Using linked administrative and census data for migration research

Annemarie Ernsten*, David McCollum, Zhiqiang Feng***, Dawn Everington*** and Zengyi Huang*****

*National Records of Scotland

**University of St Andrews

***University of Edinburgh

Migration is a core component of population change and is a symptom and cause of major economic and social phenomena. However data limitations mean that gaps remain in our understanding of the patterns and processes of mobility. This is particularly the case in relation to internal migration, which remains under researched, despite being quantitatively much more significant than international migration. Using the Scottish Longitudinal Study (SLS), this paper evaluates the potential value of NHS GP administrative health data that can be linked into census based longitudinal studies for advancing migration research. Issues relating to data quality are considered and, using the illustrative example of migrant/non-migrant migration intensities, an argument is developed which contends that such approaches can offer novel ways of comprehending internal migration by shedding additional light on the nature of both movers *and* the moves that they make.

Keywords: administrative NHS GP health data, data linkage, internal migration, Scottish Longitudinal Study

Introduction

Migration is the key driver of population change at the local, regional and national scale, but it is also the hardest to measure and predict (Stillwell *et al.*, 2011). This is especially true of internal migration, which despite being by far the quantitatively more significant phenomenon, has been subject to much less analytical scrutiny than international migration (Bell *et al.*, 2015a). This anomaly is noteworthy since the internal dynamics of population change are of considerable conceptual, policy and commercial relevance (Smith *et al.*, 2015). For example, around 5 per cent of the population of England and Wales (2.85 million people) move between local authorities every year (Office for National Statistics, 2017c), and wide disparities in the incidence of internal migration exist on the global scale (Bell *et al.*, 2015b). As a spatially and socially selective process, internal migration is known to respond to economic change in specific ways, with an overall reduction in mobility during economic downturns (Fielding, 2012; Saks and Wozniak, 2011). However much remains to be learned about how patterns and processes of internal migration have been shaped by the great recession of 2008 and ongoing

economic uncertainty and austerity (Green and Shuttleworth, 2015). Another topic that merits attention is recent and future trends in internal migration. Recent influential concepts in migration studies, such as the age of migration (Castles *et al.*, 2014) and the new mobilities paradigm (Sheller and Urry, 2006), have led to a common assumption that the current epoch is one of unprecedented mobility (Champion and Shuttleworth, 2016a). However this stands in contrast to an emerging body of evidence that points towards declining rates of internal migration, supposedly as a consequence of economic and demographic change, but also a wider societal shift towards so-called 'secular rootedness' i.e. a more universal societal transition towards less migration that cuts across multiple population sub-groups (Cooke, 2011; Champion *et al.*, 2017).

Efforts to better understand these important issues have long been hampered by a paucity of suitable data available to researchers concerning internal mobility dynamics. Internal migration can be investigated through examination of population censuses, registers and surveys, with each approach having specific benefits and limitations (Bell *et al.*, 2015b). In the UK context the best and thus most widely used data sources for research on internal migration have been the decennial censuses and the National Health Service Central Register (NHSCR), which is a health based administrative register (Raymer *et al.*, 2011). The NHSCR, which is based on GP registrations, is used by the official statistical offices in the UK to generate estimates of internal migration flows (ONS, 2016). This source provides frequent and up to date information on moves. However it contains a number of significant flaws. It undercounts some forms of mobility (such as short distance moves and those made by young adults, especially men) and can only shed light on the origin and destination and age and sex attributes of movers (Raymer *et al.*, 2011). The decennial national census on the other hand contains a wealth of demographic information about movers, but it is infrequent and only picks up individuals that have engaged in mobility sometime in the 12 months leading up to the day of the census (through the 'address one year ago' question). Nationally representative government surveys (such as the Labour Force Survey and Understanding Society in the UKs case) often contain time and attribute rich data that can be used to study migration. However relatively limited sample sizes usually preclude detailed investigation of specific population sub-groups or at sub-national spatial scales (Stillwell *et al.*, 2010).

Due to these limitations, researchers seeking to study internal migration have been faced with opting for either time rich but attribute poor administrative data, or attribute rich but time poor census data. However recent linking of NHS GP registration data into the census based Scottish Longitudinal Study (SLS) represents a potential new avenue for migration research since it allows for study of the characteristics of both moves (via the NHSCR) and movers (through their linked census records). The novel methodological approach discussed in this article thus allows, for the first time, for a detailed analysis of the recent mobility patterns of a sizeable cohort of individuals, at detailed geographies and over a significant period. This paper describes how this methodological approach operates, and assesses the potential of such administrative data linked to census based longitudinal studies to offer advances in migration studies. The following section offers a review of how internal migration is currently researched, and considers how this can be enhanced using the innovative approach described in this paper.

Literature review

As discussed above, an enhanced understanding of the dynamics of internal migration is necessary for a range of academic and practical reasons. The aim of this paper is not to directly contribute to recent research on patterns and processes of internal migration (Fielding, 2012; Smith *et al.*, 2015; Champion, 2016; Champion *et al.*, 2017), but rather to consider how new methodological innovations can build upon this body of scholarship. As such this section focuses mainly on how internal migration has been researched, rather than the results of these endeavours *per se*.

This paper seeks to demonstrate how the conventional difficulties faced by researchers drawing on either census-based longitudinal datasets or administrative data can be negated. Champion and Shuttleworth's recent studies of long-term trends in internal migration in England and Wales (2017a and 2017b) are an apt illustration of these challenges. One of their studies uses the ONS LS (Office for National Statistics Longitudinal Study), the England and Wales sister study of the SLS, to examine changes in addresses over the ten-year periods between censuses from 1971 (2017a). This provides a high level of detail on the characteristics of moves and movers, but has the limitation of only providing this type of information at decadal intervals. Their other study (2017b) uses health administrative data to generate estimates of between area moves from 1971. This approach has the benefit of annual as opposed to decennial information.

However it contains less data on the nature of moves and movers and omits within health board area (i.e. shorter distance) mobility. The methodological perspective described in this paper seeks to overcome these limitations by allowing for elucidation of nearly all moves (via unit postcode level Scottish NHS GP registration data), calculated on an annual basis, as well as the characteristics of movers (through the census based SLS).

Whilst this study represents one of only a few to explicitly link residential information from administrative health data to a census-based longitudinal dataset, researchers have for some time explored the potential of combining data from multiple sources to better understand internal migration in the absence of a population register in the UK. For example Boden and Rees (2010) set out how various proxy measures of migration derived from administrative sources can enhance conventional ways of estimating immigration to local areas in England, resulting in the Office for National Statistics (ONS) adopting these methods in their estimates of local immigration flows. Furthermore Raymer *et al.* (2011) outline a methodology for how NHSCR data can be combined with census data to generate a synthetic database of intercensal migration patterns. This approach involved supplementing information from the NHSCR with more detailed information from the censuses using log-linear modelling procedures. A well-established important limitation of using administrative health data to assess migration is the age-sex bias in the propensity to register with a doctor after moving (Boden *et al.*, 1992; Ogilvy, 1980). Young adults, especially males are less likely to reregister after moving, or take longer to do so, than women and older people. Raymer *et al.* (2011) propose weighting procedures to overcome this challenge, whereas others have actually employed these techniques to correct undercounts in young adult male migration rates (ODPM, 2002). The potential of weighting to improve the veracity of migration statistics is an issue that is considered in this analysis. A valuable extension of Raymer *et al.*'s (2011) method offered by the approach described in this paper is that longitudinal census data are used, allowing for individuals to be followed over time, and thus for consideration of the role of age, period and cohort effects in migration trends (Findlay *et al.*, 2015). Another interesting avenue offered by the approach discussed here is that the health administrative data in Scotland is available at finer geographies than is the case elsewhere in the UK, thus enabling the incorporation of short distance mobility into analysis. As such all changes in address can be assessed as opposed to just those which span higher level administrative boundaries.

A key issue when combining census-based and administrative data is the extent to which the information on the locations and migration experiences of individuals provided by each source correspond. The effectiveness of this 'matching' process is an issue which is considered at length in this analysis. A precedent can be gleaned from existing research in this respect, as NHSCR data has been linked into the ONS LS in the past. Smallwood and Lynch (2010) tested the extent to which the census and NHSCR provide consistent information on individual's health authority of residence using the 'snapshot' of census day 2001 (Sunday 29th April). They conclude that the census-based longitudinal study and health administrative data are very good 'matches' in terms of individuals' locations, with 96 per cent of ONS LS members recorded as being in the same Health Authority area as stated in their NHSCR record (although the figure was lower for young men and students). However this was a one-off study as NHSCR derived records are not routinely uploaded into the ONS LS. Another caveat here is that, as mentioned earlier, the available England and Wales health administrative data only includes moves between Health Authority areas whereas the newly available Scottish data includes all unit postcode changes (postcode units are unique references and identify an average of 15 addresses). As such the degree of 'matching' at health areas will be considerably higher than that between unit postcode levels given the higher-level geographies involved in the former (for example the average population of local health areas in England is 350,000; National Audit Office, 2012).

Drawing on the Health Card Registration System (the Northern Irish equivalent of the NHSCR) and the Northern Ireland Longitudinal Study, Barr and Shuttleworth (2012) also consider the issue of the quality of matching between census and health administrative data (at the Super Output Level). Their approach involved focusing on movers within Northern Ireland and modelling the likelihood of (a) an exact location match in their census and health administrative records, (b) a lagged match whereby it took over a year for movers' health administrative data location to match their census location and (c) no locational match being found. The study, which excluded students, found that the residential movements of men, those in good health and residents of urban and relatively deprived areas were most likely to be missed using health administrative data, via either late or non-reporting of address changes. The research discussed in this paper explores these issues in the Scottish context. Critically, the investigation also details how weighting procedures can overcome the limitations of late and non-response in

health administrative data for migration research. The following section sets out the methodological approach employed in this analysis.

Data

The core focus of this research is the matching of information from the SLS and mobility related information held by the NHSCR, and consideration of the utility of this process for migration research. A brief description of the core datasets used in this approach, and how they are matched to each other, is provided below. This is followed by consideration of the quality of this process in terms of the extent to which certain moves and movers may be underrepresented in the data. The ability of weighting procedures to address these issues is then examined. Finally, the example of internal migration rates in Scotland by country of birth is used to illustrate the usefulness of this type of approach in migration studies.

The SLS is a large-scale linkage study based on information from the Scottish Census from 1991 onwards. The study is based on 20 semi-random birthdates. Four of these match the birthdates in the England and Wales LS and the remaining 16 are chosen randomly from the remaining 361 days in the year (362 in a leap year) using probabilities derived from the daily distribution of births. Following the removal of dummies and duplicate records, about 5.3 percent of the Scottish population is covered in the sample, equating to around 265,000-270,000 members (Hattersley *et al.*, 2007; Boyle *et al.*, 2009). Data are collected on SLS members over time and their records are continuously updated through the linkage of vital events registration and NHSCR data (Hattersley and Boyle, 2007). The linking together of records from various administrative sources over time is an integrated part of the SLS. The record linkage exercise is carried out using the National Health Service Central Register (NHSCR) database of all residents in Scotland who have registered with a NHS General Practitioner. This is the most inclusive 'register' of the population in Scotland. Names and dates of birth are two of the basic pieces of information required by the NHSCR to allow them to identify an individual in their database and then 'flag' SLS members so that they can be linked to the SLS. Linking to non-census datasets and between censuses depends on the tracing of SLS members in NHSCR. Tracing SLS members at the NHSCR is carried out using a combination of exact matching, probability matching and manual matching (Hattersley and Boyle, 2008).

The SLS only recently (in 2016) received permission to add NHSCR GP postcode historical data (starting from 1st January 2000) into the SLS. This recent development means research on internal migration in Scotland can now be carried out using health administrative data linked to census based longitudinal studies that is akin to that conducted in Northern Ireland (Barr and Shuttleworth, 2012). However the Scottish data has the important additional advantage of enabling analysis of short distance moves. The way that health administrative data are incorporated into the census-based longitudinal studies of the other parts of the UK results in only moves that span health board boundaries or Super Output Areas being recorded (see ONS, 2017b) for an illustration of UK statistical geographies). In Scotland, unit postcode level information from the Community Health Index (CHI) system is now fed into the NHSCR. Postcode level data cannot be directly assessed by researchers due to the risk of statistical disclosure, but SLS support staff can derive variables of moves such as distance of move for researchers to use which allows for analysis of short distance moves without the risk of disclosure. As such the recent linking of this data into the SLS now enables analysis of moves at postcode unit level upwards, as opposed to merely the longer distance moves that cross health administrative boundaries. It should be noted that the analysis described in this article is based on a test version of NHSCR GP postcode data, which has subsequently been revised. Whilst this affects only a small proportion of the data, the dataset that is now available to researchers is slightly different. This, and the specific sample definition and methodology used in this study, means that the results from future analyses may not exactly correspond with those described here.

Effectiveness of the matching process

The previous section has briefly described the SLS and the health administrative data that constitute the basis of this new resource for migration research. An issue of critical importance in this respect relates to the effectiveness of the matching between the two data sets. Ideally everyone should be detected as being at the same postcode on the census days (Sunday 29th April 2001 and Sunday 27th March 2011) according to their census enumeration and NHS GP data. This was the case for 85 per cent of the SLS members included in this analysis (those of working age, 16-64, at the 2011 census, N: 174,258). The fact that this is lower than the 96 per cent found by Smallwood and Lynch (2010) using the ONS LS is most likely due to the finer geographies used in this analysis (postcodes versus health authority boundaries).

Of those not matching between their census and health administrative records, a relatively small number of individuals (1.1%) do not have NHS GP data mainly because these SLS members are not traced at NHSCR. The inter-census mobility of these individuals could therefore not be detected. Another, more significant, cohort for whom researching mobility behaviour is complex consists of those who displayed a delayed match between their census record and their location according to the NHSCR. In total, 13.9 per cent of the sample used in this research were present in both datasets, but did not have an exact unit postcode match on census day. For the majority of these (9.3% of the overall sample) the census enumeration postcode was not found in the NHS GP postcode history. For the remainder (4.5% of the overall sample) the census enumeration postcode was found at a later date in the NHS GP postcode history, almost always within three years (Table 1). For the purposes of the illustrative example using country of birth data in this paper, the effective study sample is the 85 per cent of the SLS members of working age (16-64 years) at the 2011 census who had exact matches between the 2011 census and NHS GP postcodes, plus the 0.6 per cent with a delayed match within six months of census day.

Table 1: Comparison of the postcodes recorded by the 2011 Census and the NHSCR

Category	**Locational information: SLS v NHSCR at census day**	**Per cent**
Exact match	Same postcode location at census day	85.0
Delayed match: NHSCR postcode matches census postcode	Within 6 months	0.6
	Within 7-12 months	0.9
	Within 13-24 months	1.3
	Within 25-36 months	1.7
	Over 36 months	0.1
No match	In SLS and traced at NHSCR, but never a postcode match	9.3
	In SLS but not traced at NHSCR	1.1
	Total	100.0

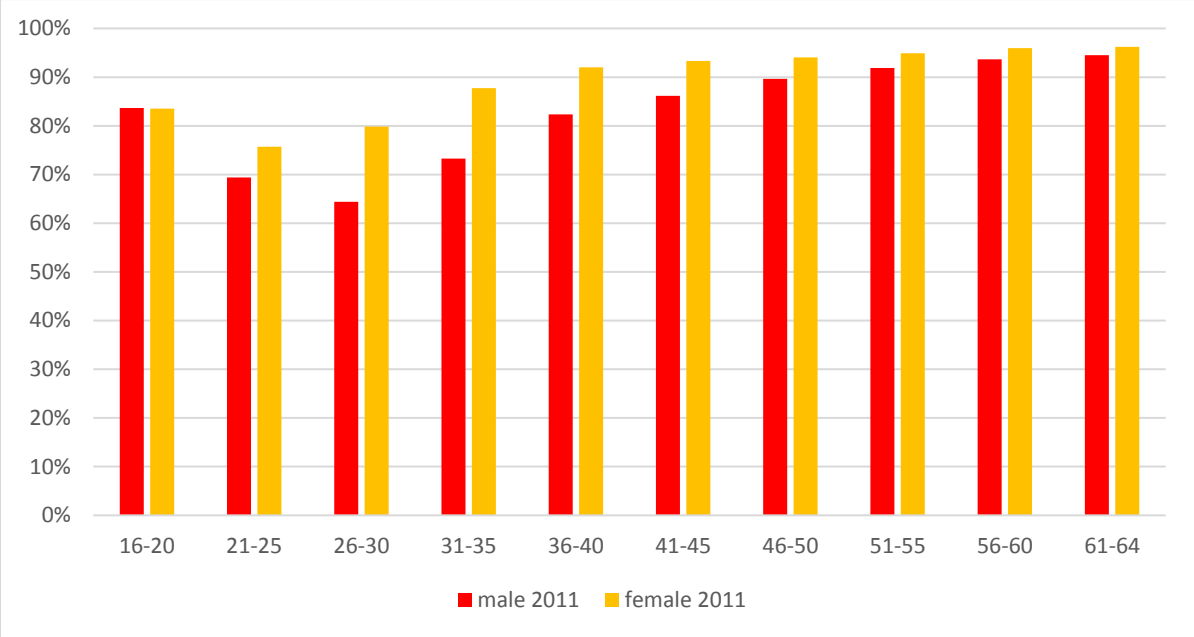
Sample: SLS study members aged 16-64 on census day 2011 (N: 174,258)
Source: SLS

Suitability for migration research, underreporting bias and weighting

Having deduced that the matching rate is high at fine geographic scales, it is reasonable to assert that the linking of administrative health data to census based longitudinal datasets represents an avenue of potential advancement in internal migration research. However, as migration is an inherently socially and spatially selective process, it is important to consider which movers and types of moves are underrepresented by this approach. This issue, and how it can be addressed through weighting procedures, is the focus of this section.

Figure 1 shows matching (up to 6 months after census date) rates by age and sex, and confirms that matching rates are poorer for younger adults (especially males) in migration estimates based on administrative health data as they are less likely to register with a new doctor after moving, or take longer to do so (Raymer *et al.*, 2011; Smallwood and Lynch, 2010).

Figure 1: SLS and NHSCR postcode matching (up to 6 months after 2011 census date) rates by age and sex



Sample: SLS study members aged 16-64 on census day 2011 (N: 174,258)

Whilst Figure 1 provides a useful illustration that some moves are likely to be underrepresented in health administrative data, inferential statistics can give a fuller account of the factors that are likely to result in one's mobility being detected in the approach discussed in this paper. Table 2 presents the results of a binary logistic model, which examined the likelihood of an individual's NHSCR GP postcode information matching their census postcode (up to 6 months after census). The model incorporated individual and place based factors, and included an interaction measure for age and sex. The results confirm expectations regarding the significant impact of age and sex on the likelihood of effective address matching between administrative health and census data. Women have much higher matching rates than men, and the mobility of those aged in their twenties is least likely to be detected using this approach. The interaction effect results indicate that men in their twenties in particular are at greatest risk of non-matching. Additionally, unpartnered individuals are less likely to be matched than those who

are partnered, and students display low matching rates relative to those in other forms of employment status, especially homemakers and the longer-term sick and disabled. In line with Barr and Shuttleworth (2012), geography matters in that residents of less deprived areas are in general more likely to be matched than those of more deprived ones, although interestingly the top and bottom quintiles are not statistically different in this respect. Finally, matching rates are relatively low in large urban areas, potentially because residents could change address without necessarily needing to change doctor (at least initially). An exception is very remote small towns and rural areas, possibly because residents of these locations utilise health services less frequently than their less geographically remote counterparts (Hine and Kamruzzaman, 2012).

Table 2: Binary logistic model, predictors of address match between census and health administrative data (up to 6 months after census day)

Variable		**OR**		s.e.
Sex	Male (ref)	1.000		
	Female	1.848	***	0.05
Age group	21-30 (ref)	1.000		
	16-20	2.891	***	0.11
	31-40	1.569	***	0.04
	41-50	3.048	***	0.09
	51-64	5.305	***	0.17
Sex*Age interaction	Male*21-30 (ref)	1.000		
	Female*16-20	0.582	***	0.03
	Female*31-40	1.356	***	0.06
	Female*41-50	1.055		0.05
	Female*51-64	0.809	***	0.04
Partnered	Yes (ref)	1.000		
	No	0.698	***	0.01
Employment status	Full time employed (ref)	1.000		
	Part time employed	1.330	***	0.03
	Unemployed	0.994		0.03
	Student	0.837	***	0.03
	Homemaker	1.689	***	0.09
	Long-term sick or disabled	1.521	***	0.06
	Other (inc. retired)	1.000		0.04
Scottish Government rural-urban Classification	Large urban areas (>125k, ref)	1.000		
	Other urban areas (10k-125k)	1.455	***	0.03
	Accessible towns (3k-10k, within 30 minutes of large settlement)	1.585	***	0.05
	Remote small towns (3k-10k, 30-60 minutes to large settlement)	1.878	***	0.12
	Very remote small towns (3k-10k, 60+ minutes to large settlement)	1.131		0.08
	Accessible rural areas (<3k, within 30 minutes of large settlements)	1.150	***	0.03
	Remote rural areas (<3k, 30-60 minutes of large settlement)	1.184	***	0.06
	Very remote rural areas (<3k, over 60 minutes to large settlement)	1.014		0.05
Scottish Index of Multiple Deprivation	Least deprived quintile (ref)	1.000		
	4th quintile	0.864	***	0.02
	Middle quintile	0.864	***	0.02
	2nd quintile	0.927	**	0.02
	Most deprived quintile	0.958		0.02
	_cons	2.428	***	0.07

Source: Authors analysis of SLS 2011. N=172,721. *p<0.05; **p<0.01; *** p<0.001

The discussion thus far has established that administrative health data linked into census based longitudinal datasets represents a potentially valuable avenue for migration research, but that such approaches suffer from a systematic underreporting of some forms of mobility. As suggested by Raymer *et al.*, (2011) weighting procedures represent a potential means of

overcoming this limitation. This analysis creates weights to demonstrate how the bias created by the underreporting of moves by particular groups can be addressed.

Weighting or data imputation are both possible responses to missing data. Weighting is normally applied in the case of unit non-response (complete absence of a record for an individual at a specific time), whereas imputation usually occurs in instances of item non-response i.e. partially missing data pertaining to an individual (Lynn, 1996). In this case weighting rather than imputation is used to adjust for those people who are missing from the matched postcode subgroup as we know that such people are not missing at random. Having up to date GP registration information will depend on several different characteristics. As was evident in Figure 1 and Table 2, the probability of postcode matching between the census and health administrative data is highly age and sex dependent. As such, weights were created to account for these factors using an inverse probability method, which multiplies units by the distribution of their age and sex defined subgroup in the total sample population divided by the distribution in the sample with matched postcode data. For example, if the subgroup accounted for ¼ of the total sample population and ⅕ of the sample with matched postcode data, then the corresponding weight would be ¼*5=1.25, increasing the presence of this group. A further factor was applied so that the total sample size of the matched postcode subsample after weighting was the same as the size of the total sample. The weights produced by this process are displayed in Table 3 and this type of approach could be of value to other researchers using the SLS for similar purposes. As can be seen, younger adult age groups have higher weights (increasing the representation of these groups), whereas older age groups have values closer to one. For completeness, weights are calculated for both the 2001 and 2011 census years (note the better matching rates in the latter) and scaled weights are included to account for variations in sample sizes in the two censuses. The 'normal' weights refer to sample members that were present in the 2001 or 2011 censuses, whereas the 'scaled' weights are for analyses involving SLS members that were present in both of these censuses. In the latter case the 2001 and 2011 columns refer to weights for sample members of a specific age and sex at the 2001 and 2011 censuses respectively.

Table 3: Illustrative example of age and sex weights used to account for response bias in the health administrative data linked to the 2001 and 2011 censuses

	Normal weights		Scaled weights	
	2001	**2011**	**2001**	**2011**
Males aged 16-20	1.290	1.191	1.303	1.178
21-30	1.595	1.469	1.612	1.454
31-40	1.376	1.258	1.390	1.245
41-50	1.227	1.128	1.240	1.116
51-64	1.164	1.072	1.176	1.061
Females aged 16-20	1.287	1.172	1.300	1.160
21-30	1.359	1.236	1.373	1.223
31-40	1.181	1.092	1.194	1.081
41-50	1.137	1.060	1.149	1.049
51-64	1.122	1.046	1.133	1.035

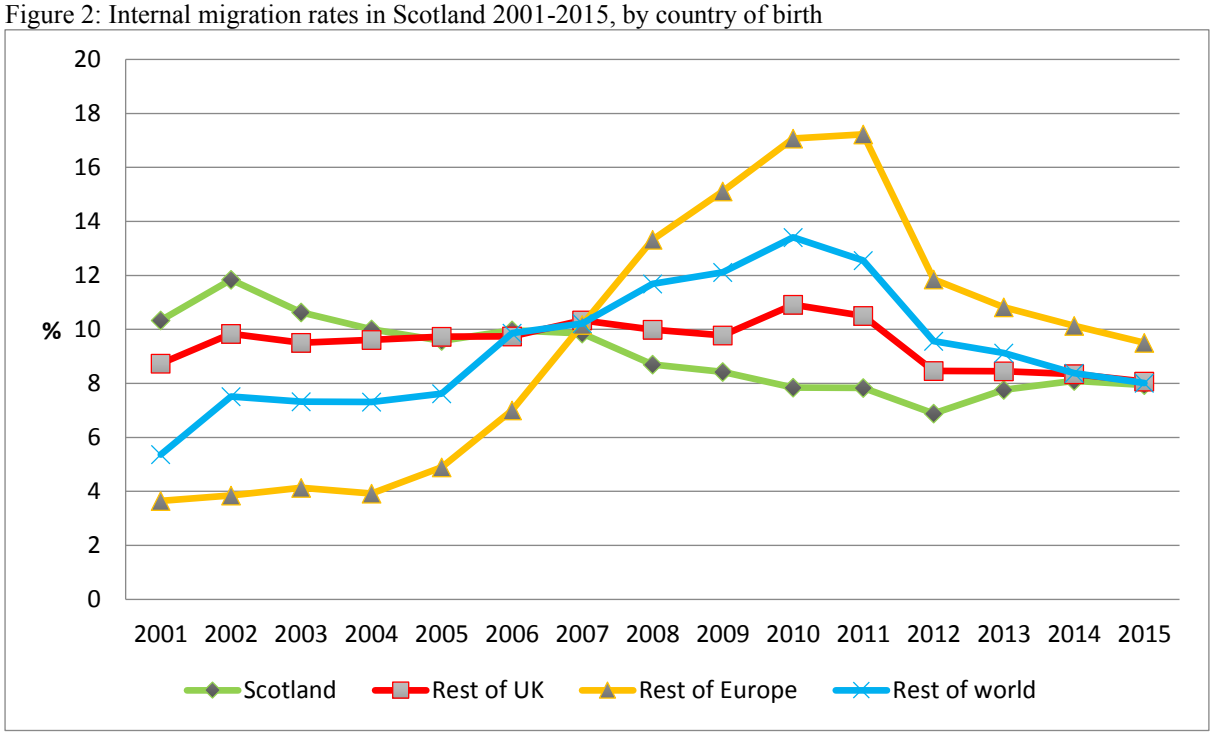
Source: Authors analysis of SLS data

Suitability for migration research, an illustrative example

Having considered the potential suitability of weighted administrative health data linked to a census-based longitudinal study for migration research, the discussion now briefly turns to an illustrative example of its application: a time-series of internal migration trends in Scotland. As touched upon earlier, much uncertainty remains with regard to how internal migration in advanced economies has responded to recent economic change (Green and Shuttleworth, 2015), as well as longer term societal shifts regarding mobility (Champion *et al.*, 2017). The approach discussed in this article can help to shed light on this issue in the Scottish context. Another pertinent issue (see Figure 2) relates to the under researched issue of the internal mobility experiences of international migrants after arriving in their host country (Jivraj *et al.*, 2012; Catney and Finney, 2012). Internal mobility patterns of international migrants may be dissimilar to those of long-term residents with similar characteristics, since the mobility trends of the former cohort may represent part of a longer period of adjustment arising from their initial entry into the country (Trevena *et al.*, 2013). The sample in this study is SLS members aged 16-64 at the 2011 census whose records were traced by NHSCR and whose location according to the census and administrative health data matched on census day or within six months of it. This equates to 151,592 individuals in 2011, with fewer before and after this date (126,755 in 2001; 138,089 in 2015) as people enter and leave the survey through birth, deaths and migration to/from Scotland. The migration rate of this sample is measured annually across 2001-2015 and is defined as the number of moves in a given year divided by the study sample population of that year.

The trends in Figure 2 suggest that Scotland, like many other countries (Champion *et al.*, 2017), is experiencing a gradual decline in rates of internal migration. Scots, who represent over four fifths of the population of Scotland, appear to be becoming gradually less mobile. However international migrants in Scotland seem to have distinct trends. The internal migration rates of EU nationals in particular exceed other groups during and following the 2008 recession. Although not presented here, by allowing for analysis of short distance moves, the research also uncovers interesting differences in residential mobilities between immigrant groups. These preliminary findings are ripe for further, more detailed, analysis. Whilst not the core focus of this paper, these trends provide an apt illustration of the intriguing research themes that the novel methodological approach described here permits.

Figure 2: Internal migration rates in Scotland 2001-2015, by country of birth



Source: Authors analysis of SLS. Sample: in SLS and of working age (16-64) in 2011, census 2011 postcode matches NHS GP postcode on census day and up to 6 months following. N=151,498.
Note the smaller sample in Figure 2 than Figure 1, which is a consequence of country of birth information being missing for some sample members

Conclusions

As detailed in this paper, the recent (2016) linking of health administrative data into the Scottish Longitudinal Study (SLS) presents a valuable opportunity to examine patterns and processes of internal migration. Whilst this innovative approach clearly offers a major

advancement in how migration can be researched in the Scottish context, it also holds lessons for demographic research more widely. Most significantly, it demonstrates the value of data linkage to aid understanding of migratory patterns and processes (ONS, 2009). Combined with ongoing advances in longitudinal datasets and analytical procedures (Findlay *et al.*, 2015), this investigation provides support for efforts to encourage enhanced linking of administrative data into major longitudinal studies. Such moves can facilitate the development of new research agendas. In the field of migration studies for example, administrative data linked to census based longitudinal studies can aid research into such pertinent issues as: the spatial-social mobility nexus (Favell and Recchi, 2012), the relationship between internal migration and business cycles (Saks and Wozniak, 2011), the internal migration of international migrants (Catney and Finney, 2012) and the question of whether there is indeed a fundamental shift towards 'secular rootedness' (Cooke, 2011) in advanced economies.

Finally, it is worth emphasising that the focus on the suitability of administrative data for assessing migration is an area that is likely to become ever more prominent as national statistics authorities begin to move away from traditional censuses. In the UK for example the next census (in 2021) will be predominantly online and will make increased use of administrative data and surveys to both enhance the statistics from the 2021 census and improve statistics between censuses (ONS, 2017a). Given the extensive use of the conventional decennial national census to research migration, the threat of its demise in the future presents some significant challenges to the research community, not least because the approach discussed in this paper would not be possible without census-based longitudinal studies. Going forward, much rests on the quality of administrative data that can inform migration research. In a best-case scenario this could be so rich that it even negates the need for census-based data. However a further challenge relates to the availability of such data. In England and Wales for example, health administrative data would theoretically allow for the same type of analysis as has been described here. However the data are unfortunately unavailable to researchers at the unit postcode scale. In the context of questions over its richness and accessibility, and in the absence of a population register, the utility of administrative data for research in the social sciences is likely to come under increasing scrutiny. These discussions will inevitably involve legal and ethical conundrums associated with data sharing, alongside practical questions relating to how researchers access and utilise these data rich resources. By shedding light on how the linkage

of administrative health data into a census-based longitudinal study can aid migration research, it is hoped that this paper contributes to these timely and significant debates.

Acknowledgements

This research was funded through the ESRC Secondary Data Analysis Initiative (grant number ES/N011430/1). We acknowledge the help provided by the staff of the Longitudinal Studies Centre - Scotland (LSCS). The LSCS is supported by the ESRC/JISC, the Scottish Funding Council, the Chief Scientist's Office and the Scottish Government. We are responsible for the interpretation of the data. Census output is Crown copyright and is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

References

Barr, P. J. and Shuttleworth, I. (2012). 'Reporting address changes by migrants: The accuracy and timeliness of reports via health card registers'. *Health and Place* 18. pp 595-604.

Bell, M. Charles-Edwards, E. Kupiszewska, D. Kupiszewski, M. Stillwell, J. and Zhu, Y. (2015a). 'Internal Migration Data Around the World: Assessing Contemporary Practice'. *Population, Space and Place* 21. pp 1-17.

Bell, M. Charles-Edwards, E. Ueffing, P. Stillwell, J. Kupiszewski, M. and Kupiszewska, D. (2015b). 'Internal Migration and Development: Comparing Migration Intensities Around the World'. *Population and Development Review* 41(1). pp 33-58.

Boden, P. Stillwell, J. and Rees, P. (1992). 'How good are NHSCR data?'. In J. Stillwell, P. Rees and P. Boden (eds) *Migration Processes and Patterns. Volume 2: Population Redistribution in the United Kingdom.* Belhaven Press. London. pp 13-27.

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

Boyle, P. J. Feijten, P. Feng, Z. Hattersley, L. Huang, Z. Nolan, J. and Raab, G. (2009). 'Cohort Profile: The Scottish Longitudinal Study (SLS)'. *International Journal of Epidemiology* 38. pp 385-392.

Castles, S. de Haas, H. and Miller, M. J. (2014). *The Age of Migration: International Population Movements in the Modern World* (5th edition). Basingstoke: Palgrave MacMillan.

Catney, G. and Finney, N. (2012). 'Minority internal migration in Europe'. In N. Finney and G. Catney (eds) *Minority internal migration in Europe*. Ashgate: Farnham. pp 1-12.

Champion, T. (2016). 'Internal migration and the spatial distribution of population'. In T. Champion and J. Falkingham (eds) *Population change in the United Kingdom*. Rowman and Littlefield. London. pp 125-142.

Champion, T. and Shuttleworth, I. (2017a). 'Are People Changing Address Less? An Analysis of Migration within England and Wales, 1971–2011, by Distance of Move'. *Population, Space and Place* 23(3).

Champion, T. and Shuttleworth, I. (2017b). 'Is Longer-Distance Migration Slowing? An Analysis of the Annual Record for England and Wales since the 1970s'. *Population, Space and Place* 23(3).

Champion, T. Cooke, T. and Shuttleworth, I. (2017). (eds). *Internal Migration in the Developed World: Are we becoming less mobile?* Routledge. Abingdon.

Cooke T. J. (2011). 'It is not just the economy: declining migration and the rise of secular rootedness'. *Population, Space and Place* 17. pp 193–203.

Favell, A. and Recchi, E. (2011). 'Social mobility and spatial mobility'. In A. Favell and V. Guiraudon (eds) *Sociology of the European Union*. Palgrave Macmillan: Houndsmills. pp 50-75.

Fielding, T. (2012). *Migration In Britain: Paradoxes of the Present, Prospects for the Future.* Edward Elgar. Cheltenham.

Findlay, A. McCollum, D. Coulter, R. and Gayle, V. (2015). 'New Mobilities Across the Life Course: a Framework for Analysing Demographically Linked Drivers of Migration'. *Population, Space and Place* 21(4). pp 390-402.

Green, A. and Shuttleworth, A. (2015). 'Labour markets and internal migration'. In D. Smith, N. Finney, K. Halfacree and N. Walford (eds) *Internal Migration: geographical perspectives and processes*. Ashgate: Farnham. pp 65-80.

Hattersley, L. and Boyle, P. (2007). The Scottish Longitudinal Study: An Introduction. SLS Technical Working Paper 1. Longitudinal Studies Centre Scotland: Edinburgh/St Andrews.

Hattersley, L. and Boyle, P. (2008). The Scottish Longitudinal Study, a technical guide to the creation, quality and linkage of the 2001 Census SLS Sample. Edinburgh/St Andrews: Longitudinal Study Centre Scotland & General Register Office for Scotland.

Hattersley, L. Raab, G. and Boyle, P. (2007). The Scottish Longitudinal Study, tracing rates and sample quality for the 1991 Census SLS sample. Edinburgh/St. Andrews: Longitudinal Studies Centre Scotland and General Register Office for Scotland.

Hine, J. and Kamruzzaman, M. (2012). 'Journeys to health services in Great Britain: An analysis of changing travel patterns 1985–2006'. *Health and Place* 18(2). pp 274-285.

Jivraj, S. Simpson, L. and Marquis, N. (2012). 'Local distribution and subsequent mobility of immigrants measured from the School Census in England'. *Environment and Planning A* 44(2). pp 491–505.

Lynn, P. (1996). 'Weighting for non-response'. *Survey and Statistical Computing*. pp 205-214.

National Audit Office. (2012). 'Healthcare across the UK: A comparison of the NHS in England, Scotland, Wales and Northern Ireland'. Report by the Comptroller and Auditor General.

ODPM. (2002). Development of a Migration Model. Report prepared by the University of Newcastle upon Tyne, the University of Leeds, and the Greater London Authority/London Research Centre. ODPM. London. Available at: https://www.researchgate.net/profile/Stamatis_Kalogirou/publication/259308761_Development_of_a_Migration_Model_Analytical_and_Practical_Enhancements/links/00b4952af09f4101aa000000/Development-of-a-Migration-Model-Analytical-and-Practical-Enhancements.pdf

Accessed 8th February 2018

Office for National Statistics. (2009). 'A summary of administrative data sources and their potential to inform statistics on migration and population'. Available at: file:///C:/Users/dm82/Downloads/statementofadminsources_tcm77-185902%20(1).pdf

Assessed 8th September 2017.

Office for National Statistics. (2016). 'NHSCR data: quality assurance of administrative data used in population statistics, Dec 2016'. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/methodologies/nhscrdataqualityassuranceofadministrativedatausedinpopulationstatisticsdec2016

Accessed 28th July 2017.

Office for National Statistics. (2017a). 'Annual assessment of ONS's progress towards an Administrative Data Census post-2021' Available at: file:///C:/Users/dm82/Downloads/annualassessment2017v04.pdf

Assessed 8th September 2017.

Office for National Statistics. (2017b). 'Historical representation of UK statistical geographies'. Available at: http://geoportal.statistics.gov.uk/datasets?q=Hierarchical%20Representation%20of%20UK%20Statistical%20Geographies%20(September%202017)&sort=name

Accessed 19th October 2017.

Office for National Statistics. (2017c). 'Population Estimates for UK, England and Wales, Scotland and Northern Ireland: mid-2016'. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/latest

Accessed 28th July 2017.

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

Raymer, J. Smith, P. and Giulietti, C. (2011). 'Combining Census and Registration Data to Analyse Ethnic Migration Patterns in England from 1991 to 2007'. *Population, Space and Place* 17. pp 73-88.

Saks, R. and Wozniak A. (2011). 'Labor Reallocation over the Business Cycle: New Evidence from Internal Migration'. *Journal of Labor Economics* 29(4). pp 697–739.

Sheller, M. and Urry, J. (2006). 'The new mobilities paradigm'. *Environment and Planning A* 38. pp 207-226

Smallwood, S. and Lynch, K. (2010). 'An analysis of patient register data in the Longitudinal Study-what does it tell us about the quality of the data?' *Population Trends* 141. pp 151-169.

Smith, D. Finney, N. Halfacree, K. and Walford, N. (2015). 'Introduction'. In D. Smith, N. Finney, K. Halfacree and N. Walford (eds) *Internal Migration: geographical perspectives and processes*. Ashgate: Farnham. pp 1-14.

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

Stillwell, J. Boden, P. and Dennett, A. (2011). 'Monitoring Who Moves Where: Information Systems for Internal and International Migration'. In J. Stillwell and M. Clarke (eds) *Population dynamics and projection methods*. Springer: London. pp 115-140.

Trevena, P. McGhee, D. and Heath, S. (2013). 'Location, location? A critical examination of patterns and determinants of internal mobility among post-accession Polish migrants in the UK'. *Population, Space and Place* 19(6). pp 671–687.