if any, research appears to have been done on the 'commute to hospital'. This data set provides the potential to investigate hospital catchment areas for different types of operation and to compute average distances to hospital for different types of treatment across the country, for example.

Negotiations are currently in progress with Pam Hughes, Dean White and Chris Roebuck at the NHS Information Centre, to provide CIDER with access to HES data on a permanent basis. It has been agreed in principal that providing CIDER with HES data that could be incorporated into WICID should be possible. As yet, negotiations relating to the outpatient dataset have not taken place, but it is envisaged that once access to the primary HES dataset is established, negotiations relating to the outpatients data set could commence.

3.7 Worker registration data

The Worker Registration Scheme (WRS), administered by the Home Office, provides a cumulative total of the number of nationals of the eight Central and Eastern European countries that joined the European Union (EU) who have registered to work in the UK. Further detail on the number of applicants to the WRS is reported quarterly

in the Accession Monitoring Report (http://www.ind.homeoffice.gov.uk/ind/en/home/about_us/reports/accession_monitoring.html).

3.8 National Insurance Number Statistics

A National Insurance Number (NINo) is allocated to each overseas national entering the UK who wishes to work or claim benefits in the UK and recorded on National Insurance Recording System (NIRS). All overseas nationals allocated a NINo are included, regardless of their length of stay in the UK. There are figures by 'year of arrival' that show arrivals subsequently allocated a NINo according to their reported arrival date into the UK. The figures by 'year of registration' are based on the date of registration onto NIRS i.e. after the NINo application and allocation process has been completed.

NINo data are extracted each year in June and Table 3.24 exemplifies data on flows by local authority area for 2005-06 that are available from the web. Similar sets of data are available for parliamentary constituencies in Britain and for Northern Ireland as a single region. The Department of Work and Pensions (DWP) have provided annual sets of data from 2001-2002 and have indicated that there would be no problem making the data available to other users, on the condition that all footnotes on the data remain intact.

Table 3.24. NINo registrations in respect of non-UK nationals, by local authority area, 2005-06

Local authority	All	Poland	India	Rep of Lithuania	Slovak Rep	South Africa
All	662,390	171,380	45,980	30,850	27,420	23,970
City of London	530	20	30	-	-	20
Barking and Dagenham	3,200	260	250	430	30	40
Barnet	8,840	1,700	500	250	610	350
Bexley	1,410	150	90	60	30	30
Brent	15,060	2,780	2,950	320	330	410
Bromley	2,130	320	180	90	60	120
Camden	7,780	560	220	70	100	210
Croydon	5,900	820	830	170	90	180
Ealing	15,250	5,200	1,380	440	230	580
Enfield	5,030	910	210	200	150	110
Greenwich	5,700	450	270	450	140	160
Hackney	7,640	1,640	160	210	120	100

Hammersmith and Fulham	9,410	1,280	120	110	170	320
Haringey	9,580	2,760	160	380	260	260
Harrow	5,070	680	1,260	130	170	100
Havering	1,100	130	100	90	50	70
Hillingdon	4,320	590	630	170	70	130
Hounslow	10,600	2,560	2,210	510	190	280
Islington	6,790	750	120	60	110	120
Kensington and Chelsea	6,460	350	130	50	80	180
Kingston upon Thames	2,830	400	200	100	70	210
Lambeth	10,460	1,920	130	200	170	270
Lewisham	6,770	920	310	360	80	290
Merton	6,470	1,150	220	180	130	1,180
Newham	14,880	1,450	2,460	2,700	150	420
Redbridge	5,320	380	970	680	40	210
Richmond upon Thames	2,960	470	80	50	100	250
Southwark	9,670	1,060	240	280	130	540
Sutton	1,680	270	120	40	30	120
Tower Hamlets	10,430	1,050	480	560	110	270
Waltham Forest	9,990	1,680	310	1,050	530	1,210
Wandsworth	11,420	1,880	270	120	250	1,480
Westminster	10,960	590	330	120	130	260

Notes: Numbers rounded to nearest 10; totals may not sum due to rounding; numbers based on 100% data from NIRS (17 June)

National Insurance Number Allocations to Overseas Nationals Entering the UK are reported annually (DWP, 2006 at www.dwp.gov.uk/asd/asd1/niall/niall_report.pdf) which contain commentary on the NINo data and, as part of the Improving Migration and Population Statistics (IMPS) project, ONS have looked at the feasibility of using NINo data to inform the estimation of international migration. In the last few years, the flow of migrant labour into the UK has increased very significantly and NINo statistics have been used to gauge some measure of this movement. Boden and Stillwell (2006), for example, identify the variations in NINo registrations by Government Office region and explore the patterns by destination within Yorkshire and the Humber for Poles and Pakistanis, the two largest groups of labour inflows.

This definition of migrants – overseas nationals allocated a NINo – differs from other published statistics in the area, such as International Migration statistics derived from the International Passenger Survey (IPS), and statistics on foreign workers derived from the Labour Force Survey (LFS). The IPS-based statistics define a migrant as someone who stays in the UK for at least a year following arrival. The LFS defines a foreign worker as someone who works but has foreign citizenship, and a foreign-born worker as anyone born outside of the UK, including British citizens. The various

definitions of migrants mean the data sources may have different numbers of migrants for the same time period.

3.9 Home Office asylum seekers and visitor switchers

The Immigration and Nationality Directorate of the Home Office has responsibility for immigration control, applications for settlement, citizenship and asylum. Consequently, the Office produces statistics on immigration control, enforcement, citizenship and asylum. Asylum applications are identified as either 'port' or 'in-country'. Port asylum seekers are those who apply at port when entering the UK. They are relatively few in number and are usually not captured in the IPS since they are detained for separate interview on arrival. In-country asylum seekers are those entering the UK who do not apply for asylum on arrival but do so once in the UK. These individuals are also unlikely to be captured as migrants in the IPS.

The Home Office also collects data on the dependents of asylum seekers (although this has not been done rigorously in the past) and these, together with the counts of principal asylum seekers and international migration data from the IPS, are used to produce estimates of Total International Migration (TIM). As with principal applicants, an allowance is made for those dependants who are not migrants because they are returned within a year.

'Visitor switchers' are visitors who enter or leave the UK intending to stay in the destination country for less than a year but who actually stay for a year or longer. For the years before 2001, estimates of visitor switcher inflows from the non-European Economic Area (non-EEA) were made from the Home Office database of after-entry applications to remain in the UK. IPS data on visitors for these years are only used to estimate visitor switcher data for individuals not covered by the available Home Office data. Since 2001, visitor switcher flows are estimated from IPS data relating to two categories of visitors: those who initially intend to stay for 6-11 months, and those who indicate that they may stay for longer than a year although intended length of stay is uncertain.

The Home Office generates estimates of asylum seekers and visitor switchers by broad origin region and local authority area of destination (Stillwell et al., 2002). These data sets would complement the results from decadal censuses on immigration numbers. Table 3.25 is a typical example of data files of asylum seekers and visitor switchers by destination and by age group.

Table 3.25 Example of data file of asylum seekers and visitor switchers by destination area and age, 1998

Asylum seekers/ visitor switchers 1998						
	Total	Total	Male Age			
	Persons	Male	0	1-4	5-9	10-14
Crown Copyright 2000						
England & Wales	46,349	31,663	46	188	117	310
England	45,873	31,373	44	186	105	300
Wales	476	290	2	2	12	10
Hartlepool UA	8	5	0	2	1	0
Middlesbrough UA	32	19	0	0	0	0
Redcar & Cleveland UA	15	9	0	1	0	0
Stockton-on-tees UA	27	16	0	0	0	1
Darlington UA	15	10	0	1	0	1
Halton UA	10	6	0	1	0	1
Warrington UA	27	16	1	2	0	0
Blackburn with Darwen UA	62	37	0	0	0	0
Blackpool UA	20	12	0	0	1	0
Kingston upon Hull, city of UA	35	21	0	0	0	0

Source: ONS Population Estimates Unit

3.10 Other sources

3.10.1 Estimates of migration between the UK and the Irish Republic

Prior to 1999, the IPS did not cover routes between the UK and the Irish Republic and, therefore, no migration estimates could be derived from this source. The coverage of routes from the British mainland started in 1999, but migration across the land border between Northern Ireland and the Irish Republic is still not included. Thus, flows between the UK and the Irish Republic are estimated using other sources and agreed between the Irish Central Statistics Office (CSO) and ONS. The Irish CSO produces estimates of migration from the UK to the Irish Republic using data from the Irish Quarterly National Household Survey (QNHS). This survey defines migrants as those who have arrived in the Irish Republic from the UK since the previous April.

The Irish CSO also provides estimates of migrants in the opposite direction, from the Irish Republic to the UK. In this case, the QNHS defines a migrant as someone who was usually living in a household in the Irish Republic in the previous April but is living in the UK at the time of the survey. Wholly emigrating households may be under-estimated using this method, so estimates of out-migration are adjusted by CSO using data from other sources including the National Health Service Central Register (NHSCR), the Census of Population and the Country of Residence Survey.

3.10.2 Home Office

The Home Office publish settlement figures and figures for grants of British citizenship that provide a measure of longer term immigration of persons subject to immigration control who are allowed to remain in the UK indefinitely. However, the definition of settlement is different to that of migration used in the IPS. Further information on settlement figures are available in the annual Command Paper Control of Immigration Statistics available from the Home Office (http://www.homeoffice.gov.uk/rds/commandpubs1.html) and further information on grants of British citizenship for 2004 is available from the Home Office Statistical bulletin (http://www.homeoffice.gov.uk/rds/pdfs05/hosb0805.pdf).

3.10.3 Eurostat

Eurostat publish tables on international migration and asylum by individual European country, the European Union as it was constituted on 1 May 2004, the former European Union 15, the Economic Monetary Union, the European Economic Area and the European Free Trade Association. ONS supply Eurostat with IPS only data for residency by age, citizenship by age and sex by age (http://epp.eurostat.cec.eu.int)

Inter-regional migration flows are also available from EUROSTAT but only for registered users with a userid and password. Once the Regions database has been accessed from the General and Regional Statistics domain, and Migration Statistics are selected, it is then possible to choose Internal Migration or International

Migration. Origin-destination flow data for individual countries can be downloaded using the Eurostat Visual Application as shown for the United Kingdom in Figure 3.11. However, the data are limited since there are 13 areas of origin but only seven areas of destination ('partner geopolitical entities') for the UK and the time series of 18 years runs from 1979 to 1996.

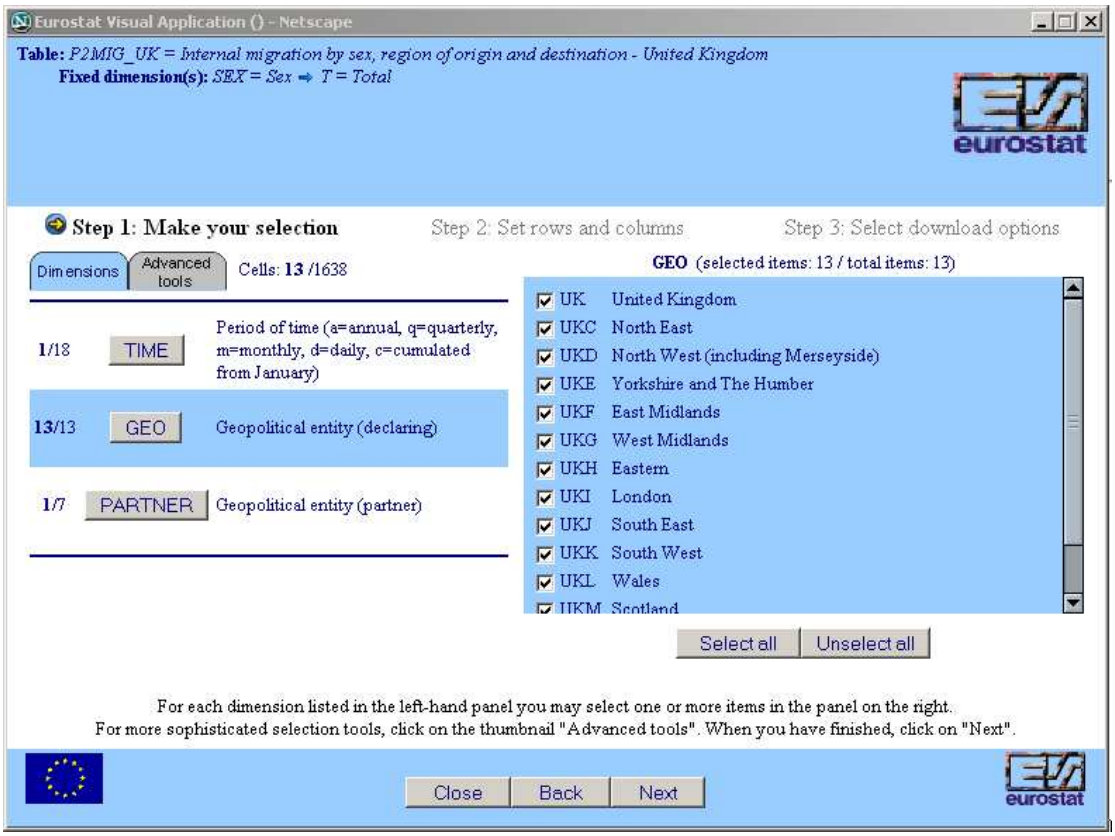

Figure 3.11: Eurostat's interface to internal migration data

3.10.4 United Nations

The United Nations Statistics Division collects and disseminates, at the international level, official national data on international migration. The United Nations Statistics Division collects statistics on international migration flows and on the stock of migrant population through the Demographic Yearbook data collection system. This system collects data from national statistical authorities using a set of questionnaires that are sent to over 230 national statistical offices. Further information about the United Nations Statistical Division is available via the following link: http://unstats.un.org/unsd.

3.10.5 Customer databases

It is likely that there are a considerable number of other sources of interaction data that are available from specific administrative sources and relating to different types of mobility. These may well be customer databases.

One example is the so-called 'arts database' used by the Arts Council which holds data collected through the process of ticket sales at a number of performing arts events. All large and medium-sized venues presenting performing arts, particularly those that receive public funding, sell the large majority of their tickets themselves and, in so doing, they record the names and addresses of customers, both for operational and marketing reasons. Audiences London is an agency that has currently (November 2006) collected five years' data for a range of major venues in London, including over 3 million households and 10 million transactions. In addition, there are similar agencies in Belfast, Birmingham, Cardiff, Edinburgh, Glasgow, Leeds, Liverpool and Manchester that have created their own regional databases. These data are going to be analysed by Orian Brook (Audiences London) in an ESRC-funded UPTAP research project in collaboration with Paul Boyle at the University of St Andrews.

Using data for each venue, it is possible to track purchases made by each household over time. By linking the postcode to a census geography, such as output areas, it will be possible identify the geodemographic characteristics of the localities from which people travel to cultural events. The existence of a home postcode also means that it is possible to calculate flows from residences to particular arts venues and to model these flows using selected geodemographic predictor variables at origin and attractiveness variables at destination.

3.10.6 Other sources

Flow data may be derived from various other sources if the resources to exploit the administrative databases were to be available. For example, residential property transactions normally, though by no means exclusively, involve migration from one house to another and the previous addresses of new owner occupiers are held by estate

agents, institutions operating in the housing market as well as the Land Registry. Likewise, changes of address are supposed to be recorded at the Drivers and Vehicle Licensing Agency (DVLA) when vehicle owners move house. Council tax records are another important administrative source of migration data which are likely to be more reliable and more comprehensive than either of the previous two sources. Data on commuting flows to entertainment events such as football matches might be derived from details of season ticket holders held by the clubs or flows to public libraries might be extracted from the computerised library systems. Clearly there are a huge number of sources that contain information about trip origins and destinations that can be geo-referenced either to a specific point or to a geographic area of some type. In the vast majority cases, the data that exist are inaccessible because of confidentiality constraints.

4 SURVEY INTERACTION DATA SOURCES

The ESRC/JISC funded Economic and Social Data Service (ESDS) provides access to a range of archived UK survey data sources. Some of these data sources include information that can be used to measure population movements over different temporal and spatial scales, both within the UK and between the UK and other countries.

4.1 Labour Force Survey

4.1.1 Labour Force Survey (Great Britain)

The quarterly Labour Force Survey (LFS) is a continuous quarterly household survey of around 53,000 households, representing around 126,000 individuals (Madouros, 2006). The main purpose of the survey is to provide information on the British labour market that can *"be used to develop, manage, evaluate and report on labour market policies"* (ONS, 2006). Each household is sampled for five successive quarters. The survey has been running since 1973, although the format has changed over the years. It was a biennial exercise until 1984 when it became an annual survey. In 1992, the survey became a continuous, quarterly exercise for which the data are also published on a quarterly basis. The LFS samples everyone in the household, regardless of age. Since 1992, the themes of the socio-demographic variables included in the LFS have remained broadly similar. The broad categories included are shown in Table 4.1.

Table 4.1: Broad categories of variable contained in the LFS

Categories
Age
Gender
Ethnicity
Educational Level
Marital status
Religion
Number of dependent children
Employment type/location/ hours worked
Health problems (including days off work sick)
Socio-economic classification
Region/country of residence/birth/origin/work

As far as interaction data are concerned, the more recent rounds of the LFS contain details of the residential and workplace movements of respondents, at the scale of Government Office Region (GOR) and Standard Region. For each respondent, their region of residence is recorded at the time of interview, as well as their region of residence three months before, and one year before. The region of place of work is also recorded for the main job and second job of each respondent. However, the region of place of work three months or one year before the interview is not recorded. LFS data are available from the ESDS in individual respondent (primary unit) data form, which allows users to create their own flow matrices, either for residential or commuting flows, through a process of cross-tabulation. Table 4.2 shows a matrix of region to region migrants for 2000-01 derived by cross-tabulating region of current residence during March-May 2001 with region of residence one year previously.

Table 4.2: An example of a simple inter-regional flow matrix for counts of individuals derived from the LFS, 2000-01

Region of usual residence / Region of res one year ago	Tyne & Wear	Rest of Northern region	Greater Manchester	Merseyside	Rest of North West	South Yorkshire	West Yorkshire	Rest of Yorks & Humberside	East Midlands	West Midlands (met county)	Rest of West Midlands	East Anglia	Inner London	Outer London	Rest of South East	South West	Wales	Strathclyde	Rest of Scotland	Northern Ireland	Total
Tyne and Wear	2589	17	0	0	0	5	1	1	2	1	0	0	2	0	3	2	0	2	3	0	2628
Rest of North East	12	3414	3	0	4	1	1	6	0	1	8	1	2	3	4	1	0	0	6	0	3467
Greater Manchester	1	0	5180	3	14	2	2	2	9	1	3	6	1	0	8	3	5	1	0	0	5241
Merseyside	2	0	7	2872	5	0	5	1	1	0	1	0	3	1	1	1	0	0	0	0	2900
Rest of North West	0	1160	20	13	4711	4	5	6	8	2	8	1	0	4	9	2	4	1	5	0	5963
South Yorkshire	0	3	0	1	4	2982	3	6	11	1	2	1	1	2	7	1	0	0	1	0	3026
West Yorkshire	0	2	0	0	5	6	4990	18	10	0	3	2	3	3	6	0	4	1	2	0	5055
Rest of Yorkshire & Humberside	0	5	0	1	1	6	12	3696	12	0	1	1	5	0	11	4	0	1	13	0	3769
East Midlands	3	1	1	4	10	5	7	13	9297	3	5	10	2	10	22	9	6	3	3	0	9414
West Midlands Metropolitan County	0	0	1	0	1	4	3	2	5	5580	32	0	2	10	6	5	3	1	3	0	5658
Rest of West Midlands	0	0	2	3	6	2	1	1	7	16	5981	2	6	4	24	16	17	0	2	0	6090
East of England	2	1	1	7	4	4	1	2	26	1	2	5094	18	24	7263	28	7	6	2	0	12493
Inner London	2	2	3	1	2	1	1	7	3	8	3	3	4851	153	63	10	5	0	7	0	5125
Outer London	2	0	3	1	0	0	1	2	9	1	3	9	56	8602	134	15	1	9	1	0	8849
Rest of South East	13	4	3	2	7	0	4	7	19	23	23	13	27	45	17563	59	12	2	3	0	17829
South West	1	2	2	0	4	5	1	7	21	2	5	3	6	13	53	11242	18	2	3	0	11390
Wales	0	0	1	1	2	0	0	0	0	0	10	3	2	3	11	23	6584	4	0	0	6644
Strathclyde	0	0	0	0	4	0	1	1	1	0	5	2	0	0	2	2	1	5135	17	0	5171
Rest of Scotland	1	6	2	0	2	0	5	3	0	4	0	2	2	3	11	2	2	12	6875	0	6932
Northern Ireland	0	0	0	0	0	0	2	0	0	0	0	0	0	0	1	1	0	0	0	4917	4921
Outside UK	14	9	13	11	14	10	24	23	21	20	16	15	69	76	130	38	20	11	36	20	590
Baby under one year	29	39	65	33	71	43	55	31	99	65	52	44	74	139	294	115	63	52	73	70	1506
Total	2671	4665	5307	2953	4871	3080	5125	3835	9561	5729	6163	5212	5132	9095	25626	11579	6752	5243	7055	5007	134661

Source: LFS, 2001

The example in Table 4.2 shows the most basic flows that can be derived from the LFS data. Similar flow matrices can be disaggregated further by any of the other variables present for individuals in the LFS (Table 4.1) although the numbers involved, particularly in the off-diagonal cells, would be correspondingly reduced.

One issue that is apparent from Table 4.2 is that regional definitions are not constant between the origin (region of residence one year ago) and destination (region of current residence). It is unclear whether these are actual geographical boundary changes or merely errors in the naming of variables. For example, the destination 'Rest of Northern Region' does not have a matching origin. The closest origin is 'Rest of North East.' This could be dismissed as a labelling error if it were not for the unusually high migration (1,160) to the Northern Region from 'Rest of the North West.' Similarly the destination 'East Anglia' does not exist as an origin. The closest origin in this case is 'East of England.' Again, this could be dismissed as a labelling error were it not for the much higher migrations (7,263) from 'East of England' to 'Rest of South East than from 'East of England' to 'East Anglia'. The metadata provided with the dataset does not give any additional information in relation to this issue. 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.

Longitudinal analysis of individual respondents is not possible with the LFS as unique identifiers that could be used for tracking specific individuals over time are not available in the data. It may be possible to observe some time series trends by referencing the year that the respondent first entered the survey, although this could prove difficult and not terribly accurate.

For respondents who were not born in the UK, information regarding their country of birth and origin is included, as well as the year of arrival in the UK. Country of residence three months ago and one year ago are also included, providing (in some cases) a timeframe in which to contextualise movements. Table 4.3 shows an example of an immigration matrix derived from the LFS, showing counts of immigrants for each of 20 regional destinations in the UK on the horizontal, and the countries from which these immigrants originated during the period 2000-2001 on the vertical. The regional destinations in this case distinguish the main metropolitan areas from the rest of each region. It can be seen from this table that there are around 70 'countries' of origin ('countries being a rather loose term as some smaller countries have been aggregated and re-classified as a larger geographical entity - 'other South America' for example). Table 4.3 shows only the countries that could be cross-tabulated with

UK regions in this year. In total, there are around 140 ‘countries’ (including some wider areas included as countries) which could be cross-tabulated with this set of regions. As with the internal flows that can be derived from the LFS, international flows can also be disaggregated by any other variable as the data are available by primary unit.

Table 4.3: An example of an immigration matrix derived from the LFS, 2000-01

Country of residence 12 months ago * Region of usual residence Crosstabulation
March to May 2001

Country of residence one year ago \ Region of usual residence	Tyne & Wear	Rest of Northern region	South Yorkshire	West Yorkshire	Rest of Yorks & Humberside	East Midlands	East Anglia	Inner London	Outer London	Rest of South East	South West	West Midlands (met county)	Rest of West Midlands	Greater Manchester	Merseyside	Rest of North West	Wales	Strathclyde	Rest of Scotland	Northern Ireland	Total
United Kingdom	2	0	1	0	0	1	0	0	1	11	1	0	0	0	2	0	3	1	2	0	25
Ireland,Republic Of	0	0	0	3	3	0	0	1	6	2	0	0	0	0	0	1	0	0	0	6	22
Channel Islands (So Stated)	0	1	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	4
Australia, Tasmania	0	0	5	3	0	1	1	5	3	13	2	1	5	3	0	0	2	0	2	4	50
Canada, Newfoundland, Nova Scotia	0	0	0	0	0	0	1	0	0	2	1	0	0	0	0	1	0	0	1	2	8
New Zealand	0	0	0	0	0	0	0	2	1	5	0	0	0	0	0	0	0	0	3	0	11
Kenya	0	0	0	0	0	0	0	0	0	3	2	0	0	0	0	1	0	0	0	0	6
Zanzibar (Tanzania)	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	4
Zimbabwe	0	0	1	2	0	0	0	0	0	7	0	1	0	0	0	0	0	0	0	0	11
Nigeria	0	0	0	0	0	0	0	0	3	0	1	0	0	0	0	0	0	0	0	0	4
Barbados	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
Jamaica	0	0	0	0	0	0	0	2	0	0	0	0	0	1	0	0	0	0	0	0	3
Belize	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Bangladesh	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
India	3	3	0	2	0	5	0	3	4	2	0	1	0	0	0	0	0	0	0	0	23
Sri Lanka	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	2
Hong Kong, Kowloon	1	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	4
Malaysia	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
Singapore	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	4
Cyprus	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	5	0	0	0	9
Mauritius	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
Libya	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
Egypt	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	5
South Africa	0	0	0	1	0	0	0	9	9	12	0	1	0	0	1	0	0	0	9	0	42
Other Africa	0	0	1	0	0	0	0	0	0	0	0	6	3	0	0	0	0	0	0	1	11
United States	1	2	0	2	7	1	3	7	6	11	6	0	4	0	5	0	3	0	3	1	62
Other South America	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	4
Pakistan	0	0	0	3	0	0	0	1	0	0	0	5	0	3	0	0	0	4	0	0	16
Burma, Myanmar	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
China	0	0	0	0	1	0	0	2	1	1	0	0	0	0	0	0	0	0	0	0	5
Japan	0	0	0	0	0	0	4	5	5	3	0	0	0	0	0	0	0	0	5	0	22
Phillipines	0	0	0	1	0	0	0	0	5	1	0	0	0	0	0	0	0	0	0	1	8
Vietnam	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
Iran	0	0	0	0	0	0	0	0	1	0	0	0	0	3	0	0	0	3	0	0	7
Other Middle East	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	3	0	1	2	0	9
Belgium	0	0	0	0	0	0	0	0	2	3	2	0	0	0	0	0	0	0	3	0	10
Denmark	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	2
France	2	0	0	0	1	1	1	4	5	14	3	2	0	0	0	0	0	0	2	0	35
Italy	0	0	0	0	0	0	0	4	3	3	0	0	0	0	0	0	1	0	0	0	11
Netherlands	1	0	0	0	0	0	0	3	0	0	0	1	0	0	0	0	0	0	0	3	8
Germany	0	0	1	0	0	0	0	0	2	2	2	0	0	0	0	1	2	1	3	0	14
Bulgaria	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
Germany	0	0	0	0	0	0	0	0	0	2	0	0	1	0	0	0	0	0	0	0	3
Poland	0	0	0	6	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	8
Austria	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	3	0	0	0	4
Switzerland	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	3	0	0	0	1	5
Greece	0	0	0	0	0	0	0	1	2	0	1	0	0	3	0	0	0	0	0	0	7
Portugal (inc Azores & Maderia)	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
Spain (inc Balearic & Canary Islands)	0	1	0	0	3	3	1	1	1	3	1	0	0	0	0	0	1	0	0	1	16
Sweden	0	0	0	0	0	0	0	2	5	3	0	0	0	0	0	0	0	0	0	0	10
(Other) Yugoslavia	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Turkey	4	0	0	0	0	2	0	0	0	1	0	0	3	0	0	0	0	0	0	0	10
Angola	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
Somalia	0	0	0	0	0	4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	5
Mexico	0	0	0	0	0	0	2	0	0	1	0	0	0	0	0	0	0	0	0	0	3
Brazil	0	0	0	0	0	0	0	1	0	2	0	0	0	0	0	0	0	0	0	0	3
Chile	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Columbia	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
Iraq	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	2
Korea	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	2
Croatia	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	4
Czech Republic	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	2
Lithuania	0	0	0	0	0	1	0	1	3	0	0	0	0	0	0	0	0	0	0	0	5
Russia (Federation Of Russian States)	0	0	0	0	0	0	0	2	0	2	0	0	0	0	0	0	0	0	0	0	4
Slovak Republic	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	3
Ukraine	0	0	0	0	0	1	1	0	0	0	0	0	0	0	2	0	0	0	0	0	4
Bermuda	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	4
Other Asia	0	0	0	1	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	3
Thailand	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	3
Former USSR etc	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
at Sea/ In The Air	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
Total	14	9	10	24	23	21	15	69	76	130	38	20	16	13	11	14	20	11	36	20	590

Source: LFS, 2001

Prior to 1992, less information was contained within the LFS in relation to migration patterns. The biennial nature of the survey early in its existence meant that between 1979 and 1983, 'usual residence one year ago' is the only variable available that could be used to measure internal residential migration. No indication of foreign residence one year ago is available for this time period. Between 1984 and 1991, when the LFS became an annual survey, the situation is similar, with the variable relating to where the respondent lived one year ago being the only measure of migration.

A major change in the structure of the quarterly frequency of the survey occurred in 2006, when the seasonal quarterly basis (March-May, June-August, September-November, December-February) of collection which had been the norm since 1992, was changed to a calendar quarterly basis (January-March, et cetera) (see Madouros, 2006, for details). This presents researchers using the data in a longitudinal fashion for time series work with additional problems.

From March-May 2005 quarter, a 'special licence' data set is also available, where the geographical scale of reference available for each individual is unitary authority (UA) or local authority district (LAD). This relates to both residence and place of work. Access to this data through the ESDS is more restricted and users are required to sign a special licence agreement and justify their need to use the data before access is granted. Primary unit data for all appropriate periods from 1975 to the present are available to download in a variety of formats from the ESDS (http://www.esds.ac.uk/findingData /lfsTitles.asp).

The availability of the LFS in Great Britain is summarized in Table 4.4 alongside the availability of the LFS in Northern Ireland.

Table 4.4: Summary of LFS availability in GB and NI, 1973-2006

Survey frequency	LFS Great Britain	LFS N Ireland
Biennial	1973, 1975, 1977, 1979, 1981, 1983	1973, 1975, 1977, 1979, 1981, 1983
Annual	1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991	1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992, 1993
Seasonal quarterly	1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005	1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006
Calendar year quarterly	2006	

Note: No primary unit data available for italicised years via ESDS

4.1.2 Labour Force Survey (Northern Ireland)

The Northern Ireland (NI) LFS is closely related to the GB LFS, with very similar variables included in the survey. It has been running for the same amount of time, with biennial coverage from 1973 to 1983, annual coverage from 1984 to 1994, and quarterly coverage from 1994 onwards. It is not clear whether these are seasonal or calendar quarters, but it is assumed that the sampling pattern follows that of the British LFS. The sample size for the NI LFS is around 8,500-9,000 individuals. Primary unit data for the NI LFS are only available from 1995 to 2000 via the ESDS. Summary results for later dates are available through the Northern Ireland Statistics and Research Agency (NISRA).

Primary unit data are available at the level of UA or LAD for two of the available years of the NI LFS: 1998 and 2000. For the other years, data are available at the regional level (Northern Ireland being the only region applicable). With this being the case, it is theoretically possible to observe intra-regional flows relating to commuting in Northern Ireland for 2000 – why this is not possible with migration is discussed later. In practice, however, closer examination of the data reveals that any hope of being able to measure intra-regional flows is false. Northern Ireland is included as a separate UA/LAD amongst the list of GB UAs/LADs. No geography smaller than Northern Ireland is available, and so any flow matrices present only Northern Ireland cross-tabulated with itself.

Flow data from the NI LFS are not available in a predictable or regular way. For the years 2000 and 1995, there is no question available in the downloadable data (this does not necessarily mean it was not asked) relating to the region or place of residence one year ago. With this being the case, it is not possible to identify migrations that have happened in the preceding year. There is, however, a question in 1995 relating to the place of residence three months ago (as there is for all other years apart from 2000). For the published years 1996 to 1999, a question relating to place of residence one year ago was present. Table 4.5 is an example of the migration flow data for Northern Ireland created from the LFS in 1999. It shows that of the 24.3% of valid cell counts (much lower than the GB LFS at 97%), only one individual lived outside of Northern Ireland one year before the survey and 28 counts were babies aged less than one year.

Table 4.5: Migration flow data for Northern Ireland, 1998-99

	Region of residence 1 year ago			Total
	Northern Ireland	Outside UK	Baby under 1 year	
Region of usual residence (Northern Ireland)	2,100	1	28	2,129

Source: NI LFS , 1999

The NI LFS has identical questions to the GB LFS regarding foreign origin. It is possible to observe flows between Northern Ireland and the same countries or small country aggregations that are present in the GB LFS. Finally, it should be noted that the NI LFS also records region (or UA/LAD for 2000) of place of work in the same way as the GB LFS for all of the available years.

Despite the some of the difficulties that are encountered when using the LFS to understand migration and commuting flows, a number of studies have used LFS data for this purpose. Forsythe (1992), Gordon (1995) and Bover (2002) have analysed inter-regional flows whilst Shields (1998), Dustmann and Faber (2005) and Dustmann et al. (2005b) have looked at immigration.

4.2 International Passenger Survey

The International Passenger Survey (IPS) is a large, multipurpose sample survey of passengers arriving at, and departing from, the main United Kingdom airports and seaports as well as those passing through the Channel Tunnel. The data collected are used in particular to produce the annual Travel Trends – A Report on the International Passenger Survey (http://www.statistics.gov.uk/statbase/Product.asp?v1nk=1391).

The principle underpinning the survey is that, without a sampling frame of travellers, time periods or sea crossings are selected and travellers are systematically sampled at fixed intervals within these periods or crossings. Heathrow, Gatwick and Manchester airports are always included in the sample and smaller airports and seaports are included each quarter according to the volume of international traffic. Details of the sampling methods are available in Appendix A of the ONS MN Series on International Migration (e.g. ONS, 2006d).

In 2004, for example, the main sample was over a quarter of a million interviews and had an overall response rate of 81%, of which approximately 1% (3,556) were migrants. Interviews are undertaken throughout the year, although statistics are compiled on a quarterly basis, with data from 1993 until the present time available from the ESDS and the ONS. Summary and selected statistics from the IPS are published by ONS through their website in the 'Travelpac' series. Primary unit data sets with accompanying documentation are available to download through the ESDS from 1993 to the present.

As a measure of migration, the IPS has three main limitations. Firstly, it does not cover all types of migration. Flows along land routes between the UK and the Irish Republic are excluded as are most asylum seekers and some of their dependants. Secondly, since the IPS is a sample survey, it is subject to a degree of uncertainty, although some of these 'errors' can be estimated. Thirdly, migration estimates are based on respondents' intentions, which may or may not become their final actions. Thus, some adjustments are required to account for 'switchers' who change their intentions.

Figures obtained from the IPS are subject to both sampling and non-sampling errors. Sampling error arises due to the variability that occurs by chance because a sample, rather than an entire population, is surveyed. Standard errors are calculated that indicate how much a sample estimate is likely to differ from the true value because of random effects. In 2004, the standard error for the estimated total 518,100 in-migrants was 3.8 per cent and for the estimated total of 310,400 out-migrants was 4.7 per cent. The problems come when the estimates are disaggregated (e.g. in-migrants by country) and it is not possible to state that a year to year change is statistically significant unless it is the major origin countries that are concerned (India, South Africa, Pakistan, Nigeria, Hong Kong).

Non-sampling error is all the remaining error that is not due to sampling error. This is very difficult to calculate. It arises because of non-response and bias will occur, for example, when passengers who do not respond to the survey have different characteristics to those who do respond.

For each passenger, the type of flow is recorded as being either 'over air' or 'over sea'. The nationality of respondents is recorded, as well as the country of residence for foreign visitors or county of residence for UK nationals. It should be noted that county of residence only features in the IPS data from 1997 onwards, meaning that emigrations measured at a geography smaller than the UK are only available from 1997. Where the respondent is a UK traveller leaving the country, the country of destination is recorded, as well as the purpose of visit. Purpose of visit is also recorded for visitors coming into the country.

A number of additional variables are recorded for each individual relating to their gender, the type of departure (alone or in a group), intended length of stay, distance of travel, level of education and point of entry or departure. As data available for download are primary unit data, it is possible to prepare migration flow matrices disaggregated by any of the other variables available in the survey. These flows can be observed between the UK and any other country as country codes are provided for every country in the world. As such the IPS data presents a picture of temporary and permanent immigrations to and emigrations from the UK. The emigration picture is

slightly more detailed in that emigrants are measured from county of residence to country of destination, whereas immigrants are only measured from country of departure and by country of usual residence. Destination of immigrants can only be crudely inferred from point of entry. Upon closer examination of the 'County' variable available for emigrants, it can be seen that these counties are further disaggregated by town. Table 4.6 is an example of the type of flow that can be derived from IPS data.

Table 4.6: Observed emigration flows from UK origins to foreign destinations, Quarter 1 in 2005

Country of Destination / Town/County of Origin	Belgium	Luxembourg	France / Corsica	Germany	Italy / Sardinia	Netherlands
Tyne & Wear - Other	0	0	8	4	4	4
Newcastle-Upon-Tyne - Tyne & Wear	0	6	28	16	10	5
Sunderland - Tyne & Wear	0	0	16	4	8	7
Whitley Bay - Tyne & Wear	0	0	3	0	1	0
Cheshire - Other	8	0	50	23	18	19
Chester - Cheshire	2	0	20	8	3	5
Manchester-Gtr-Other	9	2	86	24	28	22
Manchester GT Manchester	15	1	54	27	33	23
Lancashire - Other	4	0	25	6	10	9
Blackburn - Lancs	2	0	5	2	5	4
Blackpool - Lancs	1	0	6	3	1	1

Source: IPS 2005

The IPS has been described as the 'richest' source of information on international migration (ONS, 2005b), but as mentioned above, there are some significant issues that have been identified with using the data. The annual ONS Total International Migration (TIM) publication (ONS, 2005) outlines standard error calculations that need to be applied to the total flow estimates calculated from the sampled data. Furthermore, aggregate statistics provided by the ONS have been subjected to seasonal adjustment. The X11 ARIMA seasonal adjustment method has been applied to published statistics (ONS, 2006a and Annex D, ONS, 2006b) to 'smooth' the effect of seasonal travel in the UK, and produce quarterly information that is directly comparable. Data available from the ESDS is available by quarter in annual packages. Anyone using this primary unit data needs to be aware of seasonal adjustment issues. In addition, national estimates of international travel produced from the IPS have been created using complex variable weighting procedures. Little detailed information on

the weighting methodology is provided, although a brief overview is given in Travel Trends (ONS, 2006a).

A further problem for anyone using the primary unit data, is that whilst country and county codes are provided, look-up tables for the codes are not provided digitally. This presents initial problems with interpretation when cross-tabulating origins and destinations as only number code cross-tabulations are produced. Since 1999, the IPS began interviewing passengers on routes between the UK and the Republic of Ireland. This has created a discontinuity in the results that should be considered by anyone conducting a time series analysis.

Whilst IPS data is the primary source used by the Government to produce estimates of international migration, the Total International Migration (TIM) estimates also include data on asylum seekers from the Home Office, as well as data from the Irish Central Statistics Office (ONS, 2006c). Annual publications produced by the ONS on international migration (ONS, 2003; 2004; 2005; 2006c) have all made use of the TIM estimate and therefore the IPS. Other studies that have used the IPS to measure international migrations include Salt (2005) and Large and Ghosh (2006a; 2006b).

Attention has been drawn to problems associated with the international migration figures in the wake of the revelation that the numbers of people coming into the UK from Eastern Europe in recent years have been significantly underestimated. The report for the inter-departmental task force on migration statistics (ONS, 2006d) reviews some of the current issues and shortcomings related to current international migration estimates. Within the report there are recommendations that the IPS increases the sample of emigrants from January 2007, but perhaps more importantly for this issue, revises 'the assumptions in the mid-2006 population estimates on numbers of international migrants whose actual length of stay differs from their stated intentions.' (ONS, 2006d). It is assumed that these recommendations for the IPS will have more of an affect on the adjusted, published statistics rather than the released primary unit data.

4.3 General Household Survey and Northern Ireland Continuous Household Survey

4.3.1 General Household Survey

The General Household Survey (GHS) has been in existence since 1971, and has been conducted on an annual basis since then, with the exception of two breaks – one in 1997/98 and another in 1999/2000. The sample size changes slightly year-on-year, but it is usually between 8,000-10,000 households, which comprise around 15,000-20,000 respondents. Results are published through the ONS in summary form (http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=5756) and the ESDS in primary unit form (http://www.esds.ac.uk/findingData/ghsTitles.asp) on an annual basis. The main purpose of the study is to collect data on a range of core topics, covering household, family and individual information.

Subsequent to a review in 1997/98, the GHS has consisted of two elements – the continuous survey (which contains the household questionnaire, featuring a number of core questions in areas such as employment, education, health, income, socio-economic status and accommodation) and the trailer. The trailer questions differ between surveys, but in the past have included questions on areas such as alcohol consumption, sport and leisure activities, and contraception. The content of the trailer modules is often dictated by the government department sponsoring the trailer at the time.

The GHS has always asked a question relating, in some way, to the amount of time each respondent has lived at a current address. From this, it is possible to derive some indication of in-migration from somewhere else within Britain or immigration from outside Britain. A question relating to how many moves the respondent has made in the past five years is also included. Unfortunately, no question is included which could give a precise indication of the place of origin for internal migrants. It is only possible to determine whether an individual is currently living in a specific GOR, and that they either did or did not live there n years ago. The finest spatial unit of reference for any movement is the GOR/Standard Region scale.

Information relating to the date of arrival in the UK for respondents born elsewhere is included, as is their country of birth. From this, it is possible to infer something about international immigration. However, country of residence before moving to GB is not included as a variable, thus casting some doubt on the real origin of the migrant and limiting any conclusions that can be drawn.

GHS data available from the ESDS begin in 1971, although data for 1971 are only currently available in ASCII format, without accompanying documentation. Data from 1972 to 1982 are available in SPSS format for individuals, having been converted from household level to individual level case records. Data from 1983 to 2005 are available to download for individual cases.

It is not possible to produce an interaction matrix for internal migration from the GHS in any of these years. As indicated previously, destination is the only precise point of reference with origin specified only as 'somewhere else' for internal migrants. With immigrants, it is possible to produce an interaction matrix. However this would only be a matrix of flows from country of birth to current location, not necessarily country of origin; and these provide counts of lifetime migration. Furthermore, whilst the variable 'country of birth' exists in the dataset, 'country' in this case actually means groups of countries. There are around 20 categories of country group available. Table 4.7 shows an example of the type of flow matrix that could be produced from these data. GOR destinations are listed on the vertical axis whilst countries of birth, including the UK, are listed on the horizontal

Table 4.7: An immigration flow matrix derived from the GHS, 2004-05

GOR of Residence / Country of Birth	North East	North West	Yorks and Humber	East Midlands	West Midlands	East of England	London	South East	South West	Wales	Scotland	Total
UNITED KINGDOM	885	2259	1762	1574	1659	1789	1514	2678	1714	955	1845	18634
CARIB COMM	1	7	0	7	18	5	61	5	3	1	3	111
IND BANG	1	10	12	20	29	12	88	22	6	6	10	216
EAFR NEW COMMON	1	6	3	1	11	3	40	9	3	2	2	81
REST NEW COMMON	3	6	4	8	4	22	126	33	13	2	4	225
PAKISTAN	4	18	14	9	23	3	34	15	0	1	6	127
REST OF WORLD	26	80	66	33	76	79	315	182	58	15	79	1009
NA	0	0	5	0	2	0	8	2	1	0	0	18
Total	921	2386	1866	1652	1822	1913	2186	2946	1798	982	1949	20421

Source: GHS, 2004-05

4.3.2 Northern Ireland Continuous Household Survey

The Northern Ireland Continuous Household Survey (CHS) is related to (and indeed modelled on) the GHS in GB. However, the topics covered and continuity of the data are slightly different. Beginning in 1983, the CHS samples around 1% of the households in Northern Ireland. Covering similar general topics to the GHS, there are also variables which can be used to measure population migration. The spatial units used are of a 'finer grain' than those used in GB, with data aggregated by electoral ward as well as by district council area.

Variables allowing the monitoring of migration patterns are not included on the same regular basis as they are in the GHS. In 1983, respondents were asked how many times they had moved in the last five years – more specifically they were also asked how many times they have moved within Northern Ireland, between Northern Ireland and Eire and between Northern Ireland and GB. Unfortunately (from the point of view of comparing migratory patterns longitudinally and from a contemporary perspective) this question was not carried forward into subsequent surveys. Primary unit data are available from the ESDS annually from 1983 to the present but, other than in 1983, studying migration using the CHS is not possible.

4.4 Integrated Household Survey (formerly Continuous Population Survey)

Fieldwork for the Integrated Household Survey (IHS) is scheduled to start in January 2008. This survey will integrate several other ONS surveys including the Labour Force Survey (LFS), General Household Survey (GHS), Expenditure and Food Survey (EFS), the Annual Population Survey (formerly the NeSS Survey) and Omnibus Survey (OMN). Of these existing surveys, only the LFS and the GHS presently contain data that could be used to measure migration. The IHS proposal replaces a previous proposal for a Continuous Population Survey. This survey was only ever a proposal and did not make it 'into the field.'

The survey is to consist of a fixed core of standard questions, with a rotating set of questions in addition to those in the main core, and with a series of potentially changing topic questions. Within the core, questions will be asked relating to areas

such as gender, age, partnership status, ethnicity, religion, country of birth, living arrangements, household type and composition, family status, health, housing, income, education, employment and access to car. The country of birth question means that migration data by birthplace and place of residence may be derived. There will also be questions relating to the usual place of residence and the place of residence one year ago, from which measures of one year migration may be derived. It is not specified at what spatial scale data will be released from the IHS, but it is reasonable to assume that it will be released at very similar spatial scales to the surveys it is replacing. That is to say that GORs are the most likely scale, but there is a chance that data will be available at UA or LAD scale too if special conditions are met by users. In combining previous surveys, the IHS will cover the total number of households involved before, making it the largest continuous survey to be conducted in the UK (some 200,000 interviews per annum). The range of regular outputs from the core module of questions will be available annually through the UK data archive and the ESDS, with quarterly aggregate tables available from ONS on a rolling annual basis.

4.5 National Travel Survey (NTS)

First commissioned in 1965/66, the National Travel Survey (NTS) has, since then, provided periodic snapshots of British travel behaviour. Data are only available from 1972 onwards, with continuous data only available from 1988. From 1996, data has been published in amalgamated triennial blocks, although the most recent data has been released annually. Currently the National Travel Survey samples 16,000 addresses in Great Britain annually. Primary unit data are available to download for selected periods since 1972 from the ESDS. (http://www.esds.ac.uk/findingData/ntsTitles.asp) Summary statistics and reports are also available through the ONS and the department for transport. (http://www.dft.gov.uk/stellent/groups/dft_control/documents/contentservertemplate/dft_index.hcst?n=7216&l=3).

The smallest geographical scale for which aggregate data are made available is the GOR, despite data being collected at postcode sector level. Origins and destinations that are published for each journey are referenced only by GOR. Additional data included for all cases includes variables such as distance and frequency of journeys

made on a given travel day, mode of transport, and average annual and weekly mileage. Standard socio-demographic identifier variables are also featured, including age, gender, marital status, socio-economic group and industry of employment.

Table 4.8 shows a flow matrix between GORs. Interpretation of flow matrices from the NTS needs to be carried out carefully, as the flows represent all journeys carried out by the sample population in their given 'travel week' between origin regions (horizontal axis) and destination regions (vertical axis). As the data available from the ESDS is primary unit data, these flows can be disaggregated by any of the other variables in the survey.

Table 4.8: A regional flow matrix for all trips carried out by the sample population in the 'travel week', 2004

Trip Origin \ Trip Destination	North East	NW & Merseyside	Yorkshire & Humberside	East Midlands	West Midlands	Eastern	Greater London	South East	South West	Wales	Scotland	Total
North East	14664	41	211	10	9	7	8	5	1	3	33	14992
NW & Merseyside	42	34908	288	109	239	21	38	30	25	221	54	35975
Yorkshire & Humberside	214	268	24946	321	31	30	17	22	8	18	16	25891
East Midlands	10	117	324	20994	350	322	53	77	28	12	5	22292
West Midlands	6	235	28	345	28925	32	51	92	151	176	4	30045
Eastern	3	23	20	300	32	29476	1349	468	40	1	15	31727
Greater London	9	29	26	43	58	1358	29499	1780	68	16	10	32896
South East	5	27	28	81	86	449	1792	38487	424	14	4	41397
South West	2	27	10	25	147	30	75	420	30530	106	1	31373
Wales	6	219	17	17	171	8	17	20	107	13789	3	14374
Scotland	37	46	13	3	6	12	7	6	0	2	28971	29103
Total	14998	35940	25911	22248	30054	31745	32906	41407	31382	14358	29116	310065

Source: NTS, 2004

It should also be noted that the quality of data available for interaction is very variable between the surveys. The latest releases of the NTS (from 2002 onwards) provide data which can be used very effectively. For other years, data are much less useful. For example, for the years 1999-2001, whilst it initially appears that there would be good interaction data available, when the individual cases are examined more closely, 86% of the regions for origins and destinations are recorded as 'multipunched.' It is assumed that this means that more than one region has been entered in the questionnaire. Whatever the reason though, the effect is that the flow matrices are rendered more or less useless.

4.6 Summary

Table 4.9 is a summary of the interaction data that are available from social survey sources. Social surveys present a potentially wide-ranging pool from which to extract different types of interaction data. The principal advantages of some of these surveys are that they publish results with high frequency – often annually, but in some cases quarterly. This allows the researcher to build a time series of migration data and to identify migration trends up to the most recent quarter or annual period, thus providing valuable information with which to complement data from the decennial census of population.

Social surveys are not without some less attractive characteristics for those wishing to study population movements over time. One major drawback, shared by most of the social surveys covered here, is that the spatial resolution for published statistics tends to be the GOR (formerly Standard Regions). Such large spatial units mean that only very general patterns of movement can be observed, despite the rich variety of other attributes that can be ascribed to the individual respondents. It is only the IPS that provides details of movements on a scale below the GOR (aside from the NI CHS, which is of very little use to anyone studying population movements) but even the IPS only presents data at the level of county, and this is only for those passengers travelling out of the country.

Further drawbacks include the variability of temporal coverage, with some surveys receding further into history than others, with some surveys having intermittent coverage and with some surveys changing the time periods over which data is collected and/or disseminated.

Finally, as a consequence of the detail inherent in many of the social surveys discussed here, the sample size of the survey is often relatively small. This may well have implications for accuracy. For some surveys this is more of an issue than for others, but nevertheless, it is an issue that any research should be aware of. The National Travel Survey, for example, currently samples less than 6,000 addresses – over eight times less than the Labour Force Survey.

Table 4.9: Summary of survey data sources from which interaction data are available

Survey	LFS	LFS NI	IPS	GHS	CHS	IHS	NTS
Start date	1973	1973	1993	1971	1983	2008	1965/66
Current sample size	53,000 Households, 126,000 individuals annually.	8500- 9000 individuals annually.	250,000 passengers annually.	8000-10,000 households, 15,000-20,000 individuals annually.	4,500 households (around 1% of Northern Ireland total).	204,000 households annually.	16,000 households annually.
Current timing	Calendar quarterly sampling and release.	Calendar quarterly sampling and release.	Continual sampling, quarterly compilation, annual release.	Annual release.	Annual release.	Rolling annual release of calendar quarterly datasets.	Data collected on sample 'travel week' for study sample over course of a year. Annual release.
Main variable types covered (Variables flows can be disaggregated by).	Age, gender, ethnicity, level of education, marital status, religion, number of dependent children, employment type, sick days, socio-economic classification.	Age, gender, ethnicity, level of education, marital status, religion, number of dependent children, employment type, sick days, socio-economic classification.	Age, gender, UK port or route, type of vehicle, type of fare, purpose of visit, intended length of stay, money spent on beer, wine, spirits and cigarettes, overseas origin or destination.	Household members, household and family information, household accommodation, housing tenure, consumer durables including vehicle ownership, employment, pensions, education, health and use of health services, income.	Household members, household and family information, household accommodation, housing tenure, consumer durables including vehicle ownership, employment, pensions, education, health and use of health services, income.	Covering everything included in the LFS and GHS, with additional information on all aspects of household income and expenditure, as well as much information that is normally included in the omnibus survey.	Accessibility of public transport, access to amenities, household vehicle access, household composition and household socio-economic information, age, gender and marital status, employment, occupation and industry details, income, place of work and travel to work details.
Interaction data	GOR to GOR and International country to GOR interaction matrices possible. Disaggregation by any variable of choice. UA/LAD to UA/LAD with special permission.	GOR to GOR and International country to GOR interaction matrices possible. Disaggregation by any variable of choice. . UA/LAD to UA/LAD with special permission.	International country of origin to UK county matrices possible. Disaggregation by any variable of choice.	Very little. GOR of destination is all that can be accurately measured. Origin is either current GOR or 'elsewhere.' There is no way of telling which.	Only for 1983. NI electoral ward or council area can be origin or destination. Immigration from GB or Eire also available for this year alone.	Expected to be the same as the LFS.	GOR to GOR commuting data is available readily for most recent years. This data should be available in theory for other years too, although in practice availability is variable.

5 CONCLUSIONS

The results of the audit reported in this extensive paper demonstrate that origin-destination flow data are available from a wide-ranging set of census, administrative and survey sources, some of which were not specifically designed to provide statistical information to support research on migration or commuting directly, yet provide valuable new details about these patterns of behaviour for which that there is a considerable paucity of reliable information.

It is clear that the last three Censuses of Population have generated a number of products from which flow data can be produced. In most cases, there are online interface and extraction systems or mechanisms of assistance already in place to support users from the academic sector get access to flow data. CASWEB provides access to the sets of Key Statistics, Theme Tables, CAS and Standard Tables, whilst the SARs/CAMS/SAM data are available through the CCSR at Manchester and access to the LS and SLS is available via CeLSIUS or LSCS respectively. CIDER does not seek to replicate these census services or provide access to the data contained within these databases. The same is true for the 2001 data contained in the majority of tables commissioned by users from the ONS that are available directly from the ONS web site and which frequently refer to small subsets for different parts of the country. However, there are certain very large and complex tables where the coverage is extensive and territorially comprehensive. In these cases (e.g. flows of migrants in the 12 months before the census between districts by ethnic group and age), we recommend that it makes sense to facilitate access and extraction by incorporating these data sets into WICID.

Whilst the 2011 Census test questionnaire suggests the potential for additional interaction data relating to visitors, immigrants and moves between different home addresses becoming available in addition to the conventional migration and commuting flows, the final decision on the questions to be included in the 2011 Census and the precise nature of the data to be released post-census will not be known for some time. There are also important decisions to be made in the future about the form of disclosure control that will be applied to 2011 data. Clearly these decisions will have significant implications for the future development of WICID.

Amongst the administrative sources that have been considered in this audit, the data source that has proven most valuable to practitioners and researchers hitherto has been the NHS patient re-registration system, as evidenced by the adoption of NHSCR data into the official population estimation methodology and their use for identifying changes in the magnitude and spatial patterns of movement between censuses. Previous studies have demonstrated that, whilst measuring moves rather than migrants, the NHSCR provides a relatively reliable source of data for monitoring change year on year, even though, as the audit has demonstrated in detail, the data have their shortcomings. A lot of work was put into the preparation of a time series of NHSCR data for a consistent set of 100 zones in the UK (based on FPC areas and than FHSAs in England and Wales plus Scotland and Northern Ireland as single zones) from mid-1975 to mid-1998 by age and sex and we recommend that these flow matrices be made available in WICID for access by users. Permission has been sought from ONS for this to occur.

Since 1998, the situation with respect to NHSCR data became complicated with the transformation from FHSAs to HAs and various revisions of HA boundaries since then, together with the use of individual patient records supplied by HAs throughout England and Wales and held in a Patient Register Data System. Given the value of NHSCR data in understanding inter-censal trends in migration, our recommendation is to obtain the NHSCR data on flows between HAs from ONS since 1998, disaggregated by age and sex, and to make these data available to users through WICID, providing clear guidance on any changes in the spatial units that have taken place, year on year. ONS have confirmed their willingness to release these data although warn that flows for 1998 might not prove reliable because of problems associated with the transition from FHSAs to HAs. Whilst this recommendation would result in the availability of data for England and Wales, further work in the short term is required to negotiate with GROS and NISRA to obtain data between health areas in Scotland and in Northern Ireland respectively. In the longer term, it would be valuable to create a data set that brings these three national data sets together and which provides flows of patients between the spatial units in each

country. This proposal would require collaboration between the stakeholders and separate funding.

Among the other sources of administration data that have been assessed in the audit, two stand out as being suitable for inclusion in WICID for different reasons. These are the HESA data on student migration and the NHS hospital episode statistics. The PLASC data on journey to school also have considerable potential but significant investment is required to ensure that the attributes of individual pupils are correct and consistent from year to year and algorithms would be required to produce the flows of children between particular geographical units. Administrative data sets containing flows of individuals entering the country from overseas, such as the Worker Registration data, the NINo statistics and the Home Office data on asylum statistics and visitor switchers are all extremely useful for helping to understand trends and patterns of immigration, but these data sets are relatively simple in structure and in several cases, available directly from specific web sites or from the Government departments concerned. We conclude that the inclusion of these data sets in WICID, whilst usefully providing access through a single site, would not add significant value to these data sets or greatly facilitate their use in research and teaching. It would, however, be helpful to have some explanation of these data sets available in WICID and links to where users can either access the data or associated metadata.

Student movements from home to HEI are of considerable importance in terms of magnitude and impact on particular areas of the country. HESA collects information annually that allows for monitoring and analysis of movement patterns by age, gender, ethnicity, qualifications, disability, term-time accommodation, source of tuition fees, subject area of study and level of study. The richness of these data would provide an excellent resource for researchers, particularly in association with student data from the census. Consequently, we recommend that annual time series HESA data from 2001 be assembled in WICID for access by users at MLSOA or higher levels. The case for the inclusion of hospital episode data in WICID is rather different since there are no equivalent data in the census, the only commuting flows being those in relation to the journey to work. However, in this instance, whilst the data provide information about a subset of commuting behaviour for which we currently

know very little, they also allow detailed spatial and temporal analysis of ill-health between censuses, as well as huge potential for research on equality and effectiveness issues relating to the provision of hospital services.

Surveys are the third major type of interaction data source and whilst there are several that provide information about migration, there are very few that contain any details of commuting flows. The main advantage of survey data, especially those which are available in primary unit data form such as the IPS, GHS or NTS is that the user can cross-classify different attributes and provide cross-tabulations that complement data from the decennial census of population, as well as providing more up-to-date information. However, one of the key constraints of survey data, due largely to the fact that the records are only a small sample of some population, is the implication for analysis at spatial scales below that of the GOR. In addition, there are problems with intermittent temporal coverage as well as limitations with some of the methods used for collecting the samples associated with particular surveys, e.g. IPS. It is for these reasons, together with the fact that the results of the large surveys are available directly from web sites, that we recommend no survey data to be held in WICID, at least until the results of the Integrated Household Survey (IHS) start to become available.

One of the key messages emerging from the audit is that there is a range of data sets that provide different counts of migration because of the different methods of data collection and measurement. Information from different sources may appear inconsistent and may tell a different story about the magnitude, composition and pattern of movement. However, the idea of a 'one-stop shop' for interaction data, providing a reference resource for a researcher studying migration or commuting flows, provides the potential for better understanding of the flows involved. The idea is similar in many respects to the concept of a 'New Migrant Databank' (NMD), proposed by Rees and Boden (2006) for London, but with the potential for the whole of the UK, to facilitate comparison of internal migration datasets for a consistent set of geographies. Preparing data in this format would have the advantage of enabling a more straightforward comparison of data over time, as data is added to the databank, year-on-year, for the predefined geographies. Rees and Boden show aggregate counts

of immigrants and emigrants as well as net balances from 13 different data sources for the UK, London and its boroughs. The data sources appear as columns in a summary table, with the spatial scales appearing as rows. As might be expected, differing counts are given from each of the different sources at each of the different levels. However, by making these differences explicit and easily identifiable, users are able to reach their own conclusions on the relative usefulness of each dataset. Summary statistics in the form of averages are produced for each row of data to aid in any comparison. A very similar framework could be adopted in WICID for the measurement of internal migrants with additional summary information such as the inclusion of median, mean, maximum, minimum and range values.

The redesign of WICID needs also to incorporate the inclusion of the share of migration flows accounted for by each spatial unit as well as the count data, as suggested by Rees and Boden for the NMD. For example, for each dataset the UK would account for 100% of the data. Individual areas (such as London) within the spatial hierarchy account for differing amounts of the total flows depending on the data set used and therefore it makes sense to use relative as well as absolute magnitudes. Similarly, counts of aggregate numbers and shares of in, out and net flows should be provided, together with information about intra-area flows. As in the current WICID software, there should be functionality to allow users to examine the data sets using a variety of indicators e.g. rates, measures of turnover, churn, connectivity.

Clearly whilst the development of a web-based interaction databank must address the issues of which datasets to include, it is also necessary to consider which spatial scales are appropriate both for holding the base data sets and also for extraction. Depending on the chosen scales within the databank, some degree of data estimation may well be required. For example there would almost certainly be a desire to include NHSCR data alongside census data or data from the Higher Education Statistics Agency. Here data will be available at different and sometimes incompatible scales and work would need to be carried out to make sure that NHSCR data at HA scale is comparable with Census data at LAD scale. Indeed the creation of a consistent set of geographies for all datasets will be the key issue that will need to be addressed.

If the databank is to be produced for a range of scales for the whole country, attention must be paid to exactly how the databank is viewed and accessed by users. The current WICID system could be adapted so that users could extract the relevant sections of the databank either thematically or spatially. An example spatial hierarchy for query and extraction might be UK > Yorkshire and the Humber region > Districts/HAs within Yorkshire and the Humber. The extraction system would be flexible and allow users to extract which ever parts of the hierarchy they required, including for example, every Government Office Region and every District/HA (if those were the geographies selected for inclusion). Appropriate metadata will be required to accompany the base data sets themselves, especially where data has been derived or estimated from an original source. In terms of outputs, it would also be desirable to provide graphical and cartographic data representations where appropriate, to aid in the interpretation of the data. WICID does not offer these facilities at the moment.

Proposals have been formulated for an integrated population statistics system (IPSS) (ONS, 2003) that combines census data at individual level into a single comprehensive statistics database with survey and administrative data and will underpin the country's population and social statistics. At the heart of these proposals is a high quality address register containing information on properties and characteristics of individuals associated with these properties together with a population register, which will provide the basis for linkage with data from other sources. The results of the 2011 Census will form the basis of the information contained in the proposed system that will subsequently be updated with data from further censuses, the proposed IHS and other administrative and registration systems. This system is likely to generate interaction data on a more regular basis and it will be very important to ensure that data release is maximised without the effects of disclosure control becoming too detrimental. The experience of CIDER in handling different types of registration data in the run up to 2011 will be beneficial if and when data from the new integrated system eventually comes on line.

In summary, the recommendations emerging from this audit that relate to the data that should be included in WICID are as follows:

(1) 2001 Census: the large and more complex matrices of migration and commuting flows commissioned from ONS that have national coverage at district and sub-district spatial scales;

(2) NHSCR: annual flows, from 1975 to 1998, of NHSCR patient re-registration movements between 100 FHSA-based zones, disaggregated by age and sex; and annual flows, from 1998/99 onwards, of NHS patients movements between HAs, disaggregated by age and sex;

(3) HESA: annual flows, from 2001 onwards, of student movements between MLSOA of parental domicile and HEI, disaggregated by various characteristics; and

(4) NHS IC: annual flows, from 2001 onwards, of hospital patients from LLSOA or MLSOA of residence to hospital, disaggregated by various attributes.

References

Al-Hamad, A., Hayes, L. and Flowerdew, R. (1997) Migration of the elderly to join existing households: evidence from the Household SAR, Environment and Planning A, 29(7): 1243-55.

Boden, P. (1989) The analysis of internal migration in the UK using census and National health Service Central Register data, Unpublished PhD Thesis, University of Leeds, Leeds

Boden, P. and Stillwell, J. (2006) New migrant labour in Yorkshire and the Humber, The Regional Review for Yorkshire and the Humber, 16(3): 18-20.

Bover, O.A.M. (2002) Learning about migration decisions from the migrants: Using complementary datasets to model intra-regional migrations in Spain, Journal of Population Economics, 15(2): 357-380.

Boyle, P. (1995) Public housing as a barrier to long-distance migration, International Journal of Population Geography, 1: 147-64.

Boyle, P. and Feng, G. (2002) A method for integrating the 1981 and 1991 GB census interaction data, Computers, Environment and Urban Systems, 26: 241-256.

Bulusu, L. (1989) Migration in 1988, Population Trends, 58: 33-39

Bulusu, L. (1990) Internal migration in the United Kingdom, 1989, Population Trends, 62: 33-36.

Bulusu, L. (1991) Review of migration data sources, OPCS Occasional Paper 39, OPCS, London.

Champion, A. (2005a) Migration in the 2001 Census: what can the SAR add to our understanding? Paper presented at the SARs User Group Meeting, Royal Statistical Society, 15 July.

Champion, A. (2005b) Population movement within the UK, Chapter 6 in Chappell, R. (ed) Focus on People and Migration 2005 Edition, Palgrave Macmillan, Basingstoke, pp. 91-113.

Champion, A., Bramley, G., Fotheringham, A., Macgill, J. and Rees, P. (2003) A migration modelling system to support government decision making, Chapter 15 in Geertman, S. and Stillwell, J. (eds) Planning Support Systems in Practice, Springer, Heidelberg, pp. 269-290.

Champion, A., Coombes, M., Raybould, S. and Wymer, C. (2007) Migration and Socioeconomic Change A 2001 Census Analysis of B*ritain's Larger Cities*, Joseph Rowntree Foundation, York.

Champion, A., Fotheringham, S., Boyle, P., Rees, P. and Stillwell, J. (1998) The Determinants of Migration Flows in England: A Review of Existing Data and Evidence, DETR, London.

Chappell, R., Vickers, L. and Evans, H. (2000) The use of patient registers to estimate migration, Population Trends, 101: 19-24.

Cole, K., Frost, M. and Thomas, R. (2002) Workplace data from the census, in Rees, P., Martin, D. and Williamson, P. (eds) The Census Data System, Wiley, Chichester, pp. 269-280.

Courgeau, D. and Pumain, D. (1993) Mobilité par temps de crise, Population et Sociétés, 279: 25-41.

Dale, A. (1998) The value of the SARs in spatial and area-level research, Environment and Planning A, 30, 767-774.

Davies, E., Williamson, P. and Houldsworth, C. (2006) The Leaving of Liverpool: An Examination into the Migratory Characteristics of Liverpool, Mimeo, University of Liverpool.

Devis, T. (1984) Population movements measured by the NHSCR, Population Trends, 36: 15-20.

Devis, T. and Mills, I. (1986) A comparison of migration data from the National Health Service Central Register and the 1981 Census, OPCS Occasional Paper 35, OPCS, London.

Duke-Williams, O. (2004) The development and use of information systems for monitoring and analysing migration in Britain, Unpublished PhD Thesis, University of Leeds.

Duke-Williams, O. and Rees, P. (1993) TIMMIG: a program for extracting migration time series tables, Working Paper 93/13, School of Geography, University of Leeds.

Duke-Williams, O. and Stillwell, J. (2007) The effect of small cell adjustment on interaction data produced from the 2001 Census, Environment and Planning A.

Dustmann, C. and Faber, F. (2005) Immigrants in the British labour market, Fiscal Studies, 26(4): 423-470.

Dustmann, C., Fabbri, F. and Preston, I. (2005) The impact of immigration on the British labour market, Economic Journal, 115(507): 324-341.

Department of Work and Pensions (2006) National Insurance Number Allocations to Overseas Nationals Entering the UK 2006, DWP, London, available at www.dwp.gov.uk/asd/asd1/niall/niall_report.pdf

Elias, P. (2006) The National Strategy for Data Resources for Research in the Social Sciences, ESRC, Swindon.

Ewens, D. (2005a) The National and London Pupil datasets: An introductory briefing for researchers and research users, Data management and Analysis Group, Greater London Authority.

Ewens, D (2005b) Moving home and changing school: widening the analysis of pupil mobility, Data management and Analysis Group, Greater London Authority.

Fielding, A. (1992) Migration and social mobility – South East England as an escalator region, Regional Studies, 26(1): 1-15.

Fielding, A. (1998) Counterurbanisation and social class, in Boyle, P. and Halfacree, K. (eds) Migration in Rural Areas: Theories and Issues, John Wiley, Chichester.

Fotheringham, A., Rees, P., Champion, T., Kalogirou, S. and Tremayne, A. (2004) The development of a migration model for England and Wales: overview and modelling out-migration, Environment and Planning A, 36: 1633-1672.

Forsythe, F. (1992) The nature of migration between Northern Ireland and Great Britain - a preliminary analysis based on the Labour Force Surveys, 1986-88, Economic and Social Review, 23(2): 105-127.

Gordon, I. (1995) Migration in a segmented labour market, Transactions of the Institute of British Geographers, 20(2): 139-155.

GRO Scotland (2006) http://www.gro-scotland.gov.uk/statistics/publications-and-data/annual-report-publications/annrep/01sect5/estimating-migration.html

GRO Scotland (2007) http://www.gro-scotland.gov.uk/statistics/migration/information-on-migration.html

Harland, K. and Stillwell, J. (2007a) Commuting to school in Leeds: how useful is the PLASC?, Working Paper 07/02, School of Geography, University of Leeds, Leeds.

Harland, K. and Stillwell, J. (2007b) Using PLASC data to identify patterns of commuting to school, residential migration and movement between schools in Leeds, Working Paper 07/03, School of Geography, University of Leeds, Leeds.

HESonline (2007) http://www.hesonline.nhs.uk/ (accessed 24/1/07).

Hornsey, D. (1993) The effects of computerisation on the NHS Central Register on internal migration statistics, Population Trends, Winter, 34-36.

Horsfield, G. (2005) International migration, Chapter 7 in Chappell, R. (ed) Focus on People and Migration 2005 Edition, Palgrave Macamillan, Basingstoke, pp. 114-129.

Jones, P. and Elias, P. (2006) Administrative data as research resources: a selected audit, Draft Paper, Warwick Institute for Employment Research.

Kalogirou, S. (2005) Examining and presenting trends of internal migration flows within England and Wales, Population, Space and Place, 11(4): 283-297.

Kuijper, H. and Noordam, R. (1992) Household moves in the Netherlands in 1990, Maandstat Bevolking, 40(1): 12-7.

Large, P. and Ghosh, K. (2006a) A methodology for estimating the population by ethnic group for areas within England. Population Trends 123, 21-31.

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

Liffen, K. Maslen, S. and Price, S. (1988) HES Book, Department of Health.

Madouros, V. (2006) Impact of the LFS switch from seasonal to calendar quarters: an overview of the switch of the LFS to calendar quarters and the potential effects of this change on users, ONS, London.

Machin, S., Telhaj, S. and Wilson, J. (2006) The mobility of English school children, Discussion paper, Centre for Economic Performance.

NISRA (2005) Development of methods/sources to estimate population migration in Northern Ireland, NISRA Paper, http://www.nisra.gov.uk/archive/demography/publications/dev_est_mig.pdf (accessed 13/06/07).

Norman, P., Boyle, P. and Rees, P. (2005) Selective migration, health and deprivation: a longitudinal analysis, Social Science and Medicine, 60(12): 2755-2771.

Office of the Deputy Prime Minister (2002) Development of a Migration Model, ODPM, London. http://www.communities.gov.uk/index.asp?id=1510540

Office for National Statistics, General Register Office for Scotland, Northern Ireland Statistics and Research Agency (2006) UK Statistical Disclosure Control Policy for 2011 Census Output, November.
http://www.statistics.gov.uk/census/pdfs/SDCpolicy.pdf (accessed 1/03/07)

Ogilvy, A. (1980) Inter-regional migration since 1971: an appraisal of data from the National Health Service Register and Labour Force Surveys, OPCS Occasional Paper 16, OPCS, London.

ONS (1999) 1996-based subnational population projections England, ONS PP3 No 10, The Stationery Office, London.

ONS (2003a) Proposals for an Integrated Population Statistics System, Discussion Paper, ONS, October.

ONS (2003b) International Migration: Migrants Entering or Leaving the United Kingdom and England and Wales, 2001. Series MN no.28. Office for National Statistics, London.
http://www.statistics.gov.uk/downloads/theme_population/MN28. pdf (accessed 15/1/07)

ONS (2004a) Proposals for a Continuous Population Survey, Consultation Paper, ONS, July.

ONS (2004b) International Migration: Migrants Entering or Leaving the United Kingdom and England and Wales, 2002. Series MN no.29. Office for National Statistics, London.
http://www.statistics.gov.uk/downloads/theme_population/MN_no_29_v3.pdf (accessed 15/1/07)

ONS (2005a) The 2011 Census: Initial View on Content for England and Wales, Consultation Document, ONS, London.

ONS (2005b) International Migration: Migrants Entering or Leaving the United Kingdom and England and Wales, 2003. Series MN no.30. Office for National Statistics, London.
http://www.statistics.gov.uk/downloads/theme_population/MN_ No30_2003v3.pdf (accessed 15/1/07)

ONS (2006a) Travel Trends. A report on the 2005 International Passenger Survey, Palgrave, Macmillan. http://www.statistics.gov.uk/downloads/theme_transport/traveltrends2005.pdf (accessed 15/1/07)

ONS (2006b) Methodology for the experimental monthly index of services. http://www.statistics.gov.uk/iosmethodology/downloads/Whole_Report.pdf (accessed 15/1/07)

ONS (2006c) International Migration: Migrants Entering or Leaving the United Kingdom and England and Wales, 2004. Series MN no.31. Office for National Statistics, London. http://www.statistics.gov.uk/downloads/theme_population/MN31.pdf (accessed 15/1/07)

ONS (2006d) Report for the Inter-departmental Task Force on Migration Statistics. http://www.national-statistics.org.uk/about/data/methodology/specific/population/future/imps/updates/downloads/TaskForceReport151206.pdf (accessed 15/1/07)

OPCS (1991a) Making a population estimate in England and Wales, OPCS Occasional Paper 37, OPCS, London.

OPCS (1991b) Sub-national population projections: population projections by sex and age for standard regions, counties, London boroughs, metropolitan districts, and regional and district health authorities of England from mid-89, OPCS Series PP3 No. 8, HMSO, London.

OPCS (1995) 1993-based subnational population projections, England, OPCS Series PP3 9, London.

Platt, L., Simpson, L. and Akinwale, B. (2005) Stability and change in ethnic groups in England and Wales, Population Trends, 121: 35-46.

Poulain, M. (1996) Confrontation des statistiques de migration intra-Européennes: vers plus d'harmonisation, European Journal of Population, 9(4): 353-81.

Rees, P. (1977) The measurement of migration from census and other sources, Environment and Planning A, 9: 257-280.

Rees, P., Martin, D. and Williamson, P. (eds) (2002) The Census Data System, Wiley, Chichester.

Rees, P. and Boden, P. (2006) *Estimating London's* New Migrant Ppopulation: Stage 1 – Review of Methodology, Greater London Authority, London.

Rosenbaum, M. and Bailey, J. 1991. Movement within England and Wales during the 1980s, as measured by the NHS Central Register, Population Trends, 65: 24-34.

Salt, J. (2005) International Migration and the United Kingdom - Report of the United Kingdom SOPEMI correspondent to the OECD, 2005. http://www.geog.ucl.ac.uk/ mru/docs/Sop05fin_20060627.pdf (accessed 15/1/07)

Scott, A. and Kilbey, T. (1999) Can patient registers give an improved measure of internal migration in England and Wales?, Population Trends, 96: 44-55.

Shields, Ma, P. (1998) The earnings of male immigrants in England: evidence from the quarterly LFS, Applied Economics, 30(9): 1157-1168.

Stillwell, J. (1994) Monitoring intercensal migration in the United Kingdom, Environment and Planning A, 26: 1711-1730.

Stillwell, J. (2006a) Providing access to census-based interaction data in the UK: that's WICID, The Journal of Systemics, Cybernetics and Informatics, 4(4): 63-68.

Stillwell, J. (2006b) Using WICID (Web-based Interface to Census Interaction Data) in the classroom, The Journal of Systemics, Cybernetics and Informatics 4(6): 106-111.

Stillwell, J. and Debenham, J. (2001) The projection of population, households, housing and residential land use, Chapter 5 in Stillwell, J. and Scholten, H. (eds) Land Use Simulation for Europe, Kluwer Academic Publishers, Dordrecht, pp. 45-64.

Stillwell J. and Duke-Williams, O. (2003) A new web-based interface to British census of population origin-destination statistics, Environment and Planning A, 35: 113-132.

Stillwell J. and Duke-Williams, O. (2007) Understanding the 2001 UK census migration and commuting data: the effect of small cell adjustment and problems of comparison with 1991, Journal of the Royal Statistical Society Series A, 170(Part 2): 1-21.

Stillwell, J., Duke-Williams, O. and Rees, P. (1995) Time series migration in Britain: the context for 1991 Census analysis, Papers in Regional Science: Journal of the Regional Science Association, 74(4): 341-359.

Stillwell, J., Duke-Williams, O., Feng, Z. and Boyle, P. (2005) Delivering census interaction data to the user: data provision and software development, Working Paper 05/01, School of Geography, University of Leeds, Leeds, p. 38.

Stillwell, J., Rees, P. and Boden, P. (eds) (1992) Migration Processes and Patterns Volume 2 Population Redistribution in the United Kingdom, Belhaven Press, London.

Stillwell, J., Rees P. and Duke-Williams, O. (1996) Migration between NUTS level 2 regions in the United Kingdom, in Rees, P., Stillwell, J., Convey, A. and Kupiszewski, M. (eds) Population Migration in the European Union, Wiley, Chichester.

Stillwell, J., Rees, P., Eyre, H. and Macgill, J. (2002) Improving the treatment of international migration, Appendix 7, ODPM (2002) Development of a Migration Model, ODPM, London, pp 223-238. http://www.communities.gov.uk/index.asp?id=1510540

The Information Centre (2006) Reporting Patient Journeys: Hospital outpatient activity in 2003-04 and 2004-05, The Information Centre. (http://www.ic.nhs.uk/pubs/hesoutpatients2003to2005/hespub/file) (accessed 24/1/07).

UK Data Forum (2006) National Strategy for Data Resources in the Social Sciences, Economic and Social Research Council. http://www.esrc.ac.uk/ESRCInfoCentre/Images/National_Data_Strategy_tcm6-18160.pdf

Vignoles, A., Goodman, A., Machin, S. and McNally, S. (2006) Widening participation in higher education: a quantitative analysis, see url: http://www.tlrp.org/proj/pdf/vignoles3.pdf.

Williams, M. (2000) Migration and social change in Cornwall 1971-91, in Creeser, R. and Gleave, S. (eds) Migration in England and Wales Using the Longitudinal Study, ONS Series LS, No. 9, The Stationery Office, London, pp 30-39.

Index