

Tracking the working lives of migrants in the UK

Impro Mobil Country Report

Tim Riley

Callum Miller

Paul Bivand

Tony Wilson

June 2013

This report has been quality assured by:

Name: Tony Wilson
Position: Director of Policy
Date: June 2013

Timothy Riley

Senior Researcher

Inclusion

Third Floor

89 Albert Embankment

London

SE1 7TP

Tel: 020 7582 7221

Email: tim.riley@cesi.org.uk

Contents

Contents	3
Executive Summary	4
Introduction.....	7
Section A: Review of dataset Availability – Surveys	11
Labour Force Survey (LFS) / Annual Population Survey (APS).....	11
International Passenger Survey (IPS).....	19
Census.....	24
British Household Panel Survey/Understanding Society	30
Family Resources Survey (FRS)/EU-SILC	40
Section A: Review of dataset Availability – Administrative data	45
National Insurance Recording System (NIRS) and associated datasets.....	45
Worker Registration Scheme (WRS).....	59
Work permits and the 'Points Based System'	62
The Seasonal Agricultural Workers Scheme (SAWS).....	64
The Sectors Based Scheme (SBS)	66
Section B: Ability of survey and administrative data to track migrant workers.....	68
Who is a migrant?	68
Short and long term movers	68
Tracking foreign workers within the UK.....	69
Time of arrival in UK and work histories in previous country	72
Tracking British workers outside UK.....	73
Posted workers and multinational/supranational organisations	73
Coverage of key variables	74
Section C: Suggestions for improving data availability.....	75
Planned future changes in UK.....	75
Procedural or bilateral agreements for exchanging information on migrant workers.....	77
Recommendations to improve the ability to track migrant workers	77
Increasing access	77
Improving the quality of administrative data.....	77
Improving the quality of survey data.....	79

Executive Summary

The UK has a range of administrative data sources with which it is possible to measure the working lives of migrants. These include: the National Insurance Recording System (or NIRS, which includes data about income tax and National Insurance contributions for all people working legally in the UK); The Work and Pensions Longitudinal Study (or WPLS, which combines NIRS data with data from the Department for Work and Pensions); and the Lifetime Labour Market Database (a 1% cleansed random sample of NIRS data). Of these, the Lifetime Labour Market Database (LLMDB) is likely to be the most practical in terms of analysing the working lives of migrants.

However, access to all these administrative sources is severely restricted. Access to the LLMDB is theoretically possible, but DWP analysts report that it would require clearance from both the Department for Work and Pensions and HM Revenue and Customs. Gaining access would require a champion working in a policy role in DWP, and unlikely that data could be used away from DWP premises or (particularly) outside of the UK

As such, it may be that survey sources are most useful to understand the working lives of migrants in the UK. In particular, the British Household Panel Survey (BHPS) and its successor, Understanding Society (USoc), are household panel surveys allowing researchers to map the working lives of some respondents over a period of 20 years. However, these panel surveys have some limitations: first, sample sizes of people born outside the UK are small, restricting the possibility of subgroup analysis; second, as a panel survey with relatively static groups of respondents, it is likely they lag behind the profile of migrants in the UK (for example, the BHPS will not have captured the wave of migration that accompanied A8 accession to the EU); and third, the survey is a household survey, and therefore migrant workers living in communal establishments are not in scope.

As such, other surveys are likely to be required to more accurately understand the working lives of migrants in the UK. Fortunately, the UK has an excellent range of surveys that can help in this regard. These include the decennial Census, which allows longitudinal analysis of around 3% of the population through the Census Longitudinal Study. However, the most recent Census microdata is from the 2001 Census, so the Census Longitudinal Study will not allow analysis to be carried out within the timescales of this project.

The Labour Force Survey (LFS) and the Annual Population Survey (APS)¹ provide detailed cross sectional data about the labour market, as well as longitudinal data over the period of a year. The questions about labour market participation are particularly detailed in the LFS, and the numbers of migrants in the datasets are relatively large (around 5,500 in work in a recent LFS file).

The International Passenger Survey provides interviews with immigrants and emigrants as they enter or leave the UK. In particular, it allows us to understand migrants expectations about work at the point they migrate, and how this compares to the type of work they did in the country they migrated from.

We make the following recommendations to improve the UK's ability to track migrant workers.

- 1) Access – administrative datasets should be more easily available for academic and non-commercial research

To improve quality of administrative data:

- 2) Data on the work histories of migrants and educational attainment should be collected by the UK Government when migrants register to work in the UK. This should then be appended to administrative datasets.
- 3) Qualification levels should be appended to administrative datasets for as many residents as possible. For example, 'fuzzy matching' from Census data onto administrative data may be possible.
- 4) Better data sharing between EU members and/or some kind of harmonised administrative data collection about workers in the EU would help better understand emigrants from the UK.
- 5) EU-wide guidance to make it easier to equivalise qualifications gained in different countries in Europe would be valuable to better understand the education levels of migrant workers, as this is currently a weakness in many data sources.

Improvements to UK survey sources are also possible

- 6) The International Passenger Survey could be used to boost sample of migrants in other UK surveys such as the LFS and USoc, or even to help other EU members collect EU LFS data on emigrants to EU.

¹ The APS uses LFS data and additional boosts to provide data accurate in local areas.

- 7) One weakness of some UK surveys, which tend to focus on households, is that they do not sample those living in communal establishments like hostels and camp sites. Expanding the scope of these surveys to cover those living in communal establishments would provide better coverage of all migrants working in the UK.
- 8) Panel surveys like USoc should look to complete interviews with emigrants over the phone after they have left the UK.
- 9) Further consideration should be given to including a wage/household income in the UK Census.

Introduction

The UK collects a wealth of information about its labour force as well as specifically collecting data about migration and migrants. Net migration into the UK now accounts for a larger proportion of population growth than the net effect of births and deaths in the UK², and this has made migration increasingly prominent in UK politics.

One reason that immigration has been important in the national political discourse is that there have been concerns in some areas that larger numbers of migrants have put pressure on local services, such as healthcare. As such, a key focus has been on improving data on the stock and flow of migrants in different local areas in the UK, to ensure funding for local services is appropriate for population levels.

The National Statisticians Task Force on Migration worked on this after 2006, and in 2008 the Migration Statistics Improvement Programme³ (MSIP) was initiated. The programme had several aims, including increasing user confidence in migration statistics, increasing efficiency in the production of migration statistics and improving the quality of national and sub-national migration statistics. The MSIP was completed in 2012. Given the focus on counting the migrant stock and measuring migration flows, comparatively little was done to improve the tracking of migrant workers to better record their interaction with labour market over time. Indeed at time of writing, measuring migrant flows at both national and local authority level is still the key issue.⁴

Project aims

This project hopes to go some way to filling this void at both a national and a European level. The aims of the project are to assess the capacity of four European countries (UK, Italy, Poland and Romania) to track mobile and migrant workers over time through national data sources. The capabilities of datasets to track the working lives of both immigrants into the countries and emigrants leaving the countries were of interest, with particular interest in certain groups of migrants, such as posted workers and those working for multinational and supra-national organisations. An ideal dataset would

² The UK Statistics Authority (July 2009), *Migration statistics: the way ahead*.

³ Office for National Statistics, *Migration Statistics Improvement Programme Final Report*

⁴ See for example this early 2013 call for evidence from the UK Statistics Authority, <http://www.statisticsauthority.gov.uk/assessment/assessment/current-assessments/migration-statistics--reassessment-.html>

provide information for all workers about their entire working histories (including details of employment periods spent abroad) and also included socio-demographic personal data.

This report provides a review of the UK's capability to track migrant workers, and is complemented by three other reports produced by our project partners.

The second phase of the project will be to collect and analyse datasets from each of the four countries, to expand our understanding of migrant workers in Europe.

Methodology

The first stage of our methodology was to undertake an extensive literature review of relevant UK publications. This involved reviewing publications that reviewed the UK's capacity to measure and track migrants as well as more general information and user guides relating to individual datasets. Second, we conducted a number of telephone interviews with academics specialising in migration, as well as talking to representatives from various government Departments and Agencies responsible for different datasets relating to migrants in the UK, including at the UK Border Agency (UKBA), the Office for National Statistics (ONS), HM Revenue and Customs (HMRC) and the Department for Work and Pensions (DWP).

We have also investigated which data sources are publicly available, which are available under special license, and which can only be accessed in a secure location, usually at a government-specified location, and tried to advise on this where possible.

Data sources covered

This report will focus on UK data sources and consequently we will not formally review European wide sources such as the EU Labour Force Survey, although an ad hoc module was asked in 2008, and is planned for 2014, relating to migration and the labour market situation of migrants.⁵

Furthermore, there are a number of UK data sources we have consciously decided not to include in this review for varying reasons. First, we have chosen not to review certain administrative data sources relating to migrants that do not shed any light on their working lives. Such data sources include:

⁵ Eurostat European Labour Force Survey tool, http://epp.eurostat.ec.europa.eu/portal/page/portal/employment_unemployment_ifs/data/database

- 'Flag 4' General Practitioner (GP) registrations – This data provides a count of the number of people who have registered with a GP funded through the UK's National Health Service (NHS), 'flagged' by whether they said their previous address was outside the UK.
- Pupil Census data – The data has unique pupil number (UPN) and includes details of age, gender, ethnicity, home address and first language. This data helps understand the numbers of migrants in the education system, but is not relevant to analysis of labour market position of migrants.

We have also excluded certain survey data sources from our review that were less relevant or which did not mark migrants. In particular, we have chosen not to include the Annual Survey of Hours and Earnings (ASHE), which has detailed data on earnings and employment but has no migrant identifier and thus is of limited use for our research.

We have however, reviewed a large number of data sources, both administrative and survey, and these are listed below:

Survey

- Labour Force Survey (LFS) and the Annual Population Survey (APS);
- International Passenger Survey (IPS);
- Census;
- British Household Panel Survey (BHPS) and Understanding Society (USoc); and the
- Family Resources Survey (FRS).

Administrative

- National Insurance Recording System (NIRS) and two key extracts from NIRS, the Migrant Worker Scan (MWS) and the Lifetime Labour Market Database;
- Work and Pensions Longitudinal Study (WPLS);
- Worker Registration Scheme (WRS);
- Work permits and the Points-Based System (PBS);
- Seasonal Agricultural Workers Scheme (SAWS); and the
- Sector Based Scheme (SBS).

The rest of the report is made up of the following sections:

- Section A reviews each dataset in turn;

- Section B presents a summary of the UK's ability to track migrant workers; and
- Section C provides a number of recommendations to improve the UK's ability to track migrant workers in the future.

Section A: Review of dataset Availability – Surveys

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

The Labour Force Survey (LFS) has been running in the United Kingdom since 1973, although data has only been collected quarterly since 1992. Its aim is to provide reliable statistics on the state of the labour market in the country. It is a household survey, with respondents asked questions on a range of topics including demographics, household composition, (self-)employment status, migration and detailed questions on skills and training.

The Annual Population Survey (APS), collects annual data, with estimates published each quarter on a rolling four-quarter basis. It incorporates the main LFS data along with a number of local boosts in an attempt to provide reliable estimates at local authority level.

The collection of the LFS has recently been incorporated into the Integrated Household Survey, which is a composite survey bringing together data from a number of different social surveys into a single dataset. The IHS is, however, less relevant in understanding the working lives of migrants than the LFS or APS.

Longitudinal v cross-sectional	LFS – Repeated cross-sectional but two and five quarter longitudinal datasets can be created. APS - Cross-sectional
Coverage	All individuals resident in UK private households
Time coverage	LFS - Running from 1973 (quarterly since 1992). APS – First conducted in 2004.
Sample size	LFS – Sample of 60,000 private households per quarter from which approximately 41,000 households respond, (roughly 100,000 individuals). APS – achieves responses from 150,000 households, or 360,000 individuals in the 12 months combined data.

Collection methodology

How data is collected

The **Labour Force Survey** is based on a quarterly survey of about 41,000 responding UK private households carried out both over the telephone and face-to-face. Participant households are interviewed in five consecutive quarters, known as 'waves' and thus each survey is made up of five groups of households (with approximately 8,000 responding in each wave). In each new quarter, one fifth of the sample is refreshed, as the group completing their fifth wave is replaced by a fresh sample group who start at wave one. Thus when looking at two consecutive quarters about 80% of the addresses are common to both.

It is important to stress that it is the particular address that gets surveyed for the LFS as opposed to the individuals who were living at that address in wave one, therefore different people may be interviewed in different phases at the same address.

The LFS is based on a systematic random sample design representing about 0.16% of the Great British population, from which an additional sample of approximately 1,600 Northern Irish households, representing 0.23% of the

Northern Irish population, is added⁶. Those living in sampled households are eligible to be surveyed if they regard the sampled address as their main address, or have been living there for six consecutive months. Students are also included if their regular address is sampled and they are away studying during term time. In Great Britain, an additional sample is drawn from those living in NHS and nurses accommodation. The main sampling frame is Royal Mail's Postcode Address File (PAF), which is a list of all addresses receiving mail.

The addresses are randomly selected with some allowances made so that the same households aren't selected repeatedly. Postcode districts are also checked to ensure that sufficient geographical variance is achieved. Overall, the number of addresses selected in wave one is 16,640 although some of these addresses are either ineligible or non-respondent. Added to this are those in NHS accommodation, the NI boost, and an additional sample for those in Northern Scotland – giving a total of 17,380 addresses newly selected each quarter.

Since the introduction of the quarterly surveys in 1992 the response rate of the UK LFS has fluctuated between 74 and 84 percent.² All first wave interviews are carried out face-to-face and, with the households' permission, recall interviews are carried out over the phone, although some follow-up wave interviews are conducted face-to-face. In total, in 2010 62% of interviews were carried out via telephone and 38% face-to-face.²

The **Annual Population Survey** is an annual survey, drawing on the LFS and additional local boosts in order to provide statistics reliable at the local level. Responses are published each quarter, each including data from the previous twelve months. So as not to include data relating to the same household twice within any 4-quarter period, only Wave 1 and Wave 5 survey responses from the Main LFS are used in APS data sets. The Annual Population Survey adds data collected in the Local Labour Force Survey (LLFS), sometimes known as the LFS boost, to the main LFS data in order to provide reliable estimates at a local authority level. Each local area has a target number of economically active individuals to interview; and if this target is not achieved by the main LFS sample, a boost (through the LLFS) is required.

⁶ Office for National Statistics (2011), *Labour Force Survey User Guide, Volume 1, LFS Background and Methodology, London 2011*, http://www.esds.ac.uk/doc/7220/mrdoc/pdf/lfs_user_guide_vol1_background2011.pdf

Overall, the APS achieves responses from approximately 150,000 households with just over half of responding households coming from the main LFS sample.

The **Integrated Household Survey** is a recent move by government to increase efficiency and cut costs in social survey collection, the first dataset available is for the 12-month period from April 2009 to March 2010. From then, annual datasets will be released quarterly on a rolling basis.

The IHS is a composite survey bringing together data from a number of social surveys into a single dataset. It contains LFS and associated boosts for the APS; General Lifestyle Survey; and the Living Costs and Food Survey. Overall, there is a total sample size of 450,000 individuals. Not every individual is asked each question. A core module is asked of everyone, with the appropriate social surveys asked to their respective sample populations, and certain additional topic modules offered to selected groups of respondents (for example, a recent module asked questions about wellbeing).

The core topics include household composition, national identity, ethnicity, country of birth, health, work and education. Its relevance here is that the IHS is the collection method for attaining LFS and APS data.

Key groups captured/missed

As the LFS/APS are household surveys, some groups are more likely than others to be missed by the data collection process. For example, people living in communal establishments (such as hostels, caravan parks, or student accommodation) are missed as the survey is only completed by those living in private households. This means that some seasonal workers, including summer agricultural workers are likely to be underrepresented in the surveys. Census data shows that approximately 14% of international migrants live in communal establishments across England and Wales and thus these individuals are missed.⁷

Recent migrants are also thought to be more likely to refuse to answer the survey or provide incomplete information because of language barriers and mistrust of the interviewers – especially if their residence or work status is not entirely compliant with immigration regulations.

⁷ Improved Methods for Estimating International Migration The Use of the Labour Force Survey to Improve Estimates of International In-migration - Coverage and Quality

Sample size and composition

Approximately 40,000 households and 100,000 individuals respond to the Labour Force Survey each quarter, and each Annual Population Survey release covers approximately 150,000 households, 360,000 people.

In each LFS release, approximately 11,500 people are born outside of the UK and approximately 5,500 of them are in work. It is likely that approximately the same proportion of respondents will be born outside the UK in Annual Population Survey data as will be in LFS releases.

Time coverage / changes in methodology over time

The Labour Force Survey began in 1973, running as a biennial survey, and was started initially to satisfy Eurostat (the Statistical Office of the European Community) conditions. By 1983, it was being used to measure unemployment and was switched to an annual survey in 1984. It has since developed into a key indicator of UK labour market conditions and from 1992 onwards data has been collected on a quarterly basis.

The Annual Population Survey was first published in July 2005, and has been published on a quarterly basis since (with data reflecting data from the previous four quarters.)

The LFS has recently been incorporated into the Integrated Household Survey for collection purposes, as an independent module but the sample and questions remained the same as they were previously.

Weighting

The data is weighted to make it representative of the total resident UK population and the table below shows the sampling variability for employment levels by country of birth and nationality, when scaling up from Q4 2012 LFS data.

As the LFS is only meant to collect data on long term migrants, the calibration totals used to weigh the estimates only include long term migrants at this time.

Sampling variability of Labour Force Survey estimates of employment by country of birth and nationality, July to September 2012.

	Level	Sampling variability
UK born people in employment (000s, aged 16+)	25,420	± 181
UK born employment rate (aged 16-64)	72.2%	± 0.4
Non UK born people in employment (000s, aged 16+)	4,267	± 114
Non UK born employment rate (aged 16-64)	68.2%	± 1.1
UK nationals in employment (000s, aged 16+)	27,071	± 178
UK nationals employment rate (aged 16-64)	71.9%	± 0.4
Non UK nationals in employment (000s, aged 16+)	2,617	± 100
Non UK nationals employment rate (aged 16-64)	68.6%	± 1.4

Source: Office for National Statistics, (2012), *Labour Market Statistics*,

<http://webcache.googleusercontent.com/search?q=cache:zxOXtbNb298J:www.ons.gov.uk/ons/rel/lms/labour-market-statistics/january-2013/table-a11.xls+&cd=1&hl=en&ct=clnk&gl=uk>

Subjects covered

A copy of the full LFS questionnaire can be accessed here:

<http://www.ons.gov.uk/ons/guide-method/method-quality/specific/labour-market/labour-market-statistics/index.html>.

	Variables Recorded
Employment status	✓
Sector	✓
Employment history (work experience) ⁸	✓
Country of destination (for leavers)	
Country of birth	✓
Nationality	✓
Age	✓
Gender	✓
Qualification level (educational attainment)	✓
Date highest educational attainment was achieved ⁹	✓
Family composition	✓
Occupation	✓
Marital status	✓
Country of last residence	
Moving for work in supranational institutions	✓
Moving for work in multinational companies	
Workers posted abroad	

⁸ The LFS asks about employment 12 months ago, or longer for those whose 'last job' finished more than a year ago.

⁹ The LFS has a marker for whether qualifications were achieved in 2009 or earlier or in 2010 or later; whether apprenticeships were completed pre or post-2000; and what in-work training has been completed in the past three months and one month respectively.

Time of arrival	✓
Wage	✓

Migrants are defined, and a marker allocated to them in the survey, as those who were living abroad 12 months before being interviewed. However, the survey also captures nationality, country of birth and year of first and last arrival in the UK and whether they have lived continuously in the UK since then. Those working in supranational institutions are identifiable by Standard Occupational Classification (SOC).

In addition to the variables identified above, the Labour Force Survey also has a raft of questions examining the employment status of respondents, details of current and past employment by industry and wage bracket or benefit receipts, as well as the details of any government training schemes they may be signed up to. It also contains detailed data on education and training and highest qualification level achieved to date. The LFS asks about employment 12 months before the survey, and this may uncover what immigrants who entered the UK less than a year before being surveyed were doing in their previous country of residence.

The LFS asks respondents to self-identify their ethnicity and religion, as well as their household composition, family characteristics and accommodation details.

How the data is used

In addition to providing data from each quarterly LFS survey, the ONS merges different quarterly LFS datasets to provide longitudinal datasets for both two quarters and five quarters respectively. The five quarter longitudinal dataset is created by following a group of households, approximately 8,000, through each 'wave' of Labour Force Survey responses. Households are tracked in two consecutive quarters for the two quarter longitudinal dataset. Therefore, we can compare respondents' responses three and twelve months apart. The ONS publishes these population-weighted longitudinal datasets for each calendar quarter (the datasets are limited to people aged 16-64) and datasets are available for each quarter since 1997. Because of the resources involved in production and the size of the resultant datasets, the longitudinal datasets include only a subset of the full LFS variable set.

Public availability of the data

All Labour Force Survey and Annual Population Survey datasets are available publicly from the UK Data Service. Use of the Local Labour Force Survey (also

known as the LFS boost) is possible but only with a special license to enable access to specially commissioned tables of information.

Summary of Pros and Cons

As survey data sources, the LFS and APS have all the main advantages and disadvantages common to datasets of this type. On the plus side they contain very detailed data on the state of the UK labour market with relatively large sample sizes of migrant workers. The Annual Population Survey is more effective for looking at small geographical areas due to the larger sample size than the smaller main LFS sample. However, it lacks the longitudinal element which makes the LFS particularly useful.

One particular advantage of the LFS is that it is able to measure labour market activity of both recent migrants (by asking what respondents' address was one year ago), and longer term migrants (by asking which year respondents arrived in the UK). Also, its ability to follow individuals over time allows us to see short term movements between employment, unemployment and economic inactivity. However, longitudinal analysis is limited to the relatively short timeframe of five consecutive quarters.

On the other hand, in common with many of the United Kingdom methods of tracking migrant workers, it is only able to track immigrants and has no capacity to glean any information on emigrants. If a respondent leaves the sample prematurely (i.e. before their final wave) it is not clear whether they have moved to another UK address or emigrated out of the UK.

Moreover, because the LFS surveys private households, international migrants who live in communal establishments are not in the scope of the survey – this is estimated to represent approximately 14% of international migrants¹⁰. It is also believed that migrants may be less likely to respond to the survey.

Furthermore, as the LFS only refreshes one-fifth of its sample population each quarter, it is expected that the survey will lag behind changes in the structure of the population. However, this lag is considerably less than with a panel survey like the British Household Panel Survey (BHPS).

Key sources of information

Office for National Statistics (2011), *Labour Force Survey User Guide, Volume 1 LFS Background and Methodology London, 2011,*

¹⁰ Improved Methods for Estimating International Migration The Use of the Labour Force Survey to Improve Estimates of International In-migration - Coverage and Quality

http://www.esds.ac.uk/doc/7220/mrdoc/pdf/lfs_user_guide_vol1_background_2011.pdf

R. Beerten, Office for National Statistics (2007), *Household Survey for UK Government: Building on Survey Integration*, <http://www.fcs.gov/07papers/Beerten.VIII-A.pdf>

Office for National Statistics, *Improved Methods for Estimating International Migration. The Use of the Labour Force Survey to Improve Estimates of International In-migration - Coverage and Quality*, [Link here](#)

Office for National Statistics (February 2013), *Labour Market Flows February 2013*, http://www.ons.gov.uk/ons/dcp171766_300621.pdf

Office for National Statistics (2012), *Labour Force Survey, User Guidance 2007-12*. <http://www.ons.gov.uk/ons/guide-method/method-quality/specific/labour-market/labour-market-statistics/index.html>

International Passenger Survey (IPS)

The International Passenger Survey (IPS) is a survey asked of a sample of people who enter or leave the UK via most of the principal ports or routes. The original reason for conducting the survey was not to learn about migration but to look at overseas travel and tourism patterns.

However, the questionnaire is able to identify migrants, for whom an additional module of questions is asked. Another key feature of the IPS is that as both emigrants and immigrants are surveyed, it is a useful resource for looking at those leaving, as well as entering, the UK.

Longitudinal v cross-sectional	Repeated cross-sectional
Coverage	95% of those entering / leaving UK
Time coverage	Annually 1961
Sample size	800,000 passengers (approx. 4,000 – 5,000 migrants)

Collection methodology

How data is collected

The IPS is a face-to-face survey that covers all main air, sea and tunnel routes into the UK, the only major exceptions being the border between Northern Ireland and the Republic of Ireland. In total, 95% of people entering

or leaving the UK have the potential to be interviewed as part of the IPS, the remainder being passengers travelling at night or via routes too small or too expensive to be covered by the survey. In addition, asylum seekers are not eligible to be interviewed as part of the IPS, nor are people travelling on cruises that start or end in the UK (although numbers of the latter are relatively small).

The coverage of each port or route is dictated through a sample optimization exercise, which takes into account a) the volume and profile of passengers at each port or route and b) the multi-purpose nature of IPS. A multi-stage sampling design is employed which involves sampling a port or route on a given day and within a given period of the day (this is referred to as a 'shift'). Within a shift certain passengers passing an interview line are systematically chosen for interview at fixed intervals from a random start.¹¹

All those chosen for interview are initially asked if they are migrating to/from the UK, with any migrants identified asked to complete the migration module of the questionnaire. Of these screened for migration, some are also asked to complete certain modules of the IPS questionnaire, focussing on travel and tourism.

Sample size and composition

The International Passenger survey, in total, collects information from between 250,000 and 300,000 individuals a year, which is equal to about 1 in 500 people entering or leaving the UK.¹²

Looking specifically at migration, each year around 800,000 passengers are asked if they are migrating to or from the UK (i.e., intending to move for at least a year). From this, between 4,000 and 5,000 people are identified as long-term migrants. A paper in 2009 by the UK Statistics Authority noted that about 70% of these long-term migrants were immigrants and 30% emigrants¹³.

The overall response rate (complete and partial interviews) for the 2011 IPS was 79%. Of the 21% non-response to the IPS: 19 per cent is down to a lack

¹¹ Office for National Statistics, *International Passenger Survey Methodology*, <http://www.ons.gov.uk/ons/guide-method/method-quality/specific/travel-and-transport-methodology/international-passenger-survey/index.html>

¹² Oxford Migration Observatory, *International Passenger Survey*, - <http://migrationobservatory.ox.ac.uk/data-and-resources/data-sources-and-limitations/international-passenger-survey>

¹³ The UK Statistics Authority (July 2009), *Migration statistics: the way ahead. Report 4, July 2009*.

of available interviewers to meet the prescribed ratio of passengers surveyed, and 2 per cent is due to passengers refusing to answer the survey.¹⁴

Time coverage / changes in methodology over time

The IPS has been run since 1961 and is published monthly, quarterly and annually. Initially, the survey only operated at Heathrow and Gatwick airports and Dover port, but it has subsequently been expanded and in 1994 the Channel Tunnel was added.

The Office for National Statistics' Migration Statistics Improvement Programme (MSIP)¹⁵ led to a number of alterations to the IPS methodology. A number of questions were added to the IPS in 2009, specifically to reflect changes in migration patterns, especially with the accession of eight new countries to the EU in 2004. MSIP also led to increased coverage at regional airports, and more in depth questions about intentions of immigrants leaving the country, which are particularly useful in judging the extent to which international students remain in the UK to take up employment after they finish studying.

Weighting

The basis of the weighting of IPS survey data is that the total set of respondents interviewed at a port or route is weighted up/calibrated to passenger traffic known to have passed through that port or route in the period in question. It is then corrected for non-respondents, and those who only provide minimal detail, and finally corrected for an observed imbalance between the number of non-migrants entering and leaving the UK, to provide an estimate of the characteristics of all passengers entering or leaving the United Kingdom.

Subjects covered

The IPS, by necessity, employs a short interview of just a few minutes because of the nature of the sample, the travelling public. It covers migration intentions as well as questions relating to travel and tourism.

¹⁴ Office for National Statistics, *International Passenger Survey: Quality Information in Relation to Migration Flows*. Available from: <http://www.ons.gov.uk/ons/guide-method/method-quality/specific/population-and-migration/international-migration-methodology/index.html>.

¹⁵ The Office for National Statistics, *Migration Statistics Improvement Programme Final Report*.

The full IPS questionnaire can be found here:

<http://www.ons.gov.uk/ons/guide-method/method-quality/specific/travel-and-transport-methodology/international-passenger-survey/index.html>

	Variables Recorded
Employment status	
Sector	
Employment history (work experience) ¹⁶	✓
Country of destination (for leavers)	✓
Country of birth	
Citizenship	✓
Age	✓
Gender	✓
Qualification level (educational attainment) ¹⁷	
Date highest educational attainment was achieved	✓
Family composition	
Occupation	✓
Marital status	✓
Country of last residence	✓
Moving for work in supranational institutions	
Moving for work in multinational companies	
Workers posted abroad ¹⁸	
Time of arrival	
Wage	

Migrants are asked to self-identify themselves and their travel intentions for the purposes of the IPS. The IPS also asks questions about travel and tourism to a proportion of survey respondents, collecting data on countries visited and a breakdown of spending.

How the data is used

The International Passenger Survey is the main source of data used by the Office of National Statistics to create their estimates of Long-Term¹⁹

¹⁶ The IPS 'migrant' module of questions asks individuals for a brief description of what they used to do as their 'main job' and this could be used to build up a basic picture of employment history.

¹⁷ The IPS does have data on the age they left full time education, which could be used as a proxy for qualification level.

¹⁸ Does ask arrivals if they are 'directly employed outside of the UK'

¹⁹ I.e. migration of more than a year

International Migration (LTIM). Once the International Passenger Survey itself has been corrected for slight discrepancies during the weighting process, as described above, it is further amended to create the a LTIM estimate.

The first of these amendments is to account for the fact that people may not always stay in or out of the country for the duration they reported when asked on arrival/departure. Since January 2004, new IPS questions have been asked about the difference between respondents' intended duration of stay and their actual length of stay to make this adjustment more accurate.

In addition, an adjustment is made to take into account asylum seekers, as the IPS does not capture the vast majority of asylum seekers entering or leaving the UK. ONS attains figures from the Home Office on the number of asylum seekers and their dependants who are entering or leaving the UK.

The figure is finally altered to take into account travel between the land border between Northern Ireland and the Republic of Ireland, as these over-the-border travels will not be captured by the International Passenger Survey.

Public availability of the data

The IPS data is available on a national basis and the ONS published cross-reference tables for different variables, although restrictions apply when attempting to view the results on a more localised level. Data can be downloaded by registered researchers from the UK Data Service website going back to 1993²⁰.

Summary of Pros and Cons

The main advantage to the IPS is that it collects relatively rich information on the reason for migration and on migrant characteristics. In addition, the IPS is the only detailed source of information about emigration from the UK. This allows us to build up a picture of emigrants' characteristics as well as providing us with more accurate data on actual duration of stay (as leavers report actual time spent in the UK.)

However, despite the very high total number of interviews conducted as part of the IPS, only a relatively small number of migrants take part, generally between 4,000 and 5,000. Moreover, most of these migrant interviews are conducted with immigrants rather than emigrants, so even though the IPS is

²⁰UK Data Service, *International Passenger Survey list of datasets*, <http://www.esds.ac.uk/findingData/ipsTitles.asp>

an invaluable source about emigration from the UK, the sample of emigrants is relatively small.

Another drawback to using the IPS is that it primarily asks about migration intentions, and migrants' intentions and expectations are not always matched in reality. Whilst this is taken into account during the weighting process, the adjustments are unlikely to be flawless.

The final drawback is that its coverage is not comprehensive as it does not take into account, for example, migration between Northern Ireland (the UK) and the Republic of Ireland, or migration through certain smaller ports and airports.

Overall, the IPS is a useful tool to provide a cross-sectional look at immigrants' and emigrants' characteristics over time, however there is no capacity to track these migrants once they enter or leave the UK.

Key sources of information

Office for National Statistics, *International Passenger Survey Methodology*, <http://www.ons.gov.uk/ons/guide-method/method-quality/specific/travel-and-transport-methodology/international-passenger-survey/index.html>

Oxford Migration Observatory, *Data Sources and Limitations, The International Passenger Survey*, <http://migrationobservatory.ox.ac.uk/data-and-resources/data-sources-and-limitations/international-passenger-survey>.

UK Statistics Authority, (2009), *Migration Statistics: the Way Ahead, Report 4*.

UK Statistics Authority (2012), *Monitoring Review Online: Robustness of the International Passenger Survey*.

The Office for National Statistics, *International Passenger Survey: Quality Information in Relation to Migration Flows*.

Census

Brief description

The most comprehensive source of data on the characteristics of the population in the UK is the Census. It is collected every ten years and aims to achieve complete coverage of the UK population in all parts of the country. In practice, it cannot achieve full coverage but responses rates are well above

90%. The Census collects data on individuals and households on ethnicity, country of birth and resident address one year ago.

Longitudinal v cross-sectional	Cross-sectional but longitudinal dataset also available
Coverage	All UK residents
Time coverage	A census in some form has been collected since 1801
Sample size	In 2011 – 56.1 million.

Collection methodology

The data collection process for the Census aims to achieve complete coverage across all areas of the UK, and across all individuals in those regions.²¹ Everyone in the country is legally obliged to complete the census, and thus response rates are far above comparative rates for other surveys.

The Census questionnaire is posted to all households in the UK, accompanied by an information leaflet and a pre-paid envelope for return by post. Households are given the option of completing and returning their questionnaire via post or completing it online. Their response information is uploaded onto the questionnaire tracker system to identify non-responding households for follow-up by the collection team who provide advice and guidance on the doorstep. As Census completion is a legal obligation for UK residents, legal proceedings can be initiated with those who refuse to comply.

When an address is identified as a communal establishment, Census questionnaires are hand-delivered and responses collected by special enumerators.

Key groups captured/missed

Whilst the census aims to achieve complete coverage, in reality some people are missed, and they are likely to be the more mobile groups of a population.

As with the majority of UK data sources, the Census has no data on emigration as it takes a snapshot of people usually resident in the UK at a point in time.

²¹ For more detail on the collection process see: (<http://www.ons.gov.uk/ons/guide-method/census/2011/how-our-census-works/how-we-took-the-2011-census/how-we-collected-the-information/index.html>)

Sample size and composition

The Census achieved 94% response rate of estimated usual residents and 95% household response rate. In 2011, this equated to a total of 56.1 million respondents.²²

Time coverage / changes in methodology over time

Since its introduction in 1801, when it was only intended to be used to count the population, the Census has become increasingly complex and enables us to gain a better insight into the characteristic of the population.

Subjects covered

	Variables Recorded
Employment status	✓
Sector	✓
Employment history (work experience) ²³	✓
Country of destination (for leavers)	
Country of birth	✓
Citizenship	✓
Age	✓
Gender	✓
Qualification level (educational attainment)	✓
Date highest educational attainment was achieved	
Family composition	✓
Occupation	✓
Marital status	✓
Country of last residence	
Moving for work in supranational institutions ²⁴	✓
Moving for work in multinational companies	
Workers posted abroad	
Time of arrival	✓
Wage	

In the national census that took place in the UK on 27 March 2011, a 32-page questionnaire containing 56 questions was completed by all respondents. Of these 56: 14 were about the household and 42 were for each individual

²² Office for National Statistics (December 2012), Response rates in the 2011 Census.

²³ Has details of the last job an individual held if not currently employed but nothing more detailed. In addition, the longitudinal dataset could be used to build up a decennial employment history.

²⁴ Has SIC code data which has a classification (9900) for extraterrestrial organisations and bodies.

member of the household to complete. This included questions about work, education, national identity, citizenship, ethnic background, language, health, religion, marital status and address a year ago.

Seven new questions were included in 2011: two about the household (number of bedrooms and type of central heating), and five about the residents (passports held, national identity, year of entry to the U.K and intended length of stay for recent arrivals, main language and whether or not they have a second residence). The new questions relating to migration were added as a result of work completed during the Migration Statistics Improvement Programme²⁵.

With the introduction of the question about intended length of stay, the Census is now able to differentiate between short-term migrants and temporary workers and long-term migrants. Whilst clearly there are possible discrepancies between intended length of stay and actual length of stay, the responses will no doubt provide a fuller understanding of long and short term migration. The introduction of a question on year of (last) entry to the U.K will further improve our understanding of migration stock and flow.

A copy of the questionnaire can be found here:

<http://www.ons.gov.uk/ons/guide-method/census/2011/how-our-census-works/how-we-took-the-2011-census/how-we-collected-the-information/questionnaires--delivery--completion-and-return/2011-census-questions/index.html>

How the data is used

The Census data is released in both aggregated and individualised forms. The aggregate data is available at present for the 2011 UK Census but the individualised datasets, outlined below, that are created have not yet been released. The individualised datasets available are:

- The Samples of Anonymised Records (SARs); and
- The ONS Longitudinal Study (LS).

Samples of Anonymised Records (SARs)

The Samples of Anonymised Records (SARs) are a family of datasets containing individualised person-level records that have been anonymised to prevent individual identification. The files contain a full range of census topic

²⁵ The Office for National Statistics, *Migration Statistics Improvement Programme Final Report*.

data on individuals but are designed to ensure that sample members cannot be identified.

The SARs have the advantage of much larger sample sizes than are typical in alternative survey data sources. For example, the 2001 Individual SAR contains 3% of UK census records, equating to 1.84 million cases. The largest file, the 2001 Small Area Microdata (SAM), is a 5% file containing nearly three million cases.

The SARs files contain data from one census only (1991 or 2001 or 2011 in due course) but contain data on all census topics for individuals. This contrasts with other individual level (or microdata) census products such as the Office for National Statistics Longitudinal Study, which links individual data records over time. However, unlike the Longitudinal Study, most SARs files can be downloaded from the UK Data Service rather than requiring access from a secure setting.

The ONS Longitudinal Study (LS)

The LS is a data set comprising linked census and publicly held event records (such as births and deaths) for 1% of the population of England and Wales. It selects people born on one of four days throughout the year, about 950,000 individuals. The LS is useful for longitudinal and cross-sectional studies of census and events data. It contains:

- All information from census returns since 1971, including data on occupation, economic activity, housing, ethnicity, age, sex, marital status and qualifications and education;
- Events data on births & deaths, fertility, mortality, health, migration and cancer registrations; and
- Geographic data.

Its main limitation is the relatively restricted access afforded to external researchers when using the Longitudinal Study. To obtain the data, a number of application forms must be completed with the process usually taking 6 to 8 weeks.

Public availability of the data

Data from the 2011 Census is released gradually in between July 2012 and October 2013. The main aggregate data on migration and ethnicity was

released in the second release of data in December 2012.²⁶ Data will be gradually released on more targeted geographical areas up until March 2013.

The microdata for the 2011 census has not yet been released and thus the microdata we have available is 14 years old. The release of the Longitudinal Study is expected in November 2013.

The SARs files are available to download from the UK Data Service website but access to the ONS Longitudinal Study data is much more restricted.

Summary of Pros and Cons

The main advantage of the Census is that it covers well over 90% of UK residents, and that it contains key variables on the topics of work and migration. Improvements to the 2011 Census as a result of the Migration Statistics Improvement Programme mean that the long and short term migrants can be differentiated for the first time in the data.

The main drawback to the census is that, as it is only carried out once every ten years, and therefore it cannot capture always capture recent migration patterns, such as recent expansions of the European Union. The latest individualised data that is available is from the 2001 Census, and individualised data for the 2011 Census will not be available until the end of 2013. Nevertheless, the ONS Longitudinal Study in particular would be a powerful tool to track the working lives of migrants.

Another con of the Census is that, whilst response rate is very high, it is possible that migrants, are more likely not to respond, particularly if English is not their first language.

Key sources of information

Office for National Statistics (2012), *Response rates in the 2011 Census*.

University College London, *About the ONS Longitudinal Study*,
http://www.ucl.ac.uk/celsius/about_ls

²⁶ Aggregate data on country of birth, ethnicity and other variables on main language are now publicly available here:
<http://neighbourhood.statistics.gov.uk/dissemination/Download1.do;jsessionid=nrVkRCQCDJbkCWHRtnTpvm1L0Rc4mnPH17mmpqmJRy7hQhqhTpPd!940645360!1359122498386?&nsjs=true&nsck=true&nssvg=false&nswid=1346>

The Office for National Statistics, *Migration Statistics Improvement Programme Final Report*

Office for National Statistics, *2011 Census: Guidance and Methodology*,
<http://www.ons.gov.uk/ons/guide-method/census/2011/index.html>

British Household Panel Survey/Understanding Society

Brief description

The British Household Panel Survey (BHPS), which ran for 18 waves between 1991 and 2008, was a multi-purpose household survey whose unique value lay in the fact it followed the same representative sample of individuals, the panel, over a period of years.

The UK Household Longitudinal Survey, more commonly known as Understanding Society (USoc), is a larger household panel survey which started in 2009. The second wave of Understanding Society also incorporated the BHPS panel into its sample as well the new respondents sampled in wave 1. Taken as a whole, it is the largest study of its kind in the world. As of April 2013, data was available for the first two waves of Understanding Society.

The Understanding Society questionnaire covers topics including income and earnings, health, disability, education, migration history, nationality, employment, childcare and other care responsibilities, life satisfaction, community and leisure amongst others.

Longitudinal v cross-sectional	Longitudinal
Coverage	All individuals resident in UK private households ²⁷
Time coverage	BHPS: 1991 to 2008. USoc, 2009 to present
Sample size	In waves one and two of Understanding Society, interviews were achieved with approximately 30,000 households and approximately 50,000 individuals. In BHPS, there were 14,419 individual respondents in the most recent wave , wave 18, and 18,867 in wave 11 when the sample was at its peak.

Sample size and composition

British Household Panel Survey

The initial sample for Wave One of the BHPS consisted of 8,167 addresses drawn from the Post Office's Postcode Address File (PAF), which is a list of all addresses receiving mail. Interviews were attempted at all private households found at these addresses (subject to selection where multiple households were found). All individuals resident in respondent households became part of the longitudinal sample. The sample for the subsequent waves consists of all adults in all households containing at least one member who was resident in a household interviewed at Wave One, regardless of whether that individual had been interviewed in Wave One. In addition, a number of households where no contact had been made in Wave One were approached for interview in Wave Two after confirmation that no household moves between waves had taken place.

The wave 1 panel consisted of some 5,500 households and 10,300 individuals with additions of 1,500 households in each of Scotland and Wales in 1999, and 2,000 households in Northern Ireland in 2001, making the panel suitable for UK-wide research.

The sample sizes available within the BHPS limits the power of analysis that can be undertaken. When the BHPS reached its maximum size (18,867) at

²⁷ 'Residents' here refers to those living in the sampled addresses who regard that location as their main address.

wave 11, there were 832 respondents not born in the UK in the sample. By the final wave of the BHPS (with 14,419 respondents overall), this had fallen to 597 respondents not born in the UK. In total, there were 471 respondents included in both datasets.

Understanding Society

In the first wave of the Understanding Society survey there were three main samples from which a target sample of 40,000 UK households was hoped to be achieved. The three samples were the General Population (GP) sample; the Ethnic Minority Boost (EB); and the Innovation sample. The GP sample included almost 50,000 households, the EB had around 45,000 households, and the innovation sample contained approximately 1,500 households.

The Innovation sample, of approximately 1,500 households, was a forerunner to the main USoc survey and was conducted twelve months before the start of wave 1. It was used as a chance to test the fieldwork process and learn from the findings before the main stages of the survey were undertaken.

The GP sample in Britain came from the Royal Mail's Postal Address File (PAF), from which a total of 47,520 addresses were selected from 2,640 areas or sample points. In Scotland, England and Wales the sample was drawn in two stages. The first stage involved selecting the Primary Sampling Units (PSUs) - postcode sectors or groups of postcode sectors. The second stage selected Delivery Points (addresses) within each PSU. The PSUs were ordered, and selected using systematic random sampling with a fixed interval, and addresses were selected from amongst these using a similar technique. Of these 47,500 households, over 26,000 interviews were achieved with households.²⁸

The Northern Irish sample was drawn separately from the Land and Property Services Agency's list of domestic properties and was an unclustered sample from which 2,395 addresses were selected.

The Ethnic Minority Boost was included with the aim of interviewing at least 1,000 adults from each of five selected target communities: Indian; Pakistani; Bangladeshi; Black Caribbean; and Black African. This was done by sampling addresses known, from Census and Annual Population Survey data, to contain relatively high proportions of target groups. Overall, around 45,000 households were selected for the EMB sample. However, over three-quarters

²⁸ Understanding Society (November 2012), *UK Household Longitudinal Study, Waves 1-2, 2009-2011, User Manual*, https://www.understandingsociety.ac.uk/system/uploads/asset/000/000/004/original/User_manual_Understanding_Society_Waves_1_2.pdf?1359115559

(77%) of these addresses proved ineligible²⁹, and interviews were achieved with only 4,000 households³⁰.

In wave one, a total of 93,712 addresses were issued to interviewers from across the General Population and Ethnic Minority Boost samples, and interviews were completed in 30,000 households and approximately 51,000 individuals.³¹

It has been estimated from wave one of Understanding Society that 11% of the UK population was born outside the UK.³²

The second wave of USoc included all productive addresses from wave 1, as well as incorporating the British Household Panel Survey sample. In total, the interviewers managed to achieve interviews with approximately 30,000 households: 21,000 coming from the General Population sample, 2,500 coming from the Ethnic Minority Boost sample, and approximately 6,500 households from the British Household Panel Survey sample³³. As with wave one, this resulted in approximately 50,000 interviewed individuals (12,000 from BHPS, 37,000 from General Population sample, and 5,500 from the Ethnic Minority Boost).

Of all GP sample members who completed the individual interview at wave 1 – and excluding those known to have died by the time of wave 2 – 75.4% were interviewed again at wave 2, and proxy interviews were conducted on behalf of a further 1.9%. The remaining 22.7% were not interviewed: 13.3% refused, 6.0% were not contacted and 3.5% could not be located. Re-interview rates were lower in the Ethnic Minority Boost Sample, where 63.4% of wave 1 interviewees were interviewed again, plus 3.1% with proxy interviews.³⁴ Of all persons who completed an individual interview at wave 18 of BHPS, 79.4% also completed the Understanding Society wave 2 interview, with a further 1.3% having a proxy interview completed on their behalf.

²⁹ R.Boreham, D.Boldysevaite National Centre for Social Research, (January 2012) – *UK Household Longitudinal Study Wave 1 Technical Report, January 2012*,
https://www.understandingsociety.ac.uk/system/uploads/assets/000/000/010/original/Wave_1_Technical_Report.pdf?1354035538

³⁰ Understanding Society (November 2012), *UK Household Longitudinal Study, Waves 1-2, 2009-2011, User Manual*,
https://www.understandingsociety.ac.uk/system/uploads/assets/000/000/004/original/User_manual_Understanding_Society_Waves_1_2.pdf?1359115559

³¹ Understanding Society user manual

³² Understanding Society, *Understanding Society: Findings 2012*.

³³ User manual understanding society waves 1 and 2.

³⁴ Lynn, Burton, Nandi et al (January 2012) *An Initial Look at Non-Response and Attrition in Understanding Society*.

Data Collection methodology

British Household Panel Survey

For wave one of the British Household Panel Survey, sampled addresses were initially sent out information leaflets, with interviewers following-up the leaflets by visiting addresses within a week. A minimum of six visits were made at each address before it was considered a non-contact.

The first wave of BHPS interviewing was conducted as a pencil and paper questionnaire but from Wave 9 of the BHPS (in 1999), the survey moved from a pen-and paper (PAPI) mode of data collection to a Computer Assisted Personal Interview (CAPI) mode of collection³⁵. This represents the most significant methodological shift in the life of the BHPS.

Between waves, correspondence was maintained with respondents and an extensive database was kept up-to-date with changing details of respondents' locations.

Understanding Society

Data collection methods were relatively similar across both waves of Understanding Society, although telephone interviewing was available in wave 2, and both waves made use of Computer Assisted Interviewing. Data collection instruments at wave 1 included:

- 15 min (average) Computer Assisted Personal Interview (CAPI) household questionnaire;
- 30 min (average) Individual Adult CAPI questionnaire for all aged 16+;
- 7 min (average) Adult Paper Self-Completion questionnaire for all aged 16+;
- 10 min (average) Youth Paper Self-Completion questionnaire for all aged 10-15; and,
- 10 min (average) CAPI proxy questionnaire.

The main difference in wave 2, was that Computer Assisted Telephone Interviews (CATI) was used, when agreed by the respondent, to increase efficiency of data collection.

The responses were collected in 'waves', each lasting two years. The waves overlap so that respondents are interviewed annually, meaning that someone

³⁵ Taylor, Brice, Buck & Prentice-Lane, University of Essex (June 2010), *British Household Panel Survey, User Manual, Volume A: Introduction, Technical Report and Appendices*.

could be interviewed in wave one at the same time as someone else is being interviewed for wave two.

Before the bulk of interviews are completed within the GP and EM samples, piloting is conducted with the Innovation sample.

Key groups captured/missed

As with the Labour Force Survey, some groups are less likely to be involved in the survey than others. It is thought that recent migrants are more likely to refuse to answer the survey or provide incomplete information because of language barriers and mistrust of the interviewers – especially if their residence or work status is not entirely compliant with immigration regulations. However, the presence of the Ethnic Minority Boost, which is specifically designed to boost engagement amongst such individuals, should go at least some of the way to countering this.

Moreover, as a household survey, people living in communal establishments (such as hostels or caravan parks) are not eligible to be sampled. This means that some migrant groups, such as seasonal workers/ summer agricultural workers, are underrepresented in the survey. Census data has shown that approximately 14% of international migrants live in communal establishments across England and Wales and thus these individuals are missed.³⁶ Another key group not captured here is students living in halls of residences, who do not have a UK resident parent, (if they did the parent would answer the survey as a proxy if the household was selected in the sample).

Subjects covered

Full copies of the BHPS waves 1-18 questionnaires can be downloaded from here:

https://www.iser.essex.ac.uk/bhps/documentation/pdf_versions/index.html.

Copies of the Understanding Society questionnaires can be downloaded from here:

<https://www.understandingsociety.ac.uk/documentation/mainstage/questionnaires>.

³⁶ Improved Methods for Estimating International Migration The Use of the Labour Force Survey to Improve Estimates of International In-migration - Coverage and Quality

	Understanding Society	British Household Panel Survey
Employment status	✓	✓
Sector	✓	✓
Employment history (work experience) ³⁷	✓	✓
Country of destination (for leavers)		
Country of birth	✓	✓
Citizenship	✓	✓
Age	✓	✓
Gender	✓	✓
Qualification level (educational attainment)	✓	✓
Date highest educational attainment was achieved	✓	✓
Family composition	✓	✓
Occupation	✓	✓
Marital status	✓	✓
Country of last residence	✓	✓
Moving for work in supranational institutions ³⁸	✓	✓
Moving for work in multinational companies		
Workers posted abroad		
Time of arrival	✓	✓
Wage	✓	✓

Across different waves of the British Household Panel Survey the questions individuals were asked varied year on year. There was a core set of questions asked of adults every year about the following aspects of their lives:

- make-up of the household
- housing conditions
- residential mobility
- education and training
- health and usage of health services
- employment
- socio-economic values
- income from employment, benefits and pensions

A separate module of questions was asked of young people, which included:

³⁷ Details of a respondent's employment are available for each wave of the study they took part in.

³⁸ Has SIC code data which has a classification (9900) for extraterrestrial organisations and bodies.

- attitudes to smoking and other health-related matters
- time spent watching TV and on computer
- relationships with parents and friends
- getting into trouble
- school work
- attitudes to jobs

In addition, further modules of questions were asked in particular waves³⁹ to provide more information on life history. These included questions about:

- marriage
- cohabitation
- children and parenting
- job history
- wealth and assets
- health measures
- ageing, retirement and quality of life
- neighbourhood and friends

Like BHPS, Understanding Society has a core module of questions asked in each wave as well as ad hoc modules asked varying frequency. The content plan has been designed up to wave ten, with materials already designed for waves three and four.⁴⁰ For example, ethnic identity questions are asked in waves two, five and eight and questions on Britishness asked in waves three and six.

Moreover, a specific module of questions relating to migration history in wave one of USOc, which provides a greater level of detail than is available under any other UK data source. This module asked about all the previous countries an individual has lived in and their time of (first and last) arrival in the UK.

Public Availability of Data

To access the full datasets, one must register with the UK Data Service before a download is possible, although the data is available to all researchers who can demonstrate public interest.

³⁹ Understanding Society, *British Household Panel Survey: A guide to how it is used in Understanding Society*. <https://www.understandingsociety.ac.uk/about/bhps-in-understanding-society>

⁴⁰ Understanding Society (November 2011), *Long term questionnaire content plan*. https://www.understandingsociety.ac.uk/system/uploads/assets/000/000/018/original/Long_Term_Content_Plan_Nov2011-3.pdf?1355920157

How data is used

The Understanding Society survey data is used by a range of researchers to help understand a large range of topics ranging from well-being and sports participation to employment and perceived racial discrimination.⁴¹

Weighting

British Household Panel Survey data can be used to perform longitudinal analysis. The longitudinal respondent weight (LRWGHT) selects cases who gave a full interview at all waves from the original BHPS sample. In each dataset, LRWGHT is prefixed by a letter referring to the wave it relates to, such that the longitudinal weight for wave 2 is bLRWGHT, and the weight for wave 18 is rLRWGHT. At each wave the longitudinal weights are calculated to take account of previous wave respondents lost through sample attrition. Thus the longitudinal weight at any wave will be the product of the sequence of attrition weights accounting for losses between each adjacent pair of waves as well as the initial respondent weight at wave one.⁴²

At wave 18, the final wave of BHPS, 67.5% of respondents, 9737 out of a total of 14,419, had a weight=0 (rLRWGHT). This means that only a third of respondents at wave 18 had responded in all previous 18 waves. However, this does not take into account the incremental increases in size of the BHPS sample, as householders in Northern Ireland, Scotland and Wales were added after wave 1. The BHPS reached its maximum size at Wave 11. To look at individuals who responded in each wave since wave 11, one should use the set of weights called LRWTUK1. From this we can see that 10,229 of the 14,419 respondents in wave 18, 71%, gave answers in all the previous 8 waves.

Longitudinal datasets can also be made between the two waves of Understanding Society data so far released. There are five types of longitudinal weights: enumerated persons; proxy or main interview; main interview; self-completion; and "Extra-five minutes" interview. Each is based on the corresponding Wave 1 cross-sectional weight, with an additional adjustment for non-response at Wave 2.⁴³

⁴¹ Some initial findings can be found at <https://www.understandingsociety.ac.uk/research/publications/findings/2012>.

⁴² British Household Panel Survey Taylor, Brice, Buck & Prentice-Lane, *User Manual Volume A: Introduction, technical report and appendices*.

⁴³ Understanding Society (November 2012), *UK Household Longitudinal Study: Waves 1-2, 2009-2011 User Manual*.
http://www.esds.ac.uk/doc/6614/mrdoc/pdf/6614_user_manual_waves1-2.pdf

Summary of Pros and Cons

The main advantage of the BHPS/Understanding Society datasets is that they have the ability to track individuals' working histories over time 20 years. Not only do these longitudinal studies run over an extended period of time, they also contain rich data on employment and certain personal demographic characteristics.

On the other hand, because the BHPS/USoc surveys private households, international migrants who live in communal establishments are not in the scope of the survey – this is estimated to represent approximately 14% of international migrants⁴⁴. It is also believed that migrants may be less likely to respond to the survey.

Moreover, the sample sizes available within the BHPS limits the power of analysis that can be undertaken. When the BHPS reached its maximum size (18,867) at wave 11, there were 832 respondents not born in the UK in the sample. By the final wave of the BHPS (with 14,419 respondents overall), this had fallen to 597 respondents not born in the UK. In total, there were 471 respondents included in both datasets (not all who would have been in work). On the other hand, Understanding Society, given its higher sample size and specific Ethnic Minority boost, surveys larger numbers of migrants.

Moreover, panel surveys are slower to respond to changes in makeup of the general population than other surveys, especially in terms of migration. This is because for a new migrant to enter the sample in a panel survey it requires them to move in the same residence as a current sample member, whereas for other surveys like the Labour Force Survey they are only required to move into a private UK address.

Another problem, consistent with many UK data sources is that these surveys do not however have any ability to track UK citizens who emigrate, unless they return to the same UK address.

Key sources of information

UK Household Longitudinal Study – Wave 1 Technical Report, January 2012, R.Boreham, D.Boldysevaite, National Centre for Social Research, https://www.understandingsociety.ac.uk/system/uploads/assets/000/000/010/original/Wave_1_Technical_Report.pdf?1354035538

⁴⁴ Improved Methods for Estimating International Migration The Use of the Labour Force Survey to Improve Estimates of International In-migration - Coverage and Quality

Understanding Society W2: Technical Report, August 2012, R.Boreham, NatCen.

https://www.understandingsociety.ac.uk/system/uploads/assets/000/000/011/original/Wave_2_Technical_Report.pdf?1354035554

Understanding Society: Findings 2012,

<https://www.understandingsociety.ac.uk/system/uploads/assets/000/000/024/original/Understanding-Society-Findings-2012.pdf?1355227235>

Researchers guide to using the British Household Panel Survey and other international, longitudinal datasets, Working Paper No. 5, A.Johnson, Analytical Services Directorate for the Department for Work and Pensions, <http://research.dwp.gov.uk/asd/asd5/WP5.pdf>

An Initial Look at Non-Response and Attrition in Understanding Society, Peter Lynn, Jonathan Burton, Olena Kaminska, Gundi Knies and Alita Nandi, January 2012.

Family Resources Survey (FRS)/EU-SILC

Brief description

The Family Resources Survey (FRS) was launched in October 1992 to provide the Department for Work and Pensions (DWP) with data on individuals' incomes and state support receipt, housing tenure, savings and investments, disability and caring responsibilities, occupation and employment, and pension participation.

Longitudinal v cross-sectional	Both
Coverage	All living in private addresses
Time coverage	FRS has run since 1992 and will be used to provide EU-SILC from 2012.
Sample size	25,000 households, falling to 20,000 as of 2011.

How data is collected

The Family Resources Survey uses a sample drawn from the Royal Mail's Postal Addresses File of private addresses in the UK. It uses a systematic, random sampling procedure and interviews are carried out jointly on behalf of the DWP by interviewers from ONS and NatCen Social Research. All interviews

are conducted face-to-face and the average duration of an interview in 200/11 was just under an hour and a half.

The data is collected and released annually through the UK Data Service.

Key groups captured/missed

The FRS takes its sample from the list of private households in the UK and therefore individuals living in communal establishments are excluded from the sample. This includes individuals in nursing or retirement homes, for example, meaning that figures relating to the most elderly individuals may not be representative of the United Kingdom population, as many of those at this age will have moved into homes where they can receive more frequent help.

Other groups likely to be living in communal establishments and possibly missed by the Family Resources Survey are students living in communal establishments and individuals living in hostels or caravan parks, for example. Recent migrants are likely to be over-represented in these last two groups and may be overlooked by the survey.

Sample size and composition

From April 2011 the target achieved sample size for Great Britain will reduce by 5,000 households to around 20,000 households. This sample size will be retained and completely refreshed each year. Previously to this, the Family Resources Survey sampled approximately 25,000 Great British households.

Subjects covered

	Variables Recorded
Employment status	✓
Sector	✓
Employment history (work experience) ⁴⁵	✓
Country of destination (for leavers)	
Country of birth	✓
Citizenship	✓
Age	✓
Gender	✓
Qualification level (educational attainment)	✓
Date highest educational attainment was achieved	
Family composition	✓
Occupation	✓
Marital status	✓

⁴⁵ The FRS asks details about last job an individual had before their current one, as well as asking how many years they have been employed for since they left full-time education.

Country of last residence	
Moving for work in supranational institutions	✓
Moving for work in multinational companies	
Workers posted abroad	
Time of arrival	
Wage	✓

The questionnaire is divided into three parts. The first part is the household schedule which is addressed to one person in the household (usually the household reference person, although other members are encouraged to be present) and mainly asks household level information, such as relationships of individuals to each other, tenure and housing costs. Next is the individual schedule which is addressed to each adult in turn and asks questions about employment, benefits and tax credits, pensions, investments and other income. Information on children in the household is collected by proxy from a responsible adult. A final section asks about savings and possible investment decisions.

The primary function of the Family Resources Survey (FRS) is to collect information on household income received from all sources, including wages and salaries, state benefits, payable tax credits, private (occupational and personal) pension schemes, and investments. However, in the early sections of the individual questionnaires, there are questions about country of origin, nationality and ethnicity, but nothing on when a migrant may have arrived in the UK.

The latest questionnaire instructions for the FRS can be found here: http://www.esds.ac.uk/doc/7085/mrdoc/pdf/frs_2010_11_question_instructions.pdf.

How the data is used

Since 2005, the UK has been required to gather data on income and living conditions, known as EU Statistics on Income and Living (EU-SILC). Initially this data was collected via the General Household Survey, however following a review it was decided that using the Family Resources Survey would be more appropriate. This has the benefit of ensuring that EU and national poverty estimates for the UK are based on the same source.

As of April 2012, the FRS will be used to satisfy the cross-sectional requirements of the EU-SILC survey, and the longitudinal requirements as of 2015. The delay in using the FRS to meet longitudinal requirements is due to

the fact that FRS respondents have to be followed over a number of years to create a longitudinal dataset.

The Office for National Statistics (ONS) will retain the responsibility for longitudinal EU-SILC requirements, and completed FRS interviews will form the sample frame for a follow-up survey which ONS will use to meet EU-SILC longitudinal requirements. The ONS will select half of the FRS sample as a first wave sample in a four year panel dataset. The remaining three waves will be composed of follow-up interviews to the nine thousand cases eligible for follow up from the FRS. The follow up interviews are proposed to be conducted by telephone (although ONS is exploring using a mixed mode approach) and will be limited to only those questions necessary to meet the requirements of EU-SILC. When the design is embedded the EU-SILC study will be a four wave longitudinal survey with a sample size of approximately nine thousand cases per wave (minus attrition).

Between 2012 and 2015, the longitudinal EU-SILC dataset will use data originating from both the General Household Survey and the Family Resources Survey.

Unlike general EU-SILC data however, the Family Resources Survey allows us to look at citizenship by individual country of origin as opposed to just inside or outside the EU.

Public availability of the data

Data available to registered users of the UK Data Service.

Summary of Pros and Cons

The FRS contains detailed data on incomes and state support receipt, savings and investments but relatively little information on migration and there is no way of knowing when an individual arrived in the UK.

Moreover, it is unlikely that the level of detail about different types of income will be particularly important when analysing the working lives of migrants, and other survey, including LFS/APS and BHPS/USoc as basic wage information.

Key sources of information

V. Christian (DWP) & T. Howe (ONS) (2011), *Developments to the Family Resources Survey, FRS User meeting.*

Department for Work and Pensions, (2012). *Family Resources Survey 2010/11 Report*

Department for Work and Pensions (2011), *Introduction to the Family Resources Survey 2009-2010, UK Data Archive Study Number 6886.*
http://www.esds.ac.uk/doc/6886/mrdoc/pdf/frs_2009_10_introduction_to_the_family_resources_survey.pdf

Section A: Review of dataset Availability – Administrative data

National Insurance Recording System (NIRS) and associated datasets

Brief description

The National Insurance Recording System (NIRS) is a series of databases held by HM Revenue and Customs (HMRC) containing all income tax and National Insurance contributions registered for individuals over an extended period of time. It also contains data on age, gender, nationality, country an immigrant arrived from and when but not their route of entry into the UK.

The Department for Work and Pensions (DWP) uses NIRS data in three different forms: the Lifetime Labour Market Database (LLMDB), the Migrant Worker Scan (MWS), and the Work and Pensions Longitudinal Study (WPLS).

The **Lifetime Labour Market Database** is a 1% random sample of National Insurance Recording System data, and includes some additional welfare benefit data held by DWP as well as anonymised data held on the NIRS. It is also held in multiple datasets.

The **Migrant Worker Scan** is an extract of NIRS of all adult overseas nationals who register for a National Insurance Number (NINo) in the UK. A NINo is required by overseas nationals if they intend to work or claim benefit in UK. The only data variables available from the MWS are age, gender, nationality and when and where an individual registered for a NINo.

The **Work & Pensions Longitudinal Study** (WPLS) is the collective name for a group of datasets owned by the Department for Work & Pensions (DWP). The data are contained in different tables, which, if linked, form a good record of an individual's work, benefit, and pension history. All individuals who have had contact with the Department for Work and Pensions (DWP) since 1998 are included in the datasets.

DWP analysts note that raw NIRS data requires considerable cleaning before it is useable. DWP uses of NIRS data does have some limitations, in that limited data about hours worked is available, and there is less complete coverage of self-employed workers or workers earning wages lower than the threshold to pay income tax or national insurance.

National Insurance Recording System (NIRS)

Longitudinal v cross-sectional	Longitudinal
Coverage	All residents with a NINo
Time coverage	NIRS1 started in 1948, NIRS2 started in 1997
Sample size	UK resident population

In the UK, if an individual wants to register to work legally or claim benefits they must have a National Insurance Number (NINo). Whereas people born in the UK are automatically given a NINo just before they turn 16 years of age, immigrants typically apply for a NINo when they start interacting with the state, either by paying taxes or claiming welfare benefits.

NIRS comprises a series of databases which are held by HM Revenue and Customs (HMRC) and hold comprehensive longitudinal data on all residents' interaction with the labour market. NIRS includes data on some, but not all, welfare benefits.

There have in fact been 2 NIRS databases. NIRS1 ran from 1948 to 1995/6, and NIRS2 started in 1996/7 until the present day. In 1997, a migration exercise was performed when data was transferred from NIRS1 to NIRS2, with only cases meeting certain criteria migrated. The archiving exercise means that the new NIRS2 will be deficient of records prior to 1984 for certain groups, primarily older women.⁴⁶ A number of problems were experienced during this process.

Because individuals need to produce their NINo in every interaction with the system, NIRS effectively tracks individuals throughout their working lives. It has high levels of accuracy and relatively low levels of attrition (data is only not collected when individuals neither pay taxes nor claim any benefit nor receive retirement pension for more than 12 months).

However, NIRS data is not collected specifically to understand employment, but for tax purposes. DWP staff interviewed as part of this project said that considerable data cleaning and processing was required before NIRS data could be used to understand labour market issues as part of the Work and

⁴⁶Department for Work and Pensions (January 2013), *Methodology Statement for the Department for Work and Pensions L2 Publications*.
http://statistics.dwp.gov.uk/asd/l2_methodology.pdf

Pensions Longitudinal Study, the Lifetime Labour Market Database, or the Migrant Worker Scan. As such, data processed by the DWP is more likely to be of use for the purposes of tracking the working lives of migrants.

Moreover, different tables of information have particular limitations; for example, some elements of NIRS do not cover self-employed workers, and some employees earning below the threshold to pay income tax are not covered (particularly those working fewer than 16 hours a week). As NIRS forms the basis of other administrative datasets explained in this section, these limitations of NIRS also apply to the following datasets.

Given the scale of the NIRS databases, it is not always clear which variables are included. However, it is clear that qualification is not available.

Variables Recorded	NIRS
Employment status	✓
Sector	
Employment history (work experience)	✓
Country of destination (for leavers)	
Country of birth	
Nationality	✓
Age	✓
Gender	✓
Qualification level (educational attainment)	
Date highest educational attainment was achieved	
Family composition	
Occupation	
Marital status	
Country of last residence	✓
Moving for work in supranational institutions	
Moving for work in multinational companies	
Workers posted abroad	
Time of arrival	✓
Wage	✓

The Migrant Worker Scan (MWS)

Longitudinal v cross-sectional	Cross-sectional
Coverage	All overseas nationals NINo registrations
Time coverage	From January 2002.
Sample size	c.650,000 individuals register a year

The Migrant Worker Scan is an extract of NIRS provided by HMRC to the Department of Work and Pensions (DWP), including basic data for all overseas nationals who register for a NINo. The dataset contains data on age, gender, nationality, and address and date of registration. In the year to September 2011, 690,000 overseas nationals registered for NINOs an increase of 11% on the year to September 2010.⁴⁷ The MWS is a cumulative extract that was first created in 2000.

The MWS is the basis of a DWP tabulation tool used to show where new migrants register for NINOs, found here:

http://research.dwp.gov.uk/asd/asd1/niall/index.php?page=nino_allocation.

The data is released by year of registration onto the system, although there is no method of knowing whether an individual registered immediately on entry to the UK or after an indeterminate period of time, and displayed in aggregated tables. The MWS has no information about employment or length of stay, or any information about non-migrant workers. As such it is primarily of use in measuring the flow of migrants to different local areas. It could also be used to append a migrant flag to other datasets which contain NINo identifiers.

⁴⁷ Office for National Statistics, (February 2012), *Migration Statistics Quarterly Report February 2012*. <http://www.ons.gov.uk/ons/rel/migration1/migration-statistics-quarterly-report/february-2012/msqr.html>

Variables Recorded	MWS
Employment status	
Sector	
Employment history (work experience)	
Country of destination (for leavers)	
Country of birth	
Nationality	✓
Age	✓
Gender	✓
Qualification level (educational attainment)	
Date highest educational attainment was achieved	
Family composition	
Occupation	
Marital status	
Country of last residence	
Moving for work in supranational institutions	
Moving for work in multinational companies	
Workers posted abroad	
Time of arrival	✓
Wage	

The Lifetime Labour Market Database (LLMDB or L2)

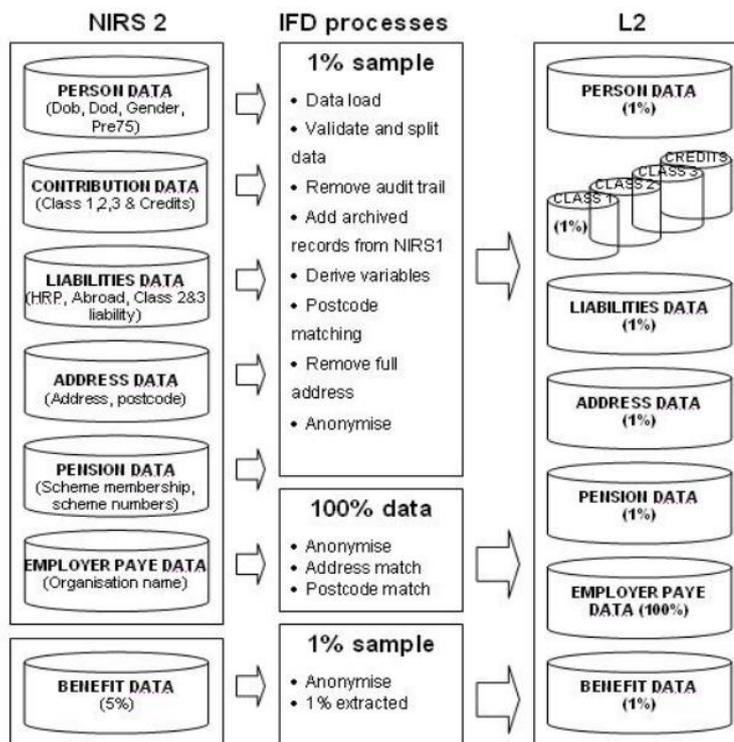
Source	LLMDB
Longitudinal v cross-sectional	Longitudinal
Coverage	1% sample of all with a NINo
Time coverage	Summaries from 1948-74, Detailed 1975 onwards
Sample size	Over 700,000 individuals

The Lifetime Labour Market Database (known as LLMDB or L2) is a random 1% sample of individuals with a National Insurance Number (NINo), comprising all people with a NINo whose last two digits are 14 (the last 2

digits can be 00-99). NIRS information is provided by HMRC to the DWP and this forms the basis of the LLMDB. As the same sample is drawn each year, it can be used to track up to 700,000 individuals over varying periods of time.

The DWP clean the NIRS data, and append data some additional DWP data, in particular, data about certain 'non-contributory' welfare benefits that are not covered by NIRS. Moreover, some additional employer data is added. The chart below visualises this process.⁴⁸

Methodology Statement for the Department for Work and Pensions L2 publications



Lifetime Labour Market Database (LLMDB) extracts are taken in January and May and generally take around 2 months to load and process. The L2 extract is generally provided at least 9 months after the last tax year being analysed.

It is thought about 10% of the sample, approximately 70,000 individuals, are likely to be migrants.⁴⁹

⁴⁸ For a detailed explanation of how the LLMDB is built see the annex of: Department for Work and Pensions (January 2013), *Methodology Statement for the Department for Work and Pensions L2 Publications*: http://statistics.dwp.gov.uk/asd/l2_methodology.pdf

⁴⁹ Sara Lemos, University of Leicester (October 2011), *Mind the gap? A detailed picture of the immigrants-native earnings gap in the UK using longitudinal data between 1978 and 2006*. <http://www.le.ac.uk/ec/research/discussion/documents/dp11-38.pdf>

The LLMDB tables contain well over 400 variables, including date of birth, date of death, age, gender, address, nationality, country which immigrants arrived from, immigrants' entry date, immigrants' age at entry, number of jobs in the year, annual earnings per job, type of employment (employee or self-employed), number of weeks employed (unemployed) in the year, dates of spells of unemployment, dates of spells of receipt of benefits, benefit type, pension contributions, pension entitlements.

The LLMDB does not contain information on education, or the immigrants' entry route (work permit, student visa, family reunification, etc.), on their household status or on their departure date. It also does not include occupational or sector data about employment, although DWP staff considered that it may be possible to append this in future.

When an individual leaves the sample, (i.e., they don't claim benefits or pay tax for 12 months) it is impossible using this dataset to distinguish whether or not an individual has left the labour market or left the country.⁵⁰

⁵⁰ Sara Lemos, University of Leicester (October 2011), *Mind the gap? A detailed picture of the immigrants-native earnings gap in the UK using longitudinal data between 1978 and 2006.* <http://www.le.ac.uk/ec/research/discussion/documents/dp11-38.pdf>

Variables Recorded	LLMDB/L2
Employment status	✓
Sector	*
Employment history (work experience)	✓
Country of destination (for leavers)	
Country of birth	
Nationality	✓
Age	✓
Gender	✓
Qualification level (educational attainment)	
Date highest educational attainment was achieved	
Family composition	
Occupation	
Marital status	
Country of last residence	✓
Moving for work in supranational institutions	
Moving for work in multinational companies	
Workers posted abroad	
Time of arrival	✓
Wage	✓

*It is unclear whether sector is included in the database. DWP analysts thought not, although one document suggest it does.⁵¹

A summary of DWP usage of LLMDB can be found here:

http://statistics.dwp.gov.uk/asd/l2_methodology.pdf.

⁵¹ Paul Jones & Peter Elias, National Data Strategy (October 2006), *Administrative data as research resources: A selected audit, Draft version 2.0*, <http://www.rss.org.uk/uploadedfiles/documentlibrary/864.pdf>

The Work and Pensions Longitudinal Study (WPLS)

Longitudinal v cross-sectional	Longitudinal
Coverage	All those having some engagement with the DWP. More information is available in England and Wales than Scotland and Northern Ireland.
Time coverage	Those involved in the system since 1998, with some details from 1975.
Sample size	In 2005, it stored details on 27 million individuals.

The WPLS, introduced in January 2005, is the collective name for a group of datasets owned by the Department for Work & Pensions. It has the ability to link social security benefit and welfare-to-work programme information held by DWP, with NIRS employment records and Pay As You Earn (PAYE) tax details from HMRC. It stores details on the labour market status of all individuals who have had contact with the DWP from 1998 onwards. The linking of data within the WPLS databases is done using National Insurance Numbers (NINOs) and a range of other variables relating to individual characteristics using fuzzy matching algorithms.⁵²

There are five primary databases included in the WPLS:

- The National Benefits Database (NDB). This database has 1 row per NINO and shows different spells on different social security benefits. Data is available from 1999 until present day.
- Data based on P45/P46 forms,⁵³ which employees are given when they start and cease being employed by a company. This provides data about spells in employment, but no data about types of work (sector, occupation, hours) or any data about self-employment. Data is available from 1998 until present day.

⁵² For an example of the type of fuzzy matching algorithms the DWP uses, see Samantha Poole, Department for Work and Pensions, *'Fuzzy' matching algorithm and steps to develop it*.

<https://docs.google.com/viewer?a=v&q=cache:m0uqt6pqunsJ:www.ons.gov.uk/ons/about-ons/user-engagement/events/past-events/fifteenth-gss-methodology-conference--1-july-2010/samantha-poole.pdf+&hl=en&gl=uk&pid=bl&srcid=ADGEESH2xVoMtNYDPcnbXBhJvSxXAnJoTiMJn4u0WJgkAZYqQTB-0k9pFefL7e04OIEmW0aMr4ZfH6xX5xGQR0Es3aefwIVXJff5PI8maeg7-I-TtTzcfDI31eJTDZv4GvCJksnYSd1&sig=AHIEtbQI7bdNfh2mJu9Yo5vdwSX6TlvpBA>

⁵³ For more information see: <https://www.gov.uk/payee-forms-p45-p46-p60-p11d/p45>

- Data based on P60 forms,⁵⁴ which employees are given at the end of every financial year. This provides details of income over a financial year, and can be (imperfectly) matched to spells of employment reported in P45/P46 forms. Again, this data does not include data about sector, occupation, or hours of work or any data about self-employment. Data is available from 2003 until present day.
- Data based on Tax Credit payments. Working Tax Credits are paid to people working in low income work of 16 hours per week or more. Data supplied to DWP therefore includes income, hours per week, and also includes data for some self-employed people. Data is available from 2003 until present day.
- Self Assessed (including self employed) data. This includes some information about wages of self-employed people, and their business costs. Data is available from 1998 until present day.

In addition, data held on DWP's Labour Market System (LMS), including details engagement with the DWP employment service (Jobcentre Plus), and privately provided Welfare to Work programmes could be matched to the WPLS.

The main WPLS datasets do not include a migrant marker flag. However, this could be appended to the WPLS from the Migrant Worker Scan (assuming legal clearances to do this are in place).

In 2005, the WPLS stored data on 27 million individuals, which includes around:

- 34 million separate benefit claims;
- 5 million periods on New Deals or other work-related activity;
- Around 3 million Pension Credit claims;
- Nearly 13 million periods on State Pension (RP);
- 2 million Attendance Allowance spells;
- 3 million spells on DLA; and
- 143 million separate linked employment records (and a further 146m unlinked HMRC employment records)

⁵⁴ For more information see: <https://www.gov.uk/payee-forms-p45-p46-p60-p11d/p60>. Whilst P60 forms are provided to workers, the same information is submitted to HMRC by employers in P14 forms.

It is not clear exactly what data is available in the WPLS, but the table below presents data that definitely can be derived. Other data can be matched to the WPLS from other sources: for example, more demographic variables of those who had had dealings with Jobcentre Plus could be appended from the LMS, and a migrant marker could be appended from the Migrant Worker Scan. Qualification level is not included in the dataset, and nor is sector or occupation data.

	Variables Recorded
Employment status	✓
Sector	
Employment history (work experience)	✓
Country of destination (for leavers)	
Country of birth	
Citizenship	
Age	✓
Gender	✓
Qualification level (educational attainment)	
Date highest educational attainment was achieved	
Family composition	
Occupation	
Marital status	
Country of last residence	
Moving for work in supranational institutions	
Moving for work in multinational companies	
Workers posted abroad	
Time of arrival	
Wage	✓

In summary, the Work and Pensions Longitudinal Study provides for every single DWP client:

- The Benefit(s) or Pension they first claim and any subsequent ones.
- Any help or interventions they receive from Jobcentre Plus, e.g. referrals to a job interview, entry to a New Deal Programme, etc.;

- Whether they go into work when they leave benefit;
- Whether they return to benefit – work and benefit history;
- Key information about their income, including savings and investments;
- Key personal details, e.g. age, sex, where they live, and ethnicity.
- Housing/Council Benefit and Tax credits for DWP clients.

Public availability of the data

Access to these data sources is extremely tightly controlled, even for DWP/HMRC analysts. Every use of administrative data within DWP requires a business case agreed at very senior levels and/or approval by an ethics committee, and legislation is often needed for different Government Departments to share their data.

To gain access to more useable, DWP cleansed NIRS data would require approval both from DWP and HMRC. Within DWP, a sponsor working in a policy (not statistics or research) role would be required.

It is likely that microdata would only be available on DWP/HMRC premises. HMRC have a datalab in London⁵⁵ but income tax data is not available at the datalab (some self-assessment data for the self-employed and Tax Credit data is available). The deadline for the next round of applications is 14 June 2013. Similarly, DWP staff considered it unlikely that data could be used off-site or (particularly) outside of the UK.⁵⁶

In fact, within the UK, there is growing concern that whilst other European countries are making better use of the rich administrative data sources they hold, the UK is too restrictive and opportunities for research are being missed. As such, the Administrative Data Taskforce (ADT) was formed in late 2011 by the Economic and Social Research Council (ESRC), the Medical Research Council (MRC) and Wellcome Trust, and is working with Government to improve data availability.⁵⁷

If it was decided to try to use administrative data for this project, a detailed written application would have to be submitted to DWP and HMRC.

⁵⁵ See <http://www.hmrc.gov.uk/datalab/about.htm> for details.

⁵⁶ More details about access to WPLS data can be found here: http://statistics.dwp.gov.uk/asd/longitudinal_study/index.php?page=ic_longitudinal_study#sa

⁵⁷ http://www.esrc.ac.uk/_images/ADT-Improving-Access-for-Research-and-Policy_tcm8-24462.pdf

Summary of Pros and Cons

Administrative data based on income tax and national insurance contributions are the most powerful way of tracking the working lives of migrants within the UK. Analysis of NIRS would allow comprehensive analysis of the migrant workers labour market situation.

However, there are significant problems with each of these sources. NIRS data held by HMRC, though comprehensive for all people in work in the UK, is 'raw' and significant cleaning would be required to make it useable. DWP has done such cleaning as part of the Lifetime Labour Market Database and the Work and Pensions Longitudinal Study.

The WPLS has very wide coverage (at least 27,000,000 people as of 2005), but is seen as very unwieldy, and would require data from a number of databases to be matched to create a useable dataset. Moreover, the WPLS does not have a migrant marker, and it is unclear whether permission would be given to append this from another source (the Migrant Worker Scan).

The Migrant Worker Scan is primarily an instrument to measure migration flows, and does not hold information on either non-migrant workers, or the employment of migrant workers.

The LLMDB is likely to be the most practical source of administrative data. Whilst it is held in a number of databases, it is less unwieldy than the WPLS. It also has longitudinal data for around 700,000 people, roughly 10% of which are migrants.

However, access to these datasets is extremely restricted, and it would prove extremely difficult to gain access to the data.

Key sources of information

Paul Jones and Peter Elias (December 2006), *Administrative data as a research resource: a selected audit*

<http://eprints.ncrm.ac.uk/452/1/Admin%2520Data%2520selected%2520audit%2520Dec06.pdf>

Sara Lemos, University of Leicester, (October 2011). *Mind the Gap: What Gap? A Detailed Picture of the Immigrant-Native Earnings Gap in the UK using Longitudinal Data between 1978 and 2006*,

<http://www.le.ac.uk/ec/research/discussion/documents/dp11-38.pdf>

Richard Dickens and Abigail McKnight, Centre for Analysis of Social Exclusion (October 2008). *Changes in earnings inequality and mobility in Great Britain 1978/9-2005/6* <http://sticerd.lse.ac.uk/dps/case/cp/CASEpaper132.pdf>

Richard Dickens and Abigail McKnight, Centre for Analysis of Social Exclusion (October 2008). *Assimilation of Migrants into the British Labour Market* <http://eprints.lse.ac.uk/28244/1/CASEpaper133.pdf>

Office for National Statistics, (May 2009) *Interim report on the potential use of Department for Work and Pensions data to improve population and migration statistics*. http://research.dwp.gov.uk/asd/asd1/niall/ons_paper.pdf

Office for National Statistics (October 2009), *Feasibility report on the potential use of record level Migrant Workers Scan data to improve migration and population statistics*.

Department for Work and Pensions. (December 2010). *Work and Pensions Longitudinal Study – List of uses*. http://research.dwp.gov.uk/asd/longitudinal_study/WPLS_Uses.pdf

The Department for Work and Pensions, Neil McIvor, 'The Work and Pensions Longitudinal Study' A presentation. <http://www.eurim.org.uk/activities/pi/060223talk.pdf>

National Statistics, Department for Work and Pensions, *The modernisation of the DWP's data sources and statistical publications*. http://statistics.dwp.gov.uk/asd/asd1/stats_consultation/modstats.pdf

Helen Barnes, Elisabeth Garratt, David McLennan and Michael Noble, Department for Work and Pensions. *Understanding the worklessness dynamics and characteristics of deprived area*. <http://research.dwp.gov.uk/asd/asd5/rports2011-2012/rrep779.pdf>

Department for Work and Pensions (January 2012), *Nationality at point of National Insurance number registration of DWP benefit claimants*. <http://webcache.googleusercontent.com/search?q=cache:3jvfc7rdQD4J:www.dwp.gov.uk/docs/mb1.pdf+when+did+the+work+and+pensions+longitudinal+study+data+back+to&cd=25&hl=en&ct=clnk&gl=uk>

Office for National Statistics, (February 2012), *Migration Statistics Quarterly Report February 2012*. <http://www.ons.gov.uk/ons/rel/migration1/migration-statistics-quarterly-report/february-2012/msqr.html>

Worker Registration Scheme (WRS)

Following the accession of the A8 countries⁵⁸ to the European Union, the UK allowed citizens of those countries relatively unrestricted access to its labour market. The Government did, however, temporarily put in place the Worker Registration Scheme (WRS) in order to regulate and measure access to the labour market and measure eligibility for benefits.⁵⁹

The scheme ran from accession in 2004 until April 2011 and was administered by the UK Border Agency (UKBA), the Department for Work and Pensions (DWP), HM Revenue and Customs (HMRC) and The Department for Communities and Local Government (DCLG).

Longitudinal v cross-sectional	Cross-sectional
Coverage	A8 workers excluding self-employed
Time coverage	May 2004 – April 2011
Sample size	Over 1.1 million workers registered over the course of the scheme ⁶⁰

Details of the scheme

During the period in which the Worker Registration Scheme was in place, individuals from A8 countries were allowed to take up work with any employer in the UK, but were required to register with the scheme and pay a registration fee to the Home Office. Upon registration, applicants had to provide information about their employer (which was used to derive information on sector of employment and locality), occupation and wages, planned duration of stay, details of dependants and a basic demographic profile including nationality and age. A8 nationals were required to re-register for subsequent jobs (paying a fee each time they changed jobs or took up an additional job) until they had worked a total of 12 out of 13 months.

Whilst all employees were required to register on the scheme, it was thought that the registration fee would put off a number of workers, and there is no

⁵⁸ A8 countries are Czech Republic, Hungary, Estonia, Latvia, Lithuania, Poland, Slovakia & Slovenia.

⁵⁹ Those from A8 countries moving to the UK could access Housing and Council Tax Benefit, but had to prove they had been employed for at least 12 months to claim the main unemployment benefit, Job Seeker's Allowance, and thus register on the scheme.

⁶⁰ Home Office, UK Border Agency Data (10 March 2011), 'Closure of the Worker Registration Scheme', <http://www.ukba.homeoffice.gov.uk/sitecontent/newsarticles/2011/march/24-wrs-closure>

way of knowing how many A8 workers did not register. For example, a local primary survey in Fife found that a quarter of employed A8 migrants were not registered under the WRS and 13% did not have a National Insurance Number (which is required to work in the UK).⁶¹

Self-employed workers were exempt from registering with the Worker Registration Scheme.

Size and composition

The data all pertains to workers from the A8 countries who came to the UK to find work. Between May 2004 and the end of 2010, over 1.1million A8 nationals had registered to work via the Worker Registration Scheme.⁶²

Subjects covered

	Variables Recorded
Employment status	✓
Sector	✓
Employment history (work experience)	
Country of destination (for leavers)	
Country of birth	
Citizenship	✓
Age	✓
Gender	✓
Qualification level (educational attainment)	
Date highest educational attainment was achieved	
Family composition ⁶³	✓
Occupation	✓
Marital status	
Country of last residence	
Moving for work in supranational institutions	
Moving for work in multinational companies	
Workers posted abroad	
Time of arrival	
Wage	✓

⁶¹Fife Research Co-ordination Group, (2007) *Migrant workers in Fife – Survey*.

⁶² Home Office, UK Border Agency Data (10 March 2011), 'Closure of the Worker Registration Scheme', <http://www.ukba.homeoffice.gov.uk/sitecontent/newsarticles/2011/march/24-wrs-closure>

⁶³ Has data on number of dependants only.

The scheme asks workers to self-report the number of hours they worked, the wage level they worked at, as well as whether the job was temporary or permanent.

Public availability of the data

The data is only available in aggregate tables⁶⁴ and no individual level data is available.

Summary of Pros and Cons

The main use of the Worker Registration Scheme was to measure and understand the number of A8 nationals who came to the UK as well as restricting some benefits, but the data gives us a detailed picture of the work the A8 migrants undertook.

However, there are a number of drawbacks to the WRS data series. First, it can only measure in-flows of migrant workers to the UK and even then, it is likely to under-estimate the number of immigrants from A8 countries. This is because some workers did not register and were working illegally. Moreover, the self-employed were excluded from the scheme altogether, and it is unclear how many migrant workers were self-employed.

Furthermore, workers were asked how long they planned to stay in the UK when they registered, but they may have changed their plans, and stayed for a longer or shorter period of time. Moreover, there was no requirement to de-register when an individual left the UK, so the data cannot accurately measure the stock of migrant workers at any one time.

Key sources of information

A. Green, D.Owen & D.Adam, (2008) *A resource guide on local migration statistics, for the Local Government Association.*

UK Border Agency: *Accession Monitory Report, A8 countries, May 2005 – March 2009.*

Dr Alessio Cangiano, ESRC Centre on Migration Policy and Society (COMPAS), (2010). *UK Data Sources on International Migration and the Migrant Population: A Review and Appraisal.*

⁶⁴ See EEA immigration statistics data tables (<https://www.gov.uk/government/publications/european-economic-area-data-tables-immigration-statistics-april-june-2011>)

McCollum, Cook, Chiroro, Platts, MacLeod, Findlay. Centre for Population Change Working Paper, (2012). *Spatial, sectoral and temporal trends in A8 migration to the UK 2004-2011: Evidence from the Worker Registration Scheme.*

Fife Research Co-ordination Group, (2007), *Migrant workers in Fife – Survey.*

Gov.uk, *European Economic Area data tables, Immigration Statistics.*

Work permits and the 'Points Based System'

A work permit is required by all non-European Economic Area nationals if they want to work in the UK. Currently, the Home Office offers different types of work permits through the Points Based System (PBS). The Points Based System was designed to meet the UK's need for migrant workers through a more simple and transparent scheme than operated previously.

Longitudinal v cross-sectional	Cross-sectional
Coverage	Those from outside EEA who apply to work
Time coverage	2008 – present, a more complex work permit system ran before this
Scale	In 2012, a total of 300,000 individuals were given visas under the PBS, split evenly between in-country extensions and out of country visas. ⁶⁵

The Points Based System provides visas through different 'tiers':

- Tier 1 - for highly skilled workers, such as scientists and entrepreneurs
- Tier 2 - for skilled workers with a job offer, such as teachers and nurses
- Tier 4 - for students
- Tier 5 - for temporary workers, such as musicians coming to play in a concert, and participants in the youth mobility scheme.

Tier 3 is currently suspended.

⁶⁵ Immigration statistics data tables: <http://www.homeoffice.gov.uk/publications/science-research-statistics/research-statistics/immigration-asylum-research/immigration-tabs-q4-2012/work-q4-2012-tabs?view=Binary>

There are different conditions of entry and assessment criteria for different 'tiers' of workers. For example, highly skilled workers have more lenient entry conditions than low skilled workers. Points are awarded according to highest qualification level held, English language skills and proof of maintenance funds.

The data on visas offered is only displayed in tables and no individual-level statistics are publicly available. However the tables do show total visa applications accepted, split by whether the applicant was made from outside the country or was an in-country extension of stay application, and by the 'tier' i.e. skill level of applicant, broken down by certain sub-groups.

A key point to make here is that under 'tier 2' (skilled workers with a sponsored job offer), there is a subsection for 'Intra Company Transfers,' which provides some information about the scale of migration of people working in multinational firms.⁶⁶

Availability of the data

Aggregate data is released quarterly and outlines numbers of out-of-country visas and in-country grants of extension issued for each category (tier) of worker. The most recent data available can be found on the Home Office website.⁶⁷ However, the data is not individualised and only provides aggregate level data in table form.

Summary of Pros and Cons

The main advantage of this dataset is that it provides information about the number of people from outside the EEA working in the UK, and that it gives a sense of workers moving between countries but staying at the same company.

However, it only covers individuals from non-EEA nations and has no data on departures from the UK. Moreover, not all those issued with a visa will actually use it and some may never enter the UK.

⁶⁶ For more details see: <http://www.ukba.homeoffice.gov.uk/visas-immigration/working/tier2/ict/>

⁶⁷ Immigration statistics data tables: <http://www.homeoffice.gov.uk/publications/science-research-statistics/research-statistics/immigration-asylum-research/immigration-tabs-q4-2012/work-q4-2012-tabs?view=Binary>

Key sources of information

Home Office UKBA for the National Audit Office (2011). *Immigration: the Points Based System – Work Routes*.

Alasdair Murray, Centre Forum. *Britain's point based migration system*.

Home Office, UK Border Agency, *Quick guide to the points-based system*.
<http://www.ukba.homeoffice.gov.uk/business-sponsors/points/quick-guide-pbs/>

Gov.uk, *Immigration statistics, data tables*.

The Seasonal Agricultural Workers Scheme (SAWS)

The Seasonal Agricultural Workers Scheme (SAWS) is a quota based scheme currently used to manage the number of Bulgarian and Romanian nationals taking up seasonal work in agriculture in the UK.

Longitudinal v cross-sectional	Cross-sectional
Coverage	Currently Bulgarian & Romanian nationals taking up seasonal agricultural work.
Time coverage	Introduced after World War II
Scale	21,250 places in 2012 and 2013

SAWS is a quota based scheme used to manage numbers of agricultural workers brought in to fill jobs the UK labour market is unable to. Farmers and growers who participate in the scheme can employ (for up to six months) a fixed number of overseas workers through the scheme each year. In 2012 and 2013 the quota is 21,250 places although this has fluctuated over time. The scheme is operated by nine operators, who are each given a fixed number of work cards per year to give to low-skilled workers who they are required to pay the Agricultural Minimum Wage and provide accommodation to.

The scheme, introduced immediately after WWII, is intended to address labour shortages in the UK by providing work permits to those who would otherwise not have the right to work in the UK. The scheme is now exclusively open to citizens of Bulgaria and Romania, although it previously had wider eligibility.

The quota was set at around 10,000 during the 1990s, but increased to 25,000 by 2003. In 2002, 50 per cent of SAWS workers were from the EU Accession countries, but after the accession to the EU of the A8 countries⁶⁸, the national origins of workers shifted to countries further east (e.g. to parts of the former USSR). With the expansion of the EU in May 2004, the quota was cut. The overall quota for 2007 was 16,250, of which 40 per cent (6,500) was reserved for citizens of Bulgaria and Romania (who did not have to be students), with the remainder allocated to students from non-EEA countries. From January 2008 the scheme was exclusively for citizens of Bulgaria and Romania. This was part of the seven-year transition period post-2007 (when Romania and Bulgaria joined the EU) where EU members were allowed to restrict access to their labour market to workers from the two countries.

Availability of the data

Publicly available data is limited to the country of origin of those taking up quota places, and which month they received a permit.⁶⁹ Local data from SAWS is available via a Freedom of Information request to the Home Office, who are responsible, via the UK Border Agency, for collating the data.

Summary of Pros and Cons

As an administrative data source, it has comprehensive coverage of all those who applied to work under the Seasonal Agricultural Workers Scheme although this is only a small sub-section of total migrants working within the UK. Given these migrants are more likely than others not live in private households (and more likely to live in communal establishments, such as hostels) they are unlikely to be surveyed in the LFS/APS or Understanding Society, therefore this is a useful source about these workers. However, very limited data is available.

Key sources of information

A. Green, D.Owen & D.Adam, (2008). *A resource guide on local migration statistics, for the Local Government Association*

Home Office, UK Border Agency, *Seasonal Agricultural Workers Scheme*.
<http://www.ukba.homeoffice.gov.uk/eucitizens/bulgaria-romania/saws/>

European Economic Area data tables, Immigration Statistics. [Link here](#)

⁶⁸ A8 countries – Latvia, Lithuania, Czech Republic, Slovakia, Poland, Slovenia, Hungary & Estonia.

⁶⁹ Gov.uk, *Immigration Statistics: January to March 2012, European Economic Area*. [Link here](#)

The Sectors Based Scheme (SBS)

Since 2001, the Sector Based Scheme has been used to recruit people who are not otherwise eligible to work in the UK to work in the food manufacturing sector, in which the UK experiences a skills shortage.

Longitudinal v cross-sectional	Cross-sectional
Coverage	Those working in specific jobs.
Time coverage	2002-present
Scale	Quota at 3,500 per year.

SBS positions are only available for positions that cannot be filled by UK residents. Each position must be advertised through by the UK Government (through Jobcentre Plus (JCP)) and European Employment Service (EURES). If resident applications are received, the SBS application must include details of why each individual has not been accepted. Only once this and other conditions around genuine nature of the post and certain employment conditions (for example that the job is full-time) are fulfilled will an SBS application be accepted and a letter of approval issued.

The accession of the new European Union member states in 2004 brought in a larger than expected wave of immigration to the UK. In response, the government decided to phase out the SBS by 31 December 2006. However, with the accession of Bulgaria and Romania to the EU on 1 January 2007, the government decided to extend the program to nationals of these countries. As such, since 2007 the scheme has been restricted to Bulgarian and Romanian nationals.

At present, the quota for SBS posts is 3,500 per year, but the threshold has not been met in recent years. The following table shows the number of individuals approved under the SBS since 2007⁷⁰.

Year	Sector Based Scheme applications approved
2007	1,407
2008	1,569
2009	775
2010	601
2011	787
2012	330

⁷⁰ Gov.uk, *Immigration Statistics: January to March 2012, European Economic Area*. [Link here](#)

Availability of data

Data is released showing how many individuals have been approved under the SBS each quarter, and their nationality. No individualised data is publicly available⁷¹.

Summary of Pros and Cons

The SBS is a tiny scheme, and of limited use in understanding migrant workers in the UK.

Key sources of information

Work Permits, Sector Based Scheme, UK Border Agency, 01 March 2010

Work permits, Sector Based Scheme, Guidance for Employers, from Home Office UK Border Agency. March 2010.

European Economic Area data tables, Immigration Statistics. [Link here](#)

⁷¹European Economic Area data tables, Immigration Statistics, <https://www.gov.uk/government/publications/data-tables-immigration-statistics-october-to-december-2012>

Section B: Ability of survey and administrative data to track migrant workers

The United Kingdom collects data from a wide number of sources on the state of the UK population at certain points in time. However, there are far fewer sources that track individuals over extended time periods. There is also a much greater focus on immigrant data than there is for emigrants and at present the UK cannot track workers once they have left the UK and moved abroad.

Who is a migrant?

There are a number of different definitions of a 'migrant,' and different sources mark migrants in different ways.

The most common ways of marking migrant workers is by country of birth (with those born outside of the UK being counted as migrants), and nationality (with non-UK citizen counted as migrants). Neither of these are perfect. British citizens may be born outside of the UK but raised within the UK, and non-UK citizens could be born and raised in the UK. Nevertheless, either of these markers, particularly combined with data about when the person entered the UK, are workable definitions.

The IPS, whilst including data on whether respondents are UK passport holders or not, defines migrants not by nationality or country of birth, but by whether they were resident in another country for the previous 12 months, and expecting to be resident in the UK for the next 12 months. This definition of migrants therefore can include UK and non-UK citizens, and those born within and outside of the UK.

Short and long term movers

Administrative and most survey sources do not discriminate between long term and short term migrants, but long-term migrants can be isolated by selecting those who entered the UK more than a year ago. A new question was added to the 2011 Census to allow this analysis, and it is also possible in the LFS/APS and USoc/BHPS.

The IPS does explicitly differentiate between long and short term migrants. The IPS defines immigrants/emigrants as those who report they are

entering/leaving the UK for 12 months or more. Although anyone leaving is asked how long they will stay away, the reason of their trip (including working or studying) and the date of their last arrival in the UK, only immigrants/emigrants are asked the more detailed module of questions about their migration.

Tracking foreign workers within the UK

The most effective data source for longitudinal analysis of the working lives of migrants in the UK is the **National Insurance Recording System**, which provides comprehensive data on all tax and national insurance payments made by all people working legally in the UK. NIRS is held by HM Revenue and Customs (HMRC), and needs considerable processing before it is useable.

The Department for Work and Pensions (DWP) holds a number of datasets using processed and cleansed NIRS data. The **Work and Pensions Longitudinal Study** (WPLS) holds similar data matched with data from the DWP. However, the WPLS does not have a migrant marker (although it may be possible to append one). The **Lifetime Labour Market Database** (LLMDB) is a 1% sample of NIRS held by the Department for Work and Pensions (DWP), and is consequently large enough to conduct robust analysis of the working lives of migrants. The LLMDB has no data variable for qualification level but does have comprehensive labour market data as well as data on nationality, country of birth and when an individual started engaging in work or the benefits system. It allows us to track migrants' earnings and employment activity over an extended period of time and, unlike the important longitudinal survey data, like the BHPS/USoc, it allows analysis of migrants arriving in more recent years.

Unfortunately, access to these data sources is extremely tightly controlled, even for DWP/HMRC analysts. Every use of administrative data within DWP requires a business case agreed at senior levels and/or approval by an ethics committee, and legislation is often needed for different Government Departments to share their data. In fact, within the UK, there is growing concern that whilst other European countries are making better use of the rich administrative data sources they hold, the UK is too restrictive and opportunities for research are being missed. As such, the Administrative Data Taskforce (ADT) was formed in late 2011 by the Economic and Social

Research Council (ESRC), the Medical Research Council (MRC) and Wellcome Trust, and is working with Government to improve data availability.⁷²

Given the restrictive access policies to large administrative datasets, it is likely that longitudinal survey datasets will be the best way of tracking the working lives of migrants. **Understanding Society** (USoc), and its forerunner the **British Household Panel Survey** (BHPS), do provide the opportunity for longitudinal analysis. Previous BHPS respondents are included, and easily identifiable, in the USoc sample (starting from wave two of USoc) allowing us to chart the working lives of some migrants over the space of two decades. The addition of the Ethnic Minority boost sample in Understanding Society will in future support more powerful analysis by increasing the number of migrants in the sample.

These are the best sources of information, but each has their own downsides. The lack of key data, such as qualification levels from the LLMDB restricts the analysis that can be undertaken, and access to the data is severely restricted. With a panel survey such as Understanding Society or the British Household Panel Survey, migrant respondents is relatively stable over time; the only way migrants could enter the BHPS sample was if they became resident in the same household as a BHPS sample member. The same rule applies to the Understanding Society survey. Nevertheless, this means that the profile of migrants in panel surveys is likely to lag significantly behind the migrant population, particularly in light of the recent expansions of the EU.

Furthermore, as with all surveys, even though responses are weighted to match the population of the UK, there are concerns that certain groups of migrants are underrepresented, for example, because of lower response rates amongst those with poorer English. Moreover, there was more attrition among ethnic minority respondents to USoc than in the general sample. Finally, given USoc/BHPS are household surveys, migrant workers living in communal establishments, such as hostels, will be missed. Analysis of Census data shows that approximately 14% of international migrants live in such communal establishments.

Moreover, the sample sizes available within the BHPS limits the power of analysis that can be undertaken. When the BHPS reached its maximum size (18,867) at wave 11, there were 832 respondents not born in the UK in the sample. By the final wave of the BHPS (with 14,419 respondents overall), this

⁷² http://www.esrc.ac.uk/images/ADT-Improving-Access-for-Research-and-Policy_tcm8-24462.pdf

had fallen to 597 respondents not born in the UK. In total, there were 471 respondents included in both datasets (and not all will have been in work).

Other sources of information provide different perspectives that help us understand different elements of migrant workers' lives. The most useful of these are cross sectional surveys, including:

- The **Census**, due to its scale;
- The **Labour Force Survey/Annual Population Survey (LFS/APS)**, due to its level of detail about work and its regularity;
- The **International Passenger Survey (IPS)**, due to its information about migrants' intentions when they move to the UK.

The **Census** is the most comprehensive source of data on the UK population, achieving approximately 94% coverage of all UK residents,⁷³ and asking basic questions about employment. However, data is only collected once every ten years and (a 3% sample of) microdata from the 2011 Census will only be available in late 2013, meaning the most recent datasets available are from over a decade ago. The Census Longitudinal Study (roughly 1% of Census respondents) will also become available for the 2011 Census, but not during the life of this project. Despite the introduction of questions in 2011 about intended length of stay of migrants and year of (last) entry to the UK, Census data will not be particularly useful in tracking migrants over time due to its decennial publication. However, it is useful in providing detailed snapshots of the make-up of the entire UK population every ten years. The Census does not include questions on respondents' wage.

The main advantages the **LFS/APS** have over the Census is the regularity of their publication, with the APS publishing annual data once a quarter on a rolling four-quarter basis. In addition, these surveys have much more detailed questions about respondents' work and work history. The LFS also asks more detailed questions about employment than BHPS/USoc. Clearly, as a sample survey, the raw numbers of migrants we in the survey is less than when working with Census or administrative data (although a recent quarterly LFS dataset included c.5,500 people not born in the UK who were in work). The LFS also has the capacity to track individuals over the course of a year.

With regular cross-sectional surveys, such as the LFS, a fixed number of households are randomly selected each period to be surveyed. Therefore, an increase in the number of migrants in the wider UK population will lead to an

⁷³ Response rates in the 2011 Census, December 2012

increase in the base number of migrants in the LFS sample. In this way, cross-sectional surveys should respond more immediately to changes in the make-up of the UK population than panel surveys as they rely on migrants being resident at any UK address, not just an address selected in the panel sample. However, the main drawback of using cross-sectional sources is that they have very limited scope for longitudinal analysis, although 2-quarter and 5-quarter longitudinal Labour Force Survey datasets are available.

The **International Passenger Survey** is useful because it provides data on migrants' intentions for work when they first arrive in the UK, and thus could be used to compare expectations of migrants with realities. It also provides a way of comparing the type of work respondent's did in their previous country of residence, and what they expect to do in the UK.

The **Family Resources Survey** (now used to provide EU-SILC data) has detailed data on respondents' income and investments but has relatively little on work or migration histories (although it does include country of birth). As such, other survey sources which provide more basic wage data coupled with more questions about work histories and will be of much greater use when developing a holistic picture of the lives of migrant workers.

Other administrative sources provide limited additional understanding of migrant workers. Some, such as the Worker Registration Scheme, the Sector Based Scheme and the Seasonal Agricultural Workers Scheme, only provide limited data of particular employment schemes which offer only a partial picture of migrant workers in the UK. Data for these sources can typically be broken down by age, gender, and nationality. Similarly, Home Office data on visas issued to non-EU nationals through the Points Based System, covers all immigrants entering the UK from outside the EU regardless of the sector they are hoping to be employed in, but tells us nothing about intra-European migrations.

Time of arrival in UK and work histories in previous country

Most survey datasets include questions about when immigrants entered the UK, including the Census (a new question in 2011), the LFS/APS and BHPS/Understanding Society (wave 1 of USoc also picks up data on where else an individual has lived over their lifetime). In NIRS/LLMDB a proxy that can help inform when someone arrived in the UK is when they registered for a National Insurance Number, although this is only required to work or collect benefits.

In terms of work histories, the International Passenger Survey provides the best information of all sources. The IPS allows analysis of whether migrant workers intended to work in the same occupations in the last country they lived in and in the UK, or whether they were expecting to move into a different type of work. However, most surveys have no way of capturing immigrant employment histories in other countries (except that the LFS asks about employment 12 months before the survey, and this may uncover what immigrants who entered the UK less than a year before being surveyed were doing in their previous country of residence).

Tracking British workers outside UK

Very little data is available to track British workers when they emigrate to another country. As they stop paying income tax / national insurance in the UK they are not captured on NIRS or other administrative databases, and British household surveys do not conduct fieldwork outside of the UK. When people disappear from administrative databases or surveys, it is not always clear why this is – for example, it can be hard to differentiate between non-response, death and emigration.

The main source of data on emigration is the International Passenger Survey, which asks about reasons for emigration, and expectations for work in country being emigrated to. It can, however, collect only a relatively limited amount of data, due to the time constraints of its collection. Typically of just over 1,000 emigrants are interviewed every year through the IPS.

In addition, the third wave of USoc will include a module on emigration intentions, although data for this will not be available within 2013.

Posted workers and multinational/supranational organisations

There is very limited information available about posted workers, and those working in multinational and supranational organisations. Within administrative datasets, one category of visa (intra company transfers) does tell us about the number of non-EU nationals moving within their company to the UK. However, this is of extremely limited use.

The Standard Industrial Classification (SIC) used in the UK, based on the EU NACE codes, includes a category, '9900 - Activities of extraterritorial organisations and bodies' which marks those working in supranational

institutions in datasets using SIC codes (including BHPS/USoc and LFS/APS).⁷⁴ However, it is extremely unlikely that any survey would have enough respondents in this category to allow subgroup analysis.

Coverage of key variables

Four key variables that are not always included in all datasets are qualification level, wage, and sector and occupation employed in. The table below summarises availability of these data in key datasets.

	Wage	Qualifications	Employment Sector	Occupation
Labour Force Survey	✓	✓	✓	✓
International Passenger Survey		See below		✓
Understanding Society / British Household Panel Survey	✓	✓	✓	✓
Census		✓	✓	✓
Family Resources Survey	✓	✓	✓	✓
Lifetime Labour Market Database	✓			
Work and Pensions Longitudinal Study	✓			

The IPS has the year a respondent left education, which could be used as a proxy for qualification level.

⁷⁴ This class includes: activities of international organisations such as the United Nations and the specialised agencies of the United Nations system, regional bodies etc., the International monetary Fund, the World Bank, the World Customs Organisation, the Organisation for Economic Co-operation and Development, the Organisation of Petroleum Exporting Countries, the European Communities, the European Free Trade Association etc.

Section C: Suggestions for improving data availability

Planned future changes in UK

The main development expected in the UK over the next few years relating to our ability to measure migration and track migrant workers is the introduction of the 'e-borders' programme.

e-borders is a Home Office programme designed primarily to improve national security by allowing border and security agencies to identify and target individuals of interest before they arrive or exit the UK. However, the scheme also has considerable potential for improving migration statistics. The scheme requires airlines, or other carriers, to provide Travel Document Information (TDI) also known as Advance Passenger Information (API) to Home Office officials.

API is the biographical data on a travel document i.e., name, date of birth, nationality, gender, travel document type, state of issue, number and expiry date. It would not, however, collect data on UK destination, country of birth, country of last/next residence or the main reason for travelling.

The current target is to achieve 95 per cent coverage of all passenger traffic into the UK by the end of 2014. The long-term hope is that by 2018 100% of all passenger traffic into the UK will be covered by e-borders. If this were the case then the only groups missed would be those illegally gaining entry to the UK from non-registered routes.

There are a number of possible ways in which e-borders could improve the UK's migration statistics.⁷⁵ In the long-term e-borders could be used to provide counts of both immigration and emigration, and thus could be used to measure net migration flows. However, this would not be available until 2018 at the earliest and there are some issues with using it as a migration count. There is, however, a possibility to link the data with a number of other sources in order to provide more accurate estimates of the migrant stock in the UK.

⁷⁵ ONS (2008), *Delivering statistical benefits from e-Borders*, <http://www.ons.gov.uk/ons/guide-method/method-quality/imps/latest-news/delivering-statistical-benefits-from-e-borders/index.html>

Moreover, in the future, e-Borders can be used to improve the process from which International Passenger Survey Data is weighted up to provide estimates of migration. E-Borders data would help to improve the weights used for different travel routes, and for those who arrive out-of-hours. It could also help to improve the efficiency of IPS sampling across different 'ports' throughout the UK. It is understood that discussions about how e-Borders could practically be used for this purpose are in very early stages, and that tangible improvements to IPS weighting are a long way off. The IPS also has to rely on stated intentions about people's length of stay in or out of the UK, and e-Borders could eventually replace such data with actual length of stay and improve the accuracy of switcher adjustments where actual length of stay is not possible.

However, apart from making the above improvements to the IPS, e-Borders is unlikely to significantly improve the UK's ability to understand the working lives of migrants. In 2006, it was suggested that e-borders, by linking with Home Office visa information, could be used to flag up when migrants with visas were entering the UK so that they could then be targeted by IPS interviewers to increase the number of migrant interviews within the IPS. However, after consultation with a range of stakeholders this was considered impractical and is very unlikely to happen.⁷⁶

There are two other developments that will alter the UK's ability to track migrant workers in the future.

- As touched on in our review of the Family Resources Survey (FRS), from 2015 the FRS will be used to meet the longitudinal requirements of EU-SILC. Approximately half of the FRS sample, 9,000 households, will form a panel followed by the Office for National Statistics over 4 years. However, in the follow-up interviews, they will only be asked questions required by EU-SILC. The fact the UK is planning to use the FRS to meet EU-SILC longitudinal requirements means that it is possible that the longitudinal dataset created will contain some of the detailed demographic information, such as country of birth, that is included in the FRS, but not required by EU-SILC.
- Another ongoing development is that with the fieldwork for further waves of Understanding Society ongoing, we will build up more longitudinal information on a larger number of migrant workers due to the USoc Ethnic Minority Boost. At present, the sample of migrant

⁷⁶ ONS (2008), *Delivering statistical benefits from e-Borders*, <http://www.ons.gov.uk/ons/guide-method/method-quality/imps/latest-news/delivering-statistical-benefits-from-e-borders/index.html>

workers in the BHPS samples is relatively small and thus, in the future, Understanding Society will allow us to conduct more powerful analysis of migrants' interactions with the labour market.

Procedural or bilateral agreements for exchanging information on migrant workers

Staff at the Office of National Statistics were unaware of any such agreements with other countries.

Recommendations to improve the ability to track migrant workers

In light of the findings in this document, we make a number of recommendations to improve the UK's ability to track the working lives of migrants. Our recommendations can be grouped into one of the following three themes:

- Increasing access to currently available data;
- Improving the quality of administrative data; and
- Improving the quality of survey data sources.

Increasing access

Widening access to administrative data. The United Kingdom currently collects a wealth of administrative data on residents' personal characteristics and their interactions with the labour market, but current access policies severely restrict their contribution to debates around a number of issues, including the position of migrant workers. We recommend that important datasets are made more widely available through the UK Data Service's Secure Data Service. Key datasets should be available to government contracted researchers as well as for academic and non-commercial research.

Improving the quality of administrative data

Despite the large quantities of administrative data collected in the UK, there is considerable scope to improve its quality. In particular, there are a some gaps in the data identified in this review that could be filled.

Collect data about immigrants' work history and education when they apply for a National Insurance Number (NINo) in the UK. We

suggest that more information is collected from immigrants when they register for a National Insurance Number, which is required to work or collect welfare benefits in the United Kingdom. The current process for registering for a NINo requires individuals to fill in an application form, and we recommend that this opportunity be taken to collect data that otherwise is hard to collect. This data could then be appended to other datasets which include NINo identifiers, such as the Migrant Worker Scan, the Lifetime Labour Market Database and the Work and Pensions Longitudinal Study.

The two main holes in current administrative data are:

- The employment history of immigrants before they entered the UK; and
- Their level of educational attainment.

Questions about employment history should include the most recent sector and occupation immigrants worked in, the amount of previous work experience they had, and their employment status immediately before moving to the UK. Questions on education could either ask what the immigrants' highest qualification was or, as a proxy given it is often hard to compare UK and foreign qualification, at what age they left full time education.

Adding qualification levels to administrative data for current UK residents. Qualification level is not available in UK administrative data of current residents, and adding it would help researchers analysing the labour market outcomes of different groups. This could be done by matching qualification level data from the 2011 Census to administrative datasets. Census data does not include NINos, which are used in administrative data, but matching could be completed using 'fuzzy matching algorithms,' in which a combination or concatenation of personal identifiers such as name, date of birth, gender, and post code are used to match datasets.

Guidance to help compare EU qualifications. It is common in UK datasets for non-UK qualifications to be classed in a catch-all 'other' category, as it is unclear how they relate to UK qualification levels. The EC could facilitate the production of guidance in order to relate all European qualifications to International Standard Classification of Education (ISCED), to allow EU members to better understand different qualification frameworks.

Better data sharing among EU members and/ or harmonised administrative data. Data sharing agreements among EU members would allow member states to exchange data about migrants as they move around the EU. Given the lack of data the UK has on emigrants, this could be very

valuable. If immigrants, when applying for their Social Security Number (SSN) in their new country of residence had also to provide their SSN from their previous country of residence, this would allow data to be matched.

From a research perspective, a longer term goal may be to better harmonise different member states' administrative data, and even create a EU SSN, from which a database of all European workers could be built.

Improving the quality of survey data

Whilst using administrative data to track migrant workers is preferable, given the larger amounts of data available, improving the quality of administrative data sources and gaining better access to them may not be possible, especially in the short-term. Therefore, it is still important to improve survey data sources where possible.

Use the International Passenger Survey to boost the sample of migrants in other national surveys. The International Passenger Survey could be used to boost the sample of migrants in other UK surveys, such as the Labour Force Survey and Understanding Society. This would be especially useful for a panel survey like USoc, as panels are likely to take longer to respond to changes in the wider population such as new migration patterns. This could work in practice by asking those identified by the International Passenger Survey as migrants if they were able to provide an address which would be used in future survey samples.

Addresses of emigrants could be shared with other EU members, to increase the numbers of mobile workers survey in, for example, the EU LFS.

Conducting surveys in communal establishments as well as private households. One study has suggested that 14% of migrant workers live in communal establishments such as hostels or caravan parks,⁷⁷ but important household surveys like the LFS and Understanding Society do not interview those living in such circumstances, meaning that they miss out this important group of migrant workers. Our second recommendation, therefore, is to alter the sampling procedure for household surveys such as to include communal establishments. The Census employs certain fieldwork techniques to ensure individuals living in such establishments are surveyed, and these techniques could be adopted in other surveys in the UK.

⁷⁷ Improved Methods for Estimating International Migration The Use of the Labour Force Survey to Improve Estimates of International In-migration - Coverage and Quality