

Written evidence submitted by Jonathan Portes (EUM 01)

Public Administration and Constitutional Affairs Committee

Inquiry into data sharing and the use of administrative records to provide enhanced statistics

Personal

I am a Principal Research Fellow at the National Institute of Economic Research (NIESR). I am also currently a Senior Fellow on the UK in a Changing Europe programme, funded by the Economic and Social Research Council (ESRC). The latter programme is intended to fund the provision of objective academic research and analysis in order to inform the public in the runup to the EU referendum. Neither the ESRC nor academics funded by the programme, including myself, take any position on EU membership as such.

Introduction and background

The Committee states that “We will be using the example of the use of National Insurance numbers (NINOs) for overseas adults in migration data as a case study”. My evidence explains the relationship between different sources of migration-related data, including NINOs, and how additional administrative data and analysis from HMRC and DWP, not yet released in response to FOI requests or provided to ONS, might assist in improving migration data. **Note:** I have attempted to make this note clear and not excessively long, so I have skipped some technical detail. ONS are of course the experts here, and no doubt their submission will explain in greater depth and rigour.

Migration- related data sources

The Office of National Statistics publishes quarterly data on international migration **flows** to and from the United Kingdom, based on a large sample survey, the International Passenger Survey (IPS), of people passing through UK air and sea arrival points. Respondents are asked if they intend to stay (or leave) for more than a year, the official, internationally recognised definition of an international migrant. Adjustments are subsequently made to the data to reflect people not recorded in the IPS (asylum seekers and so-called "visitor switchers").

However, there are two other important sources of information on immigration:

- the Labour Force Survey, which is a sample survey of households, intended to be broadly representative of the UK resident population (with some exceptions). This contains data on the country of birth of respondents. It therefore (together with the related Annual Population Survey) provides an estimate of the **stock** how many immigrants (defined as people born outside

the UK) are resident in the UK at any one time.¹ This data is published quarterly: the detailed underlying data is also available to researchers.

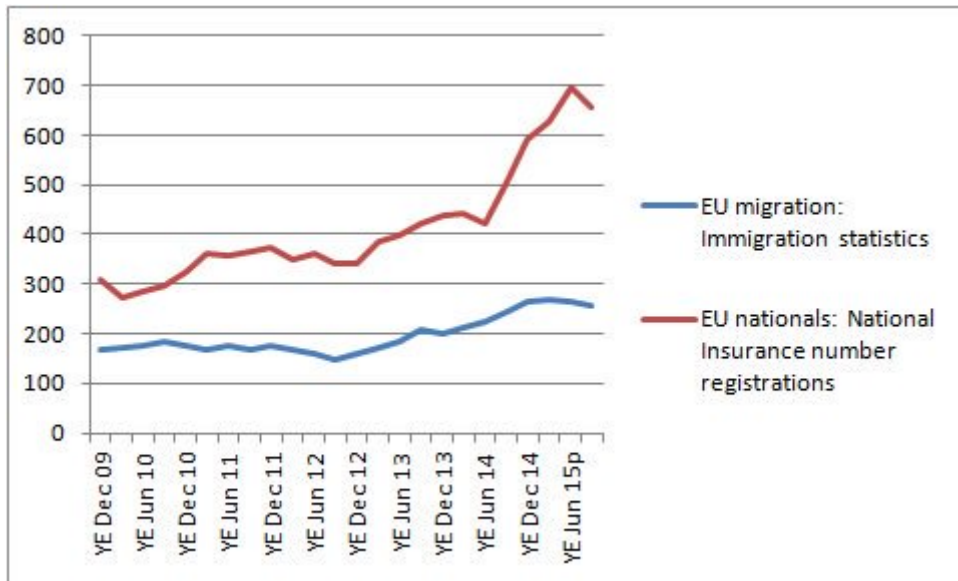
- NINO registrations. UK residents in most cases automatically receive a NINO at the age of 16. However, those moving here from abroad need to apply for one if they wish to work or claim benefits (it may also be useful in some other contexts). For adults who register for NINOs, their country of birth is recorded. Data on new NINO registrations by people born abroad is published quarterly by DWP alongside other migration statistics, including the official IPS-based migration statistics. This therefore provides a measure of **gross flows** of people from abroad registering to work (usually) in the UK. However, the underlying data is not available outside of government.

How should these data match up? The most obvious comparison is between the gross flows measured in the IPS-based migration statistics, and the gross flows recorded in the NINO registration data. While these both represent gross flows (inflows), and so are conceptually comparable, they should not necessarily match up, for at least three reasons. First, anyone coming here for less than a year is, correctly, not defined as an immigrant in the migration statistics, but some such people will register for a NINO. Second, not all immigrants will require NINOs, since they may not be active in the labour market nor seeking to claim benefits. And third, there may be timing differences, since migrants will not register immediately. These differences are well understood, and were noted by the ONS in the most recent Migration Statistics Quarterly Report.

The recent divergence

However, as has been well publicised, the differences between the two series, particularly as regards migrants from the EU, have grown very sharply in recent years, as shown in this chart. More than 2 million EU nationals have registered for NINOs over the last four years (to September 2015), while (gross) inflows, according to the migration statistics, were about 850,000.

¹ The LFS also asks about country of citizenship. However, this is not a very useful concept for looking at migration statistics, since people can arrive in the UK as an immigrant and subsequently acquire UK citizenship. But people cannot, of course, change their country of birth. So while citizenship may be relevant in other contexts (for example, in assessing integration) attempting to use it as an indicator of immigration can be misleading.



Meanwhile, while the Labour Force Survey does not provide data in a form directly comparable to either the migration statistics or the NINo registration data (since it records stocks rather than flows, although it does contain some information on year of arrival), it shows a **net** increase in the number of EU-born of working age of approximately 700,000.

With the data currently available, then, we cannot explain the recent large divergence shown in the chart. It could reflect a large rise in short-term migration, correctly not recorded in the migration statistics, nor in the LFS data; it could reflect timing differences in registration, related in particular to the ending of transitional provisions for Bulgarians and Romanians (enabling them to work legally); it could reflect unrecorded or misrecorded immigration and/or gaps in coverage of the IPS and/or LFS; or a combination of all of these factors. However, data held by government departments could shed light on the plausibility and relative weight of these possible explanations, as I discuss below.

The Prime Minister's speech and my Freedom of Information request

In a speech in November 2015, the Prime Minister claimed:

"We now know that, at any one time, around 40 percent of all recent European Economic Area migrants are supported by the UK benefit system."

At the time the Prime Minister made this statement, there was no supporting evidence for this claim. Following a complaint from FullFact, and an intervention by the UK Statistics Authority, DWP published a document setting out the basis for this calculation.

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/475765/uk-benefits-and-tax-credits-eea-migrants.pdf

It is worth reproducing in some detail DWP's explanation of the methodology:

“To identify benefit claimants the DWP working age benefit and tax credits policy simulation model– INtegrated FORecasting Model (INFORM) has been used to identify the working age benefit claiming population at a point in time. This dataset is used as the basis for dynamic modelling of the working age system. It is derived from extracts of the DWP Work and Pensions Longitudinal Study (WPLS), the Single Housing Benefit Extract (SHBE) and a sample of HMRC interim Tax Credit system. Each of these are administrative datasets from which an extract has been taken to produce a 5% sample of the working age benefit caseload... To identify the nationality of benefit claimants, the NINO of the main benefit claimant has been matched with the dataset of nationality at NINO registration (sourced from HMRCs National Insurance and Pay as you Earn System). This approach is consistent with the established approach to identify the nationality of benefit claimants in existing Official Statistics series.”

This approach was used to calculate the number of "recent EEA migrants supported by the UK benefit system" - the numerator in the 40% figure cited by the Prime Minister. However, the denominator - the total number of recent EEA migrants, supported by benefits or not - was not calculated from this data, but rather from an entirely different source, the Labour Force Survey (see above).

Having read the DWP publication carefully, it was my view:

- a) if DWP and HMRC had gone through the complex data-linking and data-matching procedures described above to establish the number of "recent EEA migrants" claiming DWP benefits and HMRC tax credits, there was no obvious technical obstacle to them doing the same for the number paying NI contributions, PAYE income tax, or self-assessment tax;
- b) this would provide a much more robust, obvious and directly comparable denominator for the Prime Minister's calculation. If it was in fact the case that significantly more NINOs registered by recent EU migrants were "active" than were shown in the LFS data, then the proportion "supported by the benefit system" would be proportionally lower. This would bring the Prime Minister's numbers more in line with those found by external researchers.
- c) that this would also provide extremely useful information that would help reconcile the NINO registration statistics and the migration statistics. In particular, if a relatively small proportion of the 2 million recent NINO registrations to EU nationals were still active, this would support the hypothesis that short-term migration, correctly not measured by the immigration statistics, had indeed increased; if on the other hand a large proportion were, it might suggest that the migration statistics were undercounting EU migration.

I therefore asked for the following information (to both DWP and HMRC)

“Using the methodology and data sources set out in

<https://www.gov.uk/government/statistics/uk-benefit-and-tax-credit-claims-by-recently-arrived-eea-migrants>

1. How many individuals are recorded in DWP and/or HMRC computer systems who were a) nationals of another EEA member country (other than the UK) at time of registration for a National Insurance Number (on the basis set out in this DWP publication) AND b) are recorded as having arrived in the last four years (using the most recent available data) and using the earlier of (a) arrival date, or; (b) NINO registration date AND c) have either (i) paid National Insurance contributions over the previous year OR (ii) paid PAYE income tax over the previous year OR (iii) claimed benefits or tax credits over the previous year (on the same basis as in the DWP publication) This question can be answered as you determine appropriate either by reference to a 100% or a smaller (presumably 5%) sample and scaled up to provide an estimate for the whole population (with error bounds in the latter case).

2. The same question but for non-EEA nationals."

It should be noted that this has been subsequently been adopted - by me and some in the media - as a working definition of "active" migrants, that is migrants who are in some sense generating activity on DWP and HMRC systems and hence may be assumed to be resident in the UK. However, the term "active" here is just my own ad-hoc definition, and has no formal official status.

HMRC responses

HMRC responded to my FOI on December 16, 2015

"However, following the General Election, there is an active negotiation process at an international level in which UK Ministers and officials are engaged to secure support from the European Commission and other Member States for changes in EU law governing EU migrants' access to benefits in the UK, in line with the Government's manifesto commitments. The information is being used to inform the development of policy options as part of the negotiation process and therefore relates to the formulation of Government policy. HMRC continues to believe that releasing information in the form requested would, at this stage, be unhelpful to the negotiation process."

HMRC's full response can be found here.

[http://www.niesr.ac.uk/sites/default/files/files/HMRC%20FOI%20reply\(1\).pdf](http://www.niesr.ac.uk/sites/default/files/files/HMRC%20FOI%20reply(1).pdf)

After I published this response on my blog, and the consequent media interest, the Head of News at HMRC sent the following note to lobby journalists:

"Clarification relating to FoI request from Jonathan Portes. An HMRC spokesperson said: "It was wrong to suggest information was withheld because of the EU renegotiation. Not all the information that has been requested is held by HMRC. The information that is held is not in a collated, publishable form. HMRC will release that data when it is properly collated early in the New Year."

This was, however, not copied to me. Since then, I have received a number of further responses (to appeals, and to different versions of the request) stating:

"The information you have requested on PAYE and National Insurance contributions paid by these individuals is held by HMRC but not in the format you have requested. We estimate that it would exceed the FOIA cost limit to collate the information into a publishable format. The FOIA cost limit, which is specified in regulations, equates to one person spending 3½ working days locating and extracting all of the information within scope of the request."

Where do we stand now?

During March, there were a number of relevant developments. In particular, Andrew Tyrie, Chairman of the Treasury Committee, wrote to Lin Homer (HMRC) asking (again) for information on the lines that I had requested; Sir Andrew Dilnot (UK Statistics Authority) wrote to John Pullinger (ONS) expressing his firm view that DWP and HMRC should give ONS access to this data so that they could incorporate it as necessary into the next publication of the migration statistics; and Bernard Jenkin, Chairman of the Public Administration and Constitutional Affairs Committee, also wrote to John Pullinger, informing him of the Committee's inquiry and asking him to provide a written submission.

On April 4, 2016, I received a further response from HMRC. While continuing to reject my request on cost grounds, it stated the following:

"The requested data on active NINOs however does not exist in one place and requires combining several HMRC datasets, that include individuals subject to income tax and National Insurance Contributions (NICs), families in receipt of tax credits and individuals in receipt of Child Benefit and then matching the data to the information on nationality from the DWP dataset of nationality at NINo registration. In addition, HMRC need to ensure that the data is not double counted in cases where NINOs are active with respect to more than one tax or benefit. Combining several datasets and hundreds of millions of lines of data takes time and that is why it has not been possible to provide the data sooner.

You will be aware that the ONS announced earlier this month that they will look at reconciling the four main sources of international migration data (International Passenger Survey, Labour Force Survey, National Insurance numbers and visas), including understanding the relationship between the International Passenger Survey estimates and National Insurance numbers. We are working closely with ONS and will be providing our data and analysis to them once it has been compiled, to allow them to combine it with their own data, analysis and quality assurance work and

thereby produce a fuller picture. Our analysis will then be published either as part of an ONS publication or alongside it."

This is obviously welcome, although somewhat belated. HMRC have now committed both to publishing analysis along the lines I originally requested, and to supplying data and analysis to ONS.

Given all this, I would make the following points, which the Committee might wish to consider:

- a) It is in my view unfortunate that to date HMRC/DWP's actions in handling the analysis of the data that they hold (on behalf of the public, I would note) appears to have been directed at averting embarrassment to the government, rather than either in informing the public debate or assisting ONS in their difficult but essential task of producing the most reliable statistics possible. This must change, and not just for this publication.
- b) it is important that any HMRC (or HMRC/DWP) publication provides as much information and data as possible. My original Freedom of Information request asked for a relatively limited set of data, as is appropriate for an FOI. However, a statistical publication should go further. As HMRC's response implicitly admits, the main work involved in preparing these statistics is collating and checking the data; once that is done, it is as easy to produce detailed statistics as it is to simply publish a few selected ones. In particular, I would suggest that any publication should, at a minimum:
 - disaggregate "active" NINOs by groups of year of arrival as far as possible (ie, not just "within the last 4 years" and "more than 4 years ago", but perhaps 4 or 5 different groupings)
 - disaggregate between different country groups (EU15, EU8, EU2, and non-EU)
 - identify separately the numbers claiming/paying DWP benefits, tax credits, NI contributions, PAYE, self-assessment, as well as the relevant overlaps
 - present the information in comprehensible, tabular form, in Excel or similar format
 - compare the statistics derived from analysis of HMRC/DWP systems with that from the LFS/APS. For example, the LFS/APS in principle will show how many "recent" EU (or non-EU) migrants currently have "employee" status; this should in principle be broadly comparable with the number recorded as paying employee NICs.
- c) it is important that ONS and HMRC/DWP work cooperatively in ensuring that ONS have access to the data and analysis they require. In other words, HMRC/DWP should be discussing with ONS throughout how the data can best be analysed, and ensuring that they produce analysis which meets ONS' needs, rather than simply providing them with a finished product.

- d) there is a degree of confusion in the public debate about what the type of analysis described above will and will not show. **It is important to note that this analysis will not in itself produce the “truth” about migration flows from the EU, for the reasons described above.** It will however provide an additional and potentially important source of evidence and data. On the basis of this additional evidence, and no doubt other data sources and analysis to which ONS have access, ONS will need to decide whether or not (and if so how) to adjust the methodology by which they produce the IPS-based migration statistics, either retrospectively or in future. This is of course a professional judgement for ONS to make independently, free from government or political pressure or other considerations;
- e) finally, this data is potentially very useful for other aspects of migration-related analysis than simply informing the migration statistics. For instance, to take an obvious example, data on the earnings trajectories (taken from tax records) of migrants is of considerable policy interest. Available data suggests that EU migrants, especially from relatively poor countries, are concentrated in relatively low-paid jobs. However, we do not know if over time they tend to move up the earnings ladder; or whether, conversely, they get “stuck”. It would be helpful if HMRC/DWP were to commit to make this data available to external researchers (in suitably anonymised form) in order to permit this type of policy-relevant analysis.

I would note that there is clearly no technical obstacle to any of my proposals above, nor is it difficult to provide the data in such a way that personal details and confidentiality are preserved. A recent IFS paper used a database produced by linking HMRC data with data from the Student Loans Company, using National Insurance numbers as the unique identifier:

<http://economics.mit.edu/files/11153>

Given the public, political and policy interest in immigration statistics, and in information on the labour market outcomes of immigrants, it is difficult to see why HMRC and DWP have not already made a similar database available in respect of non-UK nationals.

Conclusion

I welcome the government’s new commitments to make this data available: I hope the Committee will hold them to account for this commitment, and to ensure that it is done so in such a way as to maximise the information and evidence available to the public.

April 2016

Jonathan Portes

Principal Research Fellow, National Institute of Economic and Social Research

Senior Fellow, ESRC UK in a Changing Europe Programme

Appendix

I have tried to make this note self-contained. However, my earlier blogs and articles on this topic can be found at these links:

<http://www.niesr.ac.uk/blog/migrants-benefits-and-uks-renegotiation-questions-and-answers-updated#.VwLRf6QrJpl>

<http://www.niesr.ac.uk/blog/recent-eu-migration-uk-can-we-trust-government-numbers#.VwF1GaQrJpl>

<http://www.niesr.ac.uk/blog/government-admits-it-has-statistics-recent-eu-migration-and-wont-publish-them#.VwF1SqQrJpl>

<http://www.niesr.ac.uk/blog/data-eu-migration-would-be-too-expensive-collate-government-says#.VwLRxKQrJpk>

<http://www.theguardian.com/commentisfree/2016/feb/29/eu-migration-debate-worth-having-but-must-be-based-on-facts>

<http://www.niesr.ac.uk/blog/migrants-benefits-dwp-give-us-some-answers#.VwF1haQrJpk#>

I would also strongly recommend that the Committee consider the work of Michael O'Connor, and in particular the helpful explanation of some of the issues involved here:

<https://medium.com/@StrongerInNos/another-qtwtain-probably-1fb737648892#.q44dswlg3>