Published as Shuttleworth, I. and Martin, D. "People and places: Understanding geographical accuracy in administrative data from the census and healthcare systems" Environment and Planning A 2016 vol. 48 no. 3 594-610 doi: 10.1177/0308518X15618205
http://epn.sagepub.com/content/48/3/594

Abstract

Administrative systems such as health care registration are of increasing importance in providing information for statistical, research, and policy purposes. There is thus a pressing need to understand better the detailed relationship between population characteristics as recorded in such systems and conventional censuses. This paper explores these issues using the unique Northern Ireland Longitudinal Study (NILS). It takes the 2001 Census enumeration as a benchmark and analyses the social, demographic and spatial patterns of mismatch with the health register at individual level. Descriptive comparison is followed by multivariate and multilevel analyses which show that approximately 25% of individuals are reported to be in different addresses and that age, rurality, education, and housing type are all important factors. This level of mismatch appears to be maintained over time, as earlier migrants who update their address details are replaced by others who have not yet done so. In some cases, apparent mismatches seem likely to reflect complex multi-address living arrangements rather than data error.

Keywords: Administrative data, census, accuracy, spatial referencing

People and places: understanding geographical accuracy in administrative data from the census and healthcare systems

Introduction

There is increasing international interest in moving away from traditional censuses to alternative methods of population data collection. Two simultaneous trends are observable: growing interest in the use of internet census enumeration and greater use of linked administrative records and surveys (Ralphs and Tutton, 2011). Early in 2014 the National Statistician recommended to the UK government that a census should be undertaken in England and Wales in 2021 alongside increased use of administrative data and surveys, notably improving the statistics available in intercensal years (ONS, 2014). Similar models have been announced for Scotland (GRO-S, 2014) and Northern Ireland (NISRA, 2014). The government's response endorsed the recommended approach, articulating an explicit ambition that "the dual running of the decennial census and use of administrative data should not extend beyond 2021. The future should be based entirely on administrative data" (Public Administration Committee, 2014). These trends reflect both the growing costs and difficulty of achieving satisfactory census coverage and the increased opportunities afforded by improvements in the alternatives.

Countries with population registers have a strong starting point for the construction of census-like population statistics by augmenting their existing registers with additional survey and linked administrative data sources. However, in countries such as the UK which do not operate a formal population register, one of the biggest obstacles to producing population statistics by linking administrative records is that the population is not entirely accurately recorded on each data source. Mismatches between administrative data sources are unlikely to be evenly spread geographically or socially, but will be concentrated in certain places and population groups – potentially leading to misrepresentation and bias in any subsequent analysis. In the UK, health service registers provide the most comprehensive voluntary administrative list of the population (ONS, 2012a) and are thus an instance of administrative registers that potentially have a major role to play in population measurement. The primary purpose of a health register is to provide a framework for the delivery of health care services, which may extend, for example, to the management of screening programmes and the provision of denominators. Inaccuracies in a health register present two distinct challenges: not only might they impair the calculation of population and health statistics but they also have the potential to negatively impact on the delivery of services to individuals. We observe that the challenges of maintaining an accurate population list for health care organization and of generating

a statistical model of the population differ, yet they share many aspects. Despite the increasing importance of these issues, the practical details are as yet poorly understood.

This paper employs the unique Northern Ireland Longitudinal Study (NILS) (Johnston et al, 2010) to examine the relationship between the conventional census and a major administrative register in greater depth and to investigate the characteristics of individuals who are recorded at different locations in the two sources. We are specifically interested in investigating the following research questions. Firstly, how many people are reported to be in a different place in the health register to where they were enumerated in the 2001 Census? Secondly, what are the social and demographic characteristics of these mismatched individuals? Thirdly, what is the geography of this mismatch? Fourthly, are all individuals in a household captured in the census similarly matched or mismatched? Finally, how long does it take, if ever, for mismatched locations in the health register to catch up with Census locations? One principal outcome is used: the extent to which addresses are matched, as this is important for health screening but also more generally in locating people and constructing households using administrative data. Our analysis is mainly cross-sectional, looking at the location of individuals based on April 2001 data but we also look longitudinally at the speed at which address information in the health register catches up with that in the 2001 Census using sixteen later downloads of the health register data. Information is available for individuals and for multiple members of the same household, where these appear in the NILS sample. We present descriptive statistics and multi-level coefficients of the individual and ecological determinants of match and mismatch. The paper thus employs a unique dataset to develop a novel perspective on an important methodological problem. It is based on a Northern Ireland study, but the operational issues are similar to those encountered in the UK and elsewhere, while the conceptual issues raised are of international significance, both for population statistics and health service delivery. Although the specific mixture of available administrative data sources differs between countries, the desire to better integrate live administrative systems with the challenge of counting the population is widespread.

The remainder of the paper is divided into five sections. Section 2 reviews in greater depth the international interest in finding alternatives to the conventional census and the challenges which arise when attempting to use administrative registers as population data sources. Section 3 describes the data and methods used in this study, including a more detailed description of the NILS, the health registration system, the practicalities of address referencing in Northern Ireland and our analytical methods. In section 4 we present a series of results which elucidate the relationship

between the census and health register, focusing on the socioeconomically uneven pattern of mismatch between the two sources. Our discussion in section 5 explores the implications of these findings, including the transferability of our findings to other contexts, and our conclusions are presented in section 6.

Review: The international and national contexts

The 2000/1 and 2010/11 census rounds internationally have been marked by increasing numbers of nations adopting population data collection models which move away from conventional censuses, with emphasis on greater use of linked administrative data already collected by governments (Ralphs and Tutton, 2011). During the most recent round, further countries moved to the implementation of register-based censuses such as Switzerland in 2010 (Schwyn and Kauthen, 2008) and Austria in 2011 (Statistics Austria, 2014). These are both cases in which a conventional census was replaced for the first time by a population statistics system based on the linkage of multiple registers, which are able to directly provide some population characteristics and also to act as the frame for additional surveys. Benefits cited include increased frequency, speed and flexibility in the production of statistics, a favourable cost/benefit ratio and reduced burden on census respondents. Switzerland and Austria thus join countries such as Finland whose first generation of census data produced wholly from register-based sources dates back to 1990 (Myrskylá, 1991).

The UK is beginning to move in this direction, in common with countries such as Canada and New Zealand (Statistics Canada, 2012; Bycroft, 2013). Independent simultaneous censuses were conducted in each country of the UK in both 2001 and 2011 by the three national statistical organizations: the Office for National Statistics (ONS) in England and Wales, National Records of Scotland (NRS) and the Northern Ireland Statistics and Research Agency (NISRA) in Northern Ireland. Despite some differences in the detailed questions asked, collection and processing methods, these were essentially conventional censuses, undertaken to very similar designs using paper-based methods with self-completed forms and enumerator visits to households. Only in the 2011 Census were internet methods available as an option. However, administrative data were used in 2011 to adjust for missing households and individuals. An important source of these data was the health registration system.

The UK operates a state health service which is essentially free at the point of use and relies on individuals registering with a family doctor known as a general practitioner (GP). Registered individuals are given a unique health service number (unrelated to those used for other government

services). There is considerable benefit and no cost to registration and hence this register is a prime administrative data source for the augmentation of demographic data from a census, achieving greater coverage than lists maintained, for example, for electoral registration or social security purposes.

A further important consideration when comparing these different systems for recording the population is to recognise their differing treatment of persons, households and addresses. Households are differently defined in different contexts, but the term usually refers to individuals who live together and share either a physical dwelling space or living arrangements. Administrative systems such as health registration generally include information about individual persons and their addresses, but do not explicitly contain descriptions of household relationships, a feature common to many population registration systems internationally such as that in Austria (Ralphs and Tutton 2011). It may be possible to make connections between individuals who share the same address, but it must be recognised that addresses, which are textual descriptions of properties, do not always relate to single households. Indeed, one address may contain multiple households and both individuals and households may be associated with multiple addresses.

Census-taking involves an attempt to obtain information from every individual in the population. In the absence of a household list, most of the population is reached by delivering a census questionnaire to every address in an address register, whether by hand delivery or by post, the latter adopted for the first time in England and Wales in 2011 (ONS, 2012b). The success of the census is therefore dependent on having a high quality list of addresses containing households. Standard census operations already recognise the challenges of enumerating individuals who are not in households by the separate enumeration of communal establishments such as prisons, nursing homes, hospitals and student halls of residence. These have demographically diverse and 'hard-to-reach' populations.

Attempts to generate census-type data from administrative lists involves record matching due to the limited attribute detail provided in any one source, yet the ambiguities associated with each individual source (census or administrative) are compounded by the linkage process. A failure to correctly match records may be due to many types of error in either or both sources or to definitional differences between them. Locational mismatch is of particular importance as it may impair aggregation of data for publication, interpretation of residential moves, the relating of population characteristics to policy or service delivery areas or matching of record-level with

ecological data. The principle that it is desirable to locate each population member at one unique address extends beyond official statistics to health screening applications which aim at individual interventions. The address information for these programmes is usually drawn from health registers. If these are in error because people fail to update their information or have complex locational and address histories, this should be a matter for concern. Each programme for intervention has a different target demographic group (for example, screening for breast cancer or diabetes). It is therefore very important to understand more about the detailed demographics of address inaccuracies by age, gender, education and, indeed, other socioeconomic characteristics. The exploration of these patterns thus forms the focus of this paper.

Data

Northern Ireland, in common with other parts of the UK, conducts a decennial census. In 2001 a detailed self-completion questionnaire with both household and individual questions was delivered to every address appearing in NISRA's census address list or identified by enumerators in the field (NISRA, 2006). Although the questionnaire requested the names of all persons present, individual questions were addressed only to usual residents, representing the de jure population. NISRA (2006) estimate that 95% of the population were covered in completed census returns.

O'Reilly et al. (2012) report that the Northern Ireland health registration system covers almost 100% of the population but is "inflated" by 4.7%, due to individuals who have emigrated, cross-border workers or duplicate entries. This is comparable with 4.3% inflation of the England and Wales 2011 register compared to the final census estimate, but with considerable variation in the ratio for different age/sex groups and local authorities (ONS, 2012a), with over-coverage at the global level conceals under-coverage of some groups. These figures serve to indicate that patterns of coverage in both censuses and registers can display complex socioeconomic patterns which need to be properly understood by those who wish to use the data in a wider context.

This analysis is based on an extract from NILS, a 28% population record linkage sample comprising individual records matched between census enumeration, vital events and downloads, taken every six months, from the health register (Johnston et al, 2010). Most weight in the matching process is placed on individual name, gender, and date of birth. Three stages of matching are followed: exact matching, fuzzy matching, and then clerical checking (NILS 2013). In 2001, 97.1% of eligible records were matched (excluding list inflation, imputed individuals, and those with insufficient matching

information). Matching was performed on the confidential records by NISRA before the researchers accessed an anonymised research dataset in a secure laboratory.

The links with the census make it possible for researchers to use the full range of census information for individuals and households. The download matched with the 2001 Census was from April 2001 and is directly comparable with the census, taken on April 29th 2001. We focus on the population as enumerated in the census on April 29th, which excludes births and immigrants after this date. Sixteen further 6-monthly health register updates to October 2010 were analysed to investigate when mismatched records were updated. The high sample fraction allows geographical analysis using Super Output Areas (SOAs) as the base geography. SOAs were developed following the 2001 Census and provide an intermediate level of geography between the smallest census Output Areas (OAs) and larger electoral wards. There are 890 SOAs, having a mean 2001 population size of 2000 (NISRA, 2013). Two contextual indicators, the Northern Ireland multiple deprivation measure (Beatty 2004) and standard classification of settlements (NISRA 2005) are available only at SOA level. To aid reference to specific areas in Northern Ireland we include a base map in Figure 1.

The NILS and thus this analysis exclude some people. The NILS link from the census to the health register is based only on the census enumerated population – wholly imputed individuals and households are therefore absent. The analysis in this paper also excludes those who were present in the health register but who were not in the census because there are no census explanatory variables for them nor is it possible, by definition, to compare locational information from the census with the health register for this group. A proportion of this group falls under the heading of 'list inflation' – people who have died, emigrated or are otherwise no longer in Northern Ireland but who still remain on the health register. However, another fraction may be individuals who were present in Northern Ireland, and legitimately on the health register, but who may also be absent because they did not respond to the census. It is difficult, because of the nature of these groups, to put precise values on each of these fractions. As well as the 4.7% list inflation of the health register in 2001 Johnston et al (2010) report that 4.6% of individuals in the census were imputed. These real, imputed, or illusory individuals are all excluded from the analysis. It is probable that our estimates of mismatch will be conservative given the demographic and residential profiles of these hard-to-enumerate populations. We largely focus the multivariate analysis on individuals although the descriptive analysis deals with individuals and mismatches between NILS members at the same address, and we give particular attention to some groups of special interest such as students and those in communal establishments.

There are some constraints imposed by the data structure of the NILS and the census. The NILS is a sample drawn from individuals on the health register, not individuals or households in the census. There are some households in the sample with multiple NILS members but others with only one so we concentrate mainly on individuals rather than households. Other limitations arise from some NILS members not being members of households as defined by the census (for example living in communal establishments). They therefore have no associated household characteristics such as tenure or household composition, imposing some unavoidable constraints on the variable combinations in the models we report later.

The Pointer database, maintained by Land and Property Services (LPS, 2009), provides a standard address referencing frame for Northern Ireland in which each property has been assigned a Unique Property Reference Number (UPRN). This database is broadly comparable with the AddressBase database in England and Wales (Ordnance Survey, 2014) and other national address database products. Every UPRN associated with a NILS record has been converted into an anonymised property reference number, termed the XUPRN, supplied in the research extract. We use the term XUPRN to make clear the distinction from actual addresses. This design makes it possible to examine the match between XUPRNs appearing on census and health register sources for each NILS member. Possible outcomes are that the XUPRNs may be invalid in one, other or both of the data sources or, if both valid, may match or not match as is shown in Table 1. Where XUPRNs did not match, the distance between the two Pointer locations was provided. There were particular problems in County Fermanagh (see Figure 1) where there were a high proportion of invalid XUPRNs. The address system in the county in 2001 was based on traditional small geographical units known as townlands. These were incompletely incorporated in Pointer.

Methods

The vast majority of the explanatory variables and, indeed, the outcome variable are derived from the census so at this stage it is useful to reiterate that the central concern of the paper is with the enumerated 2001 census population. The census has the advantage of offering a wide range of possible explanatory variables. The approach used was exploratory and aimed to outline and to understand more about the kinds of personal background factors that were associated with matching between address information held in the health system and the census. There were few *a priori* reasons in the literature for selection of candidate variables when analysing address matching rates (but see Barr and Shuttleworth 2012) apart from the evidence on what constitutes a hard-to-

enumerate population in the census. This is partly because there have been few similar systematic assessments particularly of administrative data sources. However, the literature on under-enumeration (ONS 2012a, Carter 2009) suggested that age and gender were very likely to be important so these two variables were incorporated in the analysis. Following this, it was considered that social background, for instance education, socio-economic status (SES), and economic activity might be important so these were included, as was marital status. This was used in the multivariate analysis as an alternative to living arrangements because of an overlap of some categories with accommodation type. The literature on census under enumeration also indicates that various household variables may be important. Besides the well-known issues in dealing with communal establishments there may also be problems with other accommodation types. Therefore, tenure and accommodation type were also included for exploration. Our interest in the geography of match/mismatch suggests an investigation of between-place variation. This was done through the use of explanatory variables such as settlement band and the multiple deprivation measure, both at SOA Level.

We do not entirely exclude Fermanagh because it is an extreme example of the addressing problems likely in other rural areas but our analytical strategy recognises its unusual nature. Furthermore, simply excluding all invalid XUPRNs would remove important items of interest such as communal establishments, temporary dwellings, and addresses in multi-occupancy buildings because these often have addresses in formats which are problematic for Pointer to allocate an XUPRN. We therefore consider two population bases; all records and then only those with valid XUPRNs. Our initial descriptive results include all records but later modelling is based only on valid addresses and XUPRNs in the census and the health register, with a dummy variable for County Fermanagh.

Following descriptive analysis multivariate regression is undertaken. Since the outcome variable was binary (1=matching XUPRNs, 0=not matched XUPRNs) logistic regression was selected. A multilevel framework was used. This was appropriate because of the structure of the data with individuals nested within SOAs (Barr and Shuttleworth, 2012). This recognises that people who live in the same SOAs will more often than not share some of the same characteristics. Some consideration was given to nesting individuals in households but this was rejected because of its complexity. A series of models were estimated using MLwiN for all records and then just those with valid XUPRNs, both covering all age groups. Following this, models were specified for the population aged 16-74 with valid XUPRNs in both data sources. This more restricted group covers all records for which education and labour market variables are present and allows us to concentrate on locational

mismatch where valid address information is present. A variety of models were estimated including (and excluding) a dummy for County Fermanagh, alternating SES and economic activity, and also with and without housing tenure and housing type (it was impossible to include both variables at the same time). Only selected models are presented in full due to space constraints but comments will be offered where appropriate based on the modelling exercise in its entirety. Findings were consistent across the suite of models thereby giving considerable confidence.

Results

Descriptive results

Figure 2 shows the percentages with valid and invalid XUPRNs and those with matches/mismatches between the Census and the health register by age. The top line shows overall matching success – those records with valid XUPRNs that match in both data sources. The rates are highest at just over 80% for very young children and those aged over 50. Match rates decline rapidly after the age of 18 to a low of around 60% for those in their late twenties but then make a slow recovery as age increases. In general, the proportion of those with invalid XUPRNs in the Census, the health register, or both fluctuates at low levels for most of the life course but there is some evidence of increases for the very old. The reasons for this are unknown but might be attributed to the increased likelihood of the very old living in communal establishments which, as will be shown, are a problematic accommodation type.

Although there are no clear patterns with regard to age and invalid XUPRNs, there is evidence for geographical differences across Northern Ireland in the distribution of problem addresses. Mapping the matching results at the LSOA level shows a concentration of invalid XUPRNs in the census in the South West of Northern Ireland in County Fermanagh, a problem already noted. The distribution of invalid XUPRNs in the health register shows a more widespread distribution in the South and West. This indicates that there are address data problems in rural areas which might be attributed to the structure and perhaps non-standard format of addresses beyond the known problems in County Fermanagh. It is, however, also worthwhile noting that this geography is quite complex and the problem of invalid/null XUPRNs extends also to some urban areas. It would also be inappropriate to attach too much weight to the problems of County Fermanagh in the analysis. Just over 3% of all addresses are both invalid and in Fermanagh, so the vast majority do not fall into this special class.

Selected tabulations of individual and household characteristics for individuals with valid, invalid and matching XUPRNs, where there are large differences, are presented in Table 2 (further person,

household and labour market characteristics are shown in Tables A1 to A3 in an online Annex at http://URLtobeadvised). Table 2 shows that a change of address in the past year has a major influence on the accuracy of locational information – only 39% of those who had moved within Northern Ireland had matching XUPRNs. As expected, when considering living arrangements, there were low match rates for those in communal establishments but also lower than the average rate for those cohabiting. Across settlement bands, the open countryside has a lower match rate than all other area types, perhaps as a result of the problems in Fermanagh but also, as was discussed, more widely in some rural places. For private renters there are also much lower match rates and this extends to accommodation types such as flats, tenements, shared housing, and commercial buildings. Lower match rates are also observed for the unemployed and the self-employed. The online tables show mismatch rates increase with qualification level.

To some extent these factors are inter-related; younger people tend to be more mobile, as do those with more qualifications, and people in these categories are also more likely to be private renters. Therefore, to tease out some of these inter-relationships further, it is necessary to undertake multivariate analysis, which we present in the next section. Three additional aspects of descriptive analysis are of interest. First, we look at the average distances between the Pointer-derived grid references associated with XUPRNs in the health register and Census XUPRNs, where these differed. The median was just over 1km, the mean just under 7km. This means that around 50% of people with mismatches are misplaced by less than a kilometre. Secondly, we consider the extent to which NILS members in the same household have differing matches/mismatches. The best estimate from our analysis is that for 17% of households there were disagreements between household members with one, for example, being accurately matched and another member not matched. The third element concerns temporal lag, somethingimpossible to assess using XUPRNs. Therefore we investigated when the SOA in a health register download matched the SOA of enumeration in the 2001 Census. This showed that after three years, by the April 2004 download, 50% of mismatching individuals had updated their health register address to place them in the SOA where they were counted in the 2001 Census.

Multivariate results

The outcome variable for the models discussed in this section was MATCH (1=Census and health register XUPRN are same, 0=Census and health register XUPRN are not the same). Selected models are presented alternating housing tenure or housing type. Various specifications were explored (with and without the dummy for Fermanagh), and exchanging SES and economic activities. Models for all age groups are presented in Tables A4 and A5 in the online annex; the final models with education and labour market variables are shown in Table 3. The main focus is on the fixed effects rather than the random part of the models since Level 1 variance is constrained in the logistic specification. The majority of independent variables are statistically significant which is as expected given the sample size so most attention will be focussed on the direction and size of the effects rather than statistical significance *per se*. For all models there are log coefficients, standard errors, and T values but for ease of interpretation odds ratios for each variable (set against a value of 1 for each reference category) are also presented. Values greater than 1 indicate that a category is associated with higher odds of successful matching than the reference whereas values less than 1 associated with lower odds of successful matching.

Models 1(a) and 1(b) (Table A4) include valid and invalid XUPRNs and alternate housing tenure and type. There are many commonalities. Females always have a greater probability of matching than males and being Protestant relative to Catholic is also positive. Those with no limiting long-term illness have poorer odds of making a successful match (relative to those with no illness) as are NILS members who are single, separated, divorced, or widowed compared with those who were married. Decreased odds for NILS members who changed address in the past year are apparent as are lower odds for those resident in flats, bedsits, caravans, commercial buildings and communal establishments. There are also strong effects for age group with those falling into the younger parts of the population having lower odds for successful address matching but some differences between Models 1(a) and 1(b). The results for Models 2(a) and (b)in Table A5 for valid XUPRNs are similar to those reported above except the Fermanagh effect is smaller although in general more rural areas have greater odds of mismatching.

The final models (Models 3 and 4, Table 3) are on the base of valid XUPRNs and population aged 16-74 so they can include labour market and educational variables. Once again, housing tenure and type are used in alternative specifications. Looking across the table there are considerable similarities with regard to age, changing address the year before the Census, housing tenure and type, and gender with those reported above. There are some large effects associated with SES with

lower odds of matching for the self-employed, but higher for students, relative to the base of professionals once other variables are taken into account. Education is less important but it is interesting to note that those with the highest qualifications have smaller odds of matching.

Across the models, multiple deprivation at SOA level was statistically significant although the effect was small and so has not been a central analytical focus. Most attention has been paid to the fixed effects of the models. In relation to the random portion, in the null model at the first stage of Models 3 and 4, for instance, the Level 2 variance (on a log scale) between SOAs was 0.45 and was statistically significant. With the models fully specified this Level 2 variance fell to 0.10. It remained statistically significant with a portion still unexplained, but the fall suggests that a large part of the differences between SOAs can be attributed to the demographic and housing structures of these SOAs.

Discussion

The results have a number of implications. Our finding that 50% of mismatching people were displaced by 1km or less means that for statistical purposes, such as the estimation of small-area populations, the impact of inaccuracies might be small, especially in rural areas where spatial units are large. The spatial errors might, however, be important for other purposes such as health screening programmes where it is necessary to send information to a named person at a specific address. The descriptive results are also important when assessing the use of administrative data to recreate households since they suggest that in a substantial minority (17%) of households one person is located accurately whereas other household members as recorded in the census are located elsewhere in the healthcare data. Further work is needed on the temporal dimension of the analysis. However, the result that 50% of mismatching individuals in 2001 had updated their location by April 2004 is encouraging. These are presumably address changers who lagged in reporting their new address to their doctors. Less encouraging is the remaining 50% - these suggests a hard core of inaccuracy. Furthermore, as internal migrants update their address records, it is likely that other people move, and that some of these also are laggers. Thus, the pool of address mismatches might always be refilled by more people changing address.

We turn now to discuss the multivariate analysis results. Theoretically, address mismatch may be a function of (a) individual characteristics, (b) property and address factors, (c) institutional factors and (d) the interaction of all these. We can say nothing about institutional factors because we have no data on them but it is possible to consider individual and property/address characteristics. The

results show that mismatchers to some extent into the categories of people who are hard-to-enumerate in traditional censuses. The lower odds of successful matching for younger people, migrants, and males are very much part of this picture as are the lower odds associated with commercial buildings, flats, bedsits and communal establishments. However, there are additional characteristics associated with mismatch which are a product of the specific administrative system under review. Thus, for example, those who report better health have lower odds of matching successfully perhaps because they are less likely to visit their doctor. The results for education also indicate that it is the more educated who have lower odds of successful matching and this suggests that our observations cannot be explained only by social deprivation.

The findings with regard to lower matching success in rural areas are interesting. The addressing problems in County Fermanagh are clearly unique to Northern Ireland and raise questions about the transferability of the results. However, the other effects that are observed persist once we model this. Moreover, Northern Ireland has many features that make it a good case study. It contains a large regional city (Belfast) plus a hinterland with suburbs, commuting areas, and more remote rural areas. Arguably, it is therefore similar to large regional city and its surrounds in other geographical contexts. There are, however, limits to comparability. The small ethnic minority population of Northern Ireland in 2001, for example, led to a decision to downplay this dimension of the analysis although it may be important elsewhere.

There are also broader questions about the social meaning and practical implications of these findings. For some statistical and health purposes (for example, estimating household size and health screening programmes) it is desirable to locate each individual at a single address. Many people's lives are increasingly difficult to accurately describe in this way because of the realities of social and economic change. Long distance commuting involving multi-day trips can lead to some people having two residential locations, as can the increase in second 'holiday homes'. Likewise, fragmented family structures mean that some children spend time resident at more than one parental home. Because of these, and other trends with similar outcomes, populations are becoming more mobile and harder to geo-reference to a single address in the traditional way. The implications are that while some mismatches may be errors arising, for example, from the failure of an individual to update their address information following a move, an unknown proportion of the mismatch between the health register and the census will be what could be termed 'virtuous errors' where a person has more than one address and there is considerable ambiguity about their location in the census and the register. In this case their location may be indeterminate in both sources.

There is no easy way to be sure which type of error we are observing but it might be reasonable to comment, for example, that for cohabitees a mismatch between the census and health register might be a consequence of a dual-address individual. The reality is that a significant proportion of real complex lives will not be adequately reflected in a practical data model, however sophisticated, and that current aspirations to develop register-based census replacements are yet to adequately deal with these issues.

Some limitations to the analysis such as the lack of institutional data, for example about GPs, and the inability to deal with ethnicity have already been raised above but there are other issues inherent to the data. Chief of these is the reliance of the NILS on the census-enumerated population. Individuals missing from the census are not included in the database and there may be people who were present during the census, recorded in the healthcare data, but who did not respond to the census. Indeed, healthcare data was used by NISRA to estimate the population and to correct for census undercount of individuals and households in 2011. This limitation must be accepted since, without a census record, there is no demographic information or indeed mismatch information available but it should be acknowledged that the focus of the paper is on the census-enumerated population. We argue, however, that the analysis is robust – in 2001 around 95% of the population was enumerated. The researchers have no control over the confidential matching process, but it is estimated to have achieved a match of 97% of eligible records (NILS 2013) in 2001. There is no alternative means of gaining research access to linked administrative and census records in Northern Ireland for analysis of the type presented here.

Conclusion

This paper is the first written output of a wider programme of work. There are several ways to extend this research. The first, and most obvious, is to take advantage of the 2011 Census data that has recently been linked to the NILS to explore whether the same groups (and individuals) who had mismatching health register and census information similarly mismatch in 2011. In this context it will be interesting to assess the extent to which social composition has influenced overall address matching rates between 2001 and 2011. There has been an increase, for example, in the proportion of private renters and those with degrees, both categories associated with smaller odds of accurate matching in 2001. It might be expected, therefore, that some aspects of social change may increase the mismatch rate. However, set against this, the ageing population might be a counteracting force, and technical improvements and better address information could also act to increase match rates. A second extension would be to consider the impact of GP practice on matching/mismatching. This

requires access to more data but would be analytically possible using a multilevel cross-classified model where individuals are nested within SOAs and GP practices. Such an approach would allow institutional factors to be considered as well as individual and SOA context. A third way to extend the work would be through qualitative research on the ways that different demographic groups interact with the health system and how their demographic and geographic information are collected. All these extensions will be important to wider research and policy agendas such as the greater use of administrative data by national statistical agencies, the planning of future censuses and academic research using linked administrative data.

References

Barr P, Shuttleworth I 2012, "The accuracy of internal migration estimates using address information from health card registrations with General Practitioners: Perspectives from Northern Ireland using the Northern Ireland Longitudinal Study (NILS)", *Health and Place*, **18**, 595-604

Beatty R, 2004, "Northern Ireland multiple deprivation measure 2001 – a user's guide" NISRA Occasional Paper 20, http://www.nisra.gov.uk/deprivation/archive/deprivation_measure_2001.pdf

Beatty R, 2009, "Imputation in the 2001 Census" NILS User Forum http://www.qub.ac.uk/research-centres/NILSResearchSupportUnit/FileStore/Filetoupload,231330,en.ppt

Bycroft C, 2013, "Options for future New Zealand censuses" Wellington: Statistics New Zealand http://www.stats.govt.nz/~/media/Statistics/surveys-and-methods/methods/research-papers/topss/options-future-nz-censuses.pdf

Carter M, 2009, ``Explaining the census: investigating reasons for non-response to the ABS census of population and housing'' Institute for Social Research, Swinburne University of Technology, http://www.sisr.net/documents/Census.pdf

General Register Office - Scotland, 2014, "2021 Census" http://www.gro-scotland.gov.uk/census/2021-census/index.html

Johnston, F, Rosato, , Catney, G, 2010 "The Northern Ireland Longitudinal Study: An Introduction", NILS Working Paper 1, http://calls.ac.uk/output-entry/the-northern-ireland-longitudinal-study-an-introduction-2

Land and Property Services (2009) Pointer Version 2.1 Product Specification 2009. Belfast, Land and Property Services.

Myrskylá P, 1991, Census by Questionnaire – Census by Registers and Administrative Records: The Experience of Finland *Journal of Official Statistics* **7** (4) 457–474

Northern Ireland Longitudinal Study (2013), "NILS Data Matching Methodology", NILS Working Paper 3.0,http:// calls.ac.uk/wp-content/uploads/NILS-wp-3.0.docx

Northern Ireland Statistics and Research Agency, 2005, "Statistical classification and delineation of settlements – Februrary 2005" http://www.nisra.gov.uk/archive/demography/publications/urban_rural/ur_gaz.pdf

Northern Ireland Statistics and Research Agency, 2006, "Northern Ireland Census 2001 General Report." http://www.nisra.gov.uk/archive/census/2001/general-report.pdf Northern Ireland Statistics and Research Agency, 2012, "Population and household estimates for Northern Ireland. Quality Assurance Report." http://www.nisra.gov.uk/Census/pop_QA_2011.pdf

Northern Ireland Statistics and Research Agency, 2013, "NISRA Geography Factsheet" Belfast: Northern Ireland Statistics and Research Agency http://www.ninis2.nisra.gov.uk/public/documents/NISRA%20Geography%20Fact%20Sheet.pdf

Northern Ireland Statistics and Research Agency, 2014, "The future provision of census of population information for Northern Ireland", Belfast: Northern Ireland Statistics and Research Agency http://www.nisra.gov.uk/archive/census/2021/planning/the-future-provision-of-census-of-population-information-for-northern-ireland.pdf

Office for National Statistics, 2011, "2011 Census - Methodology for Quality Assuring the Census Population Estimates" http://www.ons.gov.uk/ons/guide-method/census/2011/the-2011-census/processing-the-information/data-quality-assurance/2011-census---methodology-for-quality-assuring-the-census-population-estimates.pdf

Office for National Statistics, 2012a "Beyond 2011: Administrative Data Sources Report: NHS Patient Register" http://www.ons.gov.uk/ons/about-ons/who-ons-are/programmes-and-projects/beyond-2011/reports-and-publications/sources-reports/beyond-2011--administrative-data-sources-report--nhs-patient-register--s1-.pdf

Office for National Statistics, 2012b, “2011 Census address register” Office for National Statistics http://www.ons.gov.uk/ons/guide-method/census/2011/how-our-census-works/how-did-we-do-in-2011-/evaluation---address-register.pdf

Office for National Statistics, 2013, “Beyond 2011: Options Report” Office for National Statistics http://www.ons.gov.uk/ons/about-ons/who-ons-are/programmes-and-projects/beyond-2011/what-are-the-options-/beyond-2011-options-report--o1-.pdf

Office for National Statistics, 2014, The Census and Future Provision of Population Statistics in England and Wales: Recommendation from the National Statistician and Chief Executive of the UK Statistics Authority http://www.ons.gov.uk/ons/about-ons/who-ons-are/programmes-and-projects/beyond-2011/beyond-2011-report-on-autumn-2013-consultation--and-recommendations/national-statisticians-recommendation.pdf

Ordnance Survey, 2014, “AddressBase products” http://www.ordnancesurvey.co.uk/business-and-government/products/addressbase-products.html

O'Reilly D, Rosato M, Catney G, Johnston F, Brolly M, 2012 “Cohort Description: The Northern Ireland Longitudinal Study (NILS)” *International Journal of Epidemiology* **41** (3) 634-641

Public Administration Committee, 2014, “Public Administration Committee - Second Special Report. Too soon to scrap the Census: Government and UK Statistics Authority Responses” Appendix 1: Government response http://www.publications.parliament.uk/pa/cm201415/cmselect/cmpubadm/601/60104.htm

Ralphs M, Tutton P, 2011, “Beyond 2011: International models for census taking: current processes and future developments” Beyond 2011 Project, Office for National Statistics http://www.ons.gov.uk/ons/about-ons/what-we-do/programmes---projects/beyond-2011/news/reports-and-publications/early-reports-and-research-papers/international-models-for-census-taking.pdf

Schwyn M, Kauthen J-P, 2008, “The Swiss Census 2010: Moving towards a comprehensive system of household and person statistics.” Neuchatel: Federal Department of Home Affairs http://www.bfs.admin.ch/bfs/portal/de/index/news/publikationen.Document.110779.html

Shuttleworth I, Barr P, 2011, “Who reports address changes through the healthcare system? Describing the characteristics of laggers and non-reporters using the Northern Ireland longitiudinal study” *Population Trends* **144** 48-54

Skinner C, Hollis J, Murphy M, 2013, “Beyond 2011: independent review of methodology” Office for National Statistics http://www.ons.gov.uk/ons/about-ons/who-ons-are/programmes-and-projects/beyond-2011/reports-and-publications/methods-and-policies-reports/beyond-2011--independent-review-of-methodology.pdf

Statistics Austria, 2014, “Census 2011 Austria. Results of the Register-based Census.” Vienna: Statistics Austria http://www.statistik.at/web_de/static/methods_of_register-based_census_in_austria_055160.pdf

Statistics Canada, 2012, “Final Report on 2016 Census Options: Proposed Content Determination Framework and Methodology Options” Statistics Canada http://www12.statcan.gc.ca/census-recensement/fc-rf/reports-rapports/census%20strategy_final%20report-strategie%20du%20recensement_rapport%20final-eng.pdf

		No XUPRN - Census and Health register	No XUPRN – Census	No XUPRN – Health register	Same XUPRN	Different XUPRN	Total
Belfast	Count	291	740	1822	54053	14254	71160
	Percentage	0.41	1.04	2.56	75.96	20.03	100.00
Fermanagh	Count	6596	1674	1053	4716	1524	15563
	Percentage	42.38	10.76	6.77	30.30	9.79	100.00
Rest of NI	Count	1705	5562	13283	285490	61254	367294
	Percentage	0.46	1.51	3.62	77.73	16.68	100.00
All Northern Ireland	Count	8592	7976	16158	344259	77032	454017
	Percentage	1.89	1.76	3.56	75.83	16.97	100.00

Table 1: Valid, invalid and matching XUPRNs by geographical area (Source: NILS extract, Project 051)

	No XUPRN - Census and Health register	No XUPRN - Census	No XUPRN – Health register	Same XUPRN	Different XUPRN	Total
Migration						
Did not move pre-census	1.94	1.52	3.49	78.90	14.16	414136
Moved within Northern Ireland	1.16	4.43	4.07	39.13	51.22	31294
Moved from outside Northern Ireland	1.70	4.92	4.36	60.24	28.78	3534
Living arrangements						
Married	1.97	1.42	3.55	78.86	14.20	169973
Remarried	0.76	1.14	2.51	81.31	14.27	9285
Cohabiting	0.86	1.97	3.37	54.05	39.74	13693
Single	1.91	1.64	3.30	75.78	17.37	202851
Separated	1.01	1.96	3.04	68.94	25.05	10782
Divorced	1.10	1.89	3.19	73.79	20.04	11020
Widowed	1.87	1.36	4.11	82.80	9.87	24799
Communal establishment	6.00	18.43	14.16	24.07	37.35	5036
Settlement Band						
A:Belfast Metropolitan Area	0.31	1.08	2.07	78.71	17.83	153664
B:Derry Urban Area	0.26	1.30	1.31	83.25	13.88	24069
C:Large town (population=18000-74999)	0.30	1.20	2.22	82.11	14.16	60992
D:Medium town (10000-17999)	0.55	1.46	3.34	75.86	18.80	27523
E:Small town (4500-9999)	0.54	1.42	2.75	77.63	17.66	27657
F:Intermediate settlement (2250-4499)	0.87	1.43	2.73	77.51	17.46	18063
G:Village (1000-2249)	1.31	1.89	4.00	73.94	18.86	18382
H:settlements(<1000)&open countryside	5.80	3.13	6.79	67.32	16.96	123667
Tenure						
Owner occupier	2.10	1.41	3.47	78.31	14.72	342238
Social rented	0.58	1.63	2.75	75.87	19.17	77224
Private rented	2.23	3.29	4.94	55.79	33.76	29499

Accommodation type						
Detached house/bungalow	3.63	2.06	4.86	74.03	15.42	193125
Semi-detached house/bungalow	0.41	0.79	2.07	80.51	16.20	125737
Terraced	0.31	0.76	2.02	80.11	16.79	112032
Flat/tenement	1.22	5.93	5.81	53.82	33.23	14311
Converted/shared house	3.15	10.05	8.22	35.06	43.53	1971
Commercial building	6.08	8.98	15.19	30.52	39.23	757
Caravan/other mobile/temporary	12.51	9.07	7.55	45.37	25.50	1047
Communal establishment	6.00	18.44	14.16	24.06	37.34	5037
Economic activity						
Employee	1.59	1.52	3.18	73.83	19.88	149409
Self-employed	3.50	2.04	6.86	67.59	20.01	26372
Unemployed	2.02	2.24	4.14	67.73	23.86	12596
Student (economically active)	1.21	2.58	2.84	74.63	18.74	6751
Retired	1.69	1.33	3.57	84.38	9.04	36359
Student (economically inactive)	1.95	3.38	4.02	70.24	20.41	17349
Home-maker	1.70	1.58	3.09	77.55	16.07	23560
Permanent sick	1.69	1.85	3.95	77.12	15.40	29452
Other	2.15	2.09	4.11	72.75	18.90	13263

Table 2: Valid, invalid and matching XUPRNs by various characteristics

	B	S.E.	T	OR	B	S.E.	T	OR
Response	**MATCH**				**MATCH**			
Constant	1.58	0.04	37.62		1.53	0.04	35.60	
Age								
18-24 (reference)				1.00				1.00
25-34	-0.34	0.02	-17.89	0.71	-0.31	0.02	-16.47	0.73
35-44	0.00	0.02	-0.10	1.00	0.04	0.02	1.95	1.04
45-54	0.35	0.02	14.58	1.42	0.41	0.02	16.96	1.50
55-64	0.63	0.03	23.15	1.87	0.69	0.03	25.70	2.00
65-74	0.70	0.03	23.20	2.01	0.77	0.03	25.80	2.17
Gender								
Male (reference)				1.00				1.00
Female	0.42	0.01	38.09	1.52	0.39	0.01	35.82	1.48
Community background								
Catholic				1.00				1.00
Protestant& other Christian	0.14	0.01	11.08	1.15	0.15	0.01	11.23	1.16
None	0.04	0.08	0.56	1.04	0.13	0.08	1.59	1.13
Other	0.00	0.04	-0.11	1.00	-0.01	0.04	-0.29	0.99
Limiting long-term illness								
Yes (reference)				1.00				1.00
No	-0.16	0.02	-10.53	0.85	-0.15	0.01	-10.93	0.86
Marital status								
Married (reference)				1.00				1.00
Single	-0.18	0.02	-11.93	0.84	-0.16	0.02	-10.40	0.86
Remarried	-0.07	0.03	-2.06	0.94	-0.08	0.03	-2.44	0.92
Separated	-0.44	0.03	-17.64	0.64	-0.45	0.03	-17.96	0.64
Divorced	-0.39	0.03	-15.72	0.68	-0.37	0.03	-14.96	0.69

Widowed	-0.21	0.03	-6.56	0.81	-0.19	0.03	-5.97	0.83
SOA deprivation								
Multiple deprivation measure	0.00	0.00	-3.00	1.00	0.00	0.00	-3.00	1.00
Settlement band								
A:Belfast (reference)				1.00				1.00
B:Derry Urban Area	0.41	0.05	7.79	1.51	0.39	0.05	7.38	1.48
C:Large town (population=18000-74999)	0.27	0.04	7.71	1.31	0.25	0.04	7.11	1.28
D:Medium town (10000-17999)	0.01	0.05	0.24	1.01	-0.01	0.05	-0.24	0.99
E:Small town (4500-9999)	-0.05	0.04	-1.09	0.95	-0.07	0.04	-1.70	0.93
F:Intermediate settlement (2250-4499)	-0.09	0.05	-1.76	0.92	-0.11	0.05	-2.33	0.89
G:Village (1000-2249)	-0.06	0.04	-1.40	0.94	-0.09	0.04	-2.14	0.91
H:settlements(<1000)&open countryside	-0.16	0.03	-5.81	0.85	-0.19	0.03	-6.93	0.82
SES								
Professional (reference)				1.00				1.00
Intermediate	0.12	0.02	6.11	1.12	0.12	0.02	6.26	1.13
Self employed	-0.14	0.02	-6.48	0.87	-0.13	0.02	-6.24	0.88
Lower supervisory	0.03	0.02	1.43	1.03	0.02	0.02	0.86	1.02
Routine	0.12	0.02	7.56	1.13	0.10	0.02	6.13	1.10
Not working	0.06	0.03	2.40	1.06	0.00	0.03	0.12	1.00
Students	0.36	0.03	13.77	1.43	0.34	0.03	13.04	1.40
Education								
No qualification (reference)				1.00				1.00
Foundation	-0.04	0.02	-2.50	0.96	-0.03	0.02	-1.56	0.98
A Level	0.00	0.02	-0.20	1.00	0.02	0.02	1.33	1.02
Degree plus	-0.10	0.02	-5.05	0.91	-0.07	0.02	-3.74	0.93
Fermanagh								
No (reference)				1.00				1.00
Yes (reference)	-0.43	0.08	-5.36	0.65	-0.45	0.08	-5.57	0.64
Address change year before the census								

No address change (reference)				1.00				1.00
Within NI	-1.75	0.02	-102.82	0.17	-1.80	0.02	-112.69	0.16
Outside NI	-0.53	0.05	-11.21	0.59	-0.60	0.05	-12.70	0.55
Housing tenure								
Owner occupied (reference)				1.00				
Social rented	-0.23	0.02	-14.25	0.80				
Private rented	-0.70	0.02	-34.75	0.50				
Housing type								
Detached dwelling (reference)								1.00
Semi-detached house/bungalow					0.01	0.01	0.57	1.01
Terraced					0.03	0.02	1.69	1.03
Flat/tenement					-1.09	0.03	-38.96	0.34
Converted/shared house					-1.25	0.07	-19.00	0.29
Commercial building					-1.47	0.11	-13.25	0.23
Caravan/other mobile/temporary					-0.62	0.11	-5.90	0.54
Communal establishment					-2.35	0.07	-34.04	0.10
Random Part								
Level 2 variation	0.09	0.01	15.67		0.10	0.01	16.17	
Level 1 variation	1	0			1	0		
Level 2 Units: SOA	890				890			
Level 1 Units: Individuals	289546				290836			

Table 3: Age 16-74 models (dependent variable=MATCH), base valid XUPRNs

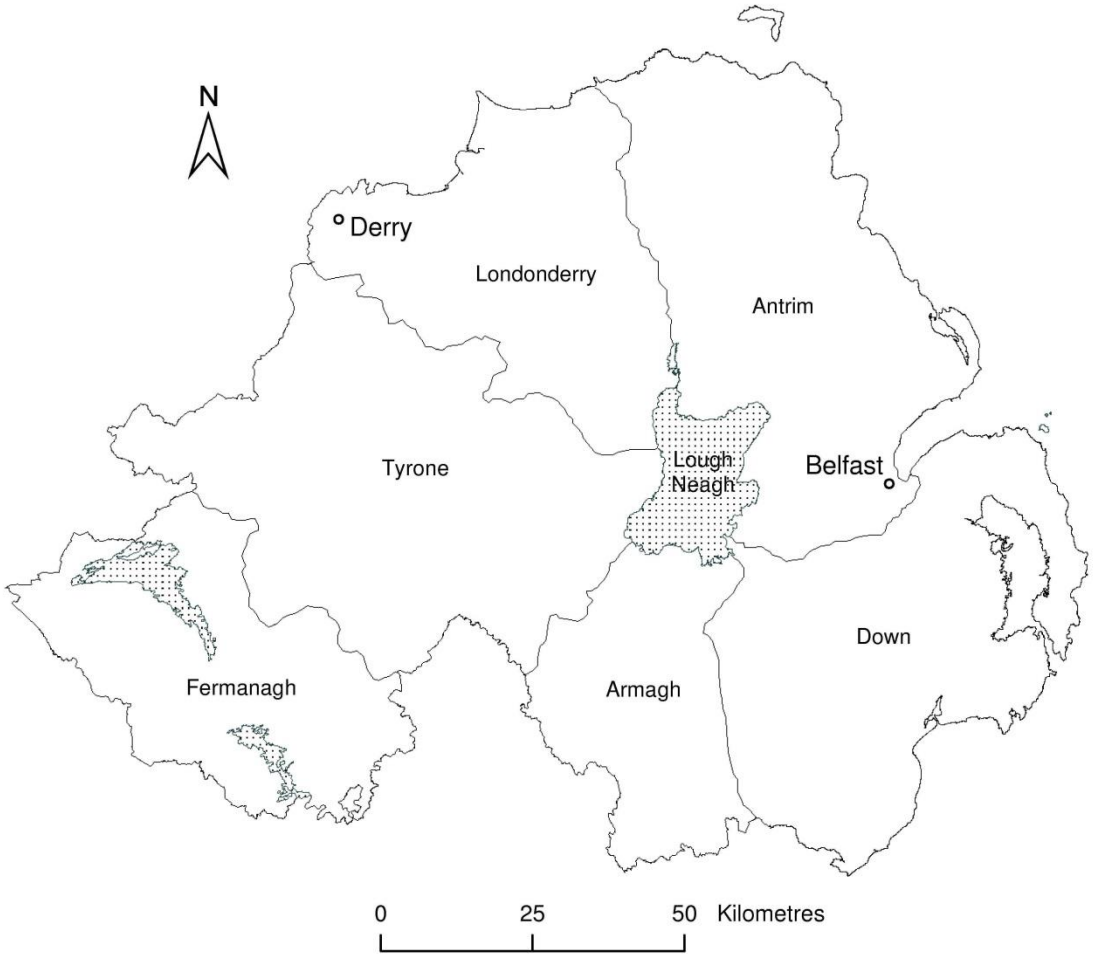

Figure 1: Base map of Northern Ireland showing counties and Belfast and Derry

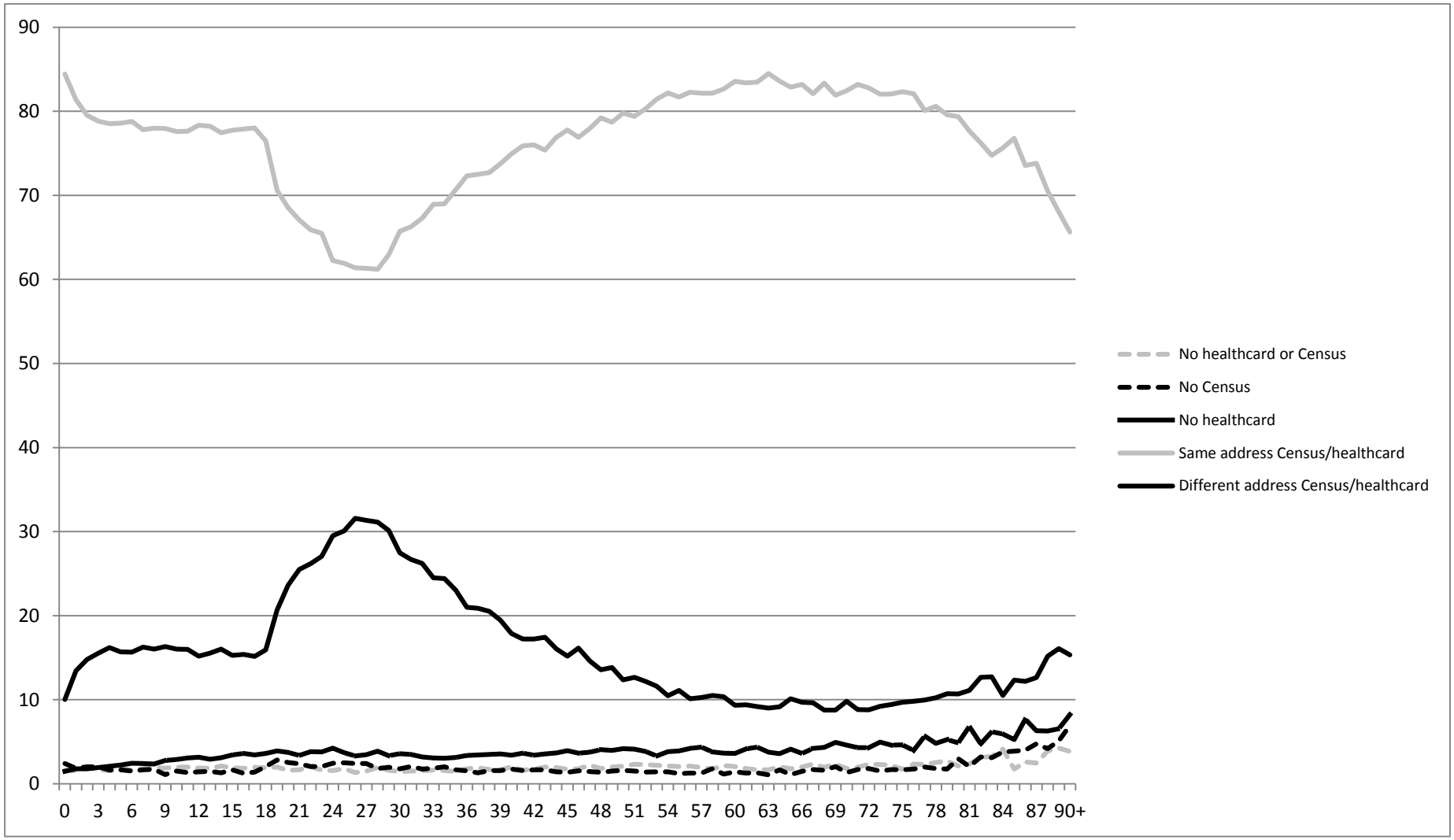

Figure 2: Geographical accuracy by age (percentages)

Online Annex of additional tables to "People and places: understanding geographical accuracy in administrative data from the census and healthcare systems"

	No XUPRN - Census and health register	No XUPRN - Census	No XUPRN – health register	Same XUPRN	Different XUPRN	Number
Community background						
Catholic	2.44	1.88	4.09	73.31	18.29	195119
Protestant	1.47	1.63	3.14	78.20	15.56	241539
None	1.48	2.40	3.26	71.30	21.57	1627
Other	1.11	1.82	2.72	75.18	19.16	10679
Limiting long-term illness						
Yes	1.94	2.04	4.06	77.91	14.06	92364
No	1.87	1.67	3.41	75.48	17.58	356600
Gender						
Male	1.99	1.76	3.89	73.59	18.77	219466
Female	1.81	1.75	3.25	77.91	15.28	234551
Migration						
Did not move pre-census	1.94	1.52	3.49	78.90	14.16	414136
Moved within Northern Ireland	1.16	4.43	4.07	39.13	51.22	31294
Moved from outside Northern Ireland	1.70	4.92	4.36	60.24	28.78	3534
Living arrangements						
Married	1.97	1.42	3.55	78.86	14.20	169973
Remarried	0.76	1.14	2.51	81.31	14.27	9285
Cohabiting	0.86	1.97	3.37	54.05	39.74	13693
Single	1.91	1.64	3.30	75.78	17.37	202851
Separated	1.01	1.96	3.04	68.94	25.05	10782
Divorced	1.10	1.89	3.19	73.79	20.04	11020
Widowed	1.87	1.36	4.11	82.80	9.87	24799
Communal establishment	6.00	18.43	14.16	24.07	37.35	5036

Settlement Band						
A:Belfast Metropolitan Area	0.31	1.08	2.07	78.71	17.83	153664
B:Derry Urban Area	0.26	1.30	1.31	83.25	13.88	24069
C:Large town (population=18000-74999)	0.30	1.20	2.22	82.11	14.16	60992
D:Medium town (10000-17999)	0.55	1.46	3.34	75.86	18.80	27523
E:Small town (4500-9999)	0.54	1.42	2.75	77.63	17.66	27657
F:Intermediate settlement (2250-4499)	0.87	1.43	2.73	77.51	17.46	18063
G:Village (1000-2249)	1.31	1.89	4.00	73.94	18.86	18382
H:settlements(<1000)&open countryside	5.80	3.13	6.79	67.32	16.96	123667

Table A1: Valid, invalid and matching XUPRNs by personal characteristics and settlement type

	No XUPRN - Census and health register	No XUPRN - Census	No XUPRN – Health register	Same XUPRN	Different XUPRN	Number
Tenure						
Owner occupier	2.10	1.41	3.47	78.31	14.72	342238
Social rented	0.58	1.63	2.75	75.87	19.17	77224
Private rented	2.23	3.29	4.94	55.79	33.76	29499
Accommodation type						
Detached house/bungalow	3.63	2.06	4.86	74.03	15.42	193125
Semi-detached house/bungalow	0.41	0.79	2.07	80.51	16.20	125737
Terraced	0.31	0.76	2.02	80.11	16.79	112032
Flat/tenement	1.22	5.93	5.81	53.82	33.23	14311
Converted/shared house	3.15	10.05	8.22	35.06	43.53	1971
Commercial building	6.08	8.98	15.19	30.52	39.23	757
Caravan/other mobile/temporary	12.51	9.07	7.55	45.37	25.50	1047
Communal establishment	6.00	18.44	14.16	24.06	37.34	5037
Household composition						
Couple with children	2.04	1.52	3.26	78.82	14.36	243279
Couple without children	1.44	1.66	3.41	71.95	21.54	44188
Single parent	1.27	1.32	2.86	74.98	19.57	57622
One person family	1.52	2.82	4.51	58.73	32.41	21773
Pensioner	1.72	1.35	3.96	83.74	9.22	44417
Other	2.30	1.68	4.32	69.79	21.90	37682

Table A2: Valid, invalid and matching XUPRNs by household characteristics

	No XUPRN - Census and health register	No XUPRN – Census	No XUPRN – health register	Same XUPRN	Different XUPRN	Number
Economic activity						
Employee	1.59	1.52	3.18	73.83	19.88	149409
Self-employed	3.50	2.04	6.86	67.59	20.01	26372
Unemployed	2.02	2.24	4.14	67.73	23.86	12596
Student (economically active)	1.21	2.58	2.84	74.63	18.74	6751
Retired	1.69	1.33	3.57	84.38	9.04	36359
Student (economically inactive)	1.95	3.38	4.02	70.24	20.41	17349
Home-maker	1.70	1.58	3.09	77.55	16.07	23560
Permanent sick	1.69	1.85	3.95	77.12	15.40	29452
Other	2.15	2.09	4.11	72.75	18.90	13263
SES						
Professional	1.55	1.58	3.49	74.46	18.91	79377
Intermediate	1.49	1.50	2.86	77.77	16.39	35279
Self-employed	3.62	2.05	6.84	68.74	18.74	27940
Lower supervisor	1.38	1.52	3.26	74.74	19.10	27815
Routine	1.69	1.50	3.20	76.97	16.64	100858
Not working	2.45	2.37	5.05	70.31	19.82	20165
Students	1.84	2.33	3.53	74.83	17.48	23677
Education						
No qualification	2.00	1.60	4.10	77.60	14.70	133758
Foundation level	1.80	1.60	3.40	72.50	20.70	53921
A level	1.70	1.80	3.50	74.20	18.80	78401
Degree and higher	1.60	1.90	3.50	71.30	21.70	49031

Table A3: Valid, invalid and matching XUPRNs by labour market and educational characteristics

	B	S.E.	OR	T	B	S.E.	OR	T
Response	**MATCH (Model 1a)**				**MATCH (Model 1b)**			
Fixed Part								
Constant	2.10	0.04		55.37	1.94	0.04		50.97
Age								
0-17 (reference			1.00				1.00	
18-24	-0.40	0.01	0.67	-28.29	-0.38	0.01	0.69	-26.79
25-34	-0.78	0.01	0.46	-56.00	-0.74	0.01	0.48	-52.57
35-44	-0.52	0.02	0.60	-32.44	-0.45	0.02	0.64	-27.94
45-54	-0.26	0.02	0.77	-14.67	-0.18	0.02	0.84	-9.72
55-64	-0.08	0.02	0.93	-4.00	0.02	0.02	1.02	1.11
65-74	-0.06	0.02	0.94	-2.86	0.05	0.02	1.05	2.14
75-84	-0.08	0.03	0.92	-3.00	0.04	0.03	1.04	1.69
>85	-0.12	0.05	0.89	-2.60	0.15	0.04	1.16	3.57
Gender								
Male (reference			1.00				1.00	
Female	0.30	0.01	1.35	37.75	0.29	0.01	1.33	35.75
Community background								
Catholic (reference)			1.00				1.00	
Protestant & other Christian	0.12	0.01	1.12	11.60	0.12	0.01	1.13	12.20
None	-0.03	0.06	0.97	-0.44	0.08	0.06	1.09	1.31
Other	0.02	0.03	1.02	0.69	0.00	0.03	1.00	0.08
Limiting long-term illness								
Yes (reference)			1.00				1.00	
No	-0.16	0.01	0.85	-14.45	-0.14	0.01	0.87	-13.09

Marital status								
Married (reference)			1.00				1.00	
Single	-0.21	0.01	0.81	-15.77	-0.18	0.01	0.84	-13.54
Remarried	-0.01	0.03	0.99	-0.39	-0.03	0.03	0.97	-0.96
Separated	-0.34	0.02	0.71	-15.45	-0.35	0.02	0.70	-15.26
Divorced	-0.33	0.02	0.72	-15.00	-0.31	0.02	0.73	-14.05
Widowed	-0.16	0.02	0.86	-7.43	-0.14	0.02	0.87	-6.71
SOA deprivation								
Multiple deprivation measure	-0.01	0.00	0.99	-6.00	-0.01	0.00	0.99	-8.00
Settlement band								
Belfast (reference)			1.00				1.00	
B:Derry Urban Area	0.37	0.06	1.44	6.02	0.34	0.06	1.40	5.54
C:Large town (population=18000-74999)	0.09	0.04	1.09	2.37	0.06	0.04	1.06	1.63
D:Medium town (10000-17999)	0.16	0.05	1.18	3.02	0.14	0.05	1.15	2.52
E:Small town (4500-9999)	-0.18	0.05	0.84	-3.91	-0.17	0.05	0.84	-3.80
F:Intermediate settlement (2250-4499)	-0.15	0.05	0.86	-3.23	-0.16	0.05	0.85	-3.45
G:Village (1000-2249)	-0.02	0.04	0.98	-0.45	-0.07	0.04	0.94	-1.68
H:settlements(<1000)&open countryside	-0.45	0.03	0.64	-15.48	-0.41	0.03	0.67	-13.97
Fermanagh								
No (reference)			1.00				1.00	
Yes	-2.17	0.09	0.11	-24.95	-2.16	0.09	0.12	-24.79
Address change year before the census								
No address change (reference)			1.00				1.00	
Within NI	-1.64	0.01	0.19	-125.85	-1.69	0.01	0.19	-129.77
Outside NI	-0.53	0.04	0.59	-13.64	-0.62	0.04	0.54	-16.37
Housing tenure								
Owner occupied (reference)			1.00					
Social rented	-0.22	0.01	0.81	-17.92				
Private rented	-0.63	0.02	0.53	-42.27				

Accommodation type								
Detached dwelling (reference)					1.00			
Semi-detached house/bungalow					0.20	0.01	1.23	18.45
Terraced					0.25	0.01	1.28	20.50
Flat/tenement					-1.08	0.02	0.34	-51.33
Converted/shared house					-1.34	0.05	0.26	-25.30
Commercial building					-1.67	0.09	0.19	-19.65
Caravan/other mobile/temporary					-0.92	0.07	0.40	-13.78
Communal establishment					-2.57	0.04	0.08	-65.95
Random Part								
Level 2 variation	0.17	0.01			0.17	0.01		
Level 1 variation	1	0			1	0		
Level 2 Units: SOA	890				890			
Level 1 Units: Individuals	441646				446605			

Table A4: All age models (dependent variable=MATCH) for all records alternating tenure and accommodation type

	B	SE	OR	T	B	SE	OR	T
	MATCH (Model 2a)				MATCH (Model 2b)			
Response								
Fixed Part								
Cons	2.15	0.04		61.31	2.06	0.04		57.17
Age								
0-17 (reference							1.00	
18-24	-0.43	0.02	0.65	-26.63	-0.40	0.02	0.67	-24.69
25-34	-0.83	0.02	0.43	-55.60	-0.78	0.02	0.46	-51.67
35-44	-0.50	0.02	0.61	-27.56	-0.42	0.02	0.65	-23.56
45-54	-0.14	0.02	0.87	-7.00	-0.06	0.02	0.95	-2.75
55-64	0.15	0.02	1.16	6.61	0.25	0.02	1.28	10.74
65-74	0.24	0.03	1.26	9.04	0.34	0.03	1.40	12.92
75-84	0.25	0.03	1.28	7.75	0.36	0.03	1.43	11.09
>85	0.27	0.06	1.31	4.68	0.55	0.05	1.73	10.58
Gender								
Male (reference							1.00	
Female	0.34	0.01	1.40	37.78	0.32	0.01	1.38	35.78
Community background								
Catholic (reference)							1.00	
Protestant & other Christian	0.14	0.01	1.15	11.58	0.14	0.01	1.15	11.50
None	0.02	0.07	1.02	0.26	0.11	0.07	1.12	1.59
Other	0.06	0.03	1.06	2.03	0.04	0.03	1.04	1.41

Limiting long-term illness								
Yes (reference)							1.00	
No	-0.15	0.01	0.86	-11.85	-0.13	0.01	0.88	-10.23
Marital status								
Married (reference)							1.00	
Single	-0.17	0.01	0.85	-12.00	-0.15	0.01	0.86	-10.50
Remarried	-0.07	0.03	0.93	-2.19	-0.08	0.03	0.92	-2.56
Separated	-0.38	0.02	0.68	-15.83	-0.40	0.02	0.67	-16.63
Divorced	-0.35	0.02	0.70	-14.71	-0.34	0.02	0.71	-14.33
Widowed	-0.19	0.03	0.83	-7.19	-0.18	0.03	0.84	-6.73
SOA deprivation								
Multiple deprivation measure	0.00	0.00	1.00	-4.00	-0.01	0.00	0.99	-6.00
Settlement band								
Belfast (reference)							1.00	
B:Derry Urban Area	0.42	0.05	1.53	8.15	0.39	0.05	1.48	7.48
C:Large town (population=18000-74999)	0.26	0.03	1.30	7.68	0.23	0.03	1.26	6.71
D:Medium town (10000-17999)	0.03	0.05	1.03	0.67	0.00	0.05	1.00	0.00
E:Small town (4500-9999)	-0.07	0.04	0.93	-1.68	-0.08	0.04	0.92	-1.98
F:Intermediate settlement (2250-4499)	-0.12	0.05	0.89	-2.69	-0.15	0.05	0.86	-3.26
G:Village (1000-2249)	-0.04	0.04	0.96	-0.90	-0.08	0.04	0.92	-2.05
H:settlements(<1000)&open countryside	-0.19	0.03	0.83	-7.23	-0.21	0.03	0.81	-7.96
Fermanagh							1.00	
No (reference)	-0.46	0.08	0.63	-5.83	-0.50	0.08	0.61	-6.27
Yes								
Address change year before the census								
No address change (reference)							1.00	
Within NI	-1.74	0.01	0.18	-123.93	-1.79	0.01	0.17	-128.07
Outside NI	-0.56	0.04	0.57	-13.61	-0.65	0.04	0.52	-15.85

Housing tenure								
Owner occupied (reference	-0.28	0.01	0.76	-21.62				
Social rented	-0.67	0.02	0.51	-42.13				
Private rented								
Housing type								
Detached dwelling (reference)							1.00	
Semi-detached house/bungalow					0.05	0.01	1.05	3.83
Terraced					0.07	0.01	1.07	4.79
Flat/tenement					-1.10	0.02	0.33	-45.63
Converted/shared house					-1.26	0.06	0.28	-21.29
Commercial building					-1.48	0.10	0.23	-15.08
Caravan/other mobile/temporary					-0.72	0.08	0.49	-8.61
Communal establishment					-2.37	0.05	0.09	-51.41
Random Part								
Level 2 variation	0.10	0.01		-42.13	0.10	0.01		-42.13
Level 1 variation	1	0			1	0		
Level 2 Units: SOA	890				890			
Level 1 units: Individuals	411622				414671			

Table A5: All age models (dependent variable=MATCH) for valid XUPRNs only alternating tenure and accommodation types