



How Does Administrative Data Compare with Census Data for the Measurement of Internal Migration? The Case of Northern Ireland, 2001–2011

Brian Foley¹ · Tony Champion² · Ian Shuttleworth³ 

Received: 8 April 2020 / Accepted: 23 November 2020 / Published online: 22 January 2021

© The Author(s) 2021, corrected publication 2021

Abstract

The paper compares and contrasts internal migration measured by healthcard-based administrative data with census figures. This is useful because the collection of population data, its processing, and its dissemination by statistical agencies is becoming more reliant on administrative data. Statistical agencies already use healthcard data to make migration estimates and are increasingly confident about local population estimates from administrative sources. This analysis goes further than this work as it assesses how far healthcard data can produce reliable data products of the kind to which academics are accustomed. It does this by examining migration events versus transitions over a full intercensal period; population flows into and out of small areas; and the extent to which it produces microdata on migration equivalent to that in the census. It is shown that for most demographic groups and places healthcard data is an adequate substitute for census-based migration counts, the exceptions being for student households and younger people. However, census-like information is still needed to provide covariates for analysis and this will still be required whatever the future of the traditional census.

Keywords Internal migration · Administrative data · Reliability · Census

✉ Ian Shuttleworth
i.shuttleworth@qub.ac.uk

Brian Foley
brian.foley@nisra.gov.uk

Tony Champion
tony.champion@newcastle.ac.uk

¹ Northern Ireland Statistics and Research Agency, Belfast, Northern Ireland

² Newcastle University, Newcastle, United Kingdom

³ Queen's University Belfast, Belfast, Northern Ireland

Introduction

Many countries across the globe, the UK (ONS 2018), the USA (US Census Bureau 2019), and New Zealand (Statistics New Zealand 2014) included, are considering supplementing or replacing their traditional household censuses with administrative data. Censuses currently provide a wide and diverse range of data and outputs to which users have become accustomed. These include population counts for a hierarchy of geographical units; attribute data on individuals and households; microdata on individuals and households; and topic files on themes such as travel-to-work and migration.

An important series of questions concerns the extent to which administrative data sources, either singly or linked together, can meet the needs and expectations of users accustomed to data from national censuses. Administrative data sources may be superior in some senses to censuses; they are updateable annually or even more frequently in contrast to national censuses which for the vast majority of countries happen only every ten years. They are already used to provide annual estimates of internal migration in the UK, for example, for local authorities by age and sex (Nisra 2014). On the other hand, administrative sources may be weaker than the census in other ways; healthcard-derived data, for instance, lack socio-demographic attribute information; it may be difficult to capture information on households and there may be biases in the information that is collected that result from the way that system captures data from people or the ways in which they engage with it. There are clearly pros and cons in the use of both census and administrative data and assessing them is made difficult by the range of themes needed by users and also by the types of data they require.

This paper makes a UK contribution to this assessment by using the Northern Ireland Longitudinal Study (NILS) to compare internal migration, measured by the Northern Ireland Census with that recorded by the country's healthcard register. The overall aim is to assess the extent to which healthcard data could provide the flow and micro- data for academic use that now comes from the census. Firstly, transition data on address changing between the 2001 and 2011 Censuses is related to events-based data from the healthcard register for a very similar 10-year period. Secondly, the analysis assesses how counts of total in-migration, out-migration and net migration for the decade between 2001 and 2011 for the country's Super Output Areas (SOAs) differ between the two sources. Finally, the socio-demographic and geographical associates of changing address, comparing the healthcard data with the census, assuming that the latter constitutes the 'gold standard', are analysed. These analyses are replicated for 2010–2011 using the census one-year ago address question as a benchmark. It is concluded that neither healthcard nor census data give a complete picture of internal migration and that they are most effectively used in combination, although this may be impossible after 2021 in the UK should the traditional census be abandoned – in which case, the rich range of covariates provided by the census will be hard to replace.

Counting Migrations and/or Migrants?

Migration is usually recorded in two ways – either as *transitions* between the places or addresses where people were living at the start and end of a given

time period or as *events* over time (Bell et al. 2002, 2015). An example of a transition is the one-year migration question in the three national censuses of the UK, which asks if a person's address at the time of the enumeration was different from that of one year ago. Migration is thus defined as a transition between addresses or geographical units. Most usually in national censuses it is this one-year transition that is captured but five-year transitions are collected in some countries and other periods in a few (Bell et al. 2002). It is also possible to compare place of enumeration in successive censuses to estimate a ten-year transition. Champion and Shuttleworth (2018), used this approach to measure 10-year migration for successive intercensal decades in order to identify long-term trends in internal migration rates in England and Wales, 1971–2011, using the Office for National Statistics Longitudinal Study (ONS LS). Migration events, on the other hand, occur every time that a person changes address or registers themselves in a different administrative area and are commonly recorded in administrative and population register data such as those in Iceland, Sweden, and the Netherlands (Bell et al. 2015). Rather than providing data on a one-off transition for a period, they allow the possibility of multiple migration events per person over the given period. Transition data undercount migration because there may be multiple migrations within a transition from one time to another; and the longer the period over which the transition is measured, the greater the undercount is likely to be.

The ability (in theory) of administrative data to count mobility fully is one of their major strengths for migration research, but there are also significant weaknesses in their use. These include a lack of attribute information. For example, in healthcard data in the UK, all that is commonly available as covariates are age and sex, and not the full range of information on personal characteristics asked for in the census. Additionally, there will be time lags in reporting address changes to healthcare professionals (Barr and Shuttleworth 2012; Champion and Shuttleworth 2017), especially by younger men (Ogilvy 1980). Furthermore, it is suspected that younger men – and transient people more generally – fail to register with health services (or for that matter to deregister) leading to uncertainties as to who is actually present at an address or in an area (Bell et al. 2015; Ernten et al. 2018; ONS 2018). There may thus be an undercount of moves with some address changes being reported late or not at all, with some people who fail to register being missed altogether. At the same time, there may be 'ghosts'; people who have left an area or even departed the country yet who remain in the administrative data as an overcount. Stillwell et al. (1994) also note that normally healthcard data remains restricted to higher-level local authority or regional geographies such as those used by Champion and Shuttleworth (2017) although this is not the case in the data assessed in this paper.

At the same time, census data also suffer some weaknesses for migration research, although their strength and the balance of importance differs from those experienced with healthcard data. Although the census aims to attain a 100% coverage of the population, in practice it suffers some degree of underenumeration. This is concentrated in certain types of household, certain types of people, and certain types of places (Shuttleworth and Martin 2016). Hard-to-enumerate groups typically include younger people (especially men), more spatially mobile people, immigrants, dwellers in communal or other establishments, and residents of socially-deprived neighbourhoods,

especially in urban areas. Attempts to deal with this problem include greater enumeration effort for groups and areas seen as being ‘hard-to-count’ and statistical modelling to impute missing individuals and households using information from the Census Coverage Survey which is a follow-up to the census. An additional problem might be recall bias. It is likely that this would be much more of a problem over five or ten year intervals rather than the one-year question currently asked in the UK Censuses. However, for highly-migratory people, it might prove to be a problem over even a one-year interval.

One particular problem with census-derived migration data is that the one-year migration question, by definition, provides information only for the year immediately prior to the census, thus yielding no information on the other nine years of an intercensal decade. This can lead to biases, notably those linked with cyclical events such as economic boom and bust. The 1981 Census (with one-year migration data 1980–81) and the 1991 Census (with similar migration data for 1990–91) took place in economic troughs, whereas the 2001 Census took place in a comparatively affluent period. This makes it hard to compare change in migration levels through time, though looking at 10-year transitions between two censuses, for example between 1981 and 1991, helps to average out these types of shorter-term ‘period’ effects.

The census, however, has considerable strengths; it collects data on a range of socio-demographic, housing and labour market variables that are known to be associated with migration (Green 2018) – single administrative data sources do not routinely have this and will be a challenge even with the linkage of multiple administrative datasets. The census can also be used to produce origin-destination matrices for a range of geographies, and microdata resources for migration analysis at household and individual level, as well as the standard aggregate statistics for a variety of output geographies. At the minute, administrative data sources cannot do all these things although the provision of flow matrices is theoretically possible as is the creation of microdata – but was impossible in this case because of NLS output rules which prevent small counts being released. The most difficult thing to provide with data like the healthcard register will be the full range of individual and household covariates captured in the census. Healthcard data, for instance, only currently provides age and sex.

The ambition of the Office for National Statistics (ONS) is to move towards an administrative data census (supplemented by regular surveys) after 2021 in England & Wales; it remains to be seen how far the statistical agencies of the devolved countries of Scotland and Northern Ireland will follow this path. ONS has developed a work package to evaluate how far administrative data can be used to provide population estimates and counts for areas (ONS 2019). These have merged various administrative data sources and considered features such as recent activity to adjust for overcount (eg people registered but no longer present). Doing this, they have approximated population counts taken from the census (ONS 2019). As was seen earlier, healthcard data are also routinely used to estimate internal migration flows between local authority areas by age and sex and, because of this, they feed into the provision of intercensal population estimates. However, there is quite a gap between this, and the provision of the full range of migration census products such as flow

and micro- data that are provided at the moment by the census and to which academics are accustomed. The purpose of the paper is thus to move towards this type of evaluation using Northern Ireland as an example – feasible here because of the range of data linkage that is routinely done – that is closely related to the rest of the UK.

Northern Ireland and the NILS

Just as in the rest of the UK, healthcard registration data are used to inform annual estimates of internal migration between various statistical geographies in Northern Ireland (NISRA 2018). It is augmented with higher education data from the Higher Education Statistics Agency (HESA) to collect information on students. It is recognised that list inflation may lead to an overcount of people in the healthcard register and, moreover, that migration flows for some population groups (primarily younger people) are likely to be underestimated because of late or non-registration. NISRA has made considerable progress in data linkage and administrative data were used in the quality assessment of the 2011 Census. Current plans (NISRA 2019) are to use administrative data to inform the design, quality and outputs of the 2021 Census, although it is unclear how far NISRA will follow ONS towards the 2031 census date.

The NILS is part of this Northern Ireland environment of data linkage. It is a census and administrative data study that routinely links census, vital events and Valuation and Land Registry data to a healthcard spine. Other information normally available includes pollution and weather data. The healthcard data start in April 2001 and are updated with downloads every six months. These report address changes as notified to healthcare professionals – in practice, usually general practitioners. As mentioned above, there can be temporal lags in reporting (see Barr and Shuttleworth 2012). The 2001 and 2011 Censuses are linked to the healthcare spine and the 1991 and 1981 Censuses are also available. The NILS sample, drawn from the healthcard register, is relatively large at approximately 28% of the Northern Ireland (NI) population, with its members selected on the basis of 104 birthdates which include the four birthdates of the England and Wales longitudinal study (ONS LS) and the 20 birthdates of the Scottish Longitudinal Study (SLS). The NILS sample is dynamic and continually changes as people with one of the 104 NILS birthdates either join the Northern Ireland population through birth or immigration or else exit it through death or leaving the country. It is also updated as more and better information becomes available that permits, for example, match rates to be improved or more address information – see below – to be added.

The sample has excellent population coverage because, given the nature of the UK healthcare system and the prominence of the National Health Service (NHS), very few people do not have an NHS registration. In constructing longitudinal databases, the same full coverage is apparent; one of the strengths of census-based longitudinal studies is their low attrition rate. In making the link to the 2011 Census, for instance, a match rate of 97% was achieved after excluding people who had died or left Northern Ireland during the intercensal decade, and those about whom there were other uncertainties (NILS 2013). In this latter group, there are coverage issues. In 2011, for example, there were more healthcard records than individuals enumerated in the census. This list inflation arises from people who leave a GP and do not de-register;

this was estimated as about 22,000 cases (NILS 2013). Likewise, there are people who are not enumerated in the census but who are imputed in census data processing. By definition, imputed people cannot be linked; this number was estimated as being about 26,000 (NILS 2013). Analysis shows that there are more imputed individuals in high-deprivation areas and among young people, especially males, reflecting known patterns of under enumeration (Shuttleworth and Martin 2016). Nevertheless, match rates even for younger men are at 90% or higher. The healthcard data record address changes every six months. They lack the full range of covariates necessary for internal migration research with only age, sex, and geography available. The census, on the other hand, has the range of individual, household and spatial covariates to which migration researchers are accustomed, but has restricted information on migration with only the one-year question. Moreover, although both data sources cover the vast majority of the population they are not 100% complete. However, they are the very best that are available and can be expected to be powerful when used in combination.

In the NILS migration can be conceptualised in different ways through the data drawn from the health register and the census. The healthcard data captures address point information from which migration is identified as a transition between two XUPRNs – anonymised unique property reference numbers. This can be used to look at all address changes or can be recompiled flexibly to measure distances moved over bands specified by researchers or between different official statistical output geographies. The data from the one-year migration question collected by the census can be used to measure all address changes over the previous 12 months as well as flows between official geographies. In addition, in the NILS it is possible to compare a person's XUPRN in 2001 with that in 2011 and thereby measure a decade-long address transition.

All the considerations set out above inform the selection of the NILS sample that we use for the analysis reported in this paper. Since the intention is to compare healthcard data with the census, only NILS members with census records were extracted. Given the aim to consider one-year moves between 2010 and 2011 but also transitions between 2001 and 2011, the sample was further restricted to NILS members with records in both these censuses. This meant that it excluded people who were born or arrived in Northern Ireland between 2001 and 2011 and people who died or who left it over the decade.

These considerations also help to fashion our research approach. The analysis starts by considering all 372,785 individuals who meet these criteria to assess overall similarities and differences between the data sets in terms of what they tell us about migration behaviour. It then concentrates on just on counts of movers into and out of SOAs for 2001–2011 and compares the geography of flows identified from the transitions between SOA of enumeration in 2001 with that of enumeration in 2011 with the healthcard data coded by SOA for this same period. Finally, it models the socio-demographic and geographical associates of making any address change between 2001 and 2011 in the two data sources to explore how far healthcard data replicate the relationships observed in census data. This latter analysis is performed using a binary logistic regression (change address at least once 2001–11: yes/no); uses 2011 Census data; and is restricted to those aged 16–74 in 2011 to cover the prime age group and include all address changers and not just those who moved between SOAs. The model includes age, sex, housing tenure, household type, NS-SEC, country of birth, urban/rural location, and social deprivation – the last two variables being specified from

location in 2011 – as independent variables. This was a pragmatic choice given the known associates of migration (Green 2018) but also one that permits analysis of how well the healthcard and census data match each other. All analyses above are replicated using the one-year ago address question and 2010–2011 healthcard data to examine moves just in that year; these results are presented in the [supplementary material](#) and are very close to those for the ten-year time interval.

Results

Comparing Census-Based Migration Transitions with Healthcard-Based Migration Events for 2001–2011

The analysis starts by looking at address changes measured as a transition in census data, over the decade between 2001 and 2011 and comparing these with the event-based migration data for this period provided by the healthcard register. From the census-derived data shown in Fig. 1, migration is conceptualised as a NILS member being enumerated at a XUPRN in the 2011 Census that is different from that in the 2001 Census. Some 211,000 – or 57% – of NILS members are at the same XUPRN in both censuses compared to some 144,000 (around 38%) being at different addresses. It is impossible to locate the remaining 5% and examine their transition status because their XUPRN was missing in either one or both censuses. This is a small point but illustrates the ‘noise’ even in official datasets, in this case the lack of property information for those who were enumerated in the census and who therefore presumably must have been located somewhere in Northern Ireland then.

Figure 2 considers address change events over the same decade using the healthcard data. As measured by these data, almost three-fifths (59%) of our NILS sample were at the same XUPRN in April 2011 as in April 2001, i.e. a slightly higher proportion than the census gives. The remaining two-fifths made at least one address change but it

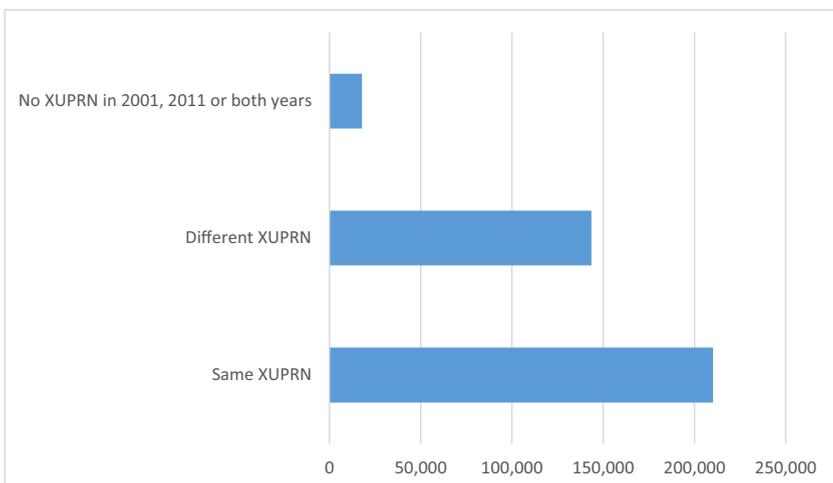


Fig. 1 Address transitions of the selected NILS sample between the 2001 and 2011 censuses. Source: NILS

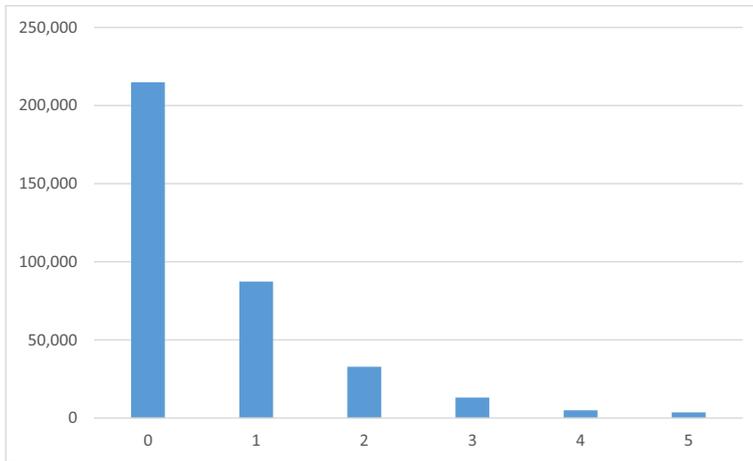


Fig. 2 Address change events of the selected NILS sample recorded by the healthcard register between April 2001 and April 2011. Source: NILS

ought to be noted that most of the movers (84% of them) made only one or two address changes – the highly mobile are a small minority. In all, there were almost 144,000 address transitions measured in the census between 2001 and 2011 compared to nearly 251,000 address-change events reported by healthcards. This is an excess of 107,000, clearly demonstrating the extent to which transition data can mask intervening migration events over this length of period.

Table 1 provides more detail on how census measures of migration – in this case transitions between XUPRNs in 2001 and 2011 – compare with the migration events reported by healthcards. Table 1 tabulates the number of healthcard address-change events for people with the same census XUPRN in 2001 and 2011, i.e. those who had made no move according to the census. While this shows considerable overlap with

Table 1 Matches and mismatches between 2001 and 2011 between the census and the Healthcard register

	Number of healthcard address changes 2001–2011		Number of healthcard address changes 2001–2011	
	Non-movers according to the census		Movers according to the census	
Number of moves	Frequency	Percentage	Frequency	Percentage
0	187,104	88.6	26,415	18.4
1	19,696	9.3	66,593	46.3
2	3278	1.6	29,600	20.6
3	726	0.3	12,511	8.7
4	190	0.1	4975	3.5
5	54	0.0	2127	1.5
6+	17	0.0	1676	1.2
Total	211,065	100.0	143,897	100.0

Source: NILS

88.6% making no move over the decade according to the healthcard register, this means that around 11.4% had made at least one move. Some NILS members might, of course, have moved back to the same address as they had in 2001, but to do this they would have had to have moved at least twice – away from the 2001 address and then back to it – but the vast majority of these cases (9.3% of the 11.4% or around 82% of this group) had only a single move recorded in their healthcard data. It is thus safe to assume that there is a mismatch between the two data sources. If this NILS sample was grossed up to the full population, this would mean that around 86,000 people would have recorded questionable migration behaviour according to this measure.

An alternative measure in Table 1 shows how many moves were reported by healthcards for those NILS members had different XUPRNs in the 2001 and 2011 Censuses. In this case, the anomalous group are those who reported no healthcard address change but were enumerated by the census at a different address in 2011 compared with that in 2001. This is 18.4% of this group, grossing up to some 94,000 of the total population. Given what is known about lags in reporting moves, and also the socio-demographic background of these lagggers (Barr and Shuttleworth 2012; Foley et al. 2018), it is very likely that a substantial proportion of this 18.4% is made up of address changers in 2010–11 who did not report their move to their GP within the year. Some will report their 2010–11 address change in later healthcard downloads but some may never report it before they make their next move. These lagggers have a strong likelihood of being young, male and healthy and also quite mobile. However, almost 82% of those with different census XUPRNs did report one or multiple moves in the healthcard register again showing that there is a substantial overlap in whom the two datasets cover.

Comparing the Geography of between-Area Migration Flows, 2001–2011

A more challenging test is to assess how similar the census and healthcard register are in capturing the geography of internal migration – an important question because this is one of the primary uses of internal migration data. Attention is given to flows between Super Output Areas (SOAs), a commonly-used statistical output geography in the UK. In Northern Ireland, SOAs are designed to have approximately 2000 residents on average and to be homogenous in housing tenure and as spatially compact as possible (Cockings et al. 2011). As they are designed to be of a standard population size, they vary in spatial extent as a function of population density being much lower in rural than in urban areas.

Using this geography, Figs. 3, 4 and 5 compare census-derived and healthcard-derived data for the absolute numbers of migrants leaving each SOA for another part of Northern Ireland, the numbers moving to each SOA from the rest of the country, and the net balance of these movements for each SOA. The census counts are derived from the address of enumeration in the 2001 and in the 2011 Censuses and the healthcard counts from transitions reported between April 2001 and April 2011. The whole SOA-to-SOA migration matrix is not analysed because of its sparsity in non-diagonal cells. This would mean a high number of counts below the NILS minimum threshold for confidentiality of 10; hence the decision was made to concentrate on the total inflow and outflow counts aggregated for each of the 890 SOAs and compared between the two data sources.

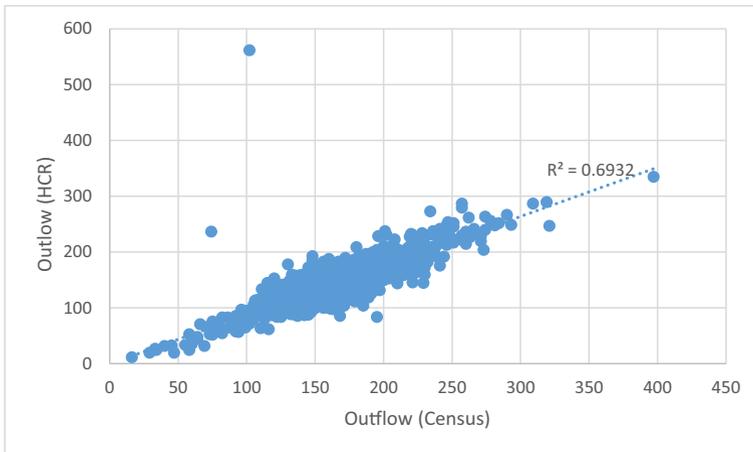


Fig. 3 Aggregate outflows between 2001 and 2011, census compared with Healthcard ($N = 890$ SOAs)

Figure 3 suggests that there is a positive and linear relationship between the counts of migrants leaving a SOA (i.e. total outflow to the rest of Northern Ireland) between the healthcard and census counts. Almost 70% of the variation in healthcard counts across SOAs is ‘explained’ by the census count. However, this is exceeded in Fig. 4 which compared the counts of inflows to each SOA. In this case the R^2 is almost 90%: there is a very close correspondence between the healthcard and census counts. There is also a close, but not perfect match, between healthcard and census counts in Fig. 5 where the two sets of net inflows are plotted. Overall, it therefore seems that the counts of in-migrants and out-migrants over the decade based on the SOA of census enumeration in 2001 and 2011 and the location recorded in the healthcard register are very close. On the other hand, the relationship breaks down, sometimes in an extreme way, in a small number of cases. This happens more markedly for the outflows shown in Fig. 3 and the net flows of Fig. 5, much less so for the inflows shown in Fig. 4. Spatial analysis shows that these SOAs are characterised by transient populations such as

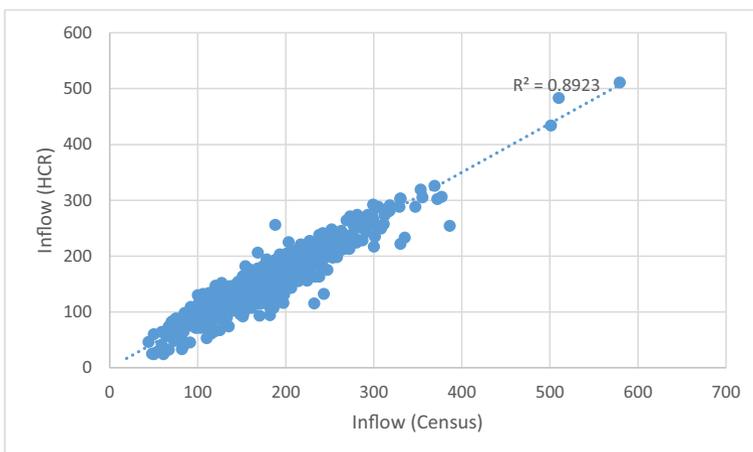


Fig. 4 Aggregate flows to SOAs between 2001 and 2011, census and Healthcard ($N = 890$ SOAs)

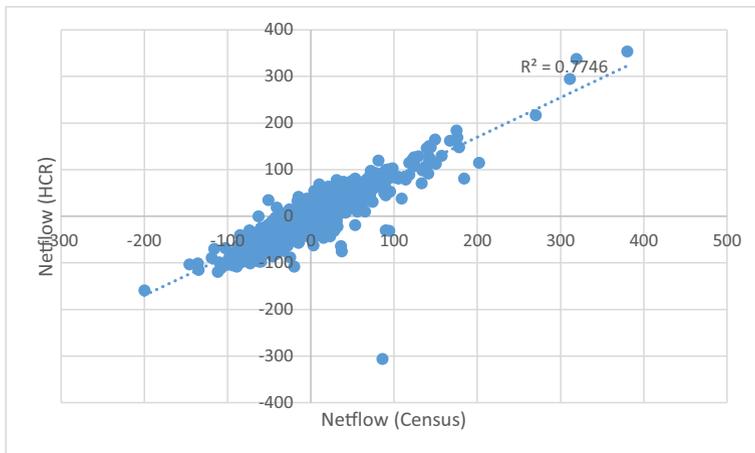


Fig. 5 Net flows between 2001 and 2011, gross migrant inflows minus gross migrant outflows, census and Healthcard register (N = 890 SOAs)

students, such as those in Stranmillis, in the South of Belfast. In this situation, the healthcard count is higher than the census count, particularly for individuals originating in this kind of SOA. Some possible reasons for this are considered later.

We have also replicated this analysis for migration based on the address one year previously in the 2011 Census and a comparison between location in the healthcard register in April 2010 and April 2011 (see [supplementary material](#)). It confirms the relationships observed for the ten-year SOA-SOA transition data except that the relationship between the net count based on the census is closer to the healthcard register for the decadal analysis than for the one-year comparison (R^2 is only 0.33 for the latter). One probable reason for this is that lags in registering some address changes via healthcards may bias annual data but that they are averaged out in a longer period.

Comparison of Census and Healthcard Data for Odds of Address Moving 2001–2011, by Type of Person and Area Lived in

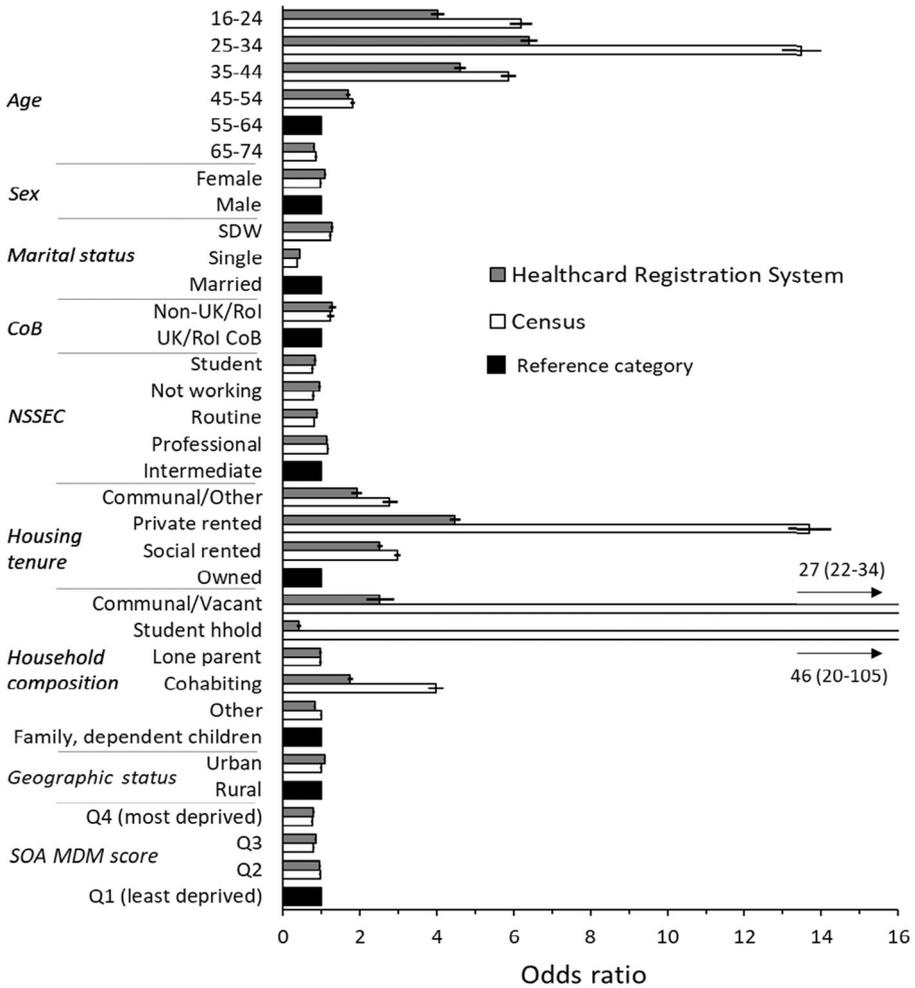
The final section of the analysis is concerned with how far the healthcard register data on address changes replicates census data in terms of the associations between independent variables and a migration outcome – in this case, changing address between 2001 and 11. This is another way to assess and to validate administrative data against the census, again on the assumption that the latter is an accurate benchmark. Looking forward, it also provides a way to consider how far the relationships between the types of independent variables often used as explanatory factors in migration analysis might be replicated in a future administrative data system. Of course, as already mentioned, the healthcard register, as the situation stands at present, has information on only age, gender and geographical location and so the analyses that researchers might want to undertake would be impossible using it alone; it is only the ability to link the census to healthcard data in the NELS environment that has made the analysis possible. It must also be noted that the independent variables

are taken from the 2011 census, so that they represent the after-move circumstances for those who moved during the decade.

Figure 6 provides a graphical summary of the odds ratios of changing address over the ten years between 2001 and 2011. The reference category for each set of independent variables is in black. The full results in tabular format are available in the Table 2 in the Appendix. The big picture is plain; in general, the odds ratios from the healthcard data are close to the census results. In showing the same patterns, relative to the respective reference categories, the results for both measurements of address changing are close to what would be expected given the literature on migration – younger people were more likely than the middle aged to have changed address, and older people less so; those who were separated, widowed and divorced in 2011 were more likely than the married to have changed address in the past decade; professionals were slightly more mobile; and there were big effects associated with housing tenure and household type, with both private and social renters being more mobile than owner occupiers in both datasets and those who were students or living in communal establishments in 2011 having been notably more mobile over the previous decade than families with dependent children, as would have been hypothesised. None of the results seemed anomalous in the sense that neither data set contradicted the other.

For some variables, however, there were large differences between the data sets in the magnitude of effects, if not their direction. In declining order of size these were for communal establishments, student households, private renters, people aged 16 to 34, and cohabiting couples. In each of these cases, the healthcard register data underestimates the size of the effect relative to the census and the relevant reference category. Excluding the massive effects for students and those living in communal establishments allows smaller but statistically significant differences to be seen. There are differences by gender; also for the non-UK born; for those not working; for those in urban locations; and also for those in the top two deprived quartiles of SOAs. These, though statistically significant, are substantively small differences and the relative propensities to have made an address change during the decade prior to 2011 are consistent between the healthcard and census data for all practical purposes. In other variables/categories such as for older age groups and country of birth, the confidence limits overlap so there is no statistically significant difference between the Census and the healthcard register. Overall, however, the types of people identified where there are large mismatches between the census and the healthcard register – for example, the young, private renters, and students – match well with the urban location of those outlying studentified SOAs seen in Figs. 3 and 5 which are characterised by these demographic groups.

The above analysis was repeated using the one-year ago address question in the Census (which was compared with location as given by the healthcard register in April 2010 and April 2011). This was done for the sake of completeness but also because this time period is nearer the demographic data collected from the 2011 Census in date. This is presented in the [supplementary material](#). The results are reassuring in the sense that they closely resemble the decadal analysis that has been presented; the largest differences between the census and the healthcard register are, again, for young people (the healthcard register appears more likely to miss younger movers than the census); private



SDW = Separated/Divorced/Widowed
 CoB = Country of birth (Rol = Ireland)
 NSSEC = Socio-economic classification
 SOA MDM = Super Output Area Multiple Deprivation Measure score (Q = quartile)

Fig. 6 Address changes in 2010–11 in the census and the healthcard register: Binary logistic regression model, $n = 340,647$. Source: NILS

renters, communal dwellers, and student households (all these being more likely to be missed by the healthcard register, it seems).

Discussion

The analyses reported above sought to go beyond those undertaken so far by UK statistical agencies to consider, for academic use, how far the geographical patterns of internal migration from the healthcard register match those from the

census benchmark and also how far the relationships between socio-demographic background and migration was maintained when using a healthcard measure of migration rather than a traditional census-based metric. The analyses reported above find a slightly mixed picture with regard to the comparability and use of migration information from administrative data sources, such as the healthcard system, in contrast to more traditional migration data from the census. The balance of the assessment depends on two issues; first, on the precision of measurement required and, second, on conceptual questions about how statistical data collection operates and what can be known (and by implication what is not and cannot be known).

Dealing with the first issue about the degree of precision with which healthcard measurements of internal migration replicate those provided by the census, our analyses have shown that in general, for most places, and demographic groups, they do. For the majority of SOAs, numbers of in-migrants are very similar in the two sources, as they are too for out-migrants. However, in a few cases there are big differences between the two datasets. Typical of these are studentified locales in places such as South Belfast. Here the relationship between census and healthcard counts breaks down. This accords with the logistic regression analysis of the socio-demographic correlates of moving in 2001–2011 and 2010–2011 where those demographic groups (younger people, private renters, cohabiters, communal dwellers) typical of these areas differ most between the two datasets. It should be remembered that these statistical outliers are geographically and demographically concentrated and are important for only a small absolute number of the population (albeit one that is important in population estimates for small areas). In all cases, the *relative* difference between variable categories and the reference category is maintained; the census and healthcard data show the same overall pattern and with large *absolute* differences for only the restricted places and demographic groups which we have identified. For the most part, differences between the census and the healthcard register are small in absolute terms and, even though statistically significant, in practical terms they do not amount to much.

However, in some cases, census hard-to-enumerate groups, such as young people aged 16–24 and students, overlap with those who do not interact with the healthcare system and efforts either through surveys, other fieldwork, or statistical modelling will be required to address these gaps in coverage. In the possible absence of a census, though, it is hard to see how these measures will be put into practice to validate estimates unless there is more administrative data linkage which incorporates a geographical element. So, the answer to the question about whether administrative data from the healthcare system can replace census data for internal migration research is a qualified ‘yes’; broadly the healthcard data tell us the same thing as the census data whether dealing with population flows or the demographic correlates of migrating. There is just the proviso that internal migration for some socio-demographic groups (and some places) will be wrongly measured.

The second issue raised above concerns how data collection operates and how a complex social reality is partially – although it is hoped as near as possible fully – captured in population-level statistics which are themselves an artefact of their

data collection and output processes. The analysis here has privileged the census as the gold standard against which the health register migration data can be compared. However, the census does not enumerate all people and households (although it covers the vast majority). It asks relatively few questions about internal migration and can be used to record transitions between only two types of periods – the year before the census or else the previous census ten years earlier. The healthcard register, in contrast, has a much higher routine population coverage which exceeds 99% and has the capability of recording multiple migration events. On the other hand, individuals lag in reporting address changes to their GPs and in a small number of cases do not report them at all. In some cases, it seems that the healthcard register and the census has the same types of problem with the same demographic groups in approximately the same places – for these it is difficult to arrive at an agreed ‘truth’, and to decide which ought to be the benchmark. Care will be needed in any future administrative data based population system to capture migration information for some of the groups who are most mobile and in whom there is the greatest statistical interest.

Each system of data collection therefore has its own strengths but also its own weaknesses. Used together, they each offer unique contrasting perspectives on a social reality which can never be fully measured; neither capture an unattainable 100% accurate picture of internal migration but each can be reasonably judged and used with an understanding of what each offers. In a world where the census continues, it is possible to imagine how administrative data could augment the migration data that are currently collected by the census; intercensal migration events if linked to census data and provided in standard statistical outputs would add considerable value. It is possible, however, to envisage a world without the census and where there are only administrative data. One scenario would be the provision of annual migration event data that could be summed over several years to create a small area-to-area flow matrix as well as migration counts but with restricted availability of covariates (perhaps only age group and sex). A more generous scenario would be where additional linkage provided more covariates and the production of census-like crosstabulations for usual statistical output geographies as well as microdata for migration and other analyses. This would not be the *same* as what is currently provided but would be comparable.

Conclusion

The paper set out to investigate how far healthcard data could begin to replace the census for academic use with a focus on spatial data and on individual-level microdata. The chief finding of the analyses that we have undertaken for Northern Ireland is that healthcard administrative data could replace the census for internal migration research in many situations and for most areas. In reporting multiple address-change events at twice-yearly intervals, administrative data are clearly superior to census data, especially if it is possible to recode finely-referenced spatial data to allow migratory distances to be analysed flexibly. This is with the caveat that special attention will need to

be paid to groups where healthcard data seems to have gaps; student and communal households, cohabiters, private renters, and adults younger than 35. These fall into the hard-to-enumerate categories that at the moment pose difficulties in censuses and other government surveys. Therefore, something akin to the Census Coverage Survey will be annually (or twice yearly given the speed of updates) needed to impute missing address changes in flow data and to calculate weights for use with administratively-derived microdata.

The greatest problem likely to be faced by migration researchers if and when the traditional census is abandoned, however, will be in providing the individual and household covariates that are currently provided by the census. Information on housing tenure, for example, is not available in the Northern Ireland healthcard register and this is also the case for other variables such as educational attainment. The analysis reported in this paper would have been impossible without the census; in a post-census future the only (partial) replacement would be the extensive routine linkage of multiple administrative datasets on social benefits, house rating valuations, and education, and the preparation of a statistical output system to provide the full gamut of current outputs including counts for official geographies, flow data as used in analyses of migration and commuting, and individual-level microdata. It is possible that a major difficulty will be in obtaining household relationship data of the kind currently provided by the census as well as information on topics such as religion and ethnicity. These are not commonly available in administrative data.

Therefore, in moving from the known world of census data to the less known world of administrative data, it is vital to stick with the familiar until researchers can be reassured that administrative data offer the same information and quality as census data, or at least that their strengths offset their weaknesses in other areas. The ability to have annual or six-monthly updates of migration events from administrative data is a clear advantage and confidence can be drawn from the way that it replicates census data for differing time periods for most of the population. However, both datasets have to grapple with hard-to-enumerate groups, especially younger adults and student households. More generally, the whole apparatus of census data collection and processing mean the census may still be more effective than the healthcard register, while its information on the correlates of migration is at present unrivalled.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12061-020-09369-w>.

Acknowledgements The help provided by the staff of the Northern Ireland Longitudinal Study (NILS) and the NILS Research Support Unit is acknowledged. The NILS is funded by the Health and Social Care Research and Development Division of the Public Health Agency (HSC R&D Division) and NISRA. The NILS-RSU is funded by the ESRC and the Northern Ireland Government. The authors alone are responsible for the interpretation of the data and any views or opinions presented are solely those of the author and do not necessarily represent those of NISRA/NILS.

Compliance with Ethical Standards

Conflict of Interest The authors declare they have no conflict of interest

Appendix

Table 2 Odds ratios, coefficients and standard errors and confidence limits underlying Fig. 6

	HCR				Census				
	Coeff	Odds Ratio	Std. Err.	[95% Conf. Interval]	Coeff	Odds Ratio	Std. Err.	[95% Conf. Interval]	
Age									
Ref: 55–64		1.000				1.000			
16–24	1.391	4.019	0.091	3.844 4.202	1.822	6.182	0.146	5.903 6.475	
25–34	1.856	6.399	0.116	6.177 6.630	2.602	13.487	0.259	12.989 14.004	
35–44	1.529	4.611	0.077	4.464 4.764	1.770	5.868	0.100	5.676 6.067	
45–54	0.530	1.699	0.026	1.649 1.752	0.592	1.808	0.028	1.753 1.864	
65–74	−0.218	0.804	0.015	0.776 0.833	−0.156	0.856	0.016	0.825 0.887	
Gender									
Ref: Male		1.000				1.000			
Female	0.095	1.100	0.010	1.081 1.118	−0.015	0.985	0.009	0.968 1.003	
Marital status									
Ref: Married		1.000				1.000			
Single	−0.824	0.439	0.007	0.426 0.452	−0.986	0.373	0.006	0.362 0.385	
Separated, Divorced, Widowed	0.246	1.279	0.019	1.242 1.317	0.204	1.226	0.019	1.189 1.264	
CoB									
Ref: UK/RoI		1.000				1.000			
Non-UK/ROI	0.248	1.282	0.049	1.189 1.382	0.210	1.234	0.048	1.142 1.333	
NS-Sec									
Ref: Intermediate		1.000				1.000			
Prof	0.134	1.144	0.013	1.118 1.170	0.157	1.170	0.014	1.143 1.197	
Routine	−0.117	0.889	0.010	0.870 0.910	−0.215	0.807	0.010	0.788 0.826	
Not working	−0.056	0.945	0.018	0.911 0.981	−0.238	0.788	0.016	0.758 0.820	
Student	−0.181	0.835	0.017	0.802 0.869	−0.273	0.761	0.016	0.731 0.794	
Housing tenure									
Ref: Owner occ		1.000				1.000			
Soc rent	0.921	2.513	0.037	2.442 2.585	1.091	2.978	0.044	2.893 3.067	
Priv rent	1.499	4.477	0.071	4.340 4.619	2.617	13.693	0.283	13.149 14.259	
CE Other	0.651	1.918	0.069	1.787 2.058	1.022	2.778	0.100	2.590 2.981	
Household composition									
Ref: Family, dep children		1.000				1.000			
Other	−0.193	0.824	0.010	0.805 0.844	−0.011	0.989	0.012	0.965 1.013	
Cohab	0.559	1.748	0.039	1.674 1.826	1.381	3.979	0.100	3.789 4.179	
Lone par	−0.033	0.967	0.016	0.936 1.000	−0.024	0.976	0.017	0.944 1.010	
Stud HH	−0.891	0.410	0.032	0.352 0.477	3.834	46.253	19.261	20.450 104.615	
CE/vacant	0.920	2.508	0.180	2.180 2.886	3.308	27.344	3.037	21.995 33.993	

Table 2 (continued)

	HCR					Census				
	Coeff	Odds Ratio	Std. Err.	[95% Conf. Interval]		Coeff	Odds Ratio	Std. Err.	[95% Conf. Interval]	
Area type										
Ref: Rural		1.000					1.000			
Urban	0.079	1.083	0.010	1.062	1.103	-0.008	0.992	0.010	0.973	1.012
Social deprivation										
Ref: 1 - least deprived		1.000					1.000			
2	-0.038	0.962	0.012	0.939	0.986	-0.034	0.967	0.012	0.943	0.991
3	-0.161	0.851	0.010	0.831	0.872	-0.226	0.798	0.010	0.778	0.818
Most deprived	-0.227	0.797	0.010	0.777	0.817	-0.276	0.759	0.010	0.739	0.779
Constant	-1.330	0.264	0.005	0.255	0.275	-1.418	0.242	0.005	0.233	0.252

Source: NLS

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Barr, P. J., & Shuttleworth, I. (2012). Reporting address changes by migrants: The accuracy and timeliness of reports via health card registers. *Health and Place, 18*(3), 595–604.
- Bell, M., Blake, M., Boyle, P., Duke-Williams, O., Rees, P., Stillwell, J., & Hugo, G. (2002). Cross-national comparison of internal migration: Issues and measures. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 165*(3), 435–464.
- Bell, M., Charles-Edwards, E., Ueffing, P., Stillwell, J., Kupiszewski, M., & Kupiszewska, D. (2015). Internal migration and development: Comparing migration intensities around the world. *Population and Development Review, 41*(1), 33–58.
- Champion, T., & Shuttleworth, I. (2017). Is longer-distance migration slowing? An analysis of the annual record for England and Wales since the 1970s. *Population, Space and Place, 23*(3), <https://doi.org/10.1002/psp.2024>.
- Champion, T., & Shuttleworth, I. (2018). 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), <https://doi.org/10.1002/psp.2026>.
- Cockings, S., Harfoot, A., Martin, D., & Hornby, D. (2011). Maintaining existing zoning systems using automated zone design techniques: methods for creating the 2011 Census output geographies for England and Wales. *Environment and Planning A, 43*(10), 2399–2418. <https://doi.org/10.1068/a43601>.
- Ermen, A., McCollum, D., Feng, Z., Everington, D., & Huang, Z. (2018). Using linked administrative and census data for migration research. *Population Studies, 72*(3), 357–367.
- Foley, B., Shuttleworth, I., & Martin, D. (2018). Administrative data quality: Investigating record-level address accuracy in the Northern Ireland health register. *Journal of Official Statistics, 34*(1), 55–81.

- Green, A. (2018). Understanding the drivers of internal migration. In T. Champion, T. Cooke, & I. Shuttleworth (Eds.), in *Internal Migration in the Developed World* (pp. 120–146). London: Routledge.
- NILS. (2013). *Linking the Northern Ireland longitudinal study to the 2011 census*. NILS Working Paper 4, Belfast: Northern Ireland Statistics and Research Agency.
- NISRA. (2014). *The use of administrative data in population estimates*. Belfast: Northern Ireland Statistics and Research Agency
- NISRA. (2018). *Methodology paper mid-year population estimates for Northern Ireland*. Belfast: Northern Ireland Statistics and Research Agency
- NISRA. (2019). *2021 census Northern Ireland proposals document*. Belfast: Northern Ireland Statistics and Research Agency
- Ogilvy, A. (1980). *Inter-regional migration since 1971: An appraisal of data from the National Health Service Central Register and Labour Force Survey*. Occasional Papers, (16), London: Office of Population Censuses and Surveys.
- ONS. (2018). *Annual assessment of ONS's progress on the administrative data census*. Titchfield: Office for National Statistics.
- ONS (2019). *Developing our approach for producing admin-based population estimates, England and Wales: 2011 And 2016*, Office for National Statistics, Titchfield.
- Shuttleworth, I., & Martin, D. (2016). People and places: Understanding geographical accuracy in administrative data from the census and healthcare systems. *Environment and Planning A: Economy and Space*, 48(3), 594–610.
- Statistics New Zealand. (2014). *An overview of Progress on the potential use of administrative data for census information in New Zealand*. Statistics New Zealand, Wellington, New Zealand: Census Transformation Programme.
- Stillwell, J., Rees, P. H., & Boden, P. (1994). Internal migration trends in the United Kingdom from National Health Service Reregistration data. *Espace, Populations. Sociétés*, 12(1), 95–108.
- US Census Bureau (2019), <https://www.census.gov/about/what/admin-data.html>, accessed November 8th 2019.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.