Working Paper 11/04

Data Sources and Estimation Methods for Sub-national Migration in the United Kingdom

Nik Lomax, John Stillwell, Paul Norman and Phil Rees

School of Geography,

University of Leeds,

Leeds LS2 9JT

Tel: 0113 343 3315

Abstract

International and internal migration flows are extremely important components of sub-national population dynamics in the United Kingdom and the three national statistics agencies have developed their own distinctive methodologies for estimating both of these component inputs to their respective mid-year population estimates at district level. This paper, which initially provides a review of these methods and the data which underpin them, attempts to identify the gaps or inconsistencies in the data sets that would enable a complete and consistent UK-wide set of flows to be estimated of both international migrants into and out of each sub-national area and internal migration between the areas concerned at different spatial scales. It then outlines selected methods that might be used to rectify the deficiencies and concludes with a series of comparisons of different parts of a matrix of flows estimated from administrative sources for mid 2000/01 with equivalent flows derived directly from the 2001 Census for 2000/01. The results of the comparison demonstrate strong correlations between the estimated and observed migration flows (apart from cross-border flows), providing supportive evidence of the value of creating a time series of sub-national migration flows within the UK over the 2000s.

Acknowledgements

The authors are grateful to Pete Large, Michelle Bowen and Alistair Davies at the Office for National Statistics, Esta Clarke, Ganka Mueller and Bart Jaron at National Records of Scotland and David Marshall and Jos Ijpelaar at the Northern Ireland Statistics and Research Agency for providing information and data. Nik Lomax is supported by an ESRC CASE studentship awarded in partnership with the Office for National Statistics.

Contents

List of Tables

List of Figures

Glossary of Terms

APS	Annual Population Survey
ASC	Annual School Census
CA	Council Area (district in Scotland)
SCHI	Scottish Community Health Index
CIDER	Centre for Interaction Data Estimation and Research
DfE	Department for Education
District	Local Authority, Unitary Authority, Council Area or Local Government District
FHA	Former Health Authority
Flag-4	The marker put on a GP registration record when the person's previous address was overseas
GOR	Government Office Region
GP	General Practitioner
HA	Health Authority
HB	Health Board
HE	Higher Education
HESA	Higher Education Statistics Authority
IPS	International Passenger Survey
L2	Lifetime Labour Market Database
LA	Local Authority
LAD	Local Authority District
LEA	Local Education Authority
LFS	Labour Force Survey
LGD	Local Government District in Northern Ireland
LTIM	Long-Term International Migration
MSIP	Migration Statistics Improvement Program
MWS	Migrant Workers Scan
MYE	Mid-Year Estimate
NHSCR	National Health Service Central Register
NICHI	Northern Irish Community Health Index
NINo	National Insurance Number

NMD	New Migrant Databank
NMGi	New Migrant Geography In
NMGo	New Migrant Geography Out
NRS	National Records of Scotland
NSAs	National Statistical Agencies
NUTS	Nomenclature of Territorial Units for Statistics
ONS	Office for National Statistics
PC	Parliamentary Constituency in Northern Ireland
PCT	Primary Care Trust
PRDS	Patient Register Data System
SAS	Census Small Area Statistics
TIM	Total International Migration
UA	Unitary Authority
UKSA	United Kingdom Statistics Authority
WICID	Web-based Interface to Census Interaction Data
WRS	Workers Registration Scheme

1. Introduction

Migration is an integral component of population change in most countries, alongside the natural change components of births and deaths. The UK Statistics Authority (UKSA) has reported that international migration is the most significant driver of population change in the United Kingdom (UK), while internal migration has a *"substantial influence on the changing level and composition of the population in local areas"* and *"sub-regional population estimates, projections and migrant information are also essential for the planning of local service delivery"* (UKSA 2009, p.1). However, for the four home countries which comprise the UK, the official national statistical agencies (NSAs) in the UK – the Office for National Statistics (ONS) in England and Wales, the National Records of Scotland (NRS) and the Northern Ireland Statistics and Research Agency (NISRA) – all recognise that migration is the most difficult component to measure, particularly the international dimension.

The need to improve internal and international migration statistics has been widely acknowledged by all three NSAs and by the report of an Interdepartmental Task Force on Migration (National Statistics, 2006) that made a series of recommendations for 'Improving Migration and Population Statistics' (IMPS), led by the ONS. In 2008, a Parliamentary Committee reviewed the inadequacy of official population statistics and its report (House of Commons Treasury Committee, 2008) resulted in the Migration Statistics Improvement Programme (MSIP), the vehicle through which the Government is intending to deliver the Task Force recommendations by 2012. UKSA (2009) reviews progress on MSIP and the adequacy of co-operation across government to deliver the planned improvements, whilst commissioned research by Rees *et al*. (2009) published in the UKSA report, provides a comprehensive summary of migration data sets, a critique of MSIP and a review of migration estimation methods.

Statistics on annual migration in the UK are compiled separately by the NSAs and fed through to ONS who compile an aggregate mid-year estimate (MYE) of the population of each local authority area (district) in the UK together with estimates of the components of change (http://www.statistics.gov.uk/statbase/product.asp?vlnk=15106). Each of the agencies is responsible for producing more detailed MYEs for the districts within its borders. There are, however, a number of availability and consistency problems associated with the international and internal migration data used in the population estimation process. These problems are identified in this paper which has the following three objectives:

(i) to outline the methods used by the NSAs for generating estimates of international and internal migration to use in their respective mid-year population estimation systems;

(ii) to review the time series data sets that are available for sub-national migration in the UK and identify the major gaps or inconsistencies that exist; and

(iii) to consider the methods that can be applied (a) to produce sub-national estimates of flows of immigrants to and emigrants from the 'rest of the world' into UK local authority districts (LADs); and (b) to produce a complete matrix of migration flows between the full set of districts within each country in the UK and also between these districts in the four home nations.

The first of these objectives provides the context in which to consider the second and third, the latter indicating that the longer-term aim is to produce a time series of annual (mid-year to mid-year) estimates that will allow the monitoring, analysis and modelling of both sub-national migration components of population change in the UK over the decade from 2001 to 2011, i.e. from one census to the next. The Special Migration Statistics (SMS) from the 2001 Census provide a complete matrix of sub-national migration flows for 2000/01 (with the exception of emigration flows from each district to the 'rest of the world'). The flow counts taking place in the year before the 2001 Census at the national level are shown in Table 1 for which a similar matrix will be available for 2010/11 from the 2011 Census in due course.

Table 1: UK migration flows, 2000/01

	Destinations					
Origins	England	Wales	Scotland	Northern Ireland	Rest of the World	Total outflows
England	5,153,436	48,248	43,675	7,899	0	5,253,258
Wales	42,614	243,851	1,546	325	0	288,336
Scotland	42,831	1,396	473,789	2,633	0	520,649
Northern Ireland	8,812	360	2,602	127,999	0	139,773
Rest of the world	360,531	9,916	28,868	7,461	0	406,776
Total inflows	5,608,224	303,771	550,480	146,317	0	6,608,792
No usual address	400,368	19,721	36,562	10,401	0	467,052

Source: 2001 Census (Special Migration Statistics Table MG101)

The diagonal elements of Table 1 represent the flows between 434 sub-national districts within each of the home nations (355 in England, 21 in Wales, 32 in Scotland and 26 in Northern Ireland) as indicated in Figure 1. This 'internal intra-national' migration

comprises almost 91% of the 6.6 million migrants whose origins and destinations were both stated in the 2001 Census returns. The off-diagonal elements are the flows between the home nations or the 'internal inter-national' flows and the 'immigration flows from the rest of the world' which account for a further 3% and 6.1% respectively of the 6.6 million migrants. In addition, the 2001 Census recorded a further 467,000 migrants at destinations within the UK whose origins were not stated on the census forms in 2001 so it is not clear whether these are internal intra-national, internal international or external immigration flows. Figure 1 shows the 434 districts that make up the sub-national geography of the UK within the 12 regions (9 English Government Office Regions (GORs) plus Wales, Scotland and Northern Ireland).

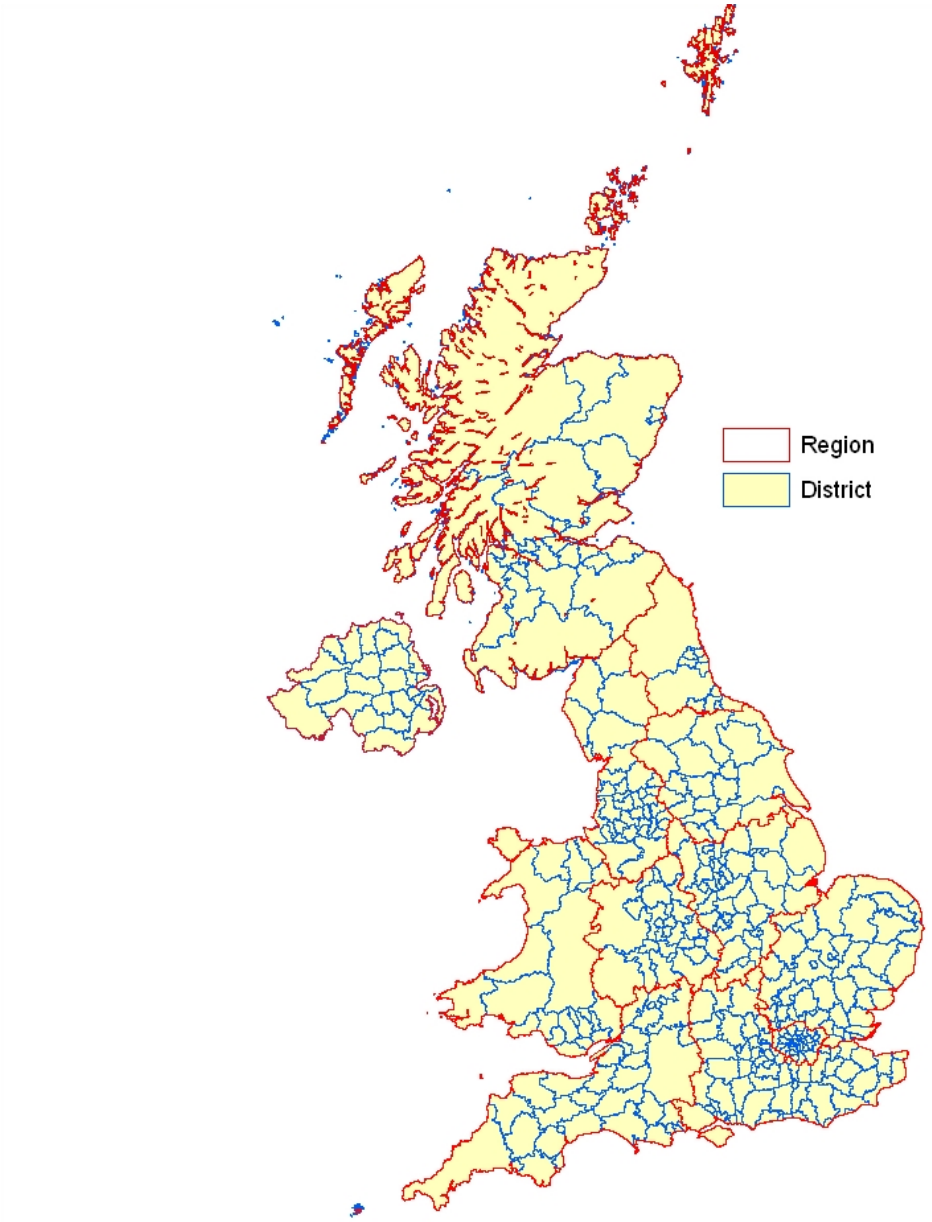


Figure 1: The spatial units for sub-national migration used in the 2001 Census

The remainder of this paper is structured into eight sections. Section 2 summarizes the cohort component method for producing MYEs. Sections 3 and 4 outline the different data sets used and the methods adopted by the NSAs for the estimation of international and internal migration respectively. Initially, we explain the methods that have been used until recently but there is an ongoing process of improvement taking place as new data sets emerge and as new methods come to the fore. Consequently, the ways in which ONS, NRS and NISRA are seeking to improve their methodologies using new or improved sources are also summarised. Section 5 is a short section that covers the availability of 'international' data on flows between England and Wales, Scotland and Northern Ireland. Section 6 addresses the availability of and inconsistencies in migration statistics at the UK level and section 7 explores various methods employed to estimate migration and missing data. Section 8 contains a discussion of the methodology required to estimate sub-national migration flows in aggregate terms for one mid-year to mid-year period (overlapping the 12 months before the 2001 Census) and these estimates are compared with an equivalent set of flows extracted from the Census. Concluding remarks are offered in section 9.

2 Cohort Component Method for Mid-Year Population Estimates

The MYEs, produced in June each year, inform resource allocation and policy decisions at national, regional and local levels. An estimate is produced at a sub-national level across the UK – for Local Authorities (LAs) and Unitary Authorities (UAs) in England and Wales, Council Areas (CAs) in Scotland and Local Government Districts (LGDs) in Northern Ireland – broken down by age and sex. The estimates are generated by a cohort component model (Figure 2) in which the migration component comprises both international and internal flows and is introduced as a net value.

Figure 2: The cohort component method for population estimation

The figures for England and Wales are produced by ONS, the Scottish estimates are produced by NRS and the Northern Ireland estimates by NISRA. ONS then compiles a UK-wide MYE assuming that "*the definition, data sources and methods used by NRS and NISRA*

are broadly consistent, providing comparable population estimates across the UK constituent countries and a coherent UK national compilation" (ONS 2010e, p.2). This is certainly the case for ageing on the population, births and deaths. The usually resident population is aged on one year from the previous MYE. Live births between 1 July of the previous year and 30 June of the reference year are added to population estimates at age zero and are allocated to the district where the mother is usually resident. Deaths during the same period are subtracted from the population of the district of residence by age at the mid-year reference point. Information on both births and deaths are disaggregated by sex (ONS 2010e; NISRA 2006; GROS 2010b). The 'other changes' comprise estimates for special populations such as armed forces and prisoners, plus any boundary changes that happen during the year. Student populations are included as part of the migration component.

However, the NSAs draw on different data sources and methods for the internal and international migration components of their respective MYEs and these are outlined in more detail in the next two sections. It is important to emphasise that sets of available data, assumptions and geographical boundaries used in the estimation of migration are constantly evolving; in this dynamic context, a definitive overview of methodology has limitations, given that historical estimates are often revised based on new information. This is considered further in sections 3.4 and 3.5. The data that are available, and that will inform future analysis in this project, are explored in more detail in section 6.

3. Estimation of the International Migration Component

3.1 National level estimation

All three national statistical agencies draw on the United Nations definition of a long-term migrant as *"someone who changes his or her country of usual residence for a period of at least a year"* (ONS 2009b, p. 3). ONS compiles an estimate of Long-Term International Migration (LTIM) for the whole of the UK as well as estimates of short-term migrants (staying for 3-12 months) and visitors to the UK (staying less than 3 months). The LTIM estimates produced by the ONS are based on the following:

- *The International Passenger Survey (IPS) flows* – international migration flows are based on a sample of 0.2% of travellers interviewed on entering or leaving the UK via the major air, sea and Channel Tunnel embarkation points. Of the 230,000 interviews conducted in 2008, 2.2% were migrants, giving a sample size of around 5,000 (ONS 2009b). The IPS is

an ‘intentions based’ survey, asking migrants where they intend to go and how long they intend to stay in the country. Reliance on the IPS has been widely criticised. Kupiszewska and Nowok (2008, p.57) found that analysis of time series IPS data showed *“strong fluctuations, compared with much smoother curves reported by, for example, the Netherlands”* where a population register is used. The limited sample size of the IPS has been emphasised by Rees (2008, p.354) who states that emigration flows at a regional scale can only be ‘guestimated’, and by Stillwell *et al.* (2010, p. 2) who suggested that even at regional scale, users are advised to *“smooth out irregularities in the data by calculating three year averages”*. A Parliamentary Select Committee (2008, p.26) concluded that the IPS was *“not fit for purpose”* as a source of migration statistics as its primary intended use is for measurement of tourist flows.

- *Migrant switcher and visitor switcher flows* – because the IPS is an intentions-based survey, an adjustment is made for ‘migrant switchers’ who state an intention to stay for over 12 months but who stay, in fact, for less than 12 months and for ‘visitor switchers’ whose intention is to stay for less than 12 months but who remain for over 12 months (ONS 2007d).
- *Northern Ireland migration flows* – since 2008, health card registrations have been used to capture the international migration flows between Northern Ireland and the rest of the world reported in the LTIM. This is because the IPS has not historically sampled air or sea ports in Northern Ireland (although Belfast is a recent addition) and the health card registration system is seen as a more accurate measure of flows (ONS 2010e). Two main criticisms of the Northern Ireland international migration estimate exist. Firstly, health card registration identifies both long-term and short-term migrants (ONS 2010d) resulting in a potential overcount of long-term immigration to Northern Ireland. Secondly, the emigration estimate derived from health card de-registrations underestimates the number of emigrants from Northern Ireland and subsequently needs to be scaled up by 50% (NISRA 2010).

These components of international migration are used to distribute migration estimates to the sub-national level, and the methodologies for this estimation process vary between England and Wales, Scotland and Northern Ireland. The data and methods used in immigration and emigration estimates at the sub-national level are outlined in more detail in the following section.

3.2 Sub-national estimation

Each of the NSAs in the UK has its own method for estimating international migration at the sub-national level. Whilst IPS data combined with migrant/visitor switcher and asylum seeker data are used for England, Wales and Scotland at the national level, these national estimates are distributed differently to the sub-national level by ONS for England and Wales and NRS for Scotland. NISRA does not use the IPS, relying instead on a distribution method based on health card registrations, as describes in section 3.1. The following sections indicate in more detail the estimation methods that the NSAs have been using for both immigration and emigration. As methods and data availability change frequently, the method below was current at September 2011. Work undertaken to develop new methods and produce new estimates is discussed in sections 3.5 and 3.6.

3.2.1 ONS estimation methods

A three-stage estimation procedure for *immigration* is as follows. The ONS first distributes the IPS estimate to the regional level (Wales plus 9 GORs in England) using the Labour Force Survey (LFS) three-year average. The LFS is a sample of 60,000 households per quarter and is seen to give a more accurate distribution of immigrants than the intention-based inflows captured by the IPS (ONS 2007c). The LFS distribution is used as a control total and the IPS estimate is then allocated to the regional level using the LFS distribution by broad age group and sex. The IPS three-year average estimate is then distributed to an intermediate geography called the 'new migrant geography' for immigration (NMGi). The NMGi is an aggregation of districts which share a boundary and have a minimum of 20 IPS contacts per year (ONS 2009a). NMGi replaced the increasingly obsolete health geography – comprising 92 former Health Authority (HA) areas in 2007. In the third stage, immigrants are allocated to the district level using a Poisson regression model which incorporates a number of covariates such as Flag-4 General Practitioner (GP) registrations and National Insurance Number (NINo) registrations of overseas immigrants and immigrant counts from the 2001 Census. The covariates vary each year as the ONS (2007f) report that fixing the covariates caused volatility in the model over time. The weighted IPS estimate is the response variable and the *"approach reduces the variability in the IPS estimates at local authority level by making use of their relationship with the predictor variables"* (ONS 2010e, p.6). The district estimates are constrained to sum to the national and regional IPS estimates.

The method used for London boroughs is slightly different; all non-students are allocated to the NMGi level using the LFS three-year average rather than using the IPS three-year average as occurs for the rest of the UK. This is because the sample size of the LFS is sufficiently large for London boroughs, but not for districts outside of London (ONS 2007b). Non-UK students are distributed to London boroughs directly without the use of the NMGi based on data supplied by the Higher Education Statistics Authority (HESA).

A similar approach is adopted for *emigration.* The IPS interview includes a sample of international emigrants at UK air, sea and Channel Tunnel embarkation points. This estimate is used at both the regional level and distributed as a three-year average at the intermediate level (NMGo) where the 'o' stands for out. NMGo areas are based on the NMGi areas with some adjustment made to account for smaller numbers of out-migrants in the IPS (ONS 2007e). The NMGo areas are larger and there are fewer individual areas in order to provide a robust sample size. A Poisson regression model is used at the district level, with the IPS direct estimate as the response variable. The specification differs from the immigration model in the selection of variables, since it includes the immigration estimate from the previous year, housing type and housing tenure. Unlike the immigration model, the covariates are fixed for each year (ONS 2010b).

Detailed data on asylum seekers is provided by the Home Office and are incorporated into the sub-national estimates of international immigration and emigration. While the data are considered high quality, they *"do not correspond directly to the standard ONS definition of a long-term international migrant"* and as such, broad assumptions are made by ONS about the proportion of asylum seeker applicants that actually correspond to the ONS definition (ONS 2011, p.2). Asylum seekers are distributed sub-nationally based on the Home Office data.

3.2.2 NRS estimation methods

The approach of the NRS, formerly the General Register Office for Scotland (GROS), to sub-national *immigration* estimation has made use of Scottish National Health Service (SNHS) patient registration data. The Scottish share of UK LTIM is initially derived using the LFS which is seen to give a more accurate distribution of international migrants than the intention based counts specified in the IPS (GROS 2010b). The Scottish allocation of LTIM is then distributed to Scottish Health Board (HB) areas using overseas inflows recorded on the Scottish National Health Service Central Register (SNHSCR) which includes an age/sex

distribution. The distribution of immigrants to districts is based on the Scottish Community Health Index (SCHI) which records the postcode of patients registering with a GP in Scotland. The SCHI gives the date of registration and a record where an individual previously resided overseas is marked IMM and classed as an international migrant move (although this field is not mandatory) (GROS 2003).

The majority of asylum seekers are assumed to be supported by the National Asylum Support Service (NASS) and as such are removed from the LTIM control totals and distributed to Glasgow, which is the only Scottish district in contact with the United Kingdom Border Agency (UKBA) (GROS 2010b). A small proportion of non-NASS asylum seekers are distributed around the rest of Scotland and all asylum seeker distributions are based on the five-year age bands provided by the ONS (GROS 2010b).

Emigration estimation is based directly on the IPS *"using averaged proportions based on international inflows, outflows to the rest of the UK and the population size of each Health Board"* (GROS 2010a, p.1). The age/sex distribution is based on the distribution of migrants to the rest of the UK, derived from the SNHSCR.

3.2.3 NISRA estimation methods

The methodology in Northern Ireland differs from the rest of the UK as NISRA does not make use of data from the IPS. Instead, health card registration data are used in both the immigration and emigration estimates. In the case of *immigration*, registration with a family doctor requires an international immigrant to apply for a health card, at which point he/she must provide information about age, place of residence and time of stay to the Business Services Organisation of Health and Social Care (HSC-BSO) in Northern Ireland (NISRA 2010). Health card data are seen as the most comprehensive source with which to estimate international migration and give an indication of intention to stay for a period of time, as registration is only possible for a migrant staying for over three months. To account for under-registration by young males, the age distribution is adjusted to be similar to that of young female migrants in the estimates (NISRA 2010). Health card registrations give detail allowing estimates to be disaggregated by age and sex.

Emigration estimates are also derived from the health card system which records de-registrations with a family doctor, and which are adjusted for young males in a similar way to the immigration estimates. The reported total is scaled up by 50% to take into account the low de-registration rate (NISRA 2010) as de-registration is not mandatory and there is little

incentive to do so. The de-registration data are combined with the data from the Central Statistics Office (CSO) Irish Quarterly National Household Survey which provides an estimate of numbers moving from Northern Ireland to the Republic of Ireland.

Immigration and emigration by asylum seekers in Northern Ireland is distributed sub-nationally using the same Home Office data used by ONS for England and Wales (ONS 2011).

3.3 Administrative sources and improvements in the estimation methods

In recent years, the NSAs have realized that administrative sources can offer new insights into migration levels at the sub-national level. There are problems, however, in estimating the entire migrant population as the administrative sources that are available measure distinct sub-groups which are not consistent. This is well illustrated by ONS' Local Area Migration Indicators (LAMI) database system (http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-222711) as reviewed by Stillwell *et al.* (2011, p.131). The LAMI system includes registrations of NINOs by overseas nationals, registrations with a GP where a person's previous address was overseas (Flag-4) and Home Office data from the Workers Registration Scheme (WRS). When these data are examined for some districts, the difference between the counts from each administrative source are substantial (Boden and Rees 2010). A NINo can be obtained by both short-term and long-term migrants, and some migrants may obtain a NINo yet never travel to the UK. Migrants may register with a doctor if they intend to stay for over three months; however this falls short of the UN definition of a long-term migrant, and undercounting of particular age groups has occurred. ONS (2007i) undertook a review of differences between the reported flows from the WRS, NINo allocations and the data from the Patient Register Data System (PRDS) for 2005-2006, finding a number of inconsistencies between expected and actual differences. So, for example, more districts recorded higher NINo registrations than PRDS registrations despite the fact that all migrants are eligible to register with a GP. ONS have been working to improve the distribution of international migrants using administrative data sources, driven by the aims of the Task Force on Migration. The first revisions occurred in 2007, with subsequent revision in 2010 and a further revision proposed for late 2011. The following section briefly outlines these changes.

3.4 ONS revisions to migrant distribution in 2007 and 2010

In 2007, ONS implemented a series of improvements which primarily impacted on the distribution of international migration to the sub-national level. It was the 2007 improvements that saw the introduction of the new migrant geography (NMG) to replace the increasingly obsolete Health Authority geography which perpetuated a *"lack of consistency across the country in terms of the numbers of LAs within each intermediate geography"* (ONS 2008, p.2). Improvements to the visitor and migrant switcher methodology (as discussed above) utilised new questions in the IPS, first asked in 2004 to improve on the assumptions made by ONS for visitor/switcher numbers. The questions asked previous migrants *"when you last arrived in* (left) *the UK, how long did you intend to stay* (away) *for?"* (ONS 2007d, p.5).

Two further improvements were implemented, utilising administrative and other data sources. First, the LFS was introduced to improve the regional distribution of migrants. This apportioned the IPS estimate to the GOR for England, and nationally to Scotland and Wales. The new method replaced distributions based on the IPS with the LFS three-year average distribution of migrants, as analysis from the 2001 Census and the LFS showed that *"the distribution of where migrants live by country and region differs from where migrants state they intend to live in the IPS"* (ONS 2007h, p.2). Second, a Poisson regression model was introduced to improve emigration estimates at the district level. The model utilised variables available at the district level such as population density, prior year immigration figures and variables derived from the 2001 Census (ONS 2007g). The 2007 improvements were applied to produce revised estimates for mid-2002 to mid-2005 and for the mid-2006 estimate going forward. The cumulative effect of the revised estimate between 2002 and 2005 was a net increase of 28,600 migrants in England and Wales (ONS 2007a).

In 2010, a series of improvements were introduced as part of the MSIP programme, most of which are currently used by the ONS for the estimation of migration and consequently have been explained already. These constituted an enhanced methodology for the estimation of international immigration and emigration, an Irish adjustment and an improved estimate of internal migration by students (covered in detail in section 4.2). The improved methodology for immigration introduced a Poisson regression model for distribution of immigrants to the district level which replaced the use of migrant distribution reported in the 2001 Census. The Poisson model for emigration was refined to include a set of fixed covariates, rather than use different variables each year (ONS 2010b). Finally, 2010

saw the ONS adopt the NISRA estimates for immigration to and emigration from Northern Ireland, derived from health card registrations. By using the health card data already utilised by NISRA, ONS sought to *"ensure consistency for users"* (ONS 2010d, p.1). ONS also ceased to use estimates of migration between the UK and the Republic of Ireland derived from the Irish Central Statistical Office, replacing this with an IPS estimate (ONS 2010d).

The revised methodology was used to update population estimates from mid-2002 to mid-2008, and from 2009 onwards. The cumulative impact of the changes to the international estimation methodology between 2002 and 2008 was an increase of 8,300 migrants for England and Wales (ONS 2010g). The improvements to the distribution of international migrants at the sub-national level were based on work carried out by Boden and Rees (2010) and informed the revision of distribution methodology currently underway at ONS.

3.5 ONS 2011 revisions to the distribution of international immigrants

Boden and Rees (2010) developed a methodology as part of the construction of the New Migrant Databank (NMD) to distribute the LTIM estimate to the district level by using administrative data sources. In light of an immigration 'reason profile' being stated in the IPS (worker, student and other), Boden and Rees (2010) allocated the estimate of Total International Migration (TIM, the precursor to LTIM) to the district level using the administrative sources that best represent the 'reason profile'. NINo allocation data were used to distribute migrants *"whose stated reason for migration was either a 'definite job' or 'looking for work'"* (Boden and Rees 2010, p. 720). Higher Education Statistics Authority (HESA) data were used to allocate immigrants who stated '*formal study*' and GP registrations were used to allocate the remaining responses of '*accompany/join'* or '*not stated*.' Boden and Rees (2010) found that there was a significant difference between the TIM estimates and their model estimates in a number of local authorities and suggested that a process of local consultation would be appropriate to understand the variances. Following the work of Boden and Rees (2010), Bijak (2010) reviewed the ONS and NMD methods, concluding that, at district level, a method based on administrative sources was preferable but highlighted differences between estimates that occur when different data sources are used (citing the use of HESA data in London as an example). The method derived by Boden and Rees (2010) coupled with the 2010 improvements discussed above have informed the latest methodological developments being undertaken by the ONS.

Based on the work of Boden and Rees, the ONS propose to allocate the IPS estimate of immigration to the sub-national level by utilising available administrative sources. The revised methodology responds to some known estimation issues in certain districts and the relative complexity of the Poisson regression model for distribution (Mcgregor 2011). The new ONS method follows the Boden and Rees (2010) method of using administrative data sources to allocate the IPS estimate based on the 'reason profile', with the main improvement being the exclusion of short-term migrants captured in the data sources. This has meant using the Migrant Worker Scan and the Lifetime Labour Market Database (known as L2), which can identify long-term and short-term migrants as opposed to all NINo registrations to overseas nationals (Mcgregor 2011). It is possible to distinguish long-term and short-term student migrants from HESA data while Flag-4 counts are still used to allocate the other flow (presumably as registration with a GP gives a good indication of intention to stay for a length of time). Data from the Department of Business, Innovation and Skills has also been made available to ONS for the distribution of Further Education Students (Swier 2011). The method will be used to create indicative LTIM estimates for mid-2006 to mid-2010 (due for release on 17 November), and will be used in the formulation of the 2011 estimates subject to benchmarking against results of the 2011 Census (Swier 2011).

3.6 NISRA and NRS use of administrative sources

The Interdepartmental Task Force on Migration has also been the catalyst for investigation of ways to improve migration statistics at NRS and NISRA. Since 2006, NISRA has undertaken research to inform the international migration statistics that they produce, drawing on estimates available from a number of different administrative sources. NISRA (2010) identifies the sources that can be used to measure sub-groups of the migrant population, both immigrants and emigrants.

In terms of immigration, workers are identified through the WRS, the number of NINo registrations and number of applications to work through the Home Office points based system. School children can be identified through the Annual School Census, which identifies a child whose first language is not English and from 2009 asked schools how many children joined the system whose previous address was outside Northern Ireland. NISRA has also assessed HESA data in order to identify higher education students who were domiciled outside of Northern Ireland. Births to mothers and fathers from outside of Northern Ireland can be counted when a new birth is registered, since the mother and father are required to

give their country of birth. Northern Ireland Housing Executive data are used to record migrant worker households applying for social housing and finally, the LFS is able to indicate the age structure of the foreign-born population.

A number of sources that measure sub-groups of the *emigrant* population are also identified by NISRA (2010). NINos issued to foreign nationals that fall out of use can be used as a proxy for worker emigration while school children emigrating are measured through the Annual School Census, which asks schools for the *"number of pupils who left Northern Ireland in the previous year"* (NISRA 2010, p.38). Finally, specific out-migration questions have been included in the Continuous Household Survey and Omnibus Survey. The immigration and emigration estimates derived from administrative data only cover sub-groups of the migrant population and are used to quality assure officially published international migration estimates which are based on data from health card registrations and de-registrations (discussed in section 3.2.3).

The NRS has also undertaken work addressing the potential of using HESA and Annual School Census data to improve the distribution of LTIM (Mueller 2011). However, to date, no revised estimates have been used or published. Rolfe and Metcalf (2009) provide a comprehensive review of data sources available to the Scottish Government. Similar to the work carried out by NISRA, they identify sources that identify migrant sub-groups highlighting that *"the data is not representative of migrants, but of a self-selecting subset of migrants"* (p.15). Rolfe and Metcalf (2009) suggest the use of the Annual School Census and HESA data to measure immigrants in education and the use of the LFS and WRS for those in employment. They also suggest that the recently launched Integrated Household Survey (IHS) could increase the availability of data on migrants, but will be limited by its sample size.

4 Estimation of Internal Intra-national flows in the UK

Across the UK, internal migration statistics are derived primarily from NHS sources. They are produced independently by ONS, NRS and NISRA and supplied to ONS for collation. ONS and NRS produce district to district tables of moves, both of which are available in the public domain. NISRA also produces internal migration statistics for their MYEs but these are currently not published. The methodologies used in each case and statistics produced are outlined in more detail.

4.1 ONS estimation method

The ONS produces a full matrix of origin-destination flows between districts in England and Wales which are estimated by combining data from the NHSCR and Patient Register Data System (PRDS). The NHSCR records movements between the former Health Authority (HA) areas in England and Wales; a download is supplied by all Primary Care Trusts (PCTs) which is then aggregated to form the NHSCR database for England and Wales. In 2006, HAs became a redundant health geography but NHSCR estimates continue to be published based on their boundaries (ONS 2010a). A review of the NHSCR using the Longitudinal Study (LS is a sample of circa 500,000 people at each census date) was carried out by Smallwood and Lynch (2010) and suggests that 95.7% of ONS LS members enumerated at the 2001 Census resided in the same area as the NHSCR record. They also found that only 1.5% of those enumerated in the 2001 Census did not appear on the NHSCR at all.

However, the NHSCR does not contain comprehensive enough detail for estimation at a lower geographical level than HA, namely there is no postcode information for patients. For this reason it is combined with the PRDS which does record the postcode of patients (ONS 2010f). At the district level, the PRDS identifies patients using a unique identification number, and a migration is recorded when a change in postcode is picked up from one yearly download to the next. Moves within a district are discarded, as are any changes that come about through boundary changes. An adjustment is then made using data from the UK Higher Education Statistics Authority (HESA) to take into account student movements as explained in more detail below. The PRDS estimates adjusted for student flows are then constrained (scaled) to the NHSCR totals to give an estimate of internal migration between districts across England and Wales. Estimates are made annually for mid-year to mid-year periods which are available in rounded form from ONS.

4.2 ONS student adjustment

The rationale behind applying the student adjustment has been set out by ONS (2010c). First, young people, particularly young men, can be slow to change their registration with a GP when they move. Second, movements of students attending higher education can be complex, including transfers to the place of study, moves during the study period and moves after completing their study programmes. Students may have two addresses, a term-time address and a home (domicile) or parental address, both of which they spend time at. For these reasons, ONS concluded that HESA data have potential to improve migration estimates for

the student population and introduced the adjustment in 2010. The focus of the adjustment is internal migration moves made by first year undergraduates and students at the end of their studies *"who did not change their GP registration when they moved"* (ONS 2010c, p.2). The adjustment consists of three calculations:

- A start of study adjustment – this is applied only to first-year undergraduate students by comparing the term-time LA to the domicile LA by single year of age and sex, and is based on the assumption that most students begin university at age 18 or 19. Where the HESA flows between domicile and term-time LA are larger than the PRDS flows (for this age group), HESA data are used. An adjustment between 2001 and 2007 was also made using the 2007/08 term time distribution since term-time address was not collected by HESA prior to academic year 2007/08. A 'flag' is used to identify a student who lives at his/her parents address during term time, and each flagged record was removed if it was a feasible distance from the campus of study (ONS 2010d).
- An end of study adjustment – as there is no source which identifies where students move to at the end of their studies, a more complex set of estimations is required as follows:
  - the number of people who end their studies each year is collected by HESA, and includes term-time address from 2007/08. For adjustment between 2002 and 2008, the 2007/08 term-time address distribution has been used;
  - the number of former students moving to a different local area after their studies is taken from 2001 Census data, using the question asking for address twelve months ago. A Census record is only used if an individual held an undergraduate degree at age 22 or a postgraduate degree at age 23. These records are used to calculate a rate for graduates leaving a district (Graduates in the Census who left the district divided by Census graduates in the LA 12 months before the 2001 Census);
  - the number of students who move but do not re-register with a GP – first the rate of students who do re-register is calculated based on moves from the PRDS for mid-2000 to mid-2001, compared to moves from the 2001 Census by sex and age of 17-28 year olds. The rate of moves not identified on the patient register is then calculated as 1 minus the above; and
  - the destination of former students not re-registering is calculated using 2001 Census data to create a matrix of district to district moves, disaggregated by sex for an individual who held an undergraduate degree at age 22, or postgraduate degree at age 23.

- A double counting adjustment – as students are likely to re-register with a GP eventually, an investigation into the amount of time this took was conducted at halls of residence at Bournemouth, Aberystwyth, Newcastle and Northumberland universities. These students were tracked over time to see how long it took to re-register. This includes both a 'start of studies' and an 'end of studies' adjustment.

The adjustment method attempts to deal with problems encountered when producing mid-year population estimates as students move to university after the mid-year reference point (30 June). Assuming students re-register with a GP when they move to university, they will be counted at their home (parents) address in the first year of their study, but their term-time address in the second. At the end of their study, the academic year (particularly for undergraduate students) often ends before the mid-year reference point, *"hence former students may be registered at a new address they have only lived at for a fraction of the mid-year to mid-year period"* (ONS 2010d, p.2).

4.3 NRS estimation method

An inter-council area (CA) matrix of flows is produced by NRS using the SNHSCR to record movements of migrants between Health Board (HB) areas combined with the Scottish Community Health Index (SCHI) which contains the postcode of people registered with an NHS doctor in Scotland, allowing for internal migration to be estimated at CA and below. As in England and Wales, the data derived from the SNHSCR are available quarterly but do not contain any address detail below the HB level. For this reason, the annually downloaded SCHI estimates are controlled to the SNHSCR totals by origin, destination, age and sex (GROS 2010b). No student adjustment is made for inter-CA flows in Scotland.

4.4 NISRA estimation method

NISRA estimate flows at Local Government District (LGD) level using the Northern Irish Central Health Index (NICHI) which records changes in address when a patient re-registers with a GP after a migration event. The accuracy of using the NICHI was investigated by NISRA by comparing results with the 2001 Census. NISRA found that the NICHI reported 35,500 inter-LGD moves while the Census recorded 37,100 moves. The age and sex breakdown showed similar patterns (NISRA 2007). In addition, a student adjustment is made, informed by administrative data sources, by removing a number of people of student

age from most LGDs and *"adding these to a small number of LGDs with centres of third level education"* (NISRA 2007, p.3).

4.5 Potential improvements to methodologies by ONS and NRS

The last sections have explained the methods that the NSAs have been using to extract and integrate data from various sources so as to generate the best estimates of internal migration for use in their cohort component models. This section introduces ways that ONS and NRS are considering to improve their internal migration methodologies using new or improved data.

4.5.1 ONS improvements

The ONS are assessing the possibility of using data held on patients in the Personal Demographics Service (PDS), a constituent part of the NHS system that make up the 'spine' of the NHS care records service (NHS 2011). By using the PDS, a wider proportion of the population of England and Wales will be covered as patient address details will be added at more points of contact with the NHS and will not rely solely on registration with a GP in England or Wales. As summarised by the Select Committee on Public Accounts (2007, p. 1), this *"provides more convenience for patients as they need only notify one authorised healthcare organisation of a change of address and this change will be available to all healthcare organisations as and when the patient records are accessed"*. The introduction of the PDS would have a significant impact on migrant estimation for hard to measure groups. Thus, for example, young males who attend an A&E department would have their address details stored even if they had not re-registered with a GP.

4.5.2 NRS improvements

The NRS is looking towards the use of a SNHSCR monthly extract which includes the postcode of all people registered with a Scottish GP. At present, the SNHSCR extract contains no address details. The effect of this change would allow patients to be tracked by postcode from the SNHSCR which is deemed to provide better coverage than the SCHI. The potential to remove reliance on the SCHI would allow the estimates to be more direct, as SCHI totals would no longer need to be constrained to SNHSCR totals and estimates would no longer need to be constrained to HB areas. Work is currently underway to assess the postcode data provided by the SNHSCR in relation to postcode data in the SCHI (Mueller 2011).

One potential advantage of being able to use SNHSCR without the reliance on SCHI is that the former is able to identify *movements* while the latter can only identify *transitions.* As highlighted by Rees and Willekens (1986), movements or events can occur multiple times within a given time period, whereas a transition compares a person's location at the beginning and end of a given time period, therefore only one person transition is measured. Transition data often miss significant migrations; for example, a person who moves from one district to another just after the start of a time period, and subsequently moves back just before the end of the time period.

5 Internal Inter-national flows in the UK

Cross-border moves between constituent countries of the UK are compiled by the ONS to inform the MYEs. These statistics are reported by the receiving country as these are seen to be more accurate than those of the sending country. The level of detail available on migration across the borders varies between the constituent countries and the following sub-sections outline the data that are available. However, flows disaggregated to anything below health geography units are not estimated and so data on cross-border flows between districts in each of the four constituent countries (internal inter-national flows) do not currently exist.

5.1 Flows in the NHSCR

Between England, Wales and Scotland, the NHSCR and SNHSCR are both able to provide counts that distinguish cross-border flows between health geographies (HBs in Scotland and HAs in England and Wales). The SNHSCR, compiled in Dumfries, records moves from HAs to HBs while the NHSCR, compiled in Southport, records the moves in the opposite direction. In addition, both are able to identify a move from Northern Ireland, but do not distinguish origins at a sub-national level (Northern Ireland is treated as a single area in the data). The data are available as counts of moves by age and sex. The register in the receiving country is adjusted first, and notification of the move is then communicated to the sending country and its register is updated. NRS send their matrix of flows from the SNHSCR to ONS who administer the data. The Audit Commission (2006), under the national duplicate registration initiative, assessed the 56 million electronic records of patients registered with a PCT or Local Health Board in England and Wales for 2004/05, identifying 185,000 records (0.3% of the population) which could be deleted based on a number of criteria, including

duplicated records, ‘gone aways’ who no longer lived at the address held and deceased persons.

5.2 Flows from the PRDS, SCHI and NICHI

The PRDS in England and Wales, the SCHI in Scotland and the NICHI in Northern Ireland record the postcode of people registered with a GP in their respective countries. These registers are also able to identify a patient who reports their previous address as being in another part of the UK, but they do not provide any level of sub-national detail of previous address. These data are available to the NSAs by age and sex.

6 Data availability and inconsistencies across the UK

The sections hitherto have been concerned with the various methodologies used by the constituent NSAs in the UK based on the international and internal migration data sets that are available and preferred in each case. In this section, we consider the availability of data across the UK. Tables 2 to 4 are metadata tables that attempt to synthesise information about data that are available at all spatial scales in the UK using a symetric origin-destination matrix. The focus of the exercise is on data that are easily obtainable or already available in the public domain as these data will be used for estimation presented in section 8. It can clearly be seen that as the level of spatial disaggregation increases, the availability and quality of data decreases. Table 2 shows flows at the national level, between the constituent countries in the UK and the rest of the world. Table 3 shows flows between ‘intermediate geographies’ in the UK – regions, NUTS1/NUTS2 areas and health geographies – and the rest of the world. Table 4 shows the flows at the most disaggregated level – between administrative districts – as well as flows to/from the rest of the world. The tables show what data are made available by the NSAs and serve to highlight the inconsistencies between data that can be obtained. Section 6.4 considers what additional data exist to augment the data that are available in the public domain.

6.1 Flows between countries in the UK and the rest of the world

The data available to measure flows between UK countries (the internal intra-national component) and between UK countries and the rest of the world can be seen in Table 2. Generally, data for flows between the constituent countries of the UK and the rest of the world are made available by the NSAs and show good consistency between sources. The

NHSCR provides aggregate UK country to country flows on a quarterly basis and are available online from March 2002 to September 2010, with a back series available from ONS on request. The flows reported in the NHSCR are provided by the receiving country.

Flows to and from the rest of the world are also well reported. The estimates of LTIM discussed above are readily available at the national level for England, Scotland and Wales from ONS and NRS. NISRA also publish on their website estimates of total immigration and emigration based on health card registrations and de-registrations at LGD level (from 2005 onwards). A back series (from 2001/02) is available on request. All international migration statistics are available for mid-year to mid-year periods and ONS LTIM tables are also available by calendar year. NISRA produce tables of aggregate flows to and from the rest of the world for mid-year 2001 to 2009. They also provide age and sex breakdown from health card registrations from mid-2008.

6.2 Flows between intermediate geographies and the rest of the world

The data available to measure flows between intermediate geographies and the rest of the world are summarised in Table 3. First, the flows between GORs and NUTS1 areas can be considered together, as English GORs are each NUTS1 areas, alongside Wales, Scotland and Northern Ireland as three separate NUTS1 areas. Data availability is good, the NUTS1-NUTS1 flows are published online by ONS on a quarterly basis between March 2002 and June 2010, with earlier tables available on request. For 2009 and 2010 the NHSCR tables are disaggregated by broad age group. No data are currently available for NUTS2 areas.

Data for flows to and from the rest of the world is sparse below the GOR and NUTS1 level of spatial disaggregation. The LTIM estimates produced by ONS include a flow between the rest of the world, English GORs and Wales, while NISRA and NRS produce national level flows. No flow data are readily available to or from the health geographies and the rest of the world.

Flow data between health geographies varies between the countries. ONS produce a HA-HA matrix of moves quarterly, starting in March 2009. This also includes flows to and from HA to/from Scotland and Northern Ireland at an aggregate level. NRS produce a HB-HB matrix of flows from 1995/96 to 2008/09. These tables also include an aggregate ‘rest of UK’ flow and a flow for armed forces. NISRA does not produce data on flows between Health and Social Service Board areas (although these are collected by the Central Services

Agency, Belfast). As the GOR flow data are based on NHSCR data, flows between FHAs in England and Wales and GORs (other than the GOR in which the FHA sits) are published by the ONS, but are only available from March 2009 to June 2010.

6.3 Flows between administrative districts and the rest of the world

The data available to measure flows between administrative districts in the UK and the rest of the world can be seen in Table 4. Firstly, the availability of data for flows between districts in each constituent country varies. Flows between LAs and UAs in England and Wales, produced by ONS are available for mid-1998 to mid-2009, and use the PRDS estimate constrained to NHSCR totals. The published tables are available at an aggregate level. Aggregate flows between CAs in Scotland are published by NRS for 2001/02 to 2008/09. Flows between LGDs in Northern Ireland, however, are not publically available.

The availability of flow data for inter-country administrative geographies also varies. Flows from England, Wales and Northern Ireland to Scottish CAs are only available as an aggregate 'rest of UK' flow from 2006/07. Prior to this, the flows from the rest of the UK were included in an aggregate 'outside Scotland' flow into each CA, which encompassed both UK and rest of the world flows. Flows into English and Welsh LAs and UAs can be derived from the rest of the UK but with no sub-national detail of the origin. Flows to LGDs in Northern Ireland from the rest of the UK are available from NISRA by request only. Internal migration statistics produced by ONS and NRS also vary as the ONS adjusts for the student population, whereas NRS does not. NISRA does apply a student adjustment to their flow estimates but this is not made publically available.

Table 2: Flows between UK countries and the rest of the world

Origin \ Destination	England	Wales	Scotland	Northern Ireland	RoW
England	N/A	ONS produces an estimate of total migration from England to Wales based on NHSCR moves. The data are available from the website for 2002 to 2010 and 2001 data are available on request.	Total migration to Scotland from England & Wales (combined) is available from NRS quarterly for 1991 to 2010 based on NHSCR moves. Additionally, the ONS NHSCR table identifies moves from England to Scotland for 2002 to 2010 (2001 on request).	NISRA supply data on total moves from England to NI to ONS who publish the UK wide matrix of moves in the NHSCR tables. ONS make the data available. 2001 data are available on request from ONS.	ONS produce LTIM estimates of emigration from England to RoW by calendar year and mid year. Data are available from 1991/92 to 2008/09. this includes revisions to 01/02- 06/07 introduced in 2010. Age and sex are also available.
Wales	ONS produces an estimate of total migration from Wales to England based on NHSCR moves. The data are available from the website for 2002 to 2010 and 2001 data are available on request.	N/A	Total migration to Scotland from England & Wales (combined) is available from NRS quarterly for 1991 to 2010 based on NHSCR moves. Additionally, the ONS NHSCR table identifies moves from Wales to Scotland for 2002 to 2010 (2001 on request).	NISRA supply data on total moves from Wales to NI to ONS who publish the UK wide matrix of moves in the NHSCR tables. ONS make the data available. 2001 data are available on request from ONS.	ONS produce LTIM estimates of emigration from Wales to RoW by calendar year and mid-year. Data are available from 1991/92 to 2008/09. This includes revisions to 01/02- 06/07 introduced in 2010. Age and sex are also available.
Scotland	ONS produce an estimate of total migration from Scotland to England based on NHSCR tables. The data are available from 2002 to 2010 and 2001 data are available on request.	ONS produce an estimate of total migration from Scotland to Wales based on NHSCR tables. The data are available from 2002 to 2010 and 2001 data are available on request.	N/A	Flows from Scotland to Northern Ireland are published by ONS based on the NHSCR for 2002 to 2010. 2001 data are available on request.	NRS produce estimates of migration from Scotland to the rest of the world by broad age group for mid 2001/02 to 2008/09. the data are based on the Scottish LTIM allocation.
Northern Ireland	ONS produce an estimate of total migration from Northern Ireland to England based on NHSCR tables. The data are available on a quarterly basis from March 2002 to June 2010 and for 2001 on request.	ONS produce an estimate of total migration from Northern Ireland to Wales based on NHSCR tables. The data are available on a quarterly basis from March 2002 to June 2010 and for 2001 on request.	NRS produce estimates of total migration in to Scotland from NI on a quarterly basis between 1991 and June 2010.	N/A	NISRA publish total migration out of NI to RoW for mid-years from 2001 to 2009.
RoW	ONS produce LTIM estimates in to England by calendar year and mid-year. Data are available from 1991/92 to 2008/09. This includes revisions to 01/02- 06/07 introduced in 2010. a disaggregation by age and sex is also available.	ONS produce LTIM estimates in to England by calendar year and mid year. Data are available from 1991/92 to 2008/09. This includes revisions to 01/02- 06/07 introduced in 2010. a disaggregation by Age and Sex is also available.	NRS produce estimates of migration to Scotland from RoW by broad age group for mid 2001/02 to 2008/09. The data are based on the Scottish LTIM allocation.	NISRA publish total migration in to Northern Ireland from RoW for mid-years 2001 to 2009. Health card registrations are available by age/sex for 2008 and 2009.	N/A

Key:	Good Data	Some Data	Poor/No Data

Table 3: Flows between regional/intermediate geographies and the rest of the world

Destination / Origin	Regional (GOR plus Wales, Scotland and NI)	NUTS1	NUTS2	Health Board/ Former Health Authority/ NI - H &	RoW
Regional (GOR plus Wales, Scotland and NI)	Flows between Scotland, Wales, Northern Ireland and GORs in England are produced by ONS based on NHSCR data for 2002 to 2009. 2001 is available on request. The data are disaggregated by broad age group for June 2009 and June 2010.	Wales, Scotland, Northern Ireland and English GORs are NUTS1 regions and flow data are produced by ONS on the same basis as the regional flows.	No Data	NHSCR tables produced by ONS provide flows from each GOR in to FHAs (within other GORs) by broad age group. Scotland and NI are included as single areas. Data are available quarterly for March 2009 - June 2010.	ONS produce LTIM estimates out of Wales and English GORs by calendar year and mid-year. Data are available from 1991/92 to 2008/09. This includes revisions to 01/02- 06/07 introduced in 2010. Data for Scotland and Northern Ireland between 2001 and 2009 are produced by NRS and NISRA respectively.
NUTS1	Wales, Scotland, Northern Ireland and English GORs are NUTS1 regions and flow data are produced by ONS on the same basis as the regional flows.	Wales, Scotland, Northern Ireland and English GORs are NUTS1 regions and flow data are produced by ONS on the same basis as the regional flows.	No Data	NHSCR tables produced by ONS provide flows from each GOR in to FHA (within other GORs) by broad age group. Scotland and NI are included as single areas and not disaggregated. Data published quarterly March 2009 - June 2010.	ONS produce LTIM estimates out of Wales and English GORs by calendar year and mid-year. Data are available from 1991/92 to 2008/09. This includes revisions to 01/02- 06/07 introduced in 2010. Data for Scotland and Northern Ireland between 2001 and 2009 are produced by NRS and NISRA respectively.
NUTS2	No Data	No Data	No Data	No Data	No Data
Health Board/ Former Health Authority/ NI - H & SSB	NHSCR tables produced by ONS provide flows from each HA to each GOR (excluding the GOR in which the HA sits) by broad age group. Scotland and NI are included as single areas. Data published quarterly March 2009 - June 2010.	NHSCR tables produced by ONS provide flows from each HA to each GOR by broad age group. Scotland and NI are included as single areas and not disaggregated. Data published quarterly March 2009 - June 2010.	No Data	ONS produce a FHA-FHA matrix available quarterly for March 2009 to June 2010 including Scotland and NI as single areas. March 2002 - Dec 2008 only show GOR level moves (including Scotland and Northern Ireland). NRS produce a HB-HB matrix 1995/96 to 2008/09. NISRA do not publish Health and SSB moves.	No Data
RoW	ONS produce LTIM estimates in to English GORs by calendar year (2 Series) and mid-year (3 Series). Data are available from 1991/92 to 2008/09. This includes revisions to 01/02- 06/07 introduced in 2010. Data for Scotland and Northern Ireland between 2001 and 2009 are produced by NRS and NISRA respectively.	ONS produce LTIM estimates in to English GORs by calendar year (2 Series) and mid-year (3 Series). Data are available from 1991/92 to 2008/09. This includes revisions to 01/02- 06/07 introduced in 2010. Data for Scotland and Northern Ireland between 2001 and 2009 are produced by NRS and NISRA respectively.	No Data	No Data	N/A

Key:	Good Data	Some Data	Poor/No Data

Table 4: Flows between districts and the rest of the world

Origin \ Destination	England and Wales (LA/UA)	Scotland (CA)	Northern Ireland (LGD)	RoW
England and Wales (LA/UA)	Full matrix of inter LA/UA flows are produced by ONS from mid 1998 to mid 2009. These use PRDS, constrained to the NHSCR totals. The flows include a student adjustment from 2002 onwards. Flows disaggregated by age and sex can be obtained on request from ONS.	The aggregate inflows for each CA from the 'rest of UK' are available for 2006/2007 to 2008/09, which is the total for E&W and NI. Prior to this, Rest of UK was included in an 'outside Scotland' total and are available from 2001.	Aggregate migration to LGDs in Northern Ireland from the rest of the UK are available on request from NISRA for 2001/02 to 2009/10.	Flows from LA/UAs to RoW are produced by ONS in the detailed components of change for mid 2002 to mid 2009 which are available from ONS by request.
Scotland (CA)	Flow data from Scotland to LA/UAs are not readily available, however a flow from the rest of the UK can be derived by using the total inflow from the rest of the UK (including England and Wales) minus the internal England and Wales moves derived from the PRDS tables. The two tables are available from 2000/01 to 2009/10 (on request).	A matrix of inter council area flows is produced by NRS. This covers 2001/02 to 2008/09.	Aggregate migration to LGDs in Northern Ireland from the rest of the UK are available on request from NISRA for 2001/02 to 2009/10.	Flows from Scottish CAs to RoW are produced by GROS. Prior to 2006/07 the RoW flow included England, Wales and NI.
Northern Ireland (LGD)	Flow data from Northern Ireland to LA/UAs are not readily available, however a flow from the rest of the UK can be derived by using the total inflow from the rest of the UK (including England and Wales) minus the internal England and Wales moves derived from the PRDS tables. The two tables are available from 2000/01 to 2009/10 (on request).	The aggregate inflows for each CA from 'rest of UK' are available since 2006/07, which is the total for E&W and NI. Prior to this, Rest of UK was included as 'outside Scotland' and these flows are available from 2001.	No data are available, although these flows are used by NISRA in the production of the MYEs.	Flows from Northern Irish LGDs to the rest of the world are available for 2001/02 to 2008/09 on request from NISRA.
RoW	Flows from RoW to LA/UA are produced by ONS in the detailed components of change for mid 2002 to mid 2009 available from ONS.	Flows from RoW to CA are available from 2002 are available, However prior to 2006/07 this included flows from England, Wales and Northern Ireland.	Flows to Northern Irish LGDs from the rest of the world are available for 2001/02 to 2008/09 on request from NISRA.	N/A

Key:	Good Data	Some Data	Poor/No Data

6.4 Supplementary data that can be made available

Two additional data sources have been identified that present the opportunity to derive more accurate estimates of cross-border flows in the UK. First, data on moves between health geographies in England, Wales and Scotland can be made available by ONS through a combined NHSCR extract. The English NHSCR used by ONS is able to record flows from Scottish HB areas into English and Welsh former HA areas. Similarly, the Scottish SNHSCR data (which are supplied by NRS to ONS) can identify flows in the opposite direction, i.e. identify the English or Welsh HAs where migrants previously resided before transferring to a Scottish HB. These flows can be disaggregated by age and sex. These cross-border flows between health geographies could provide a useful framework with which to constrain flows at the district level.

Second, the PRDS in England and Wales can report on persons residing in a district whose previous address was in Scotland and Northern Ireland (disaggregated by age and sex). Similarly, the SCHI is able to report persons residing in a district whose previous address was in England and Wales or Northern Ireland (coded as Q for England and Wales, NI for Northern Ireland) (GROS 2003). Future work could combine these two data sources and constrain the district level data to the health geography data to produce estimates of cross-border migration between England, Scotland and Wales.

A third potential data source relates to Northern Ireland, where a lack of available data has consistently presented a problem in estimation at a small scale and has led previous studies (for example, Dennett 2010) to use Northern Ireland as a single, aggregate area. The internal migration between LGDs is captured by the health card registration system and the data used by NISRA in the production of Northern Irish mid-year estimates. Gaining access to these data would go some way in resolving issues that arise when attempting to estimate internal migration in Northern Ireland.

7 Migration Estimation Methods

Section 6 has shown some of the shortcomings and deficiencies in the migration statistics when taking a UK perspective. While migration is known to be the major driver of national and sub-national population change, Raymer (2010, p. 73) stresses that *"the comparative study of migration is hindered by data availability, quality and consistency"* and advocates the use of models *"(i) to correct for the inadequacies and inconsistencies in the available*

data and (ii) to estimate the missing patterns." When presented with the challenge of creating a large dataset from different sources, creating a framework in which to understand what data are available and what gaps exist provides a useful starting point. Rees and Willekens (1986, p.21) use 'population accounts' which *"provide a useful framework for evaluating the completeness and accuracy of the available data and for estimating the missing data."* Such frameworks are employed by Raymer and Rogers (2007), Raymer (2010) and Boyle (1993) and use the known values (often aggregate flows) as a starting point to estimate the missing data, for example at the sub-national level, where data are sparse. This section of the paper reviews a number of methods used in the estimation of migration. This is addressed in two stages: first, methods used to fill the gaps are considered; and second, methods to improve the estimates by using known flow totals are explored.

7.1 Using variables to model migration expectations

Where there are no available flow data, it is necessary to populate a migration matrix using the best information that is available. The UKSA (2009, p.13), in their assessment of international migration statistics in the UK concluded that there was a need for *"a solution that relies on the integration of administrative sources and survey information"* and this has been the focus of recent methodological improvements in the distribution of international migrants to the local level. The variables that influence migration can be explored in the context of gravity and spatial interaction modelling while the development of statistical estimation techniques takes into account the models that fall under the umbrella of General Linear Models (GLM). Both take available information and use it to formulate an estimate of unknown flows.

7.2 Gravity/Spatial interaction models

In order to understand the movement of migrants, the process of push and pull factors as well as the movement itself need to be defined. The basic framework for studying origin-destination flows is summarised by Raymer and Giulietti (2010, p.281) who hypothesise that migration from origin (i) to destination (j) could be caused by *"an overall increase in levels of mobility within a country, a relative increase in the numbers leaving the origin, a relative increase in the attractiveness of the destination, an increase of the connectivity between the two"* or a combination of these factors. Of key importance amongst the determinants of migration are the 'gravity variables' where *"the characteristics of the origin may act as 'push' factors for potential out-migrants whilst the attributes of the destination reflect 'pull'*

factors that entice migrants to a particular destination" (Stillwell 2005, p.7). Also important is the frictional effect of distance. Gravity modelling provides a tool to estimate migration and at its most basic, uses the flow of population from origin to destination taking into account the size or mass of the origin/destination and distance between the two (the friction effect). To this basic model, further variables can be added, for example to represent the attractiveness of a certain destination over others (Ewing 1974). Roy and Thill (2004) provide an overview of the development of spatial interaction modelling, from its early form in the context of regional science and its use in measuring consumer behaviour at regional shopping centres through the Huff probabilistic model (1963). Raymer (2007, p.986) argues that *"the spatial interaction model is essentially a statistical form of the gravity model, which includes the factors of population size of the origin and destination regions, the distance between them, and some measure of competition or attractiveness."* The model is constantly evolving with a large number of applications being considered in the migration literature.

One of the important features of spatial interaction models is the method of calibration. Linear regression was used to derive the early gravity model parameters whereas the family of spatial interaction models derived by Wilson (1967) were based on entropy maximisation principles and optimum parameters were calibrated using mathematical methods (Stillwell 1991). More recently, several researchers have favoured new forms of statistical regression.

7.3 General linear models

An extension to the statistical modelling of migration incorporates group of models known as generalised linear models (GLMs). GLMs have been implemented in migration research due to their ability to deal with non-normal distribution, a defining characteristic of migration data and make use of variables derived from various sources that are known to influence migration decisions. Here we will focus on two of the GLM models; Poisson and log-linear regression models.

7.3.1 Poisson models

A Poisson regression model has already been used by ONS to allocate international migrants to the sub-national level by exploiting data sources that provide migrant characteristics (e.g. LFS, data derived from NHS sources and NINo registrations) and is also widely employed by academics tasked with addressing migration.

In a Poisson distribution *"it is assumed that the probability distribution underlying the migration transition data is known"* (Rogers *et al.* 2001, p. 233). As Willekens (2008, p.137) summarises, when data in a flow matrix are missing two steps are required: "*the first is to predict the missing data and the second step is to estimate the parameters of the model assuming that the data are complete"* which he terms the expectation maximisation (EM) algorithm. Using the example of an origin-destination matrix where the only available data are departures and arrivals by region, Willekens (2008) suggests that if the observations are independent then the required probability model is the Poisson model.

Lovett and Flowerdew (1989) argue that standard regression is often inappropriate for count data, and that the Poisson distribution is particularly useful where some observations have very low values, a condition afflicting migration data, especially at a disaggregated level. Boyle (1993) uses Poisson regression to model a sparse matrix (containing a large number of zeros and small flows) of ward level migration within Hereford and Worcester. He argues that Poisson regression is preferable to OLS regression models which often *"are oversensitive to flows involving very small numbers of people"* (Boyle 1993, p.1201). The method for dealing with a large number of zeros in a migration matrix is further explored by Bohara and Krieg (1996) who develop a 'zero-inflated Poisson model,' which they find reduces the under-prediction of migrants found in traditional modelling approaches. In a study of 126 labour market areas in Great Britain, Flowerdew and Aitkin (1982, p. 202) found that a Poisson regression model out-performed a log-normal model as it does not assume that error terms are normally distributed, can deal with unequal variance in error terms and "*avoids the bias of log-normal flow estimates, and can successfully handle zero flows."*

Given that the interaction between origins and destinations is a complex one, it is not unusual for the results from a gravity or spatial interaction model to be incorporated as variables in a Poisson regression. Flowerdew (2010) uses a spatial interaction model for inter-district flows from the 2001 Census to inform a Poisson regression model, while Sarra and Del Signore (2010, p.31) use a 'dynamic gravity model' calibrated using the Poisson regression model and incorporating "*environmental variables (housing, transport infrastructures, crime)"* in the modelling of internal migration flows in Poland at the NUTS-2 level.

As a predictor of migration flows, the method has been employed in a large and complex spatial interaction model using a wide range of variables for 98 zones between

England and Wales within the MIGMOD software constructed to help inform policy decisions (Rees *et al.* 2004). The application of modelling to take into account a number of complex assumptions is addressed by Fotheringham *et al.* (2001, p.902), who use clusters of locations to form a hierarchy because, they argue, the conventional model is based on the flawed assumption that "*individuals can and do evaluate all possible destinations prior to making a choice.*" They find that a newly specified competing destinations model out-performs the conventional model. Similarly, Congdon (2010, p.775) uses a Bayesian methodology to estimate migration between English LADs based on an "*extended random effects gravity model"* which links pull and push scores across all areas.

7.3.2 Log-Linear models

Rogers *et al.* (2010, p.30) advocate the use of log-linear models as *"the parameters of that model capture different features of the spatial structure of migration"* allowing for consideration of the characteristics of the origin population, the destination population and the strength of the linkages between the two. Rogers *et al.* (2003, p.67) use a log-linear regression model to "*predict migration from partial data contributed by different data sources"* by origin, destination and age, while Smith *et al.* (2010) combine PRDS data, the 2001 Census and the LFS in a log-linear model of migration by age, sex and economic activity between counties in England. Raymer *et al.* (2011) take this a stage further, producing an inter-regional age specific and sex specific model of ethnic group migration in England from 1991 to 2007 using a log-linear model to combine migration data from the 1991 and 2001 Censuses and published NHSCR tables from 1991 to 2007. Raymer *et al.* (2011, p.75-76) argue that the methodology used could be applied to higher levels of disaggregation than GOR, however this would "*require additional efforts to harmonise the Census and NHSCR data over time before combining them"*

Millington (2000) uses the age distribution in inter-county migration flows derived from the NHSCR, finding that different stimuli to migration were affected by the age distribution; young migrants fitted with labour market expectations, while elderly migrants responded most to local house price and amenity variables. Past age and spatial structures are used by Raymer and Rogers (2007) to inform log-linear estimates of age-specific migration in the USA and Mexico.

7.4 Constraining estimates to known flow totals

The previous sub-section addressed methods used to populate unknown values in a migration matrix. This section looks at ways to improve the estimates if some data is known. Aggregate flow data are widely available for migration statistics, for example flows in and out of a country or between its regions. These flows are often considered robust and as such provide a benchmark from which to address migration at a more disaggregated, sub-national scale. Dennett and Rees (2010, p.89) summarise that in theory, such methodology is fairly simple *"it is a matter of using known, complete, data values and distributions to constrain other, incomplete data, to produce estimates which adhere to the known values."* They acknowledge though, that in reality this can be harder to implement, especially where data availability varies.

A commonplace method for combining the data sources is iterative proportional fitting (IPF). The basic principle of IPF is outlined by Norman (1999, p.1) who states that "*IPF can be used to integrate disaggregated data from one source with aggregated data from another to estimate the characteristics of a population within given geographical units in a disaggregated format.*" Essentially, if the total flows at an aggregate scale are available, for example between the constituent countries of the UK, IPF can be used to adjust flows at a smaller spatial scale once they have been estimated. As summarised by Wong (1992, p.348), IPF is a *"feasible technique to re-create individual level population data if the aggregated population in each category of variables and the samples of observations are available."* Rees and Duke-Williams (1997) address the problem of suppression of small flows in the 1991 Census Special Migration Statistics. As the marginal totals were not subject to adjustment or suppression, IPF was used to adjust a number of the SMS tables which the authors found to produce a considerable improvement over the original dataset. Simpson and Tranmer (2005) use IPF to combine sample and census data for small areas and Saito (1992) uses a multi-step IPF procedure where the marginals in the table are inconsistent. Van Imhoff *et al.* (1997) use the technique in a study of multi-regional migration for Italy, the Netherlands and the UK, while Bell and Rees (2006) have used IPF to create uniform origin, destination, age and sex arrays for both Australian and British migration data.

More recently, Dennett and Rees (2010) produced a UK-wide matrix of flows at the NUTS2 (groups of LAs) scale, with totals constrained to published NUTS1 (regional) totals using 2001 Census flows. They use IPF to constrain the previously unknown NUTS2 estimate to published NUTS1 totals. Dennett and Rees (2010, p.104) suggest that *"it is*

entirely feasible that the techniques be applied to migration estimates at any other level where data from another level in the spatial hierarchy is present" The method is duly expanded on by Dennett (2010) in order to produce estimates of LA flows between England and Wales, Scotland and Northern Ireland (as a single region). Although Dennett (2010) does not use IPF technique *per se*, he does use published totals at the regional level (derived from NHSCR flows) as a constraint (both at origin and destination) for the combination of data sources at a coarser geographical scale; PRDS flows in England and Wales, flows between Council Areas in Scotland and 2001 derived Census distributions across the UK as well as the 2001 Census to derive age and sex flows. Dennett (2010, p.40) suggests that a "*fruitful avenue of future research would be in the full investigation of IPF procedures in this particular estimation problem."*

7.5 Summary

This section of the paper has reviewed a series of statistical and mathematical methods that can be used to estimate or predict unknown flows in a migration matrix from known information. In situations where information on known marginals is available, inter-area flows may be estimated using IPF methods. Log linear models have been shown by Raymer and Giulietti (2010) to be valuable in estimating time series flow matrices disaggregated by age and ethnicity. Wilsonian spatial interaction models provide an alternative approach to estimation based on gravity variables. The data gaps in the system of interest that we have identified for the UK are such as to require the application of these methods to create a full set of flow matrices, disaggregated by age and sex over a series of time periods.

8 Towards a Sub-national Migration System for the UK

This final section outlines the long-term ambitions of the project to create a sub-national time series of migration for the UK between 2001 and 2011 and presents the methods and results for an exercise to produce a set of estimates for 2001 based on the available data outlined in section 6 and to compare these estimates with flows from the 2001 Census. This comparison is an important exercise in verification of the flow estimates that are derived from other sources.

8.1 A sub-national time series, 2001-2011

Our long-term aim is to create a UK-wide annual time series of matrices of internal and international migration estimates between 2001 and 2011 and beyond, thus providing a comprehensive picture of sub-national migration over the first decade of the 21st century and a context within which to analyse the results of the 2011 Census. This section summarizes the component flows of the conceptual matrix shown in Table 5 and presents some comparative analyses of the estimates for selected parts of the matrix for 2000/01 with data from the 2001 Census.

The UK-wide matrix incorporates three types of migration flow which are illustrated in Table 5: (i) inter-district flows within each constituent country (internal intra-national flows; cells shown in green for England and Wales, orange for Scotland and purple for Northern Ireland); (ii) international flows into and out of each district in the UK (cells shown in red) and (iii) cross-border flows between districts in each of the four constituent countries (internal inter-national flows; cells shown in white).

Table 5: Migrant flows at the district level in the UK

Origin \ Destination		England			Wales			Scotland			Northern Ireland			Rest of the World
		District 1	District ...	District 355	District 1	District ...	District 21	District 1	District ...	District 32	District 1	District ...	District 26	
England	District 1	A1	A	A	A1	A	A	D	D	D	D	D	D	F
	District ...	A	A1	A	A	A1	A	D	D	D	D	D	D	F
	District 355	A	A	A1	A	A	A1	D	D	D	D	D	D	F
Wales	District 1	A1	A	A	A1	A	A	D	D	D	D	D	D	F
	District ...	A	A1	A	A	A1	A	D	D	D	D	D	D	F
	District 21	A	A	A1	A	A	A1	D	D	D	D	D	D	F
Scotland	District 1	D	D	D	D	D	D	B1	B	B	D	D	D	F
	District ...	D	D	D	D	D	D	B	B1	B	D	D	D	F
	District 32	D	D	D	D	D	D	B	B	B1	D	D	D	F
Northern Ireland	District 1	D	D	D	D	D	D	D	D	D	C1	C	C	F
	District ...	D	D	D	D	D	D	D	D	D	C	C1	C	F
	District 26	D	D	D	D	D	D	D	D	D	C	C	C1	F
Rest of the World		E	E	E	E	E	E	E	E	E	E	E	E	-

The diagonal cells which represent moves within a district (represented as A_1, B_1 and C_1 in Table 5) are reported in the 2001 Census but not in the other data sources used for comparison. Rogers *et al*. (2010) investigate the impact of including these diagonal elements

of a migration matrix in a log-linear model, finding that a 'quasi-independence model,' which assumes that origin and destination are independent by replacing the diagonal element with zeros, produces a better fitting explanatory model of migration. These intra-district flows are not included in the analysis below, but it is important to highlight that such moves represented a large proportion of the reported migration in the 2001 Census.

The following analysis draws largely on data sources used by the NSAs to estimate migration during inter-censal years but, due to data availability and varying methodologies, some estimation has been carried out to give the best possible approximation of flows for 2000/2001. The methods and data used in the comparative analysis for each section of the matrix are explored below.

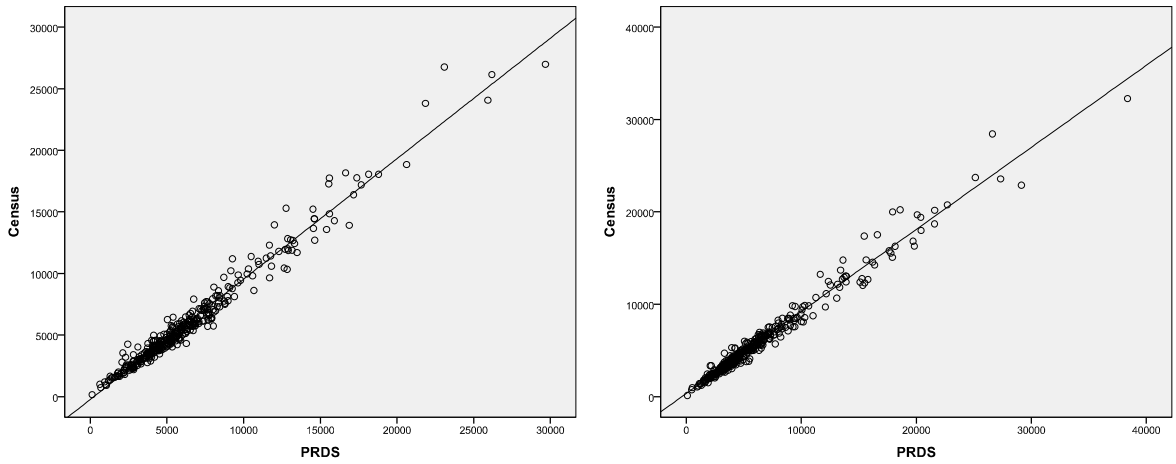
8.2 Inter-district flows within each constituent country (internal intra-national flows)

Although inter-district flows within the countries of the UK are estimated using data derived from NHS sources, a different approach was used to create equivalent flows for Scotland, Northern Ireland and England and Wales.

8.2.1 England and Wales

The estimated flows between districts in England and Wales (the green cells in Figure 5) have been derived by ONS using PRDS data. The PRDS table (adjusted to agree with the NHSCR) gives flows between districts in England and Wales for mid-year 2000/2001 and has been made available by ONS. No adjustment has been made to the supplied PRDS flows, which have been compared directly with the flows reported in the 2001 Census.

There is a strong positive correlation between the PRDS flows and flows reported in the 2001 Census. Figure 3a shows a strong positive correlation for inflows to each district ($r = 0.985$, $p<0.00$). Total outflows by district, shown in Figure 3b, are also strongly correlated ($r = 0.987$, $p<0.00$).



(a) Inflows **(b) Outflows**

Figure 3: The correlation between 2001 Census and PRDS inflows (a) and outflows (b) for each district in England and Wales from/to the rest of England and Wales

The graphs shown in figure 3 confirm that both sets of data are exhibiting the same spatial pattern of total inflow and outflow. Figure 4 shows the result of the pairwise origin-destination PRDS flows compared with Census flows. It shows that when each individual district to district flow in England and Wales is considered, the correlation is strongly positive ($r = 0.982$, $p<0.00$).

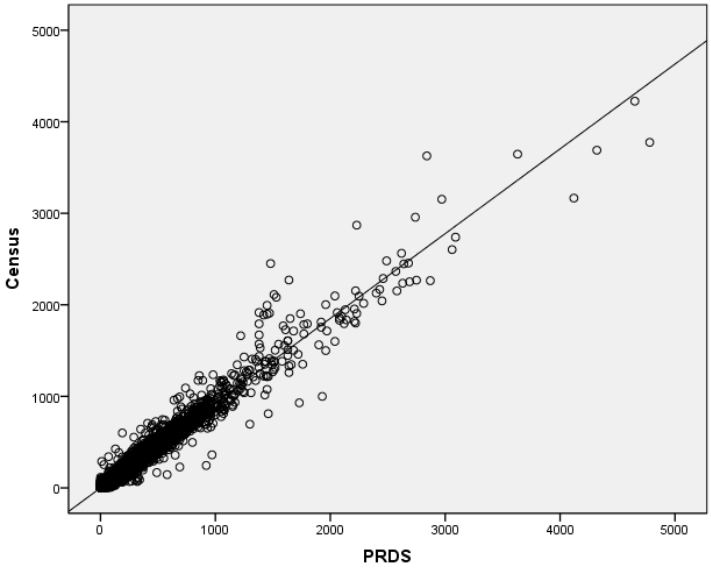
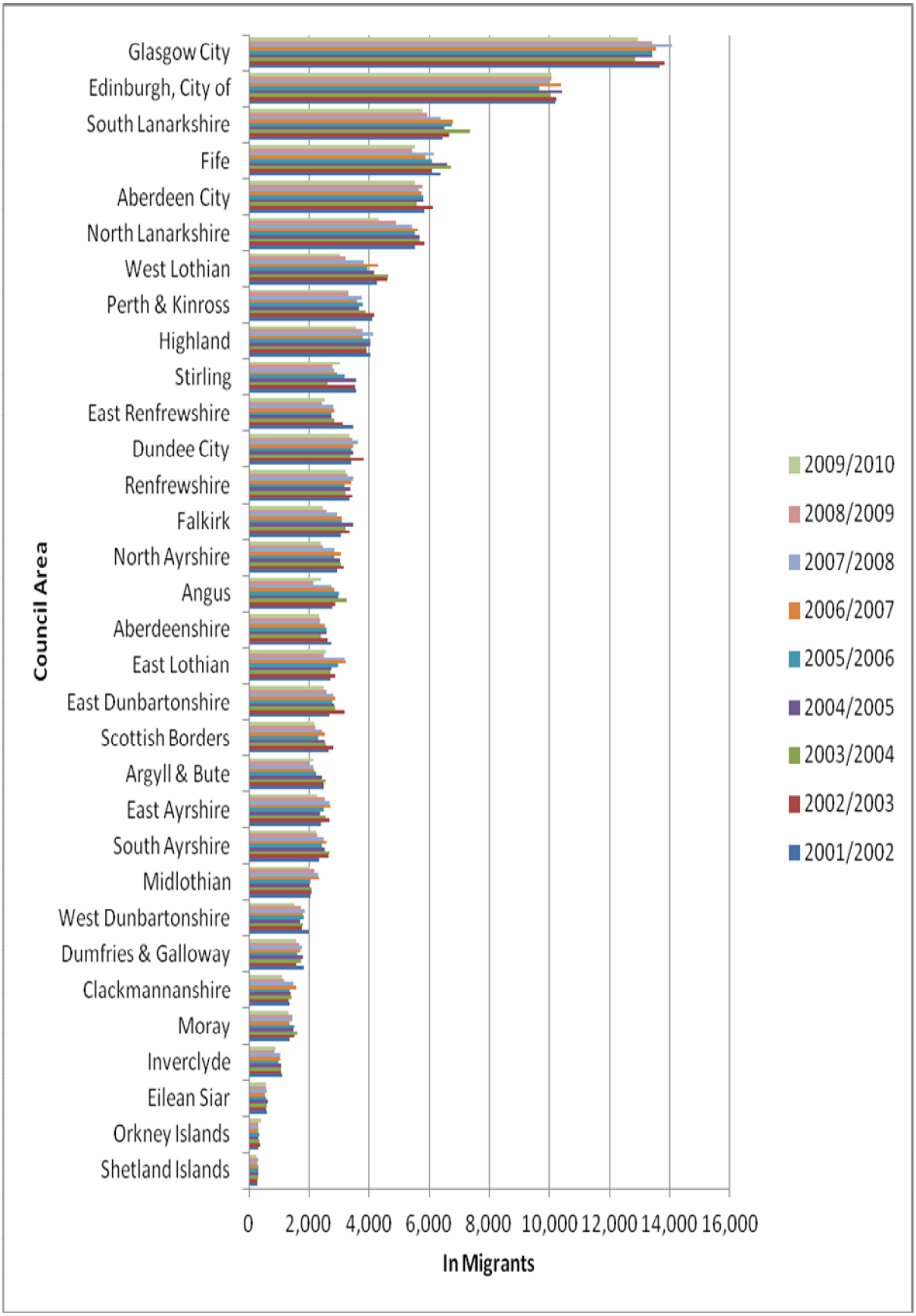


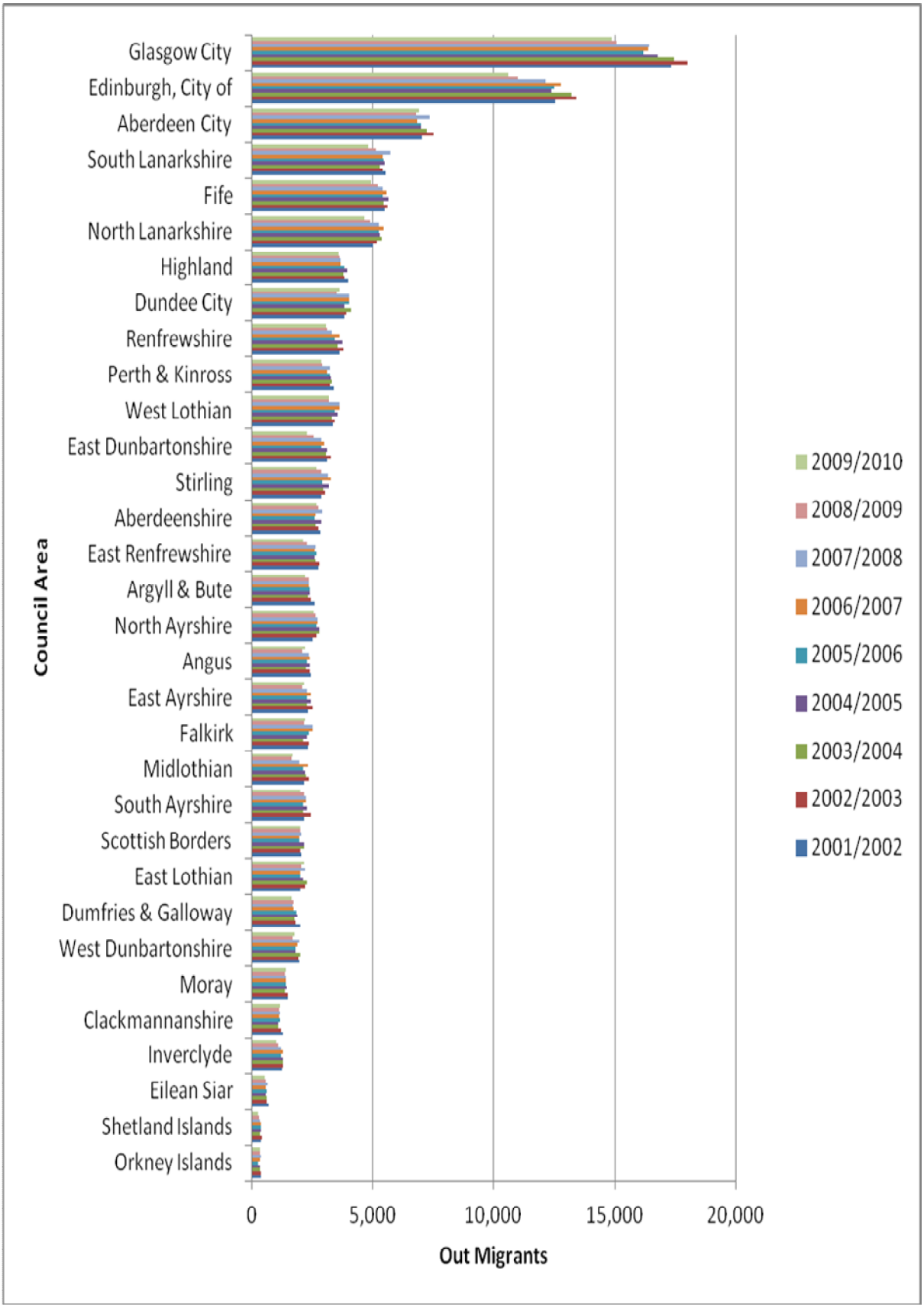
Figure 4: The correlation between pairwise district to district flows reported in the PRDS and the 2001 Census

8.2.2 Scotland

The SCHI provides details of moves between districts in Scotland. Following discussion with NRS, it has been confirmed that no SCHI data are available for 2000/01 due to a change in SCHI methodology introduced in 2001 (Mueller 2011). District level flows are available from 2001/02 onwards and an analysis of the flows between 2001 and 2010 shows remarkable consistency, both in the distribution and actual numbers of moves reported in the SCHI, which can be seen for inflows to districts from the rest of Scotland (Figure 5a) and outflows from districts to the rest of Scotland (Figure 5b).



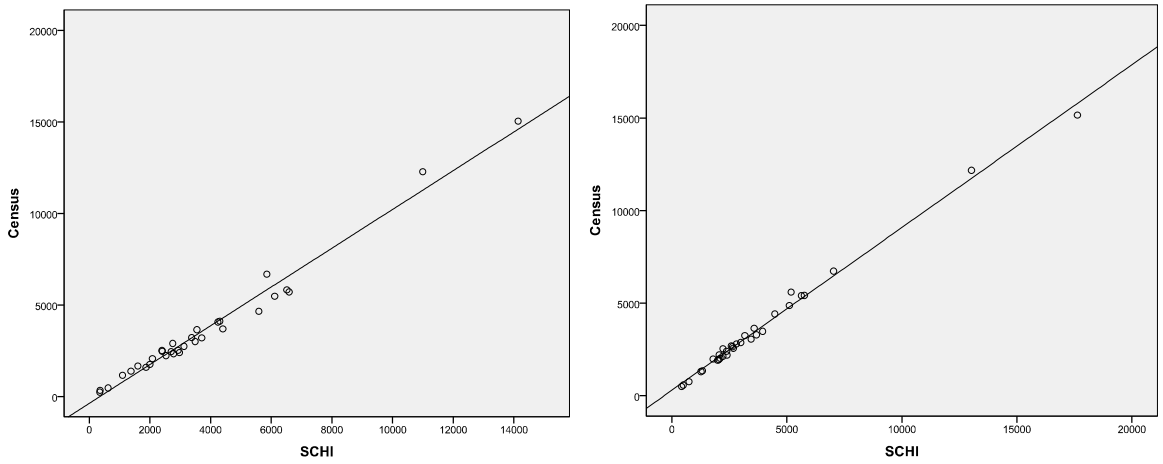
(a) Share of inflow



(b) Share of outflow

Figure 5: Share of total inflow from (a) and outflow to (b) the rest of Scotland by Council Area, 2001 – 2010 (ranked by 2001 flow)

In the absence of 2000/01 flow data and in light of the consistency of the results over time, 2001/02 flow data have been used as a proxy for 2000/01 data. Figure 6a shows a strong positive correlation between the SCHI and 2001 Census for total inflows for each district of Scotland (r = 0.989, p<0.00) while the outflows (Figure 6b) show an even stronger positive correlation (r = 0.997, p<0.00).



(a) Inflows **(b) Outflows**

Figure 6: The correlation between 2001 Census and SCHI inflows (a) and outflows (b) for each district in Scotland from/to the rest of Scotland

Figure 7 shows the pairwise district to district flows in Scotland. There is a very strong positive correlation between flows at the inter-district level estimated from the SCHI and those reported in the 2001 Census (r = 0.985, p< 0.00).

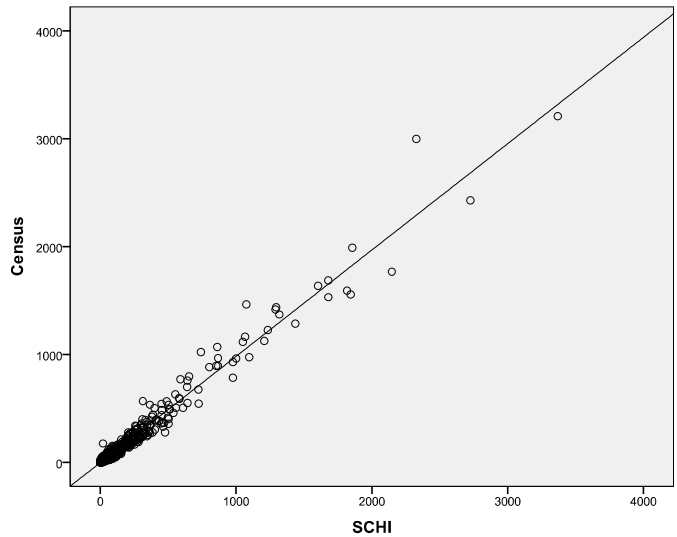
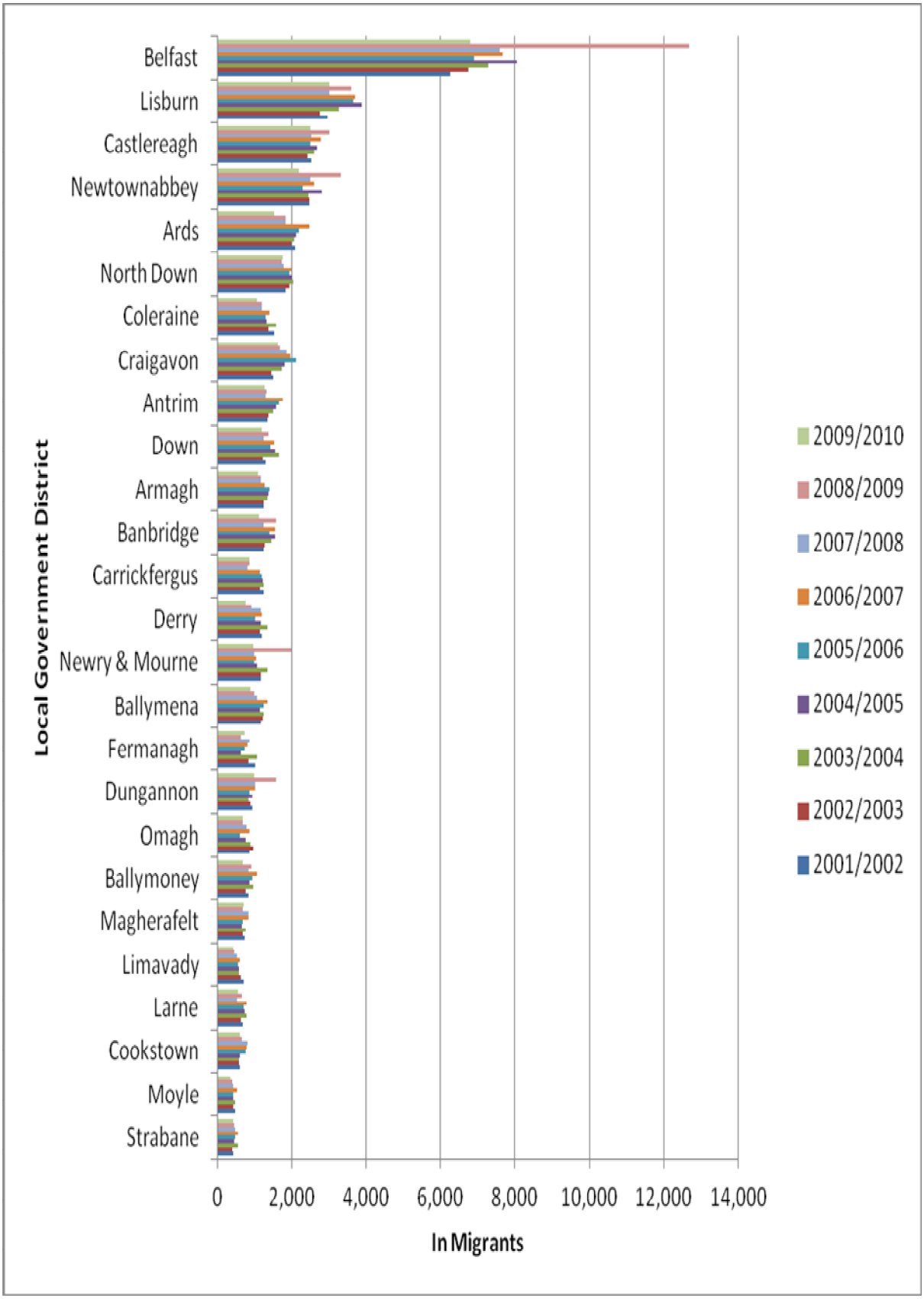


Figure 7: The correlation between pairwise district to district flows reported in the SCHI and the 2001 Census

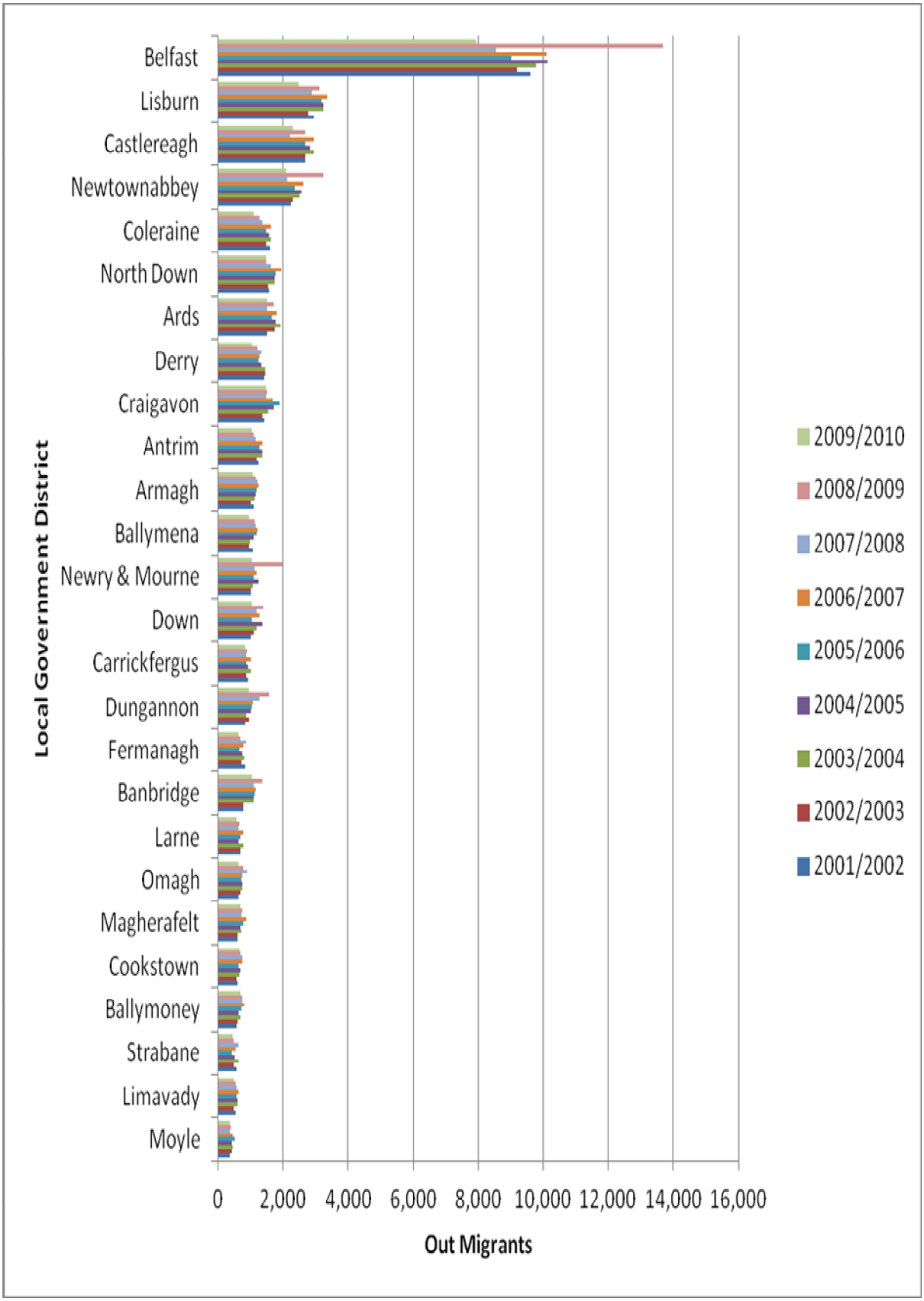
8.2.3 Northern Ireland

Inter-district flows in Northern Ireland are recorded by NISRA in the NICHI which is derived from health card re-registrations. Unfortunately, no matrix of moves at a district level was available so a process of estimation has been used to approximate flows between districts.

Total gross inflows outflows for each district to/from the rest of Northern Ireland for 2001/02 onwards were supplied by NISRA, but no table for 2000/01 was available. Similar to the process used for the Scottish data, an assessment of the data from 2001 onwards was undertaken, and for 2001 to 2010 the distribution and number of movements between each district and the rest of Northern Ireland was found to be largely consistent (except for flows to and from Belfast in 2008/09 which were far higher than other years; inflow is shown in Figure 8a and outflow in Figure 8b). In light of the consistent distribution, and largely consistent migrant numbers across time, the 2001/02 data has been used as an approximation for the 2000/01 estimate.



(a) Share of inflow



(b) Share of outflow

Figure 8: Share of total inflow from (a) and outflow to (b) the rest of Northern Ireland by LGD, 2001 – 2010 (ranked by 2001 flows)

As there was no breakdown of flows between the districts in Northern Ireland, total inflow and outflow figures were apportioned to each district to give an estimate of inter-district flows. First, the proportion of total inflow to each district from the rest of Northern Ireland was calculated as a percentage. Second, the reported total outflow from each district was apportioned between the other districts based on their proportion of total inflow from the rest of Northern Ireland. The effect was to apportion the number of out-migrants from each district by the percentage of total inflow from the whole of Northern Ireland for each district, as illustrated in Figure 9.

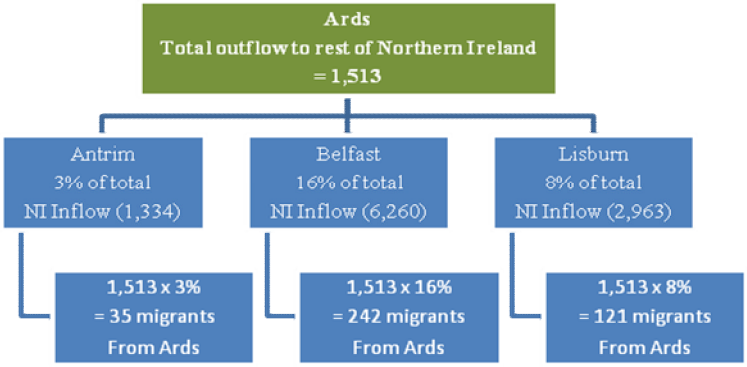
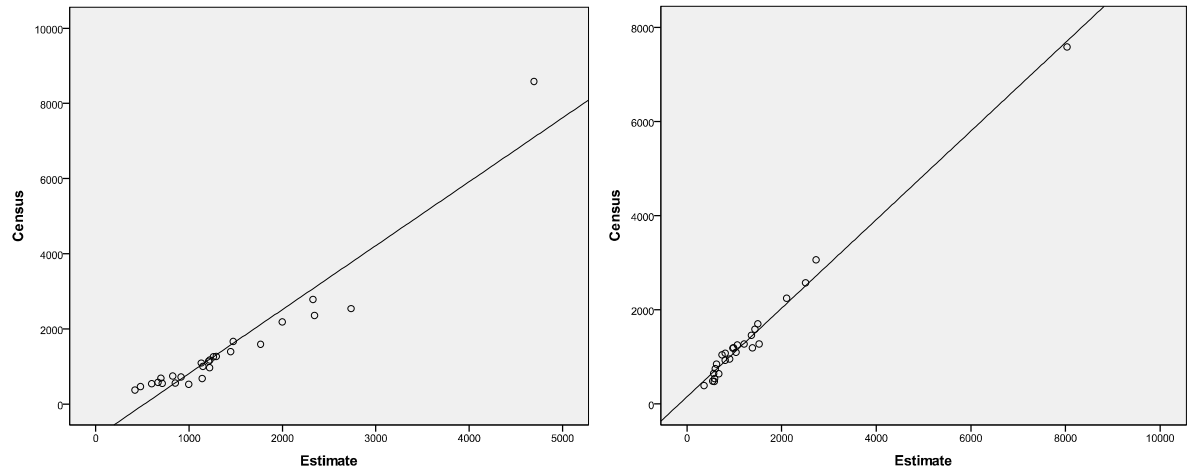


Figure 9: An illustration of the method used to apportion internal migration between districts in Northern Ireland

Figure 10 shows the correlations between the 2001 Census figures for internal migration into and out of each district and the estimated flows. The correlation between the sources for both inflows ($r = 0.951$, $p < 0.00$) seen in figure 10a and for outflows ($r = 0.994$, $p < 0.00$) seen in figure 10b are very strong and positive.



(a) Inflows **(b) Outflows**

Figure 10: The correlation between 2001 Census and NICHI inflows (a) and outflows (b) for each district in Northern Ireland from/to the rest of Northern Ireland

Figure 11 shows that when the pairwise district to district estimates are compared with the flows reported in the 2001 Census, the correlation is weaker than for total inflows and outflows, but is still fairly strong and positive ($r = 0.788$, $p<0.00$).

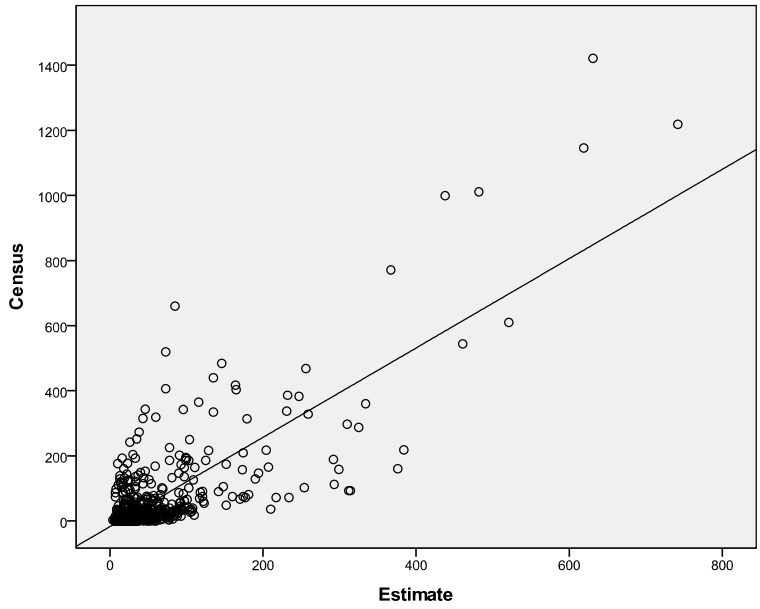


Figure 11: The correlation between pairwise district to district flows reported in the NICHI and the 2001 Census

There are significant outliers both above and below the 'best-fit' line, but the general trend is for the 2001 Census reported figure to exceed the estimated flow.

8.3 International flows into and out of each district in the UK

International flows for England, Wales and Scotland come from the estimate of LTIM as discussed in section 3. The estimate of international flows for Northern Ireland is derived from Health Card registrations to overseas nationals as discussed in section 3.2.3. The data sources and methodologies used to derive an estimate comparable to the 2001 Census reported inflows are outlined below. In this section, only the international inflows are estimated as there are no outflow figures in the Census from which to draw a comparison.

8.3.1 England, Wales and Scotland

A 2000/01 total LTIM estimate is reported for Scotland, Wales and GORs in England, but no disaggregation to districts is available for the study year. For England, Wales and Scotland the apportionment of the 2000/2001 LTIM figure was derived from 2001/02 distributions.

In England and Wales, detailed components of change data for each district are available from 2001/02 onwards, which gives the share of LTIM to each district. This distribution was used to apportion the 2000/01 LTIM estimate reported nationally for Wales and by GOR in England. The GOR figure was apportioned based on the 2001/02 distribution to groups of districts that comprised each region. Figure 12 shows that when the distributed LTIM is compared with the figures reported in the 2001 Census, there is a very strong positive correlation ($r = 0.943$, $p < 0.00$) but with some significant outliers.

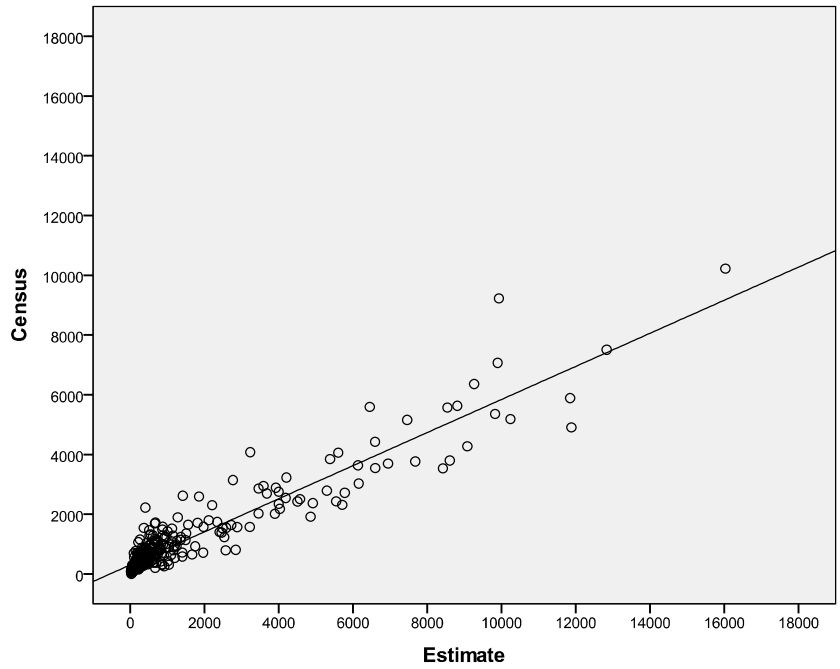


Figure 12: The correlation between immigration to districts in England and Wales derived from LTIM figures and the 2001 Census

For districts with the highest immigration from outside the UK, the estimate derived from LTIM is significantly higher than the 2001 Census derived figure. This is reflected in the total immigration figures for England and Wales; the 2001 Census reports a total of 370,447 immigrants whose previous address was outside the UK, while the total LTIM estimate is 91,557 higher at 462,004. Figure 12 shows that the distribution, if not the actual numbers, is similar so some work is required to revise the estimated immigration figures derived from LTIM. It is also feasible that the 420,089 people with 'no usual address reported' for the year before the Census represent a large proportion of the difference observed above, but it is not possible to make a definitive link.

Total inflows from outside Scotland to each district are available from 2001/02 onwards, but not for the study year 2000/01. One problem with the reported flows is that for

data between 2001/01 and 2005/06, the estimate of migration to and from districts in Scotland, the 'outside Scotland' figure, is an aggregate of both overseas migrations and migrations to/from the rest of the UK which cannot be separated out (Clarke 2011, personal communication). The separation of flows from overseas and from the rest of the UK was only reported from 2006/07 onwards, so it was necessary to use this distribution to apportion the figure for 2001/02. Figure 13 shows that between 2006/07 and 2009/10 the distribution of immigration to districts from the rest of the world (as a percentage of the total) was relatively consistent with the exception of Shetland, with 78.4% overseas immigration in 2008/09 compared with 84.6% in 2006/07. However, Shetland has a very small number of immigrants from outside Scotland across all years. The average distribution of overseas immigrants to each district was used to apportion the 2001/02 'outside Scotland' immigration figures to individual districts.

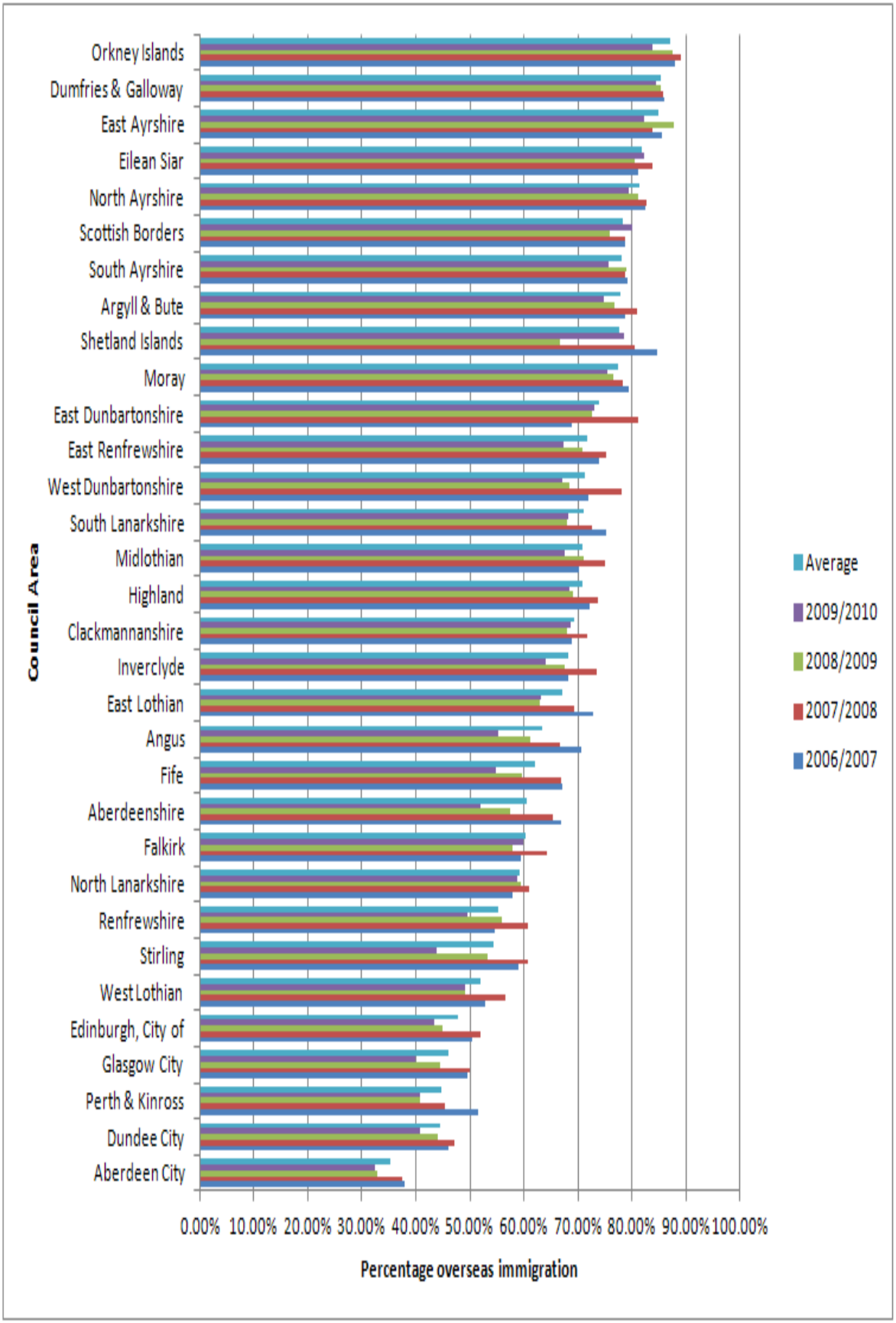


Figure 13: The proportion of 'outside Scotland' immigration which comes from overseas for 2006/07 to 2009/10 (ranked by four year average)

The correlation between the LTIM apportioned by the distribution of 2001/02 immigration and the figures reported in the 2001 Census is shown in Figure 14. There is a very strong positive correlation ($r = .982$, $p < 0.00$).

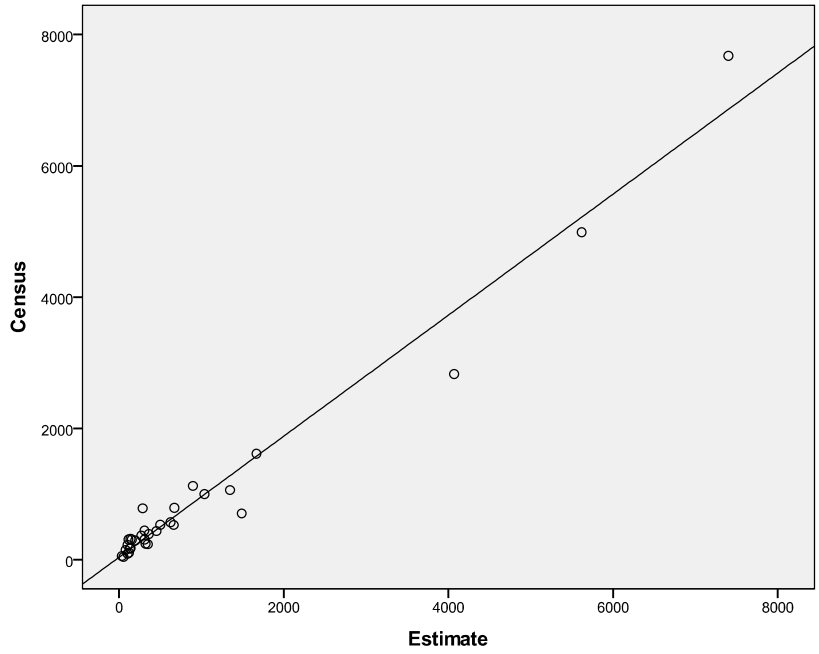


Figure 14: The correlation between immigration to districts in Scotland derived from apportioned LTIM figures and the 2001 Census

8.3.2 Northern Ireland

Although no international inflow figures are available for 2000/01, from 2001/02 onwards the total number of migrants from outside of the UK are reported by district. Figure 15 shows that between 2001/02 and 2008/09, total immigration by district from outside the UK has fluctuated, with a rise between 2001/02 and 2006/07 (with a dip in 2002/03) and a sharp decline from 2007/08 onwards. In the absence of sufficient data for 2000/01, the 2001/02 figures have been used, both because they represent a conservative estimate when compared with the rest of the time-series and because they are the closest data to the 2000/01 requirement in terms of time.

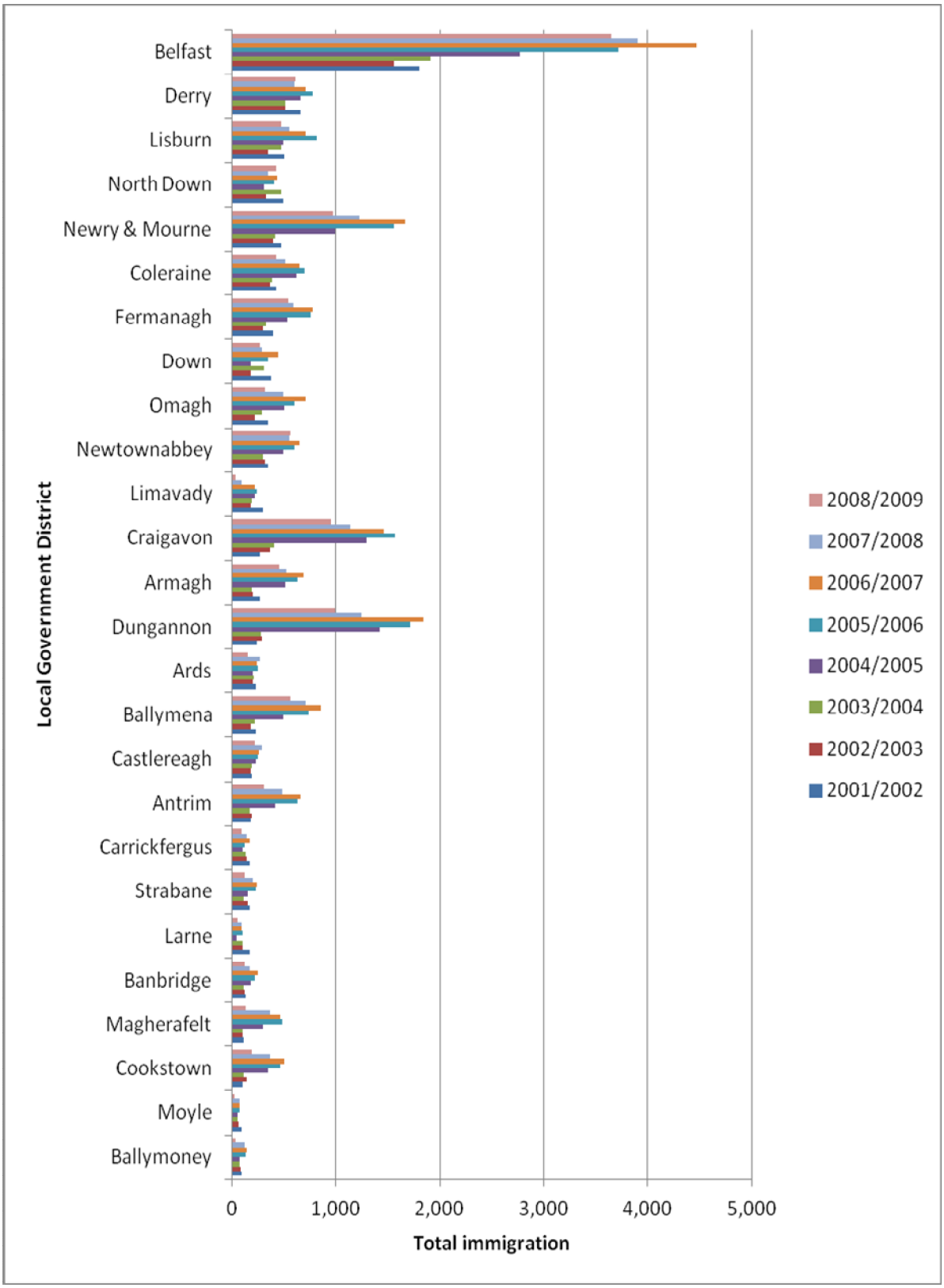


Figure 15: Total immigration from outside the UK by district reported in the health card data 2001/02–2008/09 (ranked by 2001/02 flows)

The result of the correlation between the apportioned health card data and the 2001 Census figure by district is shown in Figure 16. The correlation is a very strong positive one ($r = 0.986$, $p < 0.00$).

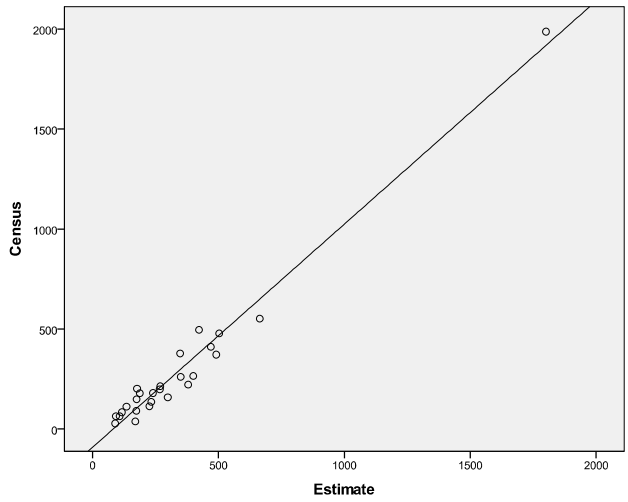


Figure 16: The correlation between immigration to districts in Northern Ireland derived from apportioned health card data and from the 2001 Census

8.3.3 All overseas immigration

Figure 17 shows that when the inflows into all districts of the UK are considered, the correlation between the estimate and the 2001 Census figure is a very strong positive one ($r = 0.939$, $p < 0.00$).

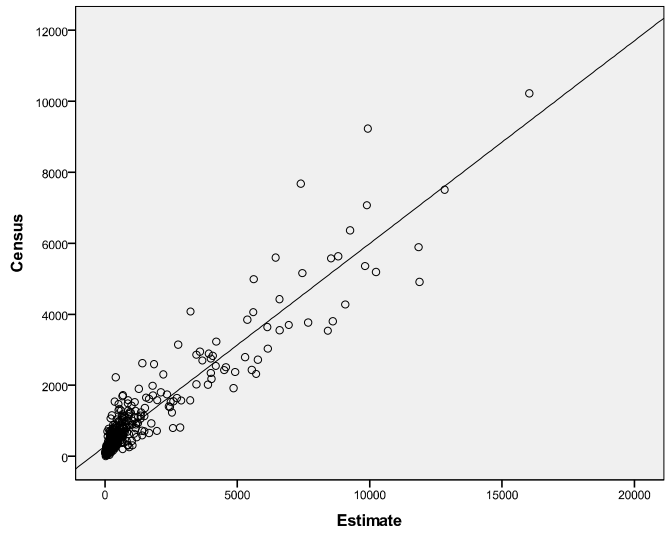
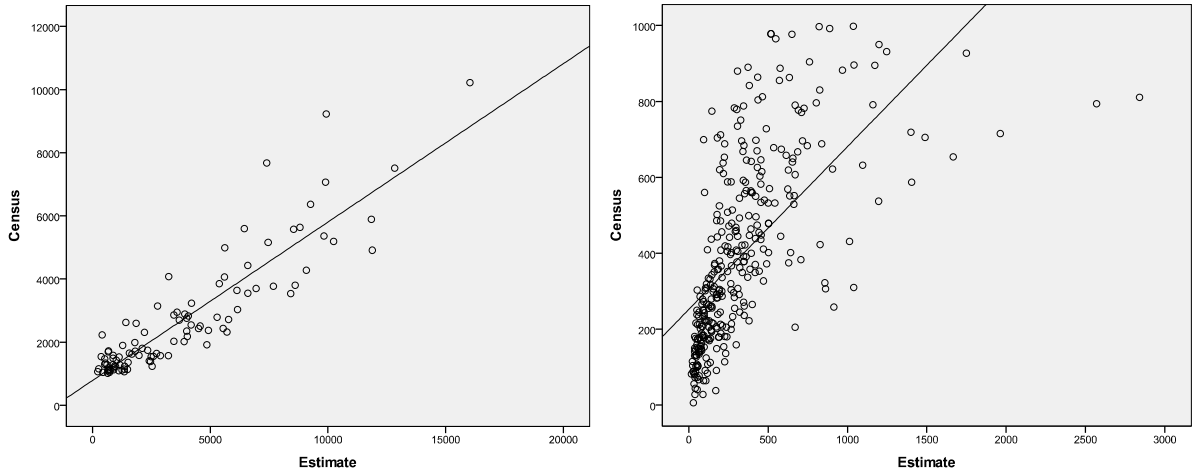


Figure 17: The correlation between estimates and the 2001 Census for immigration to all districts in the UK

Given the large variation in numbers of immigrants to UK districts, the overall picture presented above obscures the correlation variations at different scales. Figure 18a shows that when the areas with the largest overseas immigration flows are assessed separately the correlation is a strong positive one. Where the 2001 Census reported figure is over 1,000 immigrants, the r = 0.900 (p< 0.00). For districts where the overseas immigration figure is lower however (2001 Census flow under 1,000; Figure 18b) the correlation between the Census and the estimate is weaker (r = 0.636, p< 0.00).



(a) Over 1,000 immigrants **(b) Under 1,000 immigrants**

Figure 18: The correlation between estimates and the 2001 Census for immigration to all districts in the UK where the Census figure is over 10,000 (a) and under 10,000 (b)

The results seen in Figure 18 suggest that the overseas migration is a portion of the matrix that needs some improvement. The correlation for Northern Ireland and Scotland are generally good, but in England and Wales, the difference between the LTIM estimate (462,004) and the Census (370,447) is 91,557 (22%). The inconsistency between the LTIM data and Census is not echoed in Scotland, where there is a difference of only 1,131 (3.9%). The inconsistency between the estimate and the 2001 Census figure is also largely observed where the flows into a district are low (Census value under 1,000).

8.4 Cross-border flows between districts in each of the four constituent countries (internal inter-national flows)

The final and most difficult to estimate portion of the matrix is the internal inter-national flows as there is no official estimate of these flows between census years. The only source of data for UK-wide migration is the NHSCR, which provides details of flows between Wales,

Scotland, Northern Ireland and GORs in England. Figure 19 shows that the correlation between the flows reported in the NHSCR for 2000/01 and flows reported in the 2001 Census is a very strong positive one ($r = 0.999$, $p < 0.00$).

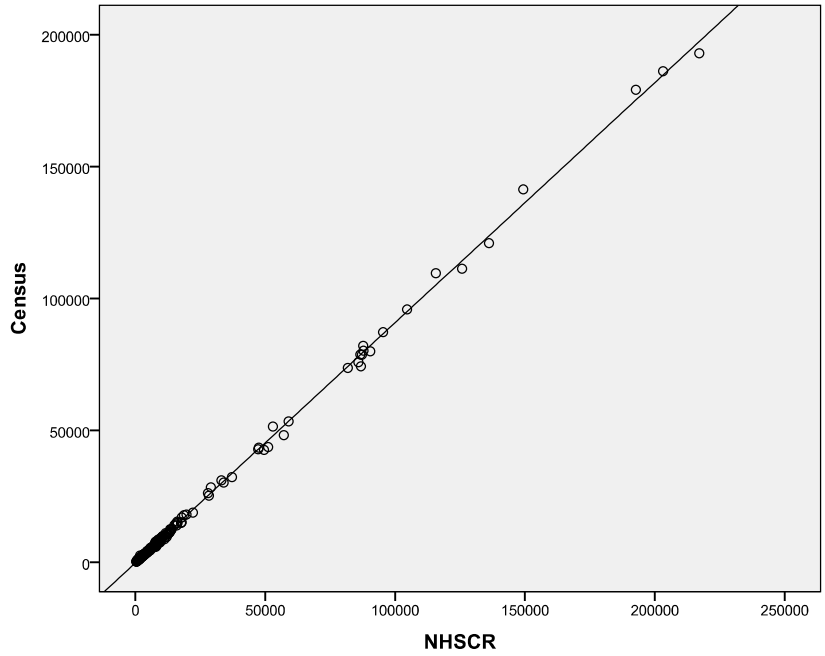


Figure 19: The correlation between country and GOR flows reported in the NHSCR and in the 2001 Census

This strong positive correlation gives the confidence to use the NHSCR as a framework within which to attempt to estimate the disaggregated district to district flows for which there are no available data. The following section outlines the methods and results of the estimation for district to district migration across the borders of the UK's constituent countries.

8.4.1 Apportioning migration between districts

The NHSCR is the only source of data that provides information on the flow of migrants across the borders of the UK, in the tables published by ONS aggregate moves between Scotland, Northern Ireland, Wales and GORs in England are reported.

At the district level, the NSAs report total numbers of migrants leaving a district for another country in the UK, but do not report the sub-national destination of that migrant, so it is only possible to deduce that a migrant has left a district for another part of the UK. Similarly, the total number of migrants entering a district from another country in the UK is reported, but there is no information on the district of origin. This section attempts to allocate

migrants leaving a district in one country to a district in the country of destination. Figure 19 shows that NHSCR provides a robust framework within which cross border district to district flows can be estimated and the NHSCR data forms the basis of the following estimates.

NHSCR totals were apportioned to sub-national geographies rather than use the total immigration and emigration reported by district in the PRDS, NICHI and SCHI to provide consistency across the UK, as the NSAs use different methods and data to estimate the cross border flows in and out of each district. These differences will be expanded on in the following sub-sections, but first the general method of estimation used will be outlined here.

The first stage was to calculate the percentage of total outflow to the rest of the UK from a district based on its proportion of total emigration from the country or GOR it sits within, as observed in the PRDS, NICHI or SCHI.

The second stage was to allocate the NHSCR country or regional flow in to a district based on the proportion of immigration that the district receives. Again this proportion was derived from the PRDS, NICHI or SCHI. Finally, the proportion of outflow from the origin district and the total inflow to the destination district were combined to estimate a district to district flow.

Figure 20 illustrates the method for estimating migration from Ards to Manchester, but the process is applicable for all cross border district to district migrations in the system. The district of Ards represents 3% of total out migration from Northern Ireland the rest of the UK (based on the NICHI). The district of Manchester receives 11.6% of the North West regional immigration from the rest of the UK (based on the PRDS). When this is apportioned between flows from Northern Ireland and Scotland, based on the distribution observed in the NHSCR for the North West, 242 migrants arrive in Manchester from Northern Ireland. So by multiplying the 242 Northern Irish Migrants by 3%, the flow from Ards to Manchester is estimated at 7.

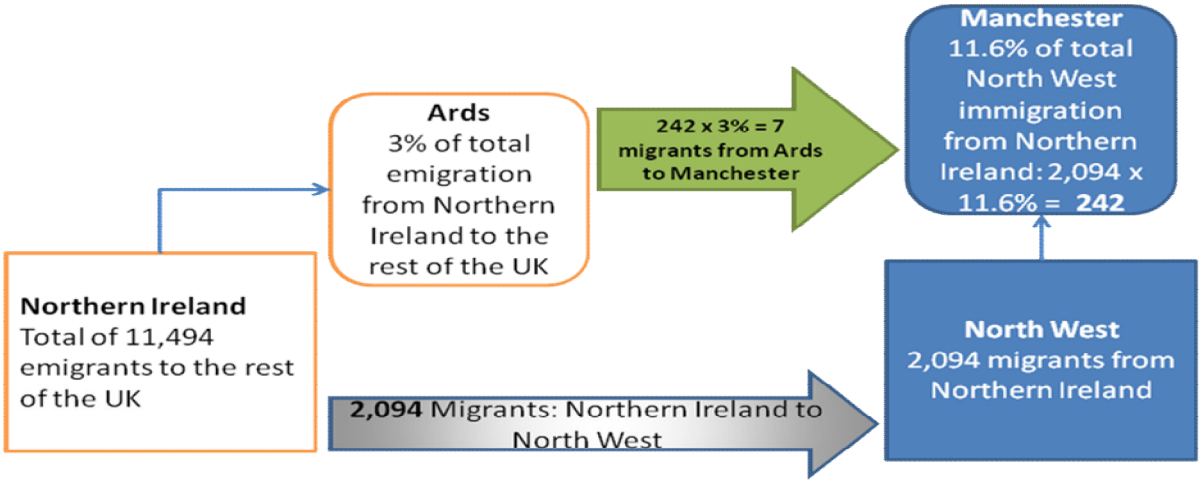
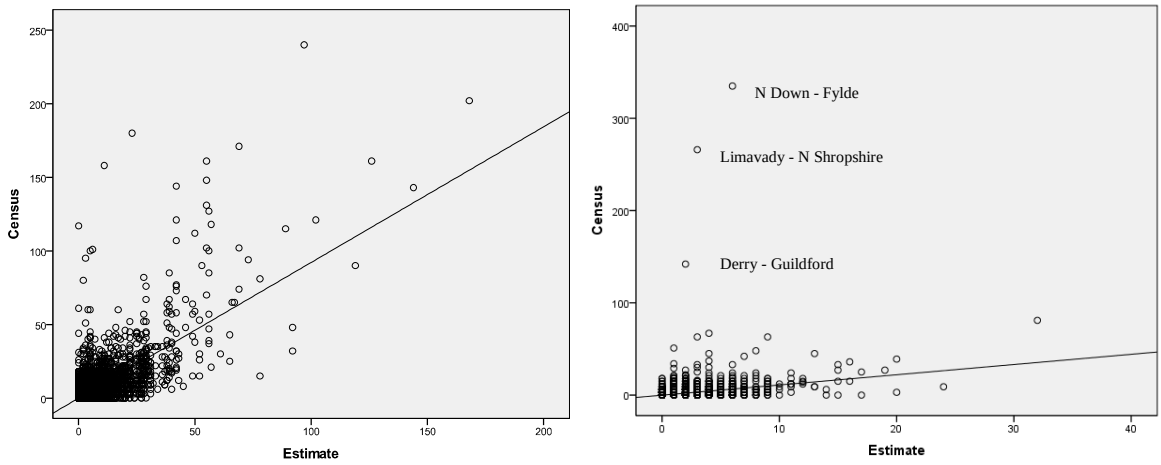


Figure 20: The process used to distribute NHSCR flows to districts in the sending and receiving country

8.4.2 England and Wales

For England and Wales, both inflows and outflows were estimated using the same two data sources. First, ONS supplied a table specifying total inflows and outflows by district to/from the rest of the UK. Second, the PRDS data for 2000/01 gives total flows between each district in England and Wales. By subtracting the England and Wales flow from the total UK flow, the flow to and from Scotland and Northern Ireland could be derived. As this total flow into/out of each district comprised Scottish and Northern Irish flows, the proportion of inflow from and outflow to each country was derived from the NHSCR table.

The results of the estimated flows to England and Wales from Scotland and Northern Ireland are shown in Figure 21. The correlation between the 2001 Census and the estimate for flows from Northern Ireland is a weak positive one ($r = 0.296$, $p < 0.00$). The correlation between the two sources for flows from Scotland is better ($r = 0.692$, $p<0.00$).



(a) From Scotland **(b) From Northern Ireland**

Figure 21: Flows from districts in Scotland (a) and Northern Ireland (b) to districts in England and Wales

Three significant outliers can be seen in the flow from Northern Ireland; North Down to Fylde (Census = 355, estimate = 6) Limavady to North Shropshire (Census = 266, estimate = 3) and Derry to Guildford (142 compared to 2). If these outliers are removed the correlation improves ($r = 0.466$, $p<0.00$).

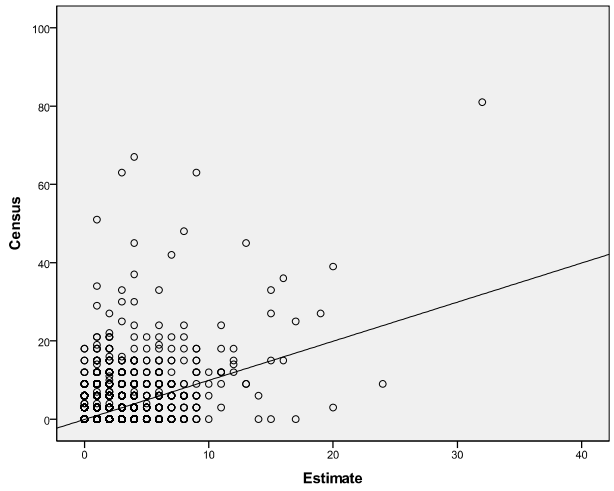
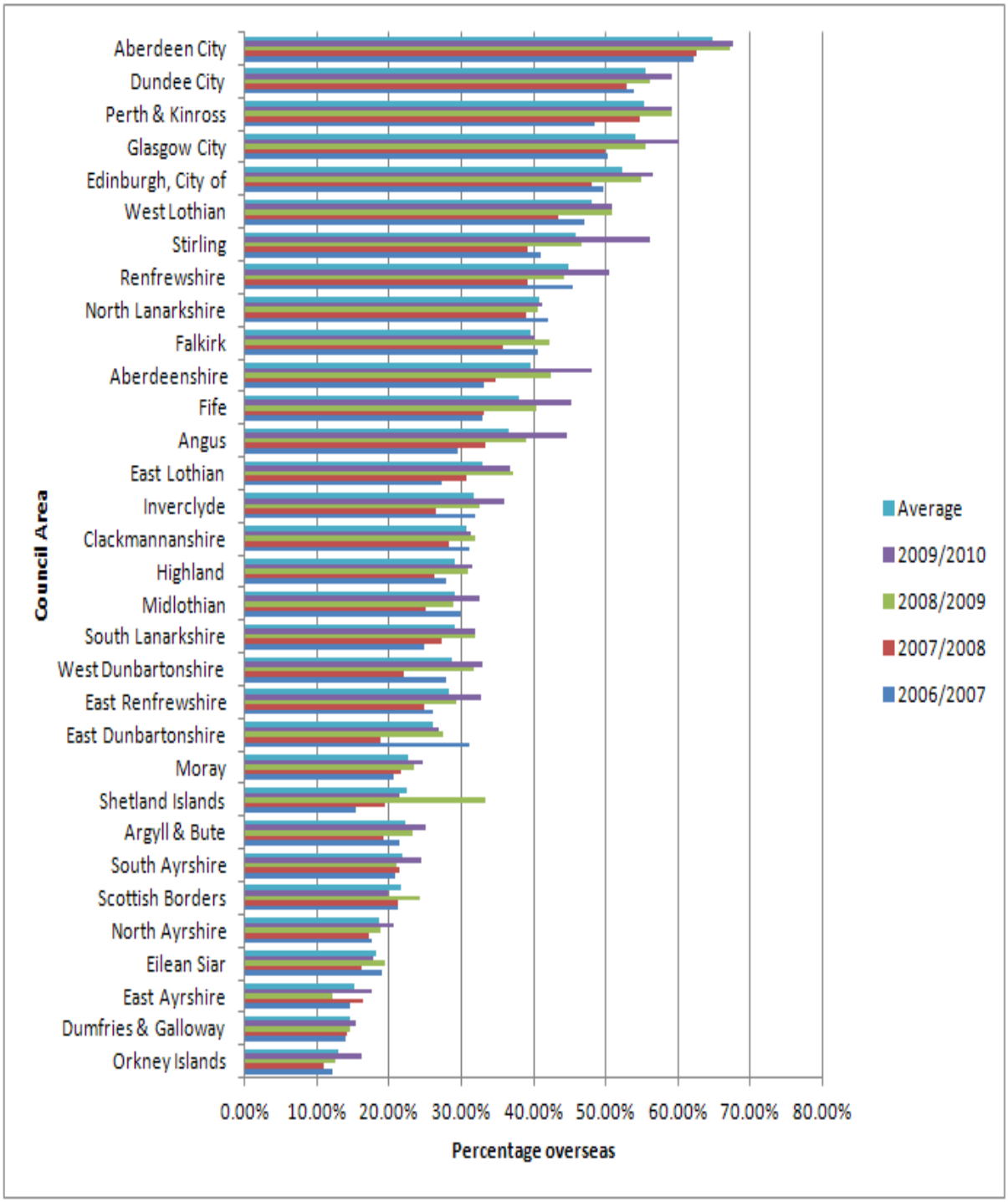


Figure 22: Flows from districts in Northern Ireland to districts in England and Wales (minus three major outliers)

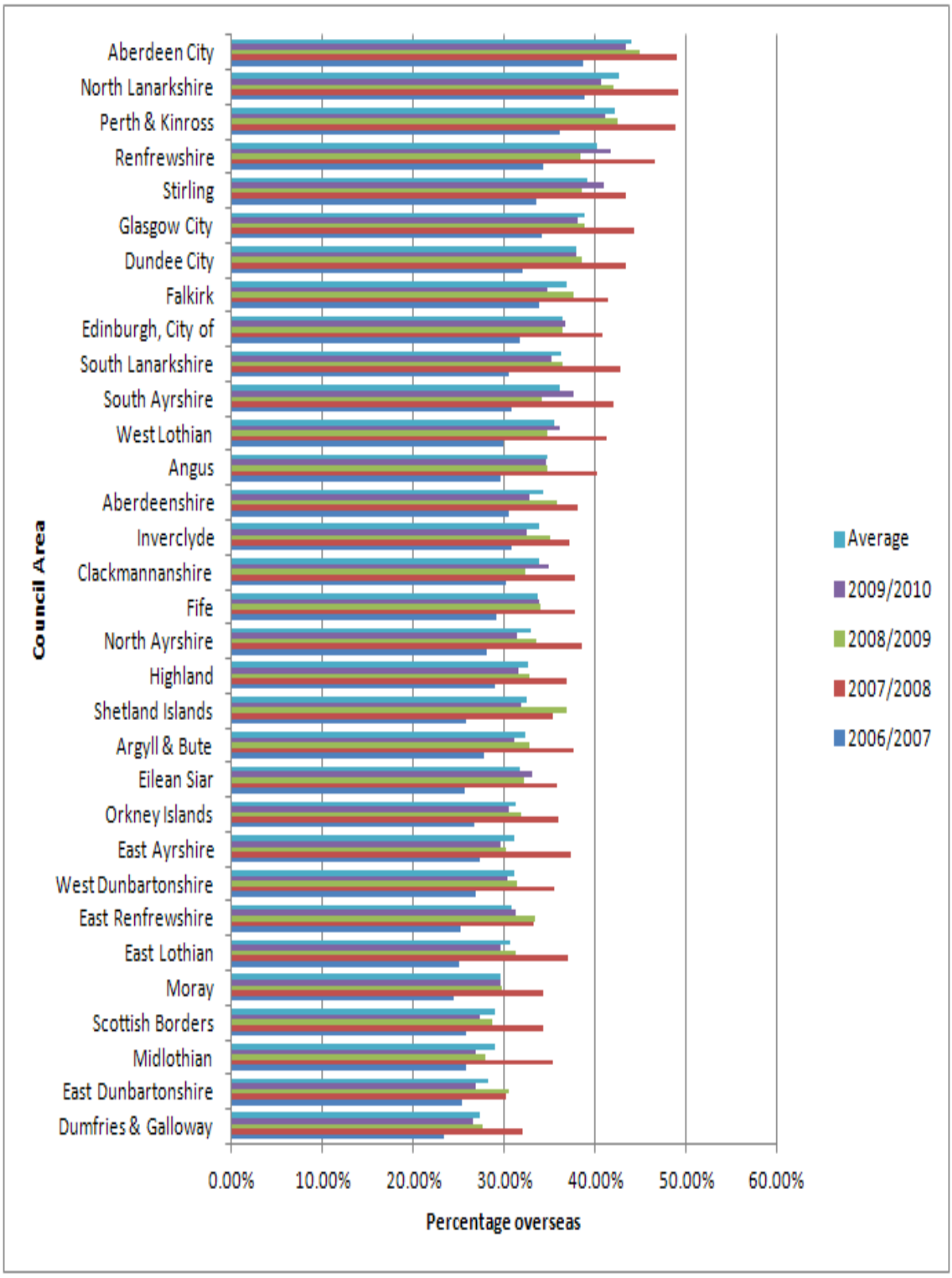
8.4.3 Scotland

For Scotland, the absence of flow data for 2000/01 means that the 2001/02 data had to be used, and similar to the problem encountered with the overseas migration estimate, prior to 2006/07 the inflows and outflows were reported as aggregate 'outside Scotland' flows

comprising both overseas and the rest of the UK. Similar to the overseas estimate, the apportionment between overseas and rest of the UK for 2006/07 onwards was applied to the 2001/02 figure to derive flows to and from England, Wales and Northern Ireland. The distribution of immigration for flows between to Scottish districts and the rest of the UK is assessed in Figure 23.



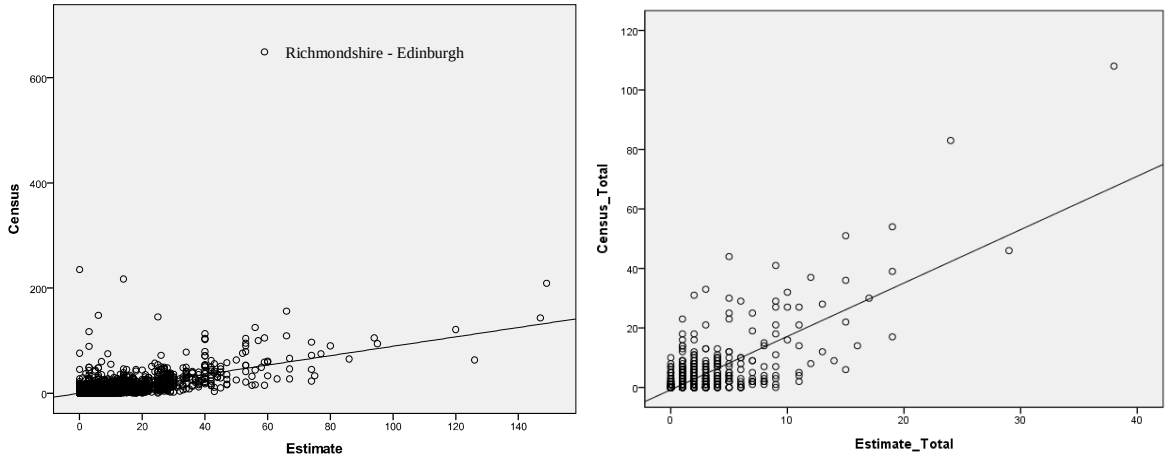
(a) Immigration



(b) Emigration

Figure 23: The proportion of 'outside Scotland' immigration from (a) and emigration to (b) the rest of the UK for 2006/07 to 2009/10 (ranked by 4 year average)

The proportion of the UK allocation of total 'outside Scotland' immigration flows is largely consistent between 2006/07 and 2009/10, with the exception of Shetland which has a very low number of migrant flows to and from Northern Ireland. The proportion of UK emigration is less consistent between 2006/07 and 2009/10, with a notably higher UK emigration in 2007/08 than in 2009/10. The average value was still used, given the small number of years for which data were available. The average inflow and outflow to/from the rest of the UK was used to apportion the 2001/02 figures. The proportions of Scottish total migration for each district reported in the 2001/02 figures were then used to apportion the flows of immigration and emigration reported in the NHSCR.



(a) From England and Wales **(b) From Northern Ireland**

Figure 24: Flows from districts in England and Wales (a) and Northern Ireland (b) to districts in Scotland

For flows from districts in England and Wales to Scotland (Figure 24a), the correlation between the 2001 Census is fairly strong and positive ($r = 0.608$, $p < 0.00$). Figure 24b shows that the correlation for flows from Northern Irish to Scottish districts is stronger ($r = 0.735$, $p < 0.00$).

A single outlier for the flow between England and Wales and Scotland skews the results; Richmondshire to Edinburgh has 649 migrations in the 2001 Census whereas the estimated figure is only 59. Figure 25 shows that when this one anomaly is removed, the correlation improves slightly ($r = 0.691$, $p < 0.00$).

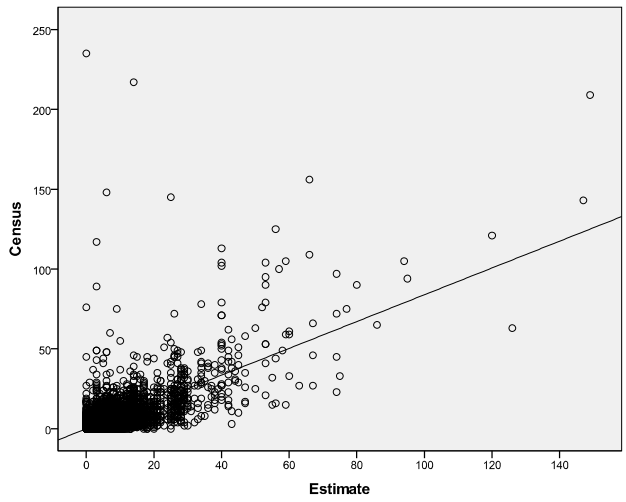
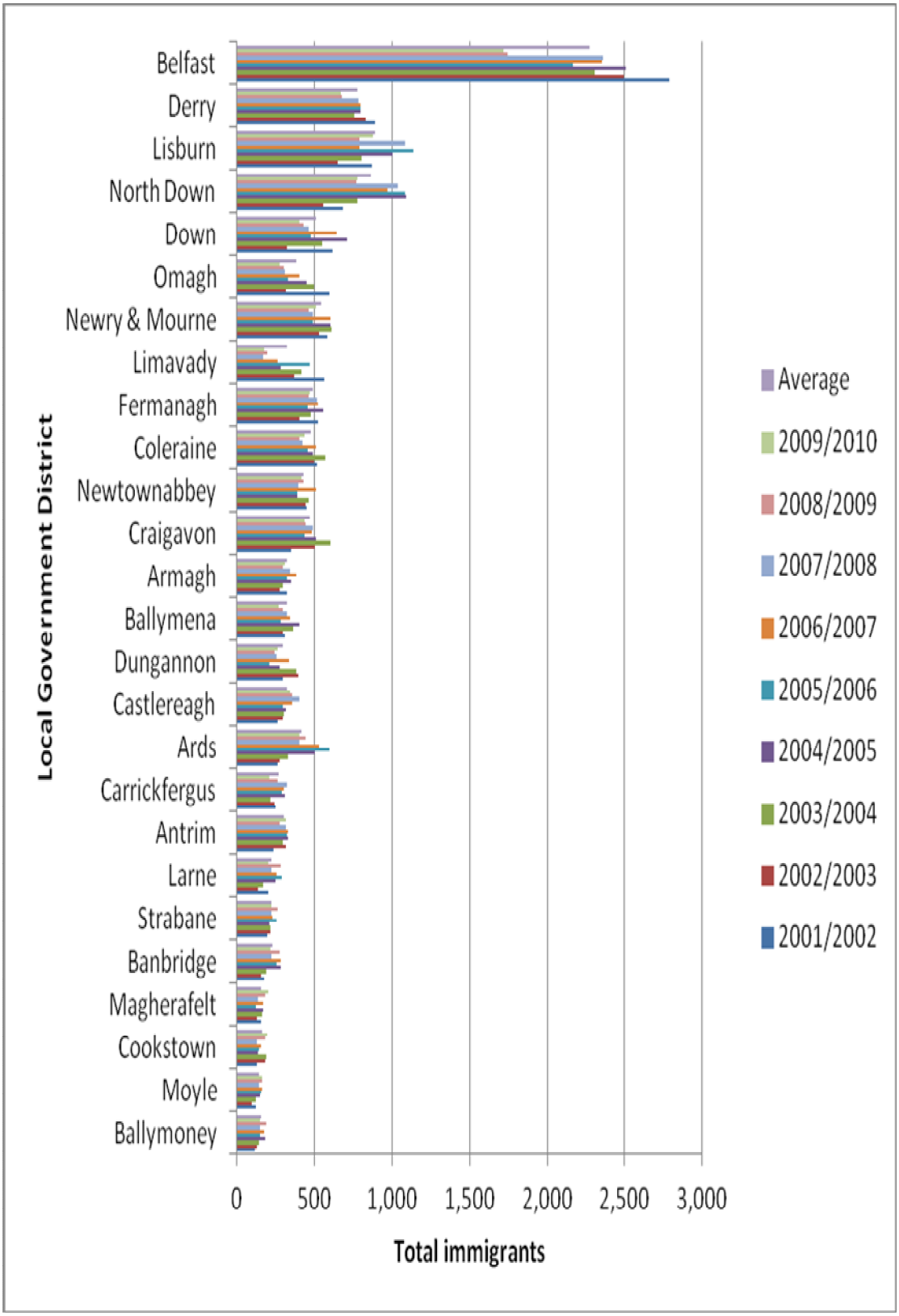


Figure 25: Flows from districts in England and Wales to districts in Scotland (minus the outlier of Richmondshire to Edinburgh)

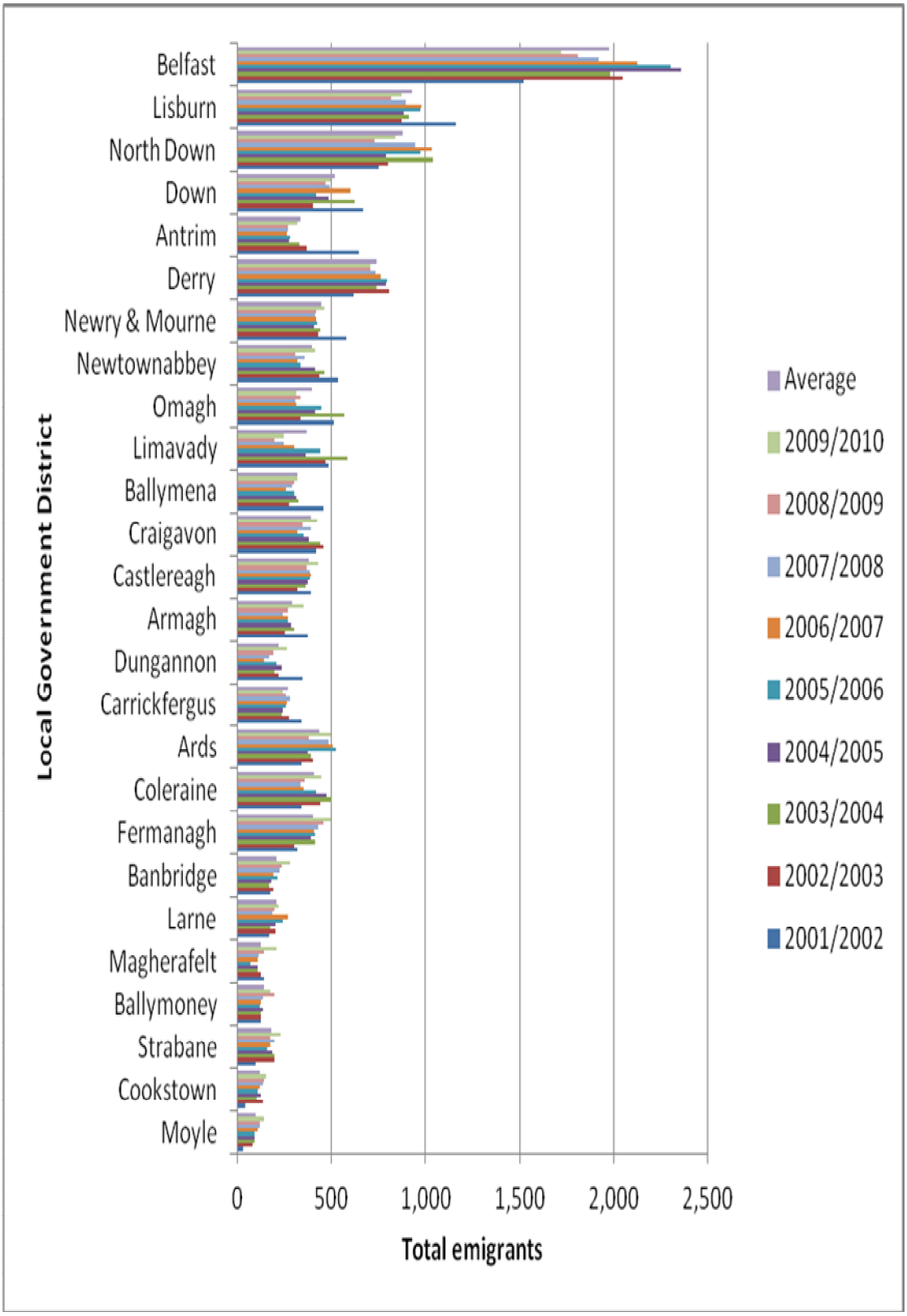
Even though the removal of the outlier displaying the largest difference improves the correlation, it is evident from Figure 25 that the estimate is still significantly different from the Census figure for a large number of outliers.

8.4.4 Northern Ireland

In Northern Ireland, the 2000/2001 data by district was not available so again the estimate of flows to and from the rest of the UK by district drew on the 2001/02 figures supplied by NISRA. Figure 26a shows that the inflow from the rest of the UK is largely consistent between 2001/02 and 2009/10, as is the outflow to the rest of the UK (Figure 26b). The 2001/02 figure has been used to distribute the NHSCR figures to districts for both inflows and outflows.


Belfast
Derry
Lisburn
North Down
Down
Omagh
Newry & Mourne
Limavady
Fermanagh
Coleraine
Newtownabbey
Craigavon
Armagh
Ballymena
Dungannon
Castlereagh
Ards
Carrickfergus
Antrim
Larne
Strabane
Banbridge
Magherafelt
Cookstown
Moyle
Ballymoney
Local Government District
0
500
1,000
1,500
2,000
2,500
3,000
Total immigrants
Average
2009/2010
2008/2009
2007/2008
2006/2007
2005/2006
2004/2005
2003/2004
2002/2003
2001/2002

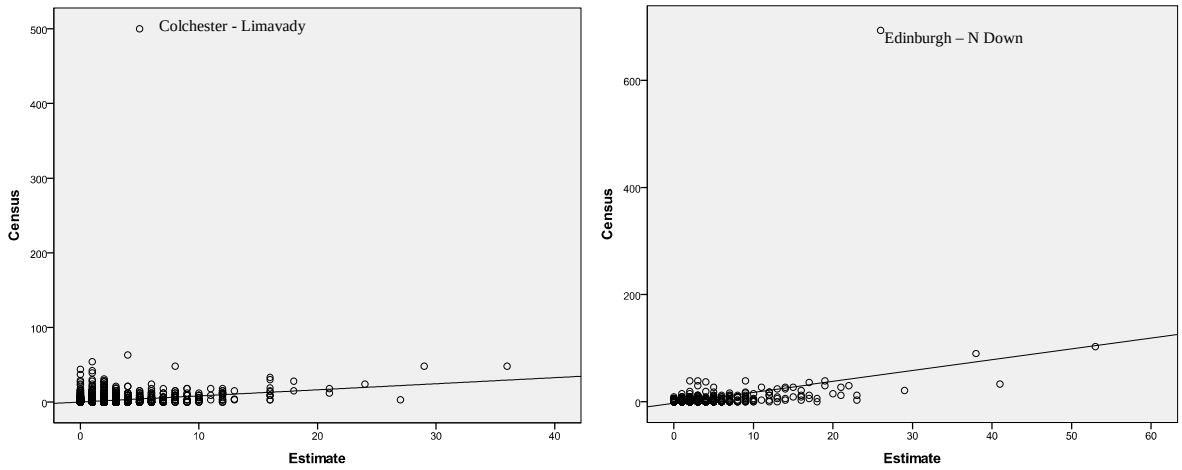
(a) Inflow



(b) Outflow

Figure 26: Share of total inflow from (a) and outflow to (b) the rest of the UK by LGD, 2001–2010 (ranked by 2001/02 flows)

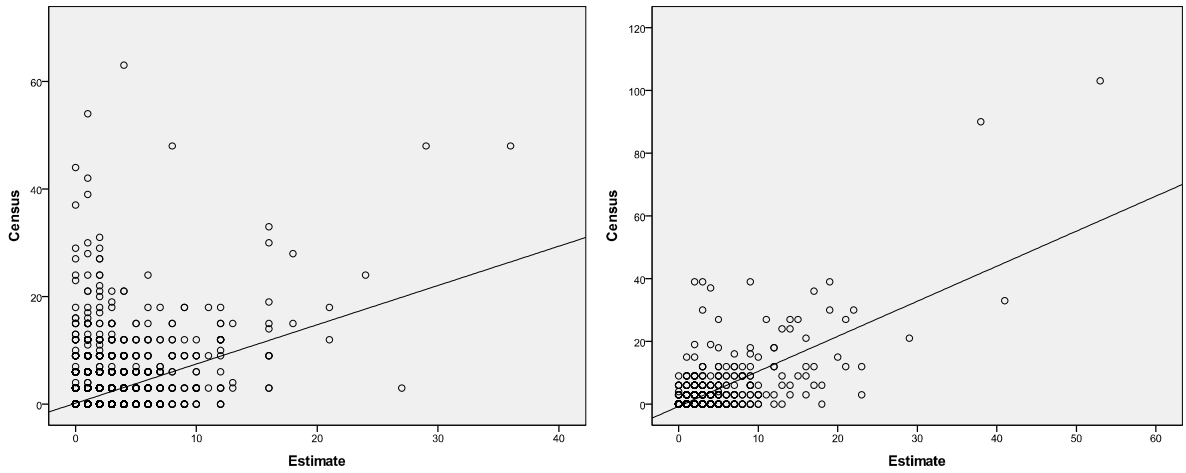
Figure 27 shows that when all flows between districts are considered, the correlation between the estimate the flows reported in the 2001 Census is a poor one, both for flows to Northern Ireland from England and Wales ($r = 0.220$, $p < 0.00$) and from Scotland ($r = 0.365$, $p < 0.00$).



(a) From England and Wales **(b) From Scotland**

Figure 27: Flows from districts in England and Wales (a) and Scotland (b) to districts in Northern Ireland

Notably, there is a single large outlier for the flows from England and Wales and one outlier for the flows from Scotland. The flow from the English district of Colchester to Limavady in Northern Ireland is hugely inconsistent, with the Census reporting 500 moves, while the estimated figure is only 5. Figure 28a shows that when this single pairwise listing is removed, the correlation improves substantially ($r = 0.403$, $p < 0.00$). Figre 28b shows dramatic improvement in the correlation between census and estimate for Scotland to Northern Ireland when the largest anomaly, the flow from Edinburgh to North Down is removed ($r = 0.689$, $p < 0.00$). The Census recorded flow between the districts is 693 while the estimate is only 26.



(a) From England and Wales **(b) From Scotland**

Figure 28: Flows from districts in England and Wales (Minus Colchester to Limavady; a) and Scotland (Minus Edinburgh to North Down; b) to districts in Northern Ireland.

By removing the single largest outlier for each flow, the correlation improves, however there are still a number of significant outliers.

8.5 All cross-border flows

Figure 29 shows that when all cross-border flows are considered, the correlation between the estimates and the 2001 reported flows is a positive one ($r = 0.567$, $p < 0.00$).

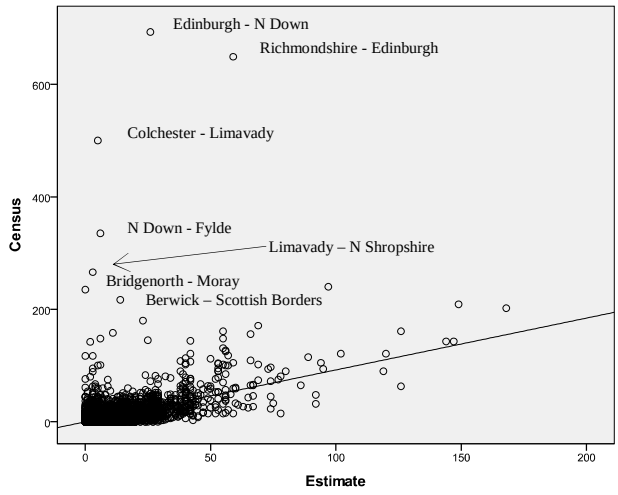


Figure 29: All flows between districts that cross the borders of the countries in the UK

The correlation includes some significant outliers, as discussed briefly above. An assessment of the largest anomalies was undertaken, and all those flows where the Census reported flow differs from the estimated flow by over 200 migrants was removed. A

difference of 200 was used as it provided a useful cut off (the majority of total flows were under 200) and identified those districts displaying very high flows in the census compared to relatively low estimated figures.

Table 6: Cross border flows where the pairwise district to district migration differs by over 200 persons between the Census and estimate

Origin Country	Origin District	Destination Country	Destination District	Census Flow	Estimate
Scotland	Edinburgh, City of	Northern Ireland	North Down	693	26
England & Wales	Richmondshire	Scotland	Edinburgh, City of	649	59
England & Wales	Colchester	Northern Ireland	Limavady	500	5
Northern Ireland	North down	England & Wales	Fylde	335	6
Northern Ireland	Limavady	England & Wales	North Shropshire	266	3
England & Wales	Bridgnorth	Scotland	Moray	235	0
England & Wales	Berwick-upon-Tweed	Scotland	Scottish Borders	217	14

Seven pairwise locations were identified (as shown in Table 6) and figure 30 shows that once these flows were removed from the analysis the correlation improves somewhat ($r = 0.698$, $p < 0.00$).

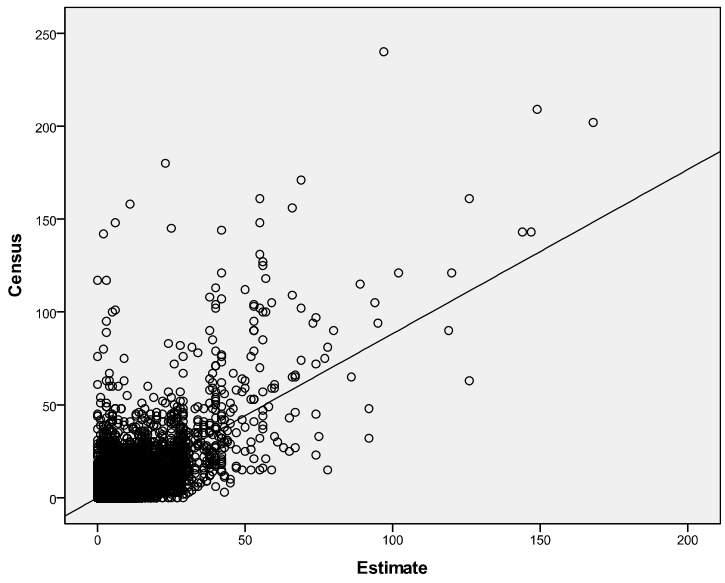


Figure 30; All flows between districts that cross the borders of the countries in the UK (minus seven pairwise flows where the Census flow exceeds the estimate by over 200)

Despite the correlation improving when the largest outliers were removed, the cross-border estimation is still weaker than the other components in the matrix.

8.5.1 Overall levels

Some of the weaker correlations between estimates and census seen in the cross border flows may be partially explained by comparing the overall levels of migration that have been used in the estimation with those reported in the 2001 Census. Table 7 shows the difference between total internal intra-national and total internal inter-national flows that have been estimated and those reported in the 2001 Census.

The internal intra-national estimates are higher than the Census for England and Wales and Scotland (140,077 and 4,994 higher respectively) while in Northern Ireland, the estimate is 1,870 lower than the Census figure. Given that the internal intra-national flows make up the bulk of all migration seen in the matrix, these differences are relatively low in percentage terms, the estimate for England and Wales is 6.05% different from the census while in Scotland the difference is 4.29% and in Northern Ireland is 5.12% (see Figure 31). This may help to explain the strong correlations between estimates and the Census for internal intra-national flows seen in section 8.2.

In contrast, some large percentage differences can be seen between the estimate and the Census for total cross-border flows, as illustrated in Figure 31. Table 7 shows that the estimated flow from Northern Ireland to Scotland (1,806) is considerably lower than the Census reported figure (2,602) which represents a percentage difference of 36.12%. The flow in the opposite direction is similarly under estimated, 2,514 compared with 2,891 which represents a 13.95% difference. The total estimated flow from England and Wales to Scotland is 8,124 higher than the census flow (53,345 compared with 45,221) which represents a difference of 16.48%. The estimate of total flow in the opposite direction is also considerably higher than the Census figure, 48,618 compared with 44,227 which represents a difference of 9.46%.

The total flows between Northern Ireland and England and Wales are the only internal inter-national flow estimates that are consistent with the Census figure. For flows from Northern Ireland the estimate is 38 lower than the Census, representing a 0.4% difference and in the opposite direction the estimate is 154 higher than the Census count, a 1.8% difference.

Table 7: A comparison between overall levels of migration reported in the 2001 Census and within and between countries

Origin \ Destination	England and Wales	Scotland	Northern Ireland
England and Wales	Census = 2,243,713 Estimate = 2,383,790 Difference = 140,077 (6.05%)	Census = 45,221 Estimate = 53,345 Difference = 8,124 (16.48%)	Census = 8,457 Estimate = 8,611 Difference = 154 (1.80%)
Scotland	Census = 44,227 Estimate = 48,618 Difference = 4,391 (9.46%)	Census = 113,824 Estimate = 118,818 Difference = 4,994 (4.29%)	Census = 2,891 Estimate = 2,514 Difference = -377 (13.95%)
Northern Ireland	Census = 9,548 Estimate = 9,510 Difference = -38 (0.40%)	Census = 2,602 Estimate = 1,806 Difference = -796 (36.12%)	Census = 37,437 Estimate = 35,567 Difference = -1,870 (5.12%)

These differences in overall levels do not explain the largest outliers seen in Table 6, where the Census flow is significantly higher than the estimate. They may, however, contribute to some of the poor correlations seen in the internal inter-national component which comprises 45,280 pairs of districts where a flow crosses a border in the UK. As Figure 30 shows, the majority of district to district flows are small, so a small level of overestimation or underestimation may well be amplified when considered across such a large number of pairwise districts.

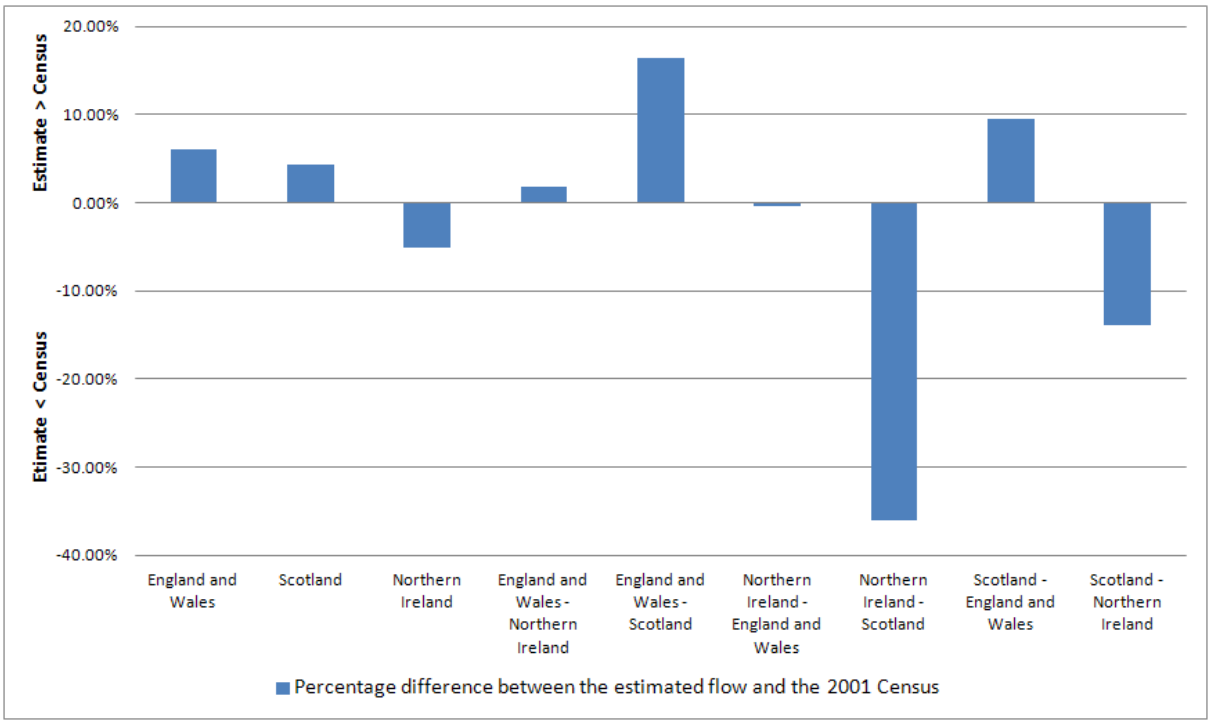
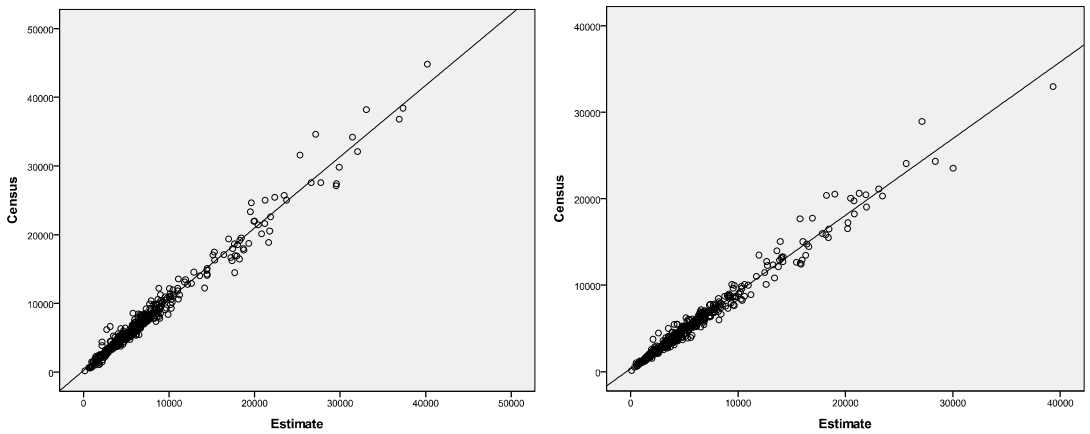


Figure 31: The percentage difference between estimated and 2001 Census reported internal intra-national and inter-national flows

It is clear from the comparison between estimated and Census reported flows that the poor correlations seen for the internal inter-national component is also due to the distribution of migrants at the district level in the estimates. This is particularly evident for flows between Northern Ireland and England and Wales, where despite the estimate and Census totals being consistent, the correlations seen in Figure 21b (Northern Ireland to England and Wales) and in Figure 27a (England and Wales to Northern Ireland) are particularly poor. Some improvement in the total flow data and distribution to districts is required for the internal inter-national portion of the matrix.

8.6 Total inflows and outflows out by district

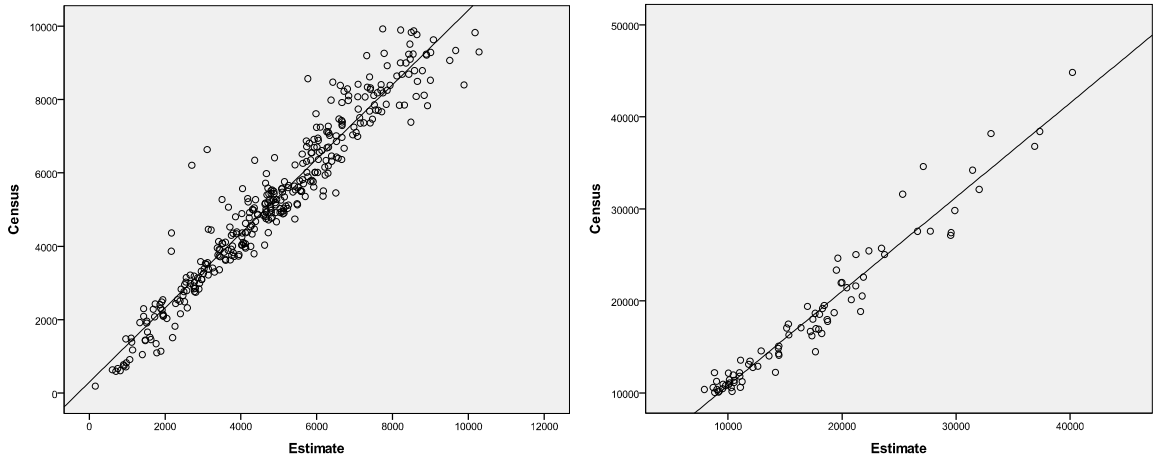
Having assessed the individual components of the matrix, as identified in Table 5, the final analysis considers the aggregate results by assessing the total flow into and out of each district in the UK.



(a) Inflows **(b) Outflows**

Figure 32: Total flows into (a) and out of (b) all districts in the UK

Figures 32a and b show the correlation between the estimate and census reported flows for total immigration and emigration for each district. The correlation for both inflow ($r = 0.989$, $p < 0.00$) and outflow ($r = 0.988$, $p < 0.00$) are strong and positive. A large number of districts exhibit a small number of migrants in and out, and when these are assessed separately the strong positive correlation seen above holds. For all inflows where the Census reported figure is under 10,000 (Figure 33a) $r = 0.962$ ($p < 0.00$) and where all Census values over 10,000 (Figure 33b) $r = 0.973$ ($p < 0.00$).

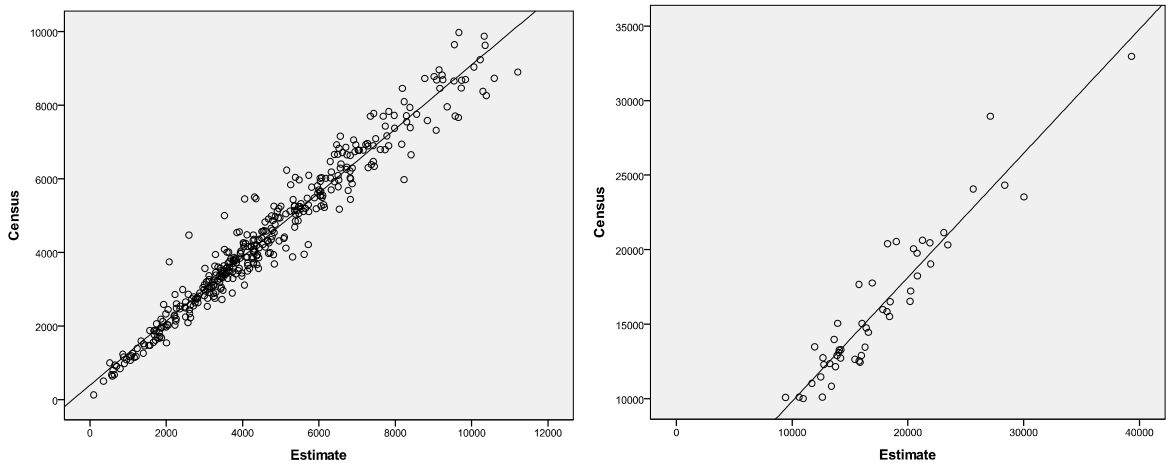


(a) Under 10,000 **(b) Over 10,000**

Figure 33: Total flows under 10,000 (a) and over 10,000 (b) into districts in the UK

The total outflow exhibits a similar story. Figure 34a shows that when all districts where the census outflow is under 10,000 are considered, the correlation between the estimate and census is strong and positive ($r = 0.976$, $p < 0.00$). The correlation when all

Census values over 10,000 is slightly lower (r = 0.946, p< 0.00) but is still strong and positive (Figure 34b).



(a) Under 10,000 **(b) Over 10,000**

Figure 34: Total flows over 10,000 (a) and under 10,000 (b) out of districts in the UK

8.7 Summary

In this section we have attempted to create a set of estimates for each component of the migration matrix which are: (i) inter-district flows within each constituent country (internal intra-national flows) (ii) international flows into each district in the UK and (iii) cross-border flows between districts in each of the four constituent countries (internal inter-national flows).

Table 8 demonstrates that the sources relied upon to estimate the internal intra-national flows show a strong correlation with the 2001 Census reported flows, both in terms of total immigration and emigration from the rest of the country in which a district is located and for the pairwise flows between districts in each country. The estimates of international immigration are strongly correlated with the 2001 Census reported figures, especially for Scotland and Northern Ireland.

The cross-border, internal inter-national component is the most difficult to estimate and the strong correlation between the NHSCR and the Census (r = 0.999) provides a robust framework from which to improve the estimation going forwards. Despite this robust framework, Table 8 shows that the estimate of internal inter-national migration is the weakest component of the matrix with low r values for all estimates of cross-border migration, although correlation does improve with the removal of the most significant outliers.

Table 8: A summary of correlations between estimates and the 2001 Census

<table>
<tr><th colspan="2" rowspan="2">Destination
Origin</th><th colspan="3">England</th><th colspan="3">Wales</th><th colspan="3">Scotland</th><th colspan="3">Northern Ireland</th><th rowspan="2">Rest of the World</th></tr>
<tr><th>District 1</th><th>District ...</th><th>District 355</th><th>District 1</th><th>District ...</th><th>District 21</th><th>District 1</th><th>District ...</th><th>District 32</th><th>District 1</th><th>District ...</th><th>District 26</th></tr>
<tr><td rowspan="3">England</td><td>District 1</td><td colspan="6" rowspan="6">Internal intra-national flows
Inflow r = 0.985
Outflow r = 0.987
Pairwise r = 0.982</td><td colspan="3" rowspan="6">Internal inter-national flows
r = 0.608
Minus one outlier r = 0.691</td><td colspan="3" rowspan="6">Internal inter-national flows
r = 0.220
Minus one outlier r = 0.403</td><td rowspan="6">N/A</td></tr>
<tr><td>District ...</td></tr>
<tr><td>District 355</td></tr>
<tr><td rowspan="3">Wales</td><td>District 1</td></tr>
<tr><td>District ...</td></tr>
<tr><td>District 21</td></tr>
<tr><td rowspan="3">Scotland</td><td>District 1</td><td colspan="6" rowspan="3">Internal inter-national flows
r = 0.692</td><td colspan="3" rowspan="3">Internal intra-national flows
Inflow r = 0.989
Outflow r = 0.997
Pairwise r = 0.985</td><td colspan="3" rowspan="3">Internal inter-national flows
r = 0.365
Minus one outlier r = 0.689</td><td rowspan="3">N/A</td></tr>
<tr><td>District ...</td></tr>
<tr><td>District 32</td></tr>
<tr><td rowspan="3">Northern Ireland</td><td>District 1</td><td colspan="6" rowspan="3">Internal inter-national flows
r = 0.296
Minus three outliers r = 0.466</td><td colspan="3" rowspan="3">Internal inter-national flows
r = 0.735</td><td colspan="3" rowspan="3">Internal intra-national flows
Inflow r = 0.951
Outflow r = 0.994
Pairwise r = 0.788</td><td rowspan="3">N/A</td></tr>
<tr><td>District ...</td></tr>
<tr><td>District 26</td></tr>
<tr><td colspan="2">Rest of the World</td><td colspan="6">International immigration
r = 0.943</td><td colspan="3">International immigration
r = 0.982</td><td colspan="3">International immigration
r = 0.986</td><td>-</td></tr>
</table>

9 Conclusions

This paper has shown that some important inconsistencies exist between the data used and the methods employed by the three national statistical agencies in the UK to produce estimates of sub-national migration. The aspiration to integrate migration estimates for the whole of the UK has some significant obstacles to overcome, and the ONS, NRS and NISRA recognise a need to work in collaboration to produce coherent estimates. This is evident through the formation of the inter-agency Migration Statistics Improvement Programme, with a primary focus on international estimates and the integration of administrative data sources, and a working group tasked with addressing the creation of a UK-wide database of migration at the sub-national level.

Our review of current practice, data sources and estimation methods provides the basis for the development of a full set of sub-national migration estimates for the UK. While various improvements in methodology to distribute the international migration estimate using administrative data sources is underway at the three statistical agencies, there is no cohesive approach and therefore the need to provide some consistency is long overdue. ONS is well advanced in its improvement programme, taking the methodology devised by Boden and Rees (2010) to distribute LTIM estimates based on the 'reason profile' stated in the IPS. A similar approach could be used to distribute the Scottish allocation of LTIM to Council

Areas. Northern Ireland presents a different challenge, with a source of detailed information provided by health card registrations and de-registrations, the potential to combine sources (such as HESA and NINo) to improve estimation exists (and has already been started by NISRA).

The results of the benchmarking exercise, using estimates compared with flows reported in the 2001 Census are encouraging. Whilst the general relationship for most data sources is one of strong positive correlation, a number of outliers need to be examined in further detail and internal inter-national flows improved on before the estimation method can be carried forward to 2010/11 for all the different types of flows concerned and the estimates disaggregated by age and sex. The inclusion of data from auxiliary administrative sources and the potential to obtain more detailed data from the national statistical agencies presents the opportunity to improve the aggregate time series database considerably, although more sophisticated estimation methods are required to disaggregate the flows in the matrix by age and sex.

References

Audit-Commission. (2006) National duplicate registration initiative. [online]. Available from: http://www.audit-commission.gov.uk/SiteCollectionDocuments/AuditCommissionReports/NationalStudies/ndri2004report.pdf.

Bell, M. and Rees, P. (2006) Comparing migration in Britain and Australia: harmonisation through use of age - time plans. *Environment and Planning A*, **38**(5), pp.959-988.

Bijak, J. (2010) *Independent review of methods for distributing international immingration estimates to regions*. Southampton: University of Southampton.

Boden, P. and Rees, P. (2010) Using administrative data to improve the estimation of immigration to local areas in England. *Journal of the Royal Statistical Society Series a-Statistics in Society*, **173**, pp.707-731.

Bohara, A. K. and Krieg, R. G. (1996) A zero-inflated Poisson model of migration frequency. *International Regional Science Review*, **19**(3), pp.211-222.

Boyle, P. (1993) A three-dimensional approach to multistream migration modelling: local-level flows in Hereford and Worcester. *Environment and Planning A*, **25**(9), pp.1279-1293.

Congdon, P. (2010) Random-effects models for migration attractivity and retentivity: a Bayesian methodology. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **173**(4), pp.755-774.

Dennett, A. (2010) *Understanding internal migration in Britain at the Start of the 21st century*. thesis, University of Leeds.

Dennett, A. and Rees, P. (2010) Estimates of internal migration flows for the UK, 2000-2007. *Population Trends*, **140**(1), pp.82-105.

Ewing, G. O. (1974) Gravity and linear regression models of spatial interaction: A cautionary note. *Economic Geography*, **50**(1), pp.83-88.

Flowerdew, R. (2010) Modelling migration with poission regression. In Stillwell, J., Duke-Williams, O. and Dennett, A., eds. *Technologies for migration and commuting analysis : spatial interaction data applications.* Hershey: Business Science Reference, pp.261-279.

Flowerdew, R. and Aitkin, M. (1982) A method of fitting the gravity model based on the Poisson distribution. *Journal of Regional Science*, **22**, pp.190-198.

Fotheringham, A. S., Nakaya, T., Yano, K., Openshaw, S. and Ishikawa, Y. (2001) Hierarchical destination choice and spatial interaction modelling: a simulation experiment. *Environment and Planning A*, **33**(5), pp.901-920.

GROS. (2003) Suppliment to PAMS (03) 10. [online]. Available from: http://www.gro-scotland.gov.uk/files/pams0310-supplement.pdf.

GROS. (2010a) Methodology used to produce GROS official migration estimates [online]. Available from: http://www.gro-scotland.gov.uk/statistics/theme/migration/methodology.html.

GROS. (2010b) Mid-year population estimates for Scotland: Methodology guide. [online]. Available from: http://www.gro-scotland.gov.uk/files2/stats/population-estimates/mid-year-pop-est-methodology.pdf.

Huff, D. (1963) A probabilistic analysis of shopping center trade areas. *Land Economics*, **39**, pp.81-90.

Kupiszewska, D. and Nowok, B. (2008) Comparability of statistics on international migration flows in the European Union. In Raymer, J. and Willekens, F., eds. *International Migration in Europe.* Chichester: Wiley and Sons, pp.41-71.

Lovett, A. and Flowerdew, R. (1989) Analysis of count data using Poisson regression. *Professional Geographer*, **41**, pp.190-198.

Mcgregor, I. (2011) *Visit to ONS Titchfield.* 1 April. Titchfield.

Millington, J. (2000) Migration and age: The effect of age on sensitivity to migration stimuli. *Regional Studies*, **34**(6), pp.521-533.

Mueller, G. (2011) *Meeting at Ladywell House.* 11 May. Edinburgh.

National-Statistics. (2006) *Report of the Inter-Departmental Task Force on Migration Statistics.* London: National Statistics.

NHS. (2011) *Spine Factsheet* [online]. [Accessed 16 May]. Available from: http://www.connectingforhealth.nhs.uk/resources/systserv/spine-factsheet.

NISRA. (2006) How are population estimates created? The methodology. [online]. Available from: http://www.nisra.gov.uk/archive/demography/population/midyear/mye_methodology.pdf.

NISRA. (2007) Development of methods/sources to estimate population migration in Northern Ireland [online]. Available from: http://www.nisra.gov.uk/archive/demography/population/migration/dev_est_mig.pdf.

NISRA. (2010) Migration statistics for Northern Ireland (2009). [online]. Available from: http://www.nisra.gov.uk/archive/demography/population/migration/Migration%20Report%202009.pdf.

Norman, P. (1999) *Putting Iterative Proportional Fitting on the researcher's desk.* Working Paper 99/3, Leeds: School of Geography, University of Leeds.

ONS. (2007a) Actual Impacts of Improved Methods for Population Statistics in 2007 [online]. Available from: http://www.statistics.gov.uk/downloads/theme_population/FinalImpactTables.xls.

ONS. (2007b) The Construction of Intermediate Geographies in London -Amendment. [online]. Available from: http://www.statistics.gov.uk/downloads/theme_population/LondonGeography.pdf.

ONS. (2007c) Improved Methods for Estimating International Migration

Geographical Distribution of Estimates of In-migration. [online]. Available from: http://www.statistics.gov.uk/downloads/theme_population/Geog_distn_in-migs.pdf.

ONS. (2007d) Improved Methods for Estimating International Migration: Estimation of People Whose Intentions Change With Respect to Their Length of Stay. [online]. Available from: http://www.statistics.gov.uk/downloads/theme_population/Mig_vis_switching.pdf.

ONS. (2007e) Improved Methods for Estimating International Migration: Geographical Distribution of Estimates of Out-migration. [online]. Available from: http://www.statistics.gov.uk/downloads/theme_population/Geog_distn_out-migs.pdf.

ONS. (2007f) Improved Methods for Estimating International Migration: Improved Local Age Distributions of International In-migrants. [online]. Available from: http://www.statistics.gov.uk/downloads/theme_population/Geog_distn_in-migs.pdf.

ONS. (2007g) Improved Methods for Estimating International Migration: Out-migration - Propensity to Migrate Models for 2002 to 2006. [online]. Available from: http://www.ons.gov.uk/ons/guide-method/method-quality/imps/archive-material/archive-background-information/improved-methods-for-population-statistics-revisions--archive-material-/propensity-to-migrate-models-used.pdf.

ONS. (2007h) Improved Methods for Estimating International Migration: The Use of the Labour Force Survey to Improve Estimates of International In-migration - Coverage and Quality. [online]. Available from: http://www.ons.gov.uk/ons/guide-method/method-quality/imps/archive-material/archive-background-information/improved-methods-for-population-statistics-revisions--archive-material-/the-use-of-the-labour-force-survey.pdf.

ONS. (2007i) *A review of the potential use of administrative sources in the estimation of population statistics.*

ONS. (2008) Improved Methods for Estimating International Migration: Geographic Areas Used in Estimation (Amendment). [online]. Available from: http://www.ons.gov.uk/ons/guide-method/method-quality/imps/archive-material/archive-background-information/improved-methods-for-population-statistics-revisions--archive-material-/geographic-areas-used-in-estimation.pdf.

ONS. (2009a) *Estimating international long-term immigration by local authority (LA).*

ONS. (2009b) *Methodology to estimate Long-Term International Migration: 1991 to 2008.*

ONS. (2010a) Estimating Internal Migration: Customer Guidance Notes. [online]. Available from: www.statistics.gov.uk/statbase/.../EstimatinginternalMigration.doc.

ONS. (2010b) Estimating international long-term emigration by local authority: update. [online]. Available from: http://www.ons.gov.uk/about-statistics/methodology-and-quality/imps/mig-stats-improve-prog/comm-stakeholders/improvements-2008-pop-est/detailed-methodology-papers/emigration-detailed-methodology-update.pdf.

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

ONS. (2010d) Improving estimates of international migration in Northern Ireland, and between the UK and Republic of Ireland. [online]. Available from: http://www.statistics.gov.uk/downloads/theme_population/Irish_methodological_changes.pdf.

ONS. (2010e) Mid-year estimates short methods guide. [online]. Available from: http://www.statistics.gov.uk/downloads/theme_population/short-methods-guide.doc.

ONS. (2010f) Summary Quality Report for Internal Migration. [online]. Available from: www.ons.gov.uk/.../mid-year-population-estimates--mye-.pdf.

ONS. (2010g) Table 2: Cumulative effect mid-2008 by type of adjustment. [online]. Available from: http://www.ons.gov.uk/about-statistics/methodology-and-

quality/imps/mig-stats-improve-prog/comm-stakeholders/improvements-2008-pop-est/tables-and-charts/t2-cumulative-effect-2008-by-type-of-adjustment.xls.

ONS. (2011) Summary quality report for long-term international migration releases. [online]. Available from: http://www.ons.gov.uk/ons/guide-method/method-quality/quality/quality-information/social-statistics/summary-quality-report-for-international-migration.pdf.

Raymer, J. (2007) The estimation of international migration flows: a general technique focused on the origin ? destination association structure. *Environment and Planning A*, **39**(4), pp.985-995.

Raymer, J. (2010) A general framework for estimating population movements. In Salzmann, T., Edmonston, B. and Raymer, J., eds. *Demographic Aspects of Migration.* Wiesbaden: VS Research, pp.73 - 96.

Raymer, J. and Giulietti, C. (2010) Analysing structures if interregional migration in England. In Stillwell, J., Duke-Williams, O. and Dennett, A., eds. *Technologies for migration and commuting analysis.* New York: Business Science Reference, pp.280-293.

Raymer, J. and Rogers, A. (2007) Using age and spatial flow structures in the indirect estimation of migration streams. *Demography*, **44**(2), pp.199-223.

Raymer, J., Smith, P. W. F. and Giulietti, C. (2011) Combining census and registration data to analyse ethnic migration patterns in England from 1991 to 2007. *Population Space and Place*, **17**(1), pp.73-88.

Rees, P. (2008) What happens when international migrants settle? Projections of ethnic groups in United Kingdom regions. In Raymer, J. and Willekens, F., eds. *International Migration in Europe.* Chichester: Wiley and Sons, pp.329-358.

Rees, P., Fotheringham, A. S. and Champion, T. (2004) Modelling migration for policy analysis. In Stillwell, J. and Clarke, G., eds. *Applied GIS and spatial analysis.* Chichester: Wiley, pp.259-296.

Rees, P. and Willekens, F. (1986) Data and accounts. In Rogers, A. and Willekens, F., eds. *Migration and settlement: A multiregional comparative study.* Dordrecht: Reidel, pp.19-58.

Rees, P. H. and Duke-Williams, O. (1997) Methods for estimating missing data on migrants in the 1991 British census. *International Journal of Population Geography*, **3**(4), pp.323-368.

Rogers, A., Little, J. and Raymer, J. (2010) *The indirect estimation of migration.* Dordrecht: Springer.

Rogers, A., Willekens, F. and Raymer, J. (2001) Modeling interregional migration flows: Continuity and change. *Mathematical Population Studies: An International Journal of Mathematical Demography*, **9**(3), pp.231 - 263.

Rogers, A., Willekens, F. and Raymer, J. (2003) Imposing Age and Spatial Structures on Inadequate Migration-Flow Datasets. *The Professional Geographer*, **55**(1), pp.56-68.

Rolfe, H. and Metcalf, H. (2009) *Recent Migration into Scotland: the Evidence Base.* Edinburgh: Scottish Government Social Research.

Roy, J. R. and Thill, J.-C. (2004) Spatial interaction modelling. *Papers in Regional Science*, **83**(1), pp.339-361.

Saito, S. (1992) A multistep iterative proportional fitting procedure to estimate cohort-wise interregional migration tables where only inconsistent marginals are known. *Environment and Planning A,* **24**(11), pp.1531-1547.

Sarra, A. L. and Del Signore, M. (2010) A Dynamic Origin-constrained Spatial Interaction Model Applied to Poland's Inter-provincial Migration. *Spatial Economic Analysis,* **5**(1), pp.29 - 41.

Select-Committee-on-Public-Accounts. (2007) *Progress with demographics in the NHSCRS since Dr Nowlan's departure* [online]. Available from: http://www.parliament.the-stationery-office.co.uk/pa/cm200607/cmselect/cmpubacc/390/6062623.htm.

Simpson, L. and Tranmer, M. (2005) Combining sample and census data in small area estimates: Iterative Proportional Fitting with standard software. *The Professional Geographer,* **57**(2), pp.222-234.

Smallwood, S. and Lynch, K. (2010) An analysis of patient register data in the Longitudinal Study – what does it tell us about the quality of the data? *Population Trends,* (141), pp.152-169.

Smith, P. W., Raymer, J. and Giulietti, C. (2010) Combining available migration data in England to study economic activity flows over time. *Journal of the Royal Statistical Society: Series A (Statistics in Society),* **173**(4), pp.733-753.

Stillwell, J. (1991) Spatial interaction models and the propensity to migrate over distance. In Stillwell, J. and Congdon, P., eds. *Migration models, macro and micro approaches.* London: Belhaven, pp.34-56.

Stillwell, J. (2005) Inter-regional migration modelling: A review and assessment. In *45th Congress of the European Regional Science Association, 23-27 August 2005, Vrije Universiteit Amsterdam, The Netherlands*.

Stillwell, J., Boden, P. and Dennett, A. (2011) Monitoring who moves where: Information systems for internal and international Migration. In Stillwell, J. and Clarke, M., eds. *Population dynamics and projection methods.* London: Springer, pp.115-140.

Stillwell, J., Duke-Williams, O. and Dennett, A. (2010) Interaction data: definitions, concepts and sources. In Stillwell, J., Duke-Williams, O. and Dennett, A., eds. *Technologies for Migration and Commuting Analysis: Spatial Interaction Data Applications.* New York: IGI Global, pp.1-30.

Swier, N. (2011) An improved method for estimating immigration to Local Authorities in England and Wales. In *British Society of Population Studies (BSPS) Annual Conference,* 7-9 September, York.

UKSA. (2009) *Migration Statistics: the Way Ahead? Report 4 (Final).* London: UKSA.

van Imhoff, E., van der Gaag, N., van Wissen, L. and Rees, P. (1997) The selection of internal migration models for European regions. *International Journal of Population Geography*, **3**(2), pp.137-159.

Wilson, A. G. (1967) A statistical theory of spatial interaction models. *Transportation Research,* **1**, pp.253-269.

Wong, D. (1992) The reliability of using the iterative proportional fitting procedure. *The Professional Geographer*, **44**(3), pp.340-348.