Migration Between Indian Reserves and Off-Reserve Areas: an Exploratory Analysis Using Census Data Linkage

Jean-Dominique Morency¹ · Patrice Dion¹ · Chantal Grondin¹

Received: 27 March 2020 / Accepted: 28 February 2021 / Published online: 7 May 2021
© Crown 2021

Abstract

New data linkages between censuses show that migration flows between Indian reserves and off-reserve areas from 2006 to 2011 and from 2011 to 2016 resulted in negative net migration for Indian reserves, meaning that—overall—more people left Indian reserves than entered them. These results differ from the portrait shown by the retrospective information from the 2011 and 2016 censuses, which indicates positive net migration for Indian reserves. A comparison of the information in the two sources revealed two types of inconsistencies that contributed to the observed differences: (1) inconsistencies in migrant status, and (2) inconsistencies in the origin location of migrants, i.e., the retrospective information about a migrant's place of residence 5 years earlier does not match the place where the migrant was enumerated in the previous census. Results from this paper suggest that there are limitations to using retrospective information on the place of residence 5 years prior to a census to derive estimates of internal migration flows for small geographic areas, such as Indian reserves. New data linkages are a source of information that can be used to validate and improve these estimates, as well as to derive alternative estimates. However, data linkages also have limitations and require careful preparation before use, particularly when it comes to calculating weights to accurately account for unlinked records.

Keywords Indian reserve · Migration · Data linkage between censuses · Weighting · Geocoding

The views and opinions expressed in this article do not necessarily reflect those of Statistics Canada.

✉ Jean-Dominique Morency
jean-dominique.morency@canada.ca

Patrice Dion
patrice.dion@canada.ca

Chantal Grondin
chantal.grondin@canada.ca

¹ Statistics Canada, Ottawa, Canada

Résumé

De nouveaux couplages de données de recensements montrent que les flux migratoires entre les réserves indiennes et les régions hors réserve entre 2006 et 2011 et entre 2011 et 2016 ont entraîné une migration nette négative pour les réserves indiennes, ce qui signifie que, dans l'ensemble, il y a eu plus de personnes ayant quitté les réserves indiennes que de personnes y étant entré. Ces résultats diffèrent du portrait dressé à partir de l'information rétrospective des recensements de 2011 et 2016, qui montre un solde migratoire positif pour les réserves indiennes. Une comparaison des informations dans les deux sources a révélé deux types d'incohérences contribuant aux différences observées: (1) des incohérences dans le statut de migrant, et (2) des incohérences dans le lieu de résidence d'origine des migrants, c'est-à-dire l'information rétrospective sur le lieu de résidence 5 ans plus tôt ne correspond pas au lieu de résidence où le migrant a été dénombré lors du recensement précédent. Les résultats de cette étude suggèrent qu'il existe des limites à l'utilisation de l'information rétrospective sur le lieu de résidence 5 ans auparavant au recensement pour estimer les flux migratoires internes pour de petites régions géographiques, comme les réserves indiennes. Les nouveaux couplages de données sont une source d'information qui peut être utilisée pour valider et améliorer ces estimations, ainsi que pour dériver des estimations alternatives. Cependant, les couplages de données ont également leurs limites et nécessitent une préparation minutieuse avant toute utilisation, en particulier lorsqu'il s'agit de calculer des poids pour tenir compte adéquatement des enregistrements non appariés.

Mots-clés Réserve indienne · Migration · Couplage de données du recensement · Pondération · Géocodage

1 Introduction

In the 1960s and 1970s, there was significant interest in studying the urbanization of Indigenous populations (Cooke & Bélanger, 2006). The movements or migrations of Indigenous people toward cities were a key part of the urbanization process. These movements toward cities were perceived to be motivated primarily by a desire to escape isolation and the difficult economic conditions of Indian reserves, and thus were seen to be an indication of various needs not been fulfilled in or outside of Indigenous communities (Falconer, 1985; Trovato et al., 1994).

Subsequent research has provided a more nuanced portrait of the phenomenon. Demographic studies of the population living on Indian reserves show that this population is growing, and that there are more people entering Indian reserves than leaving them (Amorevieta-Gentil et al., 2015). Between 2006 and 2016, the population living on reserve increased by 42,500 people.¹ The main factor for this growth is the

¹ This increase is slightly higher than the one (37,000 people) obtained by subtracting the population on Indian reserves estimated in the 2016 Census (380,000) by the one estimated in the 2006 Census (343,000) because Indian reserves are identified using a single definition (the one used in the 2016 Census) (Statistics Canada, 2018), and it excludes the population that has ever lived on an incompletely enumerated Indian reserve or on a reserve in the territories in the 2006 or 2016 Census or the 2011 National Household Survey.

high fertility rate of the Indigenous population, which has remained above the generation replacement level (Morency et al., 2018). However, internal migration is also considered to be a contributor to the population growth on Indian reserves, particularly for Registered Indians (Clatworthy & Norris, 2007, 2014; Cooke & Penney, 2019). Estimates based on the retrospective census information have shown steady, small positive net migration for Indian reserves since the 1970s (Amorevieta-Gentil et al., 2015; Clatworthy & Norris, 2007, 2014; Cooke & Penney, 2019; Norris et al., 2004; Signer, 1977). In the five quinquennial periods beginning with 1986 to 1991 until 2006 to 2011, the population of Registered Indians on Indian reserves increased by between 9000 and 15,000 people through internal migration (Clatworthy & Norris, 2007, 2014; Cooke & Penney, 2019; Norris et al., 2004). Clatworthy & Norris (2014) found that the people who migrated to Indian reserves came primarily from rural areas, but also from urban areas. Some studies suggest that a substantial portion of the migration into and out of Indian reserves is circular in nature,² i.e., that the migration is part of repetitive movements between two or more regions (Cooke, 1999). Furthermore, according to prior research, migration among Indigenous populations is influenced by several factors and triggered by multiple motivations (Cooke, 1999; Trovato et al., 1994).

Estimates of migration flows between Indian reserves and off-reserve areas in the existing literature have been derived using the retrospective information collected in the long-form census on the place of residence 5 years prior³. This is not only because the census was the only source of information available, but also since its large sample size makes it possible to measure a relatively rare event among relatively small populations. However, the recent availability of linked data from consecutive censuses provides new possibilities to study migration by comparing the place of residence of individuals enumerated in two consecutive censuses. Moreover, data linkage provides characteristics and place of residence (including Indian reserves) of individuals at two points in time, making it possible to examine both the characteristics associated with migration and the changes induced by migration.

The main goal of this paper is to study the potential uses of census data linkage in analyzing the migration flows into and out of Indian reserves, and to compare these results with the traditional retrospective approach based on a single census (long-form questionnaire).

This paper has five main sections. In the following section, we provide a description of the benefits and limitations of using census data to analyze geographic mobility (particularly at low levels of geography), and the potential of using data linkage to fill some significant data gaps. Next, the data sources and methods used to link two consecutive censuses and calculate the weights to compensate for non-linked

² Indigenous populations' attachment to a particular community may account, in part, for these circular movements, especially for those who are living or have previously lived on an Indian reserve.

³ Most census data are collected through two types of questionnaire—the short form and the long form. The short-form questionnaire asks fewer questions, but targets 100% of the Canadian population, while the long-form questionnaire contains more questions (including on mobility), but is sent only to a sample of the Canadian population living in private households.

records are explained. The strengths and limitations of using this new data source for migration analysis are then discussed, followed by an overview of the various census data linkages that were used in this paper to calculate alternative estimates of migration flows between Indian reserves and off-reserve areas. These estimates are then compared with those obtained from a single census (long-form questionnaire). The new estimates obtained from data linkage indicate that Indian reserves lost more people than they gained through internal migration between 2006 and 2016, contrary to what each individual census shows. Still, despite the striking change of sign in net migration, Indian reserves have continued growing in the period due to birth rates, and they have not experienced anything close to an exodus. The fifth section gives an overview of how data linkage was used to compare the place of residence information in the 2016 Census long-form questionnaire with that from the 2011 Census long-form questionnaire (National Household Survey [NHS]). According to the results of this analysis, various factors related to the way the information in the census long-form questionnaire is reported and processed⁴ may affect the accuracy of the information on migrant status and place of residence five years prior to the census.

2 Using the Census to Analyze Geographic Mobility

Because of the richness of its content, its large sample size, and its fine geographic detail, the census long-form questionnaire is a preferred source for analyzing geographic mobility. However, the census also has certain limitations when it comes to analyzing migration into and out of Indian reserves, and these limitations deserve attention. They are described here, and a distinction is made between those that are intrinsic to the census as a data collection tool, and those that are inherent to the use of retrospective information, such as the possibility of recall and response errors, proxy responses, and issues with geocoding an individual's reported place of residence 5 years prior to the census. While the first limitation affects the quality of the migration estimates regardless of the method used (the retrospective approach or the comparison of linked censuses), the second limitation can be avoided by using census data linkage.⁵

2.1 Limitations Inherent to the Census

Estimates of migration flows into and out of Indian reserves can be affected by how the census reaches its target population. Some individuals may not be enumerated (undercoverage), while others may be counted more than once (overcoverage). Compared with the rest of the population, highly mobile individuals and young adults are

⁴ Data processing includes data entry, data editing and imputation, and geocoding.

⁵ Linked files also have their own shortcomings, which will be described in detail in Section 3.

more likely to be missed. This could be an issue for the Indigenous population in particular, as it is generally younger and more mobile than the Canadian population as a whole (Clatworthy & Norris, 2014; Dion & Coulombe, 2008). The combined effect of undercoverage and overcoverage is referred to as census net undercoverage (since undercoverage tends to be greater). Net undercoverage is positive in most regions of Canada and is generally higher on than off Indian reserves (Clatworthy & Norris, 2007, 2014; Norris & Clatworthy, 2011).

Some individuals living on Indian reserves may also not be enumerated because census data were either not collected on their reserve or were incomplete. For example, there were 14 reserves (referred to as incompletely enumerated reserves) in the 2016 Census and 31 in the 2011 Census⁶ for which data were not available (Statistics Canada, 2017b). Any individual who lived in an incompletely enumerated reserve in the start year, the end year, or both years of a linkage cannot be linked, and is therefore excluded from the analysis. In a linkage of the 2011 and the 2016 census, this represents about 40,000 individuals from 35 reserves (Statistics Canada, 2020), or 10% of the total population living on Indian reserves.

Other types of non-sampling errors that can affect the quality of estimates from census data are misunderstanding of the questions, incorrect data capture, or non-response (Statistics Canada, 2018). Non-response, in particular, is hard to measure for specific populations, but could be relatively more prevalent among Indigenous populations for reasons such as apprehensions about government data collection, use of the data for government control, or barriers related to language or literacy and numeracy issues (Wright et al., 2020). Higher level of non-response increases the risk of non-response bias, which occurs when non-respondents differ from respondents (Statistics Canada, 2018).

The quality of migration flow estimates also depends on whether enumerated individuals are registered in the right place. For example, individuals enumerated at the wrong residence during the census may mistakenly be considered migrants. While the census data are generally thought to be of good quality in this regard, it is more challenging to accurately collect the usual place of residence for some populations, such as individuals living in more than one place (e.g., students living in an apartment during the school year and with their parents the rest of the time, or workers living in another area for work).

Lastly, the census long-form questionnaire is not distributed to individuals living in institutions. This could be non-negligible for Indigenous populations, as they tend to be overrepresented among the institutional population (Clatworthy & Norris, 2007, 2014; Norris & Clatworthy, 2011).

2.2 Limitations Inherent to the Use of Retrospective Information

Collecting information on individuals place of residence 1 or 5 years prior to the census date relies on their recollection of past events, which may be poor, especially after 5 years. Because the reference date is not linked to any specific event that could

⁶ This figure was 36 in the NHS because five reserves that had been enumerated in the 2011 Census were partially enumerated in the NHS.

provide a reference point for respondents, it could be more challenging for them to remember this information. Memory problems can affect the quality of the migration flow size estimates (e.g., a migration could be forgotten) and of the migrants' characteristics (e.g., a place of residence error or a forgotten postal code could affect the quality of the geocoding). Accurately remembering information about one's place of residence 5 years prior to the census can be more challenging for highly mobile individuals (e.g., circular migrants).

The vagueness of the concept of municipality used in the census question on the place of residence 1 or 5 years earlier may also affect the quality of responses. Statistically speaking, municipality refers to the boundaries of the census subdivision (CSD), and an Indian reserve has a distinct CSD name and code.⁷ However, there can be discrepancies between statistical rules and individuals' perception of their reality. For instance, some respondents provide a name of municipality (CSD) of origin that refers to the closest census metropolitan area (CMA) or census agglomeration (CA) instead of the name of their Indian reserve.⁸

The quality of responses can also be affected by the fact that household members' responses are often given in an indirect interview (proxy responses).⁹ It can be difficult for a person to know or remember the place of residence of each member of their household 5 years earlier, especially if they are not immediate family. This is more likely to be the case for larger households (which are more prevalent on Indian reserves). Furthermore, the effects of proxy responses are potentially greater on reserves, since data are collected through interviews with enumerators who ask the questions to the person or people in the household when they visit, compared with online questionnaires, for which the person most equipped to respond for all family members will generally do so.

Once collected, the information goes through an intensive data processing step. Geocoding—a process in which collected information is matched to a CSD within the Standard Geographic Classification¹⁰—is key for information on the place of residence 1 or 5 years prior to the census. This is a highly sophisticated process that continuously improves from census to census. However, there are circumstances that may affect its precision, for example, when the place of residence 1 or 5 years earlier is not a unique name, which is the case for a considerable number of Indian reserves. A number of reserves have names that are similar to those of neighboring off-reserve CSDs (e.g., Chilliwack 1 [on reserve] and Chilliwack [off reserve]), which can create confusion in the geocoding process, especially if both CSDs share the same postal code. The complex spelling of certain Indian reserves—which are more prone to errors—can also lead to geocoding errors. Postal codes help improve

⁷ The question on the place of residence 5 years earlier asked migrants to “specify the name of the city, town, village, township, municipality or Indian reserve of residence 5 years ago,” as well as the province or territory of residence, and the postal code.

⁸ For example, a person who lived on the Mashteuiatsh reserve (on-reserve CSD) 5 years earlier could report that they lived in Roberval (off-reserve CSD), the main agglomeration nearest to that reserve, which could lead to omitting an off-reserve migration or adding a false on-reserve migration.

⁹ Indirect interview (or proxy response) is where a person (the proxy) answers questions on behalf of another person (the sampled interview subject).

¹⁰ For more information on the Standard Geographic Classification, visit <http://www23.statcan.gc.ca/imdb/p3VD.pl?Function=getVD&TVD=314300>.

the geographic coding process, but less so in rural areas, as one postal code can often cover more than one CSD. Postal code information is also often unreported, which is most likely because of respondents' trouble remembering. Geocoding-related aspects will be addressed in greater detail in Section 5 and Appendix 2.

Overall, at large geographic scales, these obstacles are likely minor and do not significantly bias the results. Migratory flows—as measured by census data—are generally in line with the measurements from other sources, such as income tax records. However, for lower levels of geography (e.g., CSD level), these obstacles could be more problematic. Unfortunately, a number of factors can make comparisons with other data sources harder, including sources that lack variables, different methods of operationalizing migration, differences in the target populations, and different reference periods. Census data linkage provides new opportunities for validation and analysis.

3 Description of Data Files and Method

Linking files is a complex task. The two most challenging steps are linking records—which involves linking as many records as possible while avoiding false positives (i.e. two records that belong to different individuals)—and accurately weighting the file so that the linked individuals are representative of the whole population of interest (including individuals who could not be linked). Below is a brief description of the data files used in this study and the methods used for linkage and weighting.

3.1 Description of Linked Census Data Files

Every 5 years, Statistics Canada conducts the Census of Population, which is a key source of information on the Canadian population. All Canadians are asked to complete the census short-form questionnaire. A sample of the population is also asked to answer the census long-form questionnaire, which is more elaborate in terms of content. The two questionnaires are mandatory, except in 2011, when the census long-form questionnaire was replaced by the 2011 NHS, which—although it was similar in content to the previous (2006) and subsequent (2016) census long-form questionnaires—was voluntary. The sampling scheme for the long-form questionnaire has also varied over the years. The proportion of sampled households was 33% in the NHS and 25% in the 2016 Census. However, for all northern communities and Indian reserves, no sampling was used. In other words, all households were asked to complete the census long-form questionnaire.

Data linkage makes it possible to combine the data collected through these questionnaires from two consecutive years. Each combination has its advantages. For example, linkage between two consecutive census long-form questionnaires contains a large number of variables at two points in time, allowing for comprehensive multivariate analyses. Conversely, linkage between two consecutive census short-form questionnaires contains a much larger number of records. Hence, a linkage between

a census short-form questionnaire and a census long-form questionnaire provides a mix of benefits: a relatively large number of records with a large number of variables (at one point in time). In general, the objectives of a researcher will dictate which type of linkage to use.

Several different linkages built from the 2006, 2011, and 2016 census short-form questionnaires (S2006, S2011, and S2016, respectively), and the 2011 NHS and the 2016 Census long-form questionnaires (L2011 and L2016, respectively) were used for this study.¹¹ The goal was to assess the consistency of the results and make sure that they are not artifacts of the linking methodology and the weighting process (discussed in the next sections). Likewise, the linkage of S2006/S2011 was made to assess consistency overtime.¹²

In all the files considered, the geographical limits used were those defined based on the 2016 Census. The five different linkages¹³ used in this study are:

- L2011/L2016: linkage of the 2011 NHS and the 2016 Census long-form questionnaires (L2011 and L2016)
- S2011/S2016: linkage of the 2011 and 2016 census short-form questionnaires (S2011 and S2016)
- L2011/S2016: linkage of the 2011 Census long-form questionnaire (L2011) and the 2016 Census short-form questionnaire (S2016)
- S2011/L2016: linkage of the 2011 Census short-form questionnaire (S2011) and the 2016 Census long-form questionnaire (L2016)
- S2006/S2011: linkage of the 2006 and 2011 census short-form questionnaires (S2006 and S2011).

For clarity, retrospective information in a census file will be referred to with “_R” in the file name. For example, the place of residence in L2016_R is the place of residence 5 years earlier collected retrospectively in L2016.

3.2 Linkage Method

The linkage of individuals from two consecutive censuses was performed using a multistage method. This involved using personal information (given name, family name, sex, and birth date) and household information (geography [province or territory and census division], postal code, telephone number, and household composition). Not all of this information is used at each stage. For instance, some stages use the geographic

¹¹ At the time of publication of this article, these files could only be accessed through a partnership with researchers from Statistics Canada.

¹² There are other variants possible. However, the linkage and weighting process of each variant involve a lot of work, and the selection of variants used in this paper was considered appropriate for the objectives at hand.

¹³ For more information on the data sources used for the linkages, visit:
 S2016 and L2016: <http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/index-eng.cfm>
 L2011: <http://www12.statcan.gc.ca/nhs-enm/2011/ref/dict/index-eng.cfm>
 S2011: <http://www12.statcan.gc.ca/census-recensement/2011/ref/dict/index-eng.cfm>
 S2006: <http://www12.statcan.gc.ca/census-recensement/2006/ref/dict/overview-apercu/index-eng.cfm>.

information, while at other stages, geography is not considered at all in order to maximize the linkage of migrants. Only single links were accepted to avoid linking records that do not belong to the same individual.¹⁴ Because of weak linkage rates for the population living in institutions and in collective dwellings (for linkages using the census short-form questionnaire), these populations were excluded from the linkage process. Therefore, this study covers only the population living in private households at both points in time.¹⁵

Linkage rates are lower for the population living on reserve than for the entire Canadian population. For example, using the population of the start-year file (2011) as the denominator, and observing the whole population intended to be linked, the linkage rates for the population living on reserve are 42% in S2011/S2106 and 46% in L2011/S2016, in comparison to 80% and 83%, respectively, for the population living off reserve (please see Appendix 1 for important considerations on the calculation and interpretation of linkage rates).

A number of reasons can explain the differences in linkage rates observed between subgroups of the population, but incorrect or missing information could be an important factor. First, the collection method using enumerators—the predominant method used in Indigenous communities—could introduce errors in the names collected because of misspellings. Moreover, the impact of this situation could be made worse by names with a complex spelling.¹⁶ The information collected (e.g., date of birth) can also be less accurate when collected through indirect interviews. The use of enumerators on reserves can increase the chances for indirect interviews if the questions are asked only to the people present in the household at the time of their visit. Proxy responses may also be more frequent in large households, which is a situation that is more common on reserves.¹⁷

The relatively weak linkage rates for the population living on reserve are a source of concern because biases could occur if unlinked individuals have characteristics different from those of linked individuals (selection bias). In the context of this study, there is no particular reason to believe that the quality of the matches is lower among migrants than among non-migrants because the control methods—i.e., methods used to assess the quality of each linked record—are the same in both cases. Moreover, the proportion of reserve in-migrants and out-migrants is similar among

¹⁴ For a more detailed description of the methodology used to micro-match the 2011 and the 2016 censuses, please refer to Part 1 of the document entitled “Method for linking the 2011 and 2016 censuses and calculating weights for linked census files” provided as [supplementary material](#) to this article.

¹⁵ According to the 2011 and 2016 censuses, most of the population living in institutions were located off reserve (there are few institutions on reserve). The exclusion of the population living in institutions (which is proportionally higher among the Indigenous population) means that some migratory movements were not enumerated: (1) migrations from institutions located off-reserve to Indian reserves, (2) migrations from institutions located on reserve to off-reserve areas, (3) migrations from off-reserve areas to institutions located on reserve; and (4) migrations from Indian reserves to institutions located off-reserve. This exclusion can impact the global estimates of reserve in-migrants and out-migrants, as well as the reserve’s net migration. Unfortunately, there are no data available to estimate the overall impact of this exclusion on the results.

¹⁶ The use of special characters (e.g., “8” and “Ø”) in certain Indigenous languages can cause particular difficulties.

¹⁷ According to the 2016 Census, 25% of on-reserve households had five or more people, compared with only 8% of off-reserve households.

the linked and unlinked populations, which likely means that there is no bias for this specific characteristic. Lastly, linkage rates were similarly low for people living on reserve in 2011 and in 2016, and should be therefore similar for both reserve in-migrants and out-migrants. While these facts are reassuring, they do not imply a total absence of bias. The choice of an appropriate weighting strategy (described next) can minimize potential biases induced by the linkage process.

3.3 Weighting Method

Linked files are subsets of the source files they stem from.¹⁸ For example, the S2011/S2016 file is a subset of the 2011 Census file because not everyone in the 2011 Census population was linked to the 2016 Census file. People who were enumerated in the 2011 Census but were not linked to the 2016 Census file can be thought as non-respondents of a survey. Therefore, to infer results from the S2011/S2016 file for the entire 2011 Census population, a special set of weights must be computed to account for these non-respondents (i.e., non-linked records).

The weights must also take into account the fact that some people in the 2011 Census population could not be linked to the 2016 Census because they had either died or left the country before the 2016 Census.¹⁹ These people can be considered “out-of-scope” cases in a longitudinal survey sample. As a result, a special adjustment to the weights is needed to take these cases into account. The weights for the S2011/S2016 file make it possible to make inferences for the population that was alive and living in the country in both 2011 and 2016.²⁰ This is also true for other linked files used in this study. Various methods were used to identify the key variables and interaction terms to be used to create homogeneous adjustment classes depending on whether start year file is a long-form census or a short-form census. Tests were conducted to ensure that the results were not overly sensitive to the choice of variables.

In the S2011/S2016 file, all individuals have an initial weight of 1, since the S2011 file covers the entire population and every record is self-representative. This weight of 1 must first be adjusted to account for individuals in the 2011 Census population that were not linked to S2016. This process is somewhat similar to the calibration of weights for non-response in a survey. This adjustment is done among groups of people with similar characteristics (called homogeneous groups), made by combining several variables relevant to this study and available in S2011, including geography (province or territory, census division, and an on-reserve or off-reserve

¹⁸ Among the linked files used for this paper, four (L2011/L2016, S2011/S2016, L2011/S2016, and S2006/S2011) were weighted using the method outlined in this section. The fifth linked file (S2011/L2016), which was not used in this study to estimate migration flows, but only to document discrepancies between retrospective responses and linked information, was weighted using a much more simplified strategy that is outlined in Section 5.

¹⁹ Individuals living in an institution or collective dwelling in 2011 or 2016 were also excluded.

²⁰ The weighting takes into account the characteristics of individuals at the beginning and end of the period. As a result, the weighting aims to reflect the characteristics of individuals at both times, rather than at a specific moment (2011 or 2016).

indicator), age, sex, marital status, mother tongue (including Aboriginal and Inuit languages), language spoken at home (including Aboriginal and Inuit languages), household size, and type of census family. Within each homogeneous group, the weights of linked records with characteristics XYZ are inflated so that they also represent the non-linked records with the same characteristics XYZ. After this step, the sum of the weights for the linked records in the S2011/S2016 file is equal to the 2011 Census population size.²¹

The next step was to adjust the weights of the linked records, which now represented the entire 2011 Census population, to account for the fact that not all of them could be linked to S2016 because of death or emigration, i.e., some people in the 2011 Census population were no longer alive or living in Canada in 2016. To make this adjustment, estimates of the number of people in the 2016 Census population who were alive and in the country in 2011 were needed (i.e., the population that could potentially be linked), then the weights of the linked records were calibrated to these new totals. Estimates of the population that could potentially be linked between 2011 and 2016 were obtained using information collected in L2016_R. With this information, anyone in the 2016 Census population who was not born or did not live in Canada 5 years prior to the census (i.e., people who could not be linked) were removed, resulting in a weighted estimate of the population for both 2011 and 2016.²² These new totals were calibrated by province or territory of residence in 2016, place of residence on- and off-reserve, age, and sex. This made it possible for the final weights to be representative of the population that was alive and living in the country in both 2011 and 2016.²³ The same strategy was used to calculate the weights for the S2006/S2011 file.²⁴ The weights of the different linkage files are described in Table 1.

The strategy used to compute the weights for the L2011/L2016 file differs slightly from the strategy used for S2011/S2016 because it must account for the fact that not everyone was selected to respond to the NHS in 2011 and the long-form

²¹ A potential future improvement of the weighting process could be to identify individuals who died in the intercensal period using a linkage with Canadian vital statistics database and to remove them from the start year file.

²² The response categories “Born after May 10, 2011” and “Lived outside Canada” on the 2016 Census long-form questionnaire were used to identify individuals who could not be linked.

²³ In order to assess the impact of the weighting on the results, comparisons of the unweighted and weighted versions of the reserve and off-reserve out-migration rates were made using the S2011/S2016 file. The unweighted reserve out-migration rate is 158‰, while the weighted estimate is 160‰. The unweighted off-reserve out-migration rate is 10‰ while the weighted estimate is 12‰. These results suggest that the weighting does not distort significantly the outcomes.

²⁴ The population living in institutions or collective dwellings in 2016 was excluded automatically, as it is not part of the target population of L2016. However, the population living in institutions or collective dwellings (in which Indigenous people are overrepresented) in 2011 could not be excluded because this retrospective information was not collected in the census. This has two main—albeit minor—impacts on the weighting. This causes (1) a slight overestimation of the population that was living in private households in Canada on May 10, 2011, and May 10, 2016, and (2) a slight overestimation of the number of migrants (from one private household to another) who moved from an off-reserve location to an on-reserve location. Unfortunately, this problem cannot be resolved unless the required information is collected in the census.

questionnaire in 2016. Therefore, the initial weights were computed as the product of the 2011 NHS final weight and the 2016 Census long-form questionnaire final weight to reflect the inverse of the probability that an individual would be selected to complete both. The next steps (i.e., making adjustments for non-linked records and for deaths and emigration) are similar to those used for the S2011/S2016 file, except that, in the L2011/L2016 file, a much broader range of characteristics could be considered when adjusting for non-linked individuals. As a result, more complex methods (logistic regression and clustering techniques) were used to generate homogeneous response groups.

The distribution of the adjusted weights for the L2011/L2016 file (Table 1) shows wide variations by place of residence in 2011 and 2016, primarily because of the different collection strategies used in Indigenous communities. This has important implications, as an incorrect link between two records can have a major impact on the quality of the results if its weight is high.

The strategy used to weight the L2011/S2016 is similar to what was described above, but is not presented here to avoid redundancy.²⁵

3.4 Limitations of Linked Files

A large number of complex procedures must be carried out before using linked data to analyze migration flows. This is particularly true for Indigenous populations and populations living on Indian reserves because of inherent difficulties, such as relatively low response and linkage rates, as well as the use of different collection strategies. In addition to the limitations inherent to census data (described in Section 2.1), linked census files have their own limitations. However, while linked census files could be used to evaluate the quality of the retrospective information on the place of residence 5 years prior to the collection date (Section 5), no data file could be used to highlight errors in the places where individuals were enumerated, wrong or missed record linkages, or remaining selection biases that could not be compensated for through the reweighting process, despite the wide range of variables used.²⁶ Therefore, despite all of the precautions taken, users should keep these considerations in mind when using estimates from linked census files.

Lastly, estimates produced from linked census files are subject to sources of uncertainty that are impossible to measure. Unfortunately, it was not possible to provide variance for estimates computed using linked census files (nor for estimates computed using retrospective census information). Because migration is a fairly rare event, counts are often relatively small and must be interpreted with caution.

²⁵ For a more detailed description of the weighting procedure, readers are invited to consult Part 2 of the document entitled “Method for linking the 2011 and 2016 censuses and calculating weights for linked census files”, which is provided as [supplementary material](#) to this article.

²⁶ Different types of adjustments were tested in the computation of weights, and it was found that these adjustments had only a small impact on the estimation and did not affect the overall results. However, these tests were limited to the information available from the census data files.

4 Main Results

This section provides the estimates of migration flows into and out of Indian reserves computed using census data linkages for the quinquennial periods of 2006 to 2011 and 2011 to 2016. The range of sources makes it possible to verify the robustness of these estimates, as each file has its own strengths and limitations (described in Section 3). Listed first are the totals for Canada and each province, for large geographical areas following a rural–urban gradient, and then by age group. These estimates are compared with those obtained using retrospective information from the census. At the end of this section, an additional comparison with an alternative data source is presented.

4.1 Comparison of Results from Linked Files with Those Obtained from a Single Census

According to the retrospective information collected by the 2016 Census long-form questionnaire (L2016_R), between 2011 and 2016 Indian reserves in Canada had a net gain of about 10,600 people through internal migration (Table 2) which contributed to a 3.4% increase of their population during that period.²⁷ In contrast, estimates from the S2011/S2016, L2011/L2016, and L2011/S2016 files show net losses of 18,900, 20,400, and 18,700 people, respectively, which contributed to a decrease of between 5.4 and 5.9% of their population. The differences found were mainly in the number of out-migrants, as estimates of out-migrants from the linked files were more than double those from the 2016 Census. Estimates of in-migrants were fairly consistent across all sources.²⁸ Similar results were obtained for the 2006-to-2011 period by comparing estimates from L2011 with those from S2006/S2011 (Table 3). According to the L2011_R file, at the national level, Indian reserves had a net migration gain of 11,700 people, compared with a net loss of about 14,000 people in the S2006/S2011 file.

The results for the population of Registered Indians were similar (see Table 2). According to the 2016 Census (L2016_R), in the quinquennial period from 2011 to 2016, Indian reserves gained 6500 people through internal migration (which contributed to an increase of 2.4% of the Registered Indian population on reserve during this 5-year period) compared with net losses of 21,400 and 25,100 in L2011/L2016 using respectively the Registered Indian status information in 2016 and 2011, and a loss of 24,000 people in L2011/S2016 using the Registered Indian status information in 2011 (which contributed to a decrease of between 6.9 and 7.8% of the Registered Indian population). While it seems like it should not matter whether the information about Registered Indian status in 2011 or 2016 is used, the population may—in fact—differ slightly for a few reasons. For example, some people may have registered

²⁷ Results presented in this section exclude the population living in the territories or on an incompletely enumerated Indian reserve at the beginning or end of the period studied.

²⁸ Section 5 contains more information on the factors that can contribute to the differences observed between linked and cross-sectional census files. It is shown in this section that the seemingly consistent results between sources for the number of in-migrants actually hide some significant discrepancies.

Table 1 Description of weights in the S2011/S2016, L2011/L2016, L2011/S2016, and S2006/S2011 files

PLACE OF RESIDENCE		WEIGHTS				
S2011/S2016						
S2011	S2016	Average	Lowest (min.)	Lowest (5%)	Highest (95%)	Highest (max.)
Off reserve (1)	Off reserve (1)	1.2	0.3	1.0	1.5	137.5
Off reserve (1)	On reserve (1)	1.3	0.3	0.9	2.1	18.1
On reserve (1)	Off reserve (1)	2.4	0.7	1.0	5.2	56.1
On reserve (1)	On reserve (1)	2.3	0.4	1.0	4.9	47.7
L2011/L2016						
L2011	L2016	Average	Lowest (min.)	Lowest (5%)	Highest (95%)	Highest (max.)
Off reserve (3)	Off reserve (4)	22.7	0.2	4.6	57.2	695.7
Off reserve (3)	On reserve (1)	9.4	0.7	1.4	32.0	154.8
On reserve (1)	Off reserve (4)	11.6	0.7	2.1	28.2	174.8
On reserve (1)	On reserve (1)	2.7	0.4	1.3	5.3	124.6
L2011/S2016						
L2011	S2016	Average	Lowest (min.)	Lowest (5%)	Highest (95%)	Highest (max.)
Off reserve (3)	Off reserve (1)	5.8	0.5	1.4	14.3	210.9
Off reserve (3)	On reserve (1)	7.4	0.8	1.5	24.2	87.9
On reserve (1)	Off reserve (1)	3.0	0.7	1.3	6.5	45.1
On reserve (1)	On reserve (1)	2.6	0.3	1.2	5.5	42.1
S2006/S2011						
S2006	S2011	Average	Lowest (min.)	Lowest (5%)	Highest (95%)	Highest (max.)
Off reserve (1)	Off reserve (1)	1.2	0.1	1.0	1.5	75.9
Off reserve (1)	On reserve (1)	1.3	0.4	1.0	1.9	7.8
On reserve (1)	Off reserve (1)	2.0	0.4	1.1	4.0	24.3
On reserve (1)	On reserve (1)	2.0	0.2	1.1	3.8	43.4

The presence of weights less than 1 occurs because of the final calibration used to reduce the overall weights to account for deaths and emigration between the two censuses. Since the individuals who died or left the country cannot be identified, the weights of all linked individuals have been reduced which can generate weights of less than 1. Numbers in parentheses in the first two columns refer to the theoretical average weight based on the sampling plan of each data source. For instance, a value of 1 means that 100% of the population was sampled while a value of 4 means that 25% of the population was sampled.

Sources: S2011/S2106: Data linkage between the 2011 Census short-form questionnaire and the 2016 Census short-form questionnaire; L2011/L2016: Data linkage between the 2011 National Household Survey and the 2016 Census long-form questionnaire; L2011/S2016: Data linkage between the 2011 National Household Survey and the 2016 Census short-form questionnaire; S2006/S2011: Data linkage between the 2006 Census short-form questionnaire and the 2011 Census short-form questionnaire.

between 2011 and 2016, in part because of legislative changes to the *Indian Act*,²⁹ and some others may have given different responses in 2011 and 2016 regarding their Registered Indian status because of response mobility (Statistics Canada, 2019).

²⁹ The inverse situation—an individual with Registered Indian status losing their status—is not theoretically possible as per the current rules of the *Indian Act*.

Table 2 Estimates of Indian reserve in-migrants, out-migrants, and net migration between 2011 and 2016, total population and Registered Indian population, for Canada and the provinces, by data source

TOTAL POPULATION												
	S2011/S2016			L2011/L2016			L2011/S2016			L2016_R		
	In	Out	Net	In	Out	Net	In	Out	Net	In	Out	Net
	thousands											
CANADA	36.1	55.0	-18.9	34.0	54.5	-20.4	34.8	53.5	-18.7	31.7	21.1	10.6
ATLANTIC PROVINCES	1.3	1.9	-0.7	1.1	1.9	-0.8	1.1	1.9	-0.8	0.9	0.6	0.3
QUEBEC	1.8	4.0	-2.2	1.4	3.7	-2.2	1.4	4.0	-2.6	2.2	2.3	-0.1
ONTARIO	3.3	7.3	-4.0	3.3	7.7	-4.4	3.3	7.4	-4.1	4.4	2.3	2.1
MANITOBA	3.3	9.3	-6.0	3.6	8.9	-5.3	3.6	8.7	-5.1	3.3	3.8	-0.5
SASKATCHEWAN	4.9	9.1	-4.2	3.8	9.6	-5.8	4.1	9.2	-5.1	4.1	4.0	0.0
ALBERTA	4.4	7.4	-3.1	4.6	6.9	-2.3	4.5	7.0	-2.6	4.9	2.7	2.1
BRITISH COLUMBIA	17.8	16.6	1.2	16.8	16.4	0.4	17.5	15.9	1.5	12.3	5.6	6.7
REGISTERED INDIAN POPULATION IN 2016												
	S2011/S2016			L2011/L2016			L2011/S2016			L2016_R		
	In	Out	Net	In	Out	Net	In	Out	Net	In	Out	Net
	thousands											
CANADA	--	--	--	18.1	39.4	-21.4	--	--	--	20.3	13.8	6.5
ATLANTIC PROVINCES	--	--	--	0.7	1.3	-0.7	--	--	--	0.6	0.4	0.2
QUEBEC	--	--	--	1.0	2.7	-1.7	--	--	--	1.5	1.6	-0.1
ONTARIO	--	--	--	2.4	6.4	-4.0	--	--	--	3.3	1.7	1.6
MANITOBA	--	--	--	3.2	8.1	-4.9	--	--	--	2.8	3.3	-0.5
SASKATCHEWAN	--	--	--	3.2	8.8	-5.6	--	--	--	3.6	3.4	0.2
ALBERTA	--	--	--	3.7	5.5	-1.8	--	--	--	3.9	1.9	2.0
BRITISH COLUMBIA	--	--	--	4.6	7.3	-2.6	--	--	--	4.7	1.6	3.1

Table 2 (continued)
REGISTERED INDIAN POPULATION IN 2011

	S2011/S2016			L2011/L2016			L2011/S2016			L2016_R		
	In	Out	Net	In	Out	Net	In	Out	Net	In	Out	Net
	thousands											
CANADA	--	--	--	16.6	41.7	--	16.5	40.5	--	--	--	--
ATLANTIC PROVINCES	--	--	--	0.6	1.4	--	0.5	1.4	--	--	--	--
QUEBEC	--	--	--	0.9	2.8	--	0.8	3.2	--	--	--	--
ONTARIO	--	--	--	2.1	6.6	--	2.1	6.0	--	--	--	--
MANITOBA	--	--	--	3.0	8.4	--	2.9	8.2	--	--	--	--
SASKATCHEWAN	--	--	--	2.9	9.0	--	3.2	8.5	--	--	--	--
ALBERTA	--	--	--	3.5	6.2	--	3.4	6.2	--	--	--	--
BRITISH COLUMBIA	--	--	--	4.4	7.9	--	4.3	7.5	--	--	--	--

Results exclude the population living in the territories or on an incompletely enumerated Indian reserve in 2011 or 2016. Individuals who moved from one reserve to another in the same province or region are not accounted for as in-migrants or out-migrants. Results are rounded to the nearest hundred. For this reason, net migration may not be equal to the difference between in-migrants and out-migrants.

Sources: L2016_R: retrospective information from the 2016 Census (long-form questionnaire); L2011/L2016: Data linkage between the 2011 National Household Survey and the 2016 Census long-form questionnaire; S2011/S2016: Data linkage between the 2011 Census short-form questionnaire and the 2016 Census short-form questionnaire; L2011/S2016: Data linkage between the 2011 National Household Survey and the 2016 Census short-form questionnaire.

Table 3 Estimates of Indian reserves’ in-migrants, out-migrants and net migration between 2006 and 2011, total population, for Canada and the provinces, by data source

	TOTAL POPULATION					
	S2006/S2011			L2011_R		
	In	Out	Net	In	Out	Net
	thousands					
CANADA	30.0	44.0	− 14.0	28.9	17.2	11.7
ATLANTIC PROVINCES	1.3	1.4	− 0.1	1.3	0.4	0.8
QUEBEC	1.6	2.3	− 0.7	2.2	1.2	1.0
ONTARIO	2.5	6.1	− 3.5	3.9	2.2	1.7
MANITOBA	3.2	6.3	− 3.1	3.0	2.9	0.0
SASKATCHEWAN	4.2	7.3	− 3.1	4.7	1.9	2.8
ALBERTA	2.9	5.8	− 2.9	2.8	2.9	− 0.1
BRITISH COLUMBIA	14.7	15.2	− 0.5	11.4	5.9	5.5

Results exclude the population living in the territories or on an incompletely enumerated Indian reserve in 2006 or 2011. Individuals who moved from one reserve to another in the same province or region are not accounted for as in-migrants or out-migrants. Results are rounded to the nearest hundred. For this reason, net migration may not be equal to the difference between in-migrants and out-migrants.

Sources: L2011_R: retrospective information from the 2011 National Household Survey; S2006/S2011: data linkage between the 2006 Census short-form questionnaire and the 2011 Census short-form questionnaire.

The trends described above at the national level were also observed at the provincial level.³⁰ Linked files show negative net migration on Indian reserves in all provinces except British Columbia in 2011/2016, whereas the non-linked L2011 and L2016 files (using retrospective responses) almost always show net positive migration.

Figure 1 shows the estimated net migration for Indian reserves in relation to other types of regions along an urban–rural continuum: urban CMA, urban non-CMA, and rural.³¹ All estimates computed from linked files were negative, suggesting that Indian reserves lost people through internal migration to all three types of regions, for both the population as a whole and Registered Indians. The results show that Indian reserves’ largest net losses were to rural areas or urban CMAs. For

³⁰ The territories were excluded from the analysis for two main reasons: (1) the number of (and the population living on) Indian reserves is very small and limited only to the Northwest Territories, and (2) the linkage rates were very low in the territories.

³¹ In this study, urban areas were divided in two categories: (1) urban CMA and (2) urban non-CMA. Urban CMAs are CMAs, while urban non-CMAs are CAs. For a more complete definition of CMAs and CAs, please refer to the 2016 Census Dictionary (Statistics Canada, 2018). Rural areas are regions outside CMAs and CAs.

Fig. 1 Five-year net migration flows for the total population and the Registered Indian population, by data source, Canada, 2011 to 2016. Notes: CMA stands for census metropolitan area. Results exclude the population living in the territories or on an incompletely enumerated Indian reserve in 2011 or 2016. Sources: L2016_R: retrospective information from the 2016 Census; L2011_R: retrospective information from the 2011 National Household Survey; L2011/L2016: Data linkage between the 2011 National Household Survey and the 2016 Census long-form questionnaire; S2011/S2016: Data linkage between the 2011 Census short-form questionnaire and the 2016 Census short-form questionnaire

the Registered Indian population in particular, the largest net losses were to urban CMAs. In contrast, L2016_R data show positive net migration everywhere.

Figure 2 shows Indian reserves' 5-year out-migration rates by age group for the total population and the Registered Indian population in Canada in 2011. While the different data sources show similar patterns by age group,

Fig. 2 Indian reserves' 5-year out-migration rates (per thousand), 2011 to 2016, by age group and data source. **a** Total population. **b** Registered Indian population. Sources: L2016_R: Retrospective information from the 2016 Census long-form questionnaire; L2011/L2016: Data linkage between the 2011 National Household Survey and the 2016 Census long-form questionnaire; S2011/S2016: Data linkage between the 2011 Census short-form questionnaire and the 2016 Census short-form questionnaire; L2011/S2016: Data linkage between the 2011 National Household Survey and the 2016 Census short-form questionnaire

out-migration rates estimated from linked files were much higher than in the L2016 file (at all ages). Conversely, rates of in-migration on Indian reserves were much more similar in both level and age profile (Figs. 2 and 3) from one source to another.

Fig. 3 Indian reserves' 5-year in-migration rates (per thousand), 2011 to 2016, by age group and data source. **a** Total population. **b** Registered Indian population. Sources: L2016_R: Restrospective information from the 2016 Census long-form questionnaire; L2011/L2016: Data linkage between the 2011 National Household Survey and the 2016 Census long-form questionnaire; S2011/S2016: Data linkage between the 2011 Census short-form questionnaire and the 2016 Census short-form questionnaire; L2011/S2016: Data linkage between the 2011 National Household Survey and the 2016 Census short-form questionnaire

Figure 4 shows Indian reserves' net migration rates by age group, computed from all sources. Net migration estimated using linked files was negative for all age groups except for the population aged between 50 and 75. In contrast, net migration estimated using L2016_R was positive or close to 0 for all ages.

Fig. 4 Indian reserves' 5-year net migration rates (per thousand), 2011 to 2016, by age group and data source. **a** Total population. **b** Registered Indian population. Sources: L2016_R: Retrospective information from the 2016 Census long-form questionnaire; L2011/L2016: Data linkage between the 2011 National Household Survey and the 2016 Census long-form questionnaire; S2011/S2016: Data linkage between the 2011 Census long-form questionnaire and the 2016 Census short-form questionnaire; L2011/S2016: Data linkage between the 2011 National Household Survey and the 2016 Census short-form questionnaire

4.2 Comparison with Net Migration Obtained from a Residual Approach Using Multiple Censuses

As mentioned in the Introduction, the population living on Indian reserves increased by 42,500 people between 2006 and 2016 according to the retrospective information collected in each census. This can be decomposed roughly

by estimating the contribution of the various components (fertility, mortality, and migration) that influenced this population growth. First, it can be assumed that the contribution of international migration to population growth on Indian reserves is negligible. Second, the number of births between two censuses can be estimated by counting the population younger than age 5 in the last census. It was estimated that between 2006 and 2016, there were 72,800 births on Indian reserves. By applying death rates to the population at the beginning of an intercensal period, the number of deaths on Indian reserves was estimated at around 17,200 over the same period. As a result, natural growth (i.e., the difference between the number of deaths and number of births) contributed to 131% of the observed growth (+ 55,600) between 2006 and 2016.

If these figures are assumed to be accurate and net internal migration on Indian reserves is estimated as a residual (i.e., the number required to balance total growth at 42,500), then an estimate of – 13,100 would be obtained, which is higher than the sum of the estimates from S2006/S2011 and S2011/S2016 (– 32,900), but much lower than those from L2011_R and L2016_R (+ 22,300). Given that natural growth exceeds the estimated total growth over the study period, net migration on Indian reserves could not be positive.

Once again, it is important to interpret these results with caution, as there is a lot of uncertainty with regard to census estimates of population counts on Indian reserves, particularly because of net undercoverage. This uncertainty not only affects the quality of the migration estimates obtained from linked files, but also attempts to reconcile estimates of migration with those related to observed population growth, like the one done here.

5 Analysis of Discrepancies Related to Migrant Status or Previous Place of Residence Between Retrospective Responses and Linked Census Information

In this section, the S2011/L2016 file was used to compare the retrospective information about an individual's place of residence in 2011 from the L2016_R file with their place of residence in 2011, as listed in the S2011 file.³² In theory, this information should be identical, and discrepancies between the two files could explain why the two sources provide different estimates of migration flows when the place of residence is located on an Indian reserve in one file but not in the other. Table 4 shows a summary of the discrepancies found among linked individuals in the S2011/L2016 file. These discrepancies are of two main types: those related to self-declared migrant status (migrants are those who reported living in a different municipality five years earlier in L2016_R), and those related to

³² The S2011/L2016 file was weighted using a simpler approach than the one presented in Section 3. Series of average weights were computed for the four combinations of interest: (1) on reserve in 2011 and on reserve in 2016, (2) on reserve in 2011 and off reserve in 2016, (3) off reserve in 2011 and on reserve in 2016, and (4) off reserve in 2011 and off reserve in 2016. The weights correspond to the weighted population for these four combinations from the weighted S2011/S2016 file divided by the equivalent unweighted population from the S2011/L2016 file.

migrants for whom the place of residence in 2011 does not match the retrospective information captured in 2016. Table 4 also shows the effect of each type of discrepancy on estimates of global net migration on Indian reserves, assuming that the information from the census linked files is correct.

5.1 Discrepancies Related to Migrant Status

There are two cases of migrant status-related discrepancies. The first involves individuals who declared in 2016 that they were living in the same place 5 years earlier, so they were considered non-migrants in L2016, but who were enumerated at different places of residence in S2011 and L2016 (i.e., they migrated into or out of an Indian reserve). Overall, discrepancies of this nature accounted for a difference of 7600 in the net migration on Indian reserves in S2011/L2016 when migration was measured using the linked information on place of residence in 2011 and 2016, instead of the retrospective information contained in L2016_R. The second case is the opposite of the first. It involves self-declared migrants who were enumerated at the same place of residence in S2011 and L2016. Of interest in this study among these cases are the records where retrospective information showed a migration into or out of an Indian reserve. Records exhibiting this type of discrepancy contributed to a difference of 3500 individuals in the net migration on Indian reserves between the linked and retrospective sources.

What causes these differences? There are reasons to believe that the retrospective information is at fault most of the time. One reason is that it may be difficult for respondents to remember whether they were living in the same place 5 years earlier, on a specific date that bears no particular significance to them, and this is especially true for highly mobile individuals. Some very mobile individuals, such as circular migrants, may believe that they have more than one place of residence (although they should choose one usual place of residence as per census guidelines). Another reason is that responses indicating that the respondent was living in the same place 5 years earlier help alleviate response burden, since no retrospective place of residence information needs to be added.

There is also another factor at play in the case of self-declared migrants with two identical addresses in the linked files. An examination of the coding of the retrospective information on the place of residence 5 years earlier showed that there was a bias toward coding these places as “not a reserve.” This could explain why there is a larger number of inconsistent records for migrants onto a reserve (off-reserve to on-reserve) than for migrants out of a reserve (on-reserve to off-reserve). The specific topic of geocoding is addressed in Appendix 2.

5.2 Same Migrant Status, But Discrepancies in On-Reserve and Off-Reserve Residence in 2011

The second type of discrepancy involves individuals who are self-declared migrants and have two distinct places of residences in the linked file, but whose place of residence in S2011 does not match their place of residence 5 years earlier as collected in L2016. This type of discrepancy occurs in four different forms:

Table 4 Summary of the discrepancies observed in migrant status or place of residence (on or off reserve) between the retrospective source and the linked source in the S2011/L2016 file

Respondent living in different CSDs in 2011 and 2016 (migrant status)		Place of residence in 2011		Place of residence in 2016	Number of linked individuals with a different (incoherent) place of residence on or off reserve (weighted)
L2016_R ¹	S2011/L2016 ²	L2016_R	S2011		
DISCREPANCIES RELATED TO MIGRANT STATUS					
Yes	No	Same as L2016	Off reserve	On reserve	– 18,600
Yes	No	Same as L2016	On reserve	Off reserve	<u>+26,200</u>
					Total: + 7600
No	Yes	Off reserve	Same as L2016	On reserve	+ 5300
No	Yes	On reserve	Same as L2016	Off reserve	<u>–1800</u>
					Total: + 3500
DISCREPANCIES IN PLACE OF RESIDENCE IN 2011					
No	No	On reserve	Off reserve	On reserve	– 400
No	No	Off reserve	On reserve	Off reserve	+ 16,200
No	No	Off reserve	On reserve	On reserve	+ 2300
No	No	On reserve	Off reserve	Off reserve	<u>–4500</u>
					Total: + 13,600

¹Respondent was living in the same CSD 5 years prior to the census according to L2016_R.²Respondent was living in the same CSD according to S2011 and L2016.

CSD stands for census subdivision. Cases presented are only those for which there was a discrepancy in the place of residence on- or off-reserve between the retrospective response and the observed information obtained from S2011.

Source: S2011/L2016; Data linkage between the 2011 Census short-form questionnaire and the 2016 Census long-form questionnaire

1. Migration onto a reserve:
 - a. Origin is on an Indian reserve as per S2011, but off a reserve as per L2016_R (400 cases)
 - b. Origin is off an Indian reserve as per S2011, but on reserve as per L2016_R (2,300 cases)
2. Migration off reserve:
 - a. Origin is on an Indian reserve as per S2011, but off a reserve as per L2016_R (16,200 cases)
 - b. Origin is off an Indian reserve as per S2011, but on reserve as per L2016_R (4,500 cases).

Overall, these four versions of the second type of discrepancy account for a difference of – 13,600 individuals in the net migration of Indian reserves when the place of residence obtained from S2011 is used instead of the one obtained from L2016_R.

The above discrepancies may occur for many reasons, but it is likely that a large proportion of the inconsistencies originates from L2016_R because of factors such as the provision of imprecise or erroneous retrospective information, possibly caused by respondents having trouble remembering the information, or the challenges associated with coding the retrospective information to a precise location. As mentioned above, there seems to be a bias in the geocoding where more locations are wrongly coded as off reserve (if S2011 is assumed to be accurate) than on reserve. Therefore, there are more “erroneous” migrations from outside reserves than from on reserves (see Appendix 2). This is consistent with the results shown here.

6 Conclusion

For several decades, the retrospective questions on the place of residence one year and 5 years prior to the census have been the preferred source for measuring migration flows between Indian reserves and off-reserve areas in Canada. From one census to the next, the same observation has emerged: slightly more people entered Indian reserves than left. Consistency over time and the absence of data allowing for alternative estimates have led these results to be widely used and accepted, in particular by departments, policy-makers and the scientific community.

During that period, Statistics Canada was aware of limitations related to these retrospective census questions, and that these limitations have a greater impact for smaller geographic areas.³³ For this reason, the agency put forward various initiatives during the last two censuses to improve the coding and processing operations related to these variables. However, without census data linkage, it is very difficult to assess how and to what extent these limitations could affect the quality of the estimates.

³³ See—for example—the data quality note on migration data in small areas in the Mobility and Migration Reference Guide (Statistics Canada, 2017a).

The fact that two consecutive censuses are now linked made it possible—for the first time—to obtain alternative estimates of migration flows between Indian reserves and off-reserve areas without some of the limitations inherent to traditional censuses. These new estimates reveal a somewhat different portrait for the periods from 2006 to 2011 and from 2011 to 2016: there were more out-migrants from than in-migrants to Indian reserves, leading to negative overall net migration on Indian reserves. Furthermore, comparisons of data from the two sources shed light on some specific limitations associated with the collection of retrospective information about an individual's prior place of residence, which led to an underestimation of the number of people leaving Indian reserves.

However, it is important to be cautious when interpreting these results. Although the change of sign in net migration is striking, the size of outward migration flows from Indian reserve does not in any way suggest an exodus, and Indian reserves have been continuously experiencing population growth in last decades due to high birth rates. Besides, these new estimates carry a certain dose of uncertainty. There are, indeed, factors that can affect the precision of the estimates, such as the potential linkages of two records that do not represent the same individual, the exclusion of the population living in incompletely enumerated Indian reserves, the low linkage rates on Indian reserves, and the limitations in the weighting processes. The population living in institutions (which includes many Indigenous people) also created a particular challenge, as it had low linkage rates and was excluded from the analysis as a result. However, as previously mentioned, this population could not be fully excluded from the weighting process.

That said, the results of net migration on reserves obtained from census data linkages may be more consistent with the population growth observed on reserves in the past, as demonstrated by the result of the decomposition of population growth according to the various demographic components, which shows that net migration on reserves could not be positive between 2006 and 2016.

Lastly, census linkages have served in this study as an alternative data source for evaluating the quality of the retrospective information related to the place of residence. This evaluation showed some limitations of this information such as a bias in the geocoding process favoring attribution to off-reserve locations and an underestimation of individuals having migrated. While there is no indication that these issues are specific to some population groups, these results could serve as a starting point for further investigation of the quality of retrospective census information, particularly in regard to other types of geographic areas. They could also provide an opportunity to integrate new data sources, such as data linkages and administrative data, into the coding and processing of census migration data, in particular for smaller areas and Indigenous communities. This would make it possible to further improve the quality of Canadian census data.

Appendix 1 Description of Linkage Rates

Linkage rates provide useful information, but their calculation can get complicated, and depends on what the researchers want to know about the linked data. The simplest formulation is to compute how many individuals from the start-year

file population (2011) were successfully linked to the end-year file (2016), and divide by the population from the start-year file. Intuitively, this is in line with the weighting procedure, since adjustment of the weights for non-linkage is done assuming that everyone in the 2011 population should have been linked (the weights are reduced to account for losses due to deaths and emigration between 2011 and 2016 during the calibration stage). Linkage rates computed with this approach are presented for files S2011/S2016 and L2011/S2016 in Table 5.

Note that linkage rates for the L2011/L2016 and S2011/L2016 files could also be computed in theory. However, a very high proportion of individuals from the start-year file were not selected to respond to the 2016 Census long-form questionnaire (around 75% of the population living outside an Indian reserve) and therefore cannot be linked, so that the procedure results in very low linkage rates that reflect essentially the sampling scheme of the 2016 Census (long-form).

The linkage rates in Table 5 are indicative of the importance of the weighing process to correct for non-linked individuals.³⁴ However, they cannot be interpreted as indicators of the performance of the linkage procedure. This is because the calculations make no exclusions for individuals who could not be linked. For example, in Table 5, the linkage rates tend to be lower in older age groups due mainly to the fact that the proportions of individuals in the start-year file who died between 2011 and 2016 increase with age, not to a decreasing capacity to link those who could be linked. There are, in fact, multiple reasons why an individual in the start-year file could not be linked to the end-year file, and vice versa:

- An individual in the start-year file died or left the country before the 2016 Census.
- An individual in the end-year file was not born or lived outside the country in 2011.
- An individual in the start-year (or end-year) file did not respond to the 2011 (or 2016) census (non response).
- An individual in the start-year (or end-year) file was missed by the 2011 (or 2016) census (undercoverage).
- An individual is present in both the start-year and end-year files but could not be linked due to errors in the linkage keys, wrong information having been captured, or other operational reasons.

A linkage rate that reflects how successful the linkage was for those that could effectively be linked needs to account for all the above cases in the choice of the denominator. Applying such adjustments results in a linkage rate of about 90% for the entire Canadian population and 55% for the population living on an Indian reserve. Unfortunately, not all the information required for these adjustments is available at disaggregated levels (i.e., for population groups or geographical areas).

³⁴ Readers interested to know more about the weighting procedure are invited to consult Part 2 of the document entitled “Method for linking the 2011 and 2016 censuses and calculating weights for linked census files”, which is provided as [supplementary material](#) to this article.

Table 5 Estimated linkage rates on and off reserve (percent) according to selected individual characteristics in 2011, by data linkage source

	S2011/S2016			L2011/S2016		
	Total	Off reserve	On reserve	Total	Off reserve	On reserve
TOTAL—CANADA	80.0	80.4	41.8	81.2	82.8	45.9
PLACE OF RESIDENCE						
ATLANTIC PROVINCES	81.2	81.3	60.9	82.8	83.6	64.1
QUEBEC	83.1	83.3	48.7	85.0	85.6	56.9
ONTARIO	80.9	81.0	37.6	83.4	83.9	49.9
MANITOBA	78.4	81.1	32.6	74.3	84.1	36.2
SASKATCHEWAN	77.1	79.7	33.8	73.0	81.7	39.6
ALBERTA	78.7	79.3	32.7	79.4	82.4	33.6
BRITISH COLUMBIA	74.1	74.5	54.8	75.7	76.8	56.1
TERRITORIES	49.8	49.9	0.0	41.1	41.3	0.0
SEX						
FEMALE	80.8	81.2	43.5	82.0	83.4	47.7
MALE	79.1	79.6	40.1	80.4	82.1	44.0
AGE GROUP (years)						
0 TO 10	81.4	82.4	30.9	79.9	83.7	34.6
11 TO 20	77.0	77.7	33.6	77.0	79.7	37.6
21 TO 30	75.7	76.1	41.8	77.4	79.0	46.6
31 TO 40	83.3	83.7	48.8	85.0	86.3	54.5
41 TO 50	85.7	86.0	50.9	87.4	88.5	56.8
51 TO 60	85.7	86.0	53.4	87.8	88.6	59.1
61 TO 70	82.5	82.6	55.4	84.4	85.0	58.0
71 TO 80	69.9	70.0	48.4	71.4	71.9	48.4
81 TO 90	44.2	44.3	31.6	44.8	45.0	29.3
91 +	21.7	21.8	20.4	18.8	18.9	7.2
REGISTERED INDIAN STATUS						
REGISTERED	—	—	—	47.9	59.2	44.6
NOT REGISTERED	—	—	—	83.0	83.0	60.8
INDIGENOUS IDENTITY						
FIRST NATIONS	—	—	—	50.2	64.2	44.5
METIS	—	—	—	75.4	76.0	50.3
INUIT	—	—	—	49.9	49.9	49.3
OTHER INDIGENOUS	—	—	—	74.8	77.5	57.0
NON-INDIGENOUS	—	—	—	83.3	83.4	67.1

Only the population living in a private household in 2011 was considered in the calculation of these linkage rates.

Sources: S2011/S2016: Data linkage between the 2011 Census short-form questionnaire and the 2016 Census short-form questionnaire; L2011/S2016: Data linkage between the 2011 National Household Survey and the 2016 Census short-form questionnaire

Appendix 2 Analysis of Discrepancies in Migrants' Place of Residence Information

Differences between the estimates of migration flows into and out of Indian reserves obtained from the 2016 Census and from the various 2011 and 2016 census linked files can be partially explained by inaccuracies in responses to the question in the 2016 Census on self-reported place of residence 5 years prior to the census (for respondents who identified as migrants). There are several reasons why this may occur, and they are examined briefly in this appendix.

First, while the statistical definition of a municipality is clear (it is a census subdivision), it is not always clear for respondents, despite the fact that examples are provided in the questionnaire (city, town, village, township, municipality, or Indian reserve). This may explain the recurring situation in which a respondent provides the name of the closest census agglomeration (which is often an off-reserve location) instead of the name of the municipality. For example:

- Belleville instead of Tyendinaga Mohawk Territory
- Brantford instead of Six Nations (Part) 40
- Edmonton instead of Stony Plain 135
- Fredericton instead of Devon 30
- Kelowna instead of Duck Lake 7
- Port Alberni instead of Tsahaheh 1

Second, it is also difficult to accurately assign a geographic code when a respondent provides a place of residence that is not a unique name (some distinct municipalities share the same name). In the specific case of Indian reserves, many share the same name as a municipality outside the reserve, and the only difference is that it is followed by a number. If these numbers are omitted, the location will be coded incorrectly. For example:

- Berens (off-reserve area) and Berens River 13 (Indian reserve)
- Cote (off-reserve area) and Cote 64 (Indian reserve)
- Grand Rapids (off-reserve area) and Grand Rapids 33 (Indian reserve)
- Wabasca (off-reserve area) and Wabasca 166D (Indian reserve)
- Kamloops (off-reserve area) and Kamloops 1 (Indian reserve)
- Penticton (off-reserve area) and Penticton 1 (Indian reserve).

Even though respondents have been asked since 2006 to provide the postal code of their place of residence 5 years earlier (which, in theory, should help identify a precise location), this is not always helpful because, in rural areas, a postal code can cover both on-reserve and off-reserve areas. Furthermore, respondents sometimes do not report their postal code, possibly because they were unable to remember it.

Third, the quality of the place of residence coding can also be compromised by various factors. For instance, some responses provided by respondents could be incorrect, perhaps because it is difficult for the respondent to remember their place

of residence 5 years earlier. Furthermore, proxy responses could also be a major source of imprecision and errors.

Respondent-provided information is sometimes incomplete or too vague for accurate geographic coding. For example, responses can sometimes point to more than one location, in which case the location may be incorrectly coded.

Poor transcription of the place of residence can also hinder geographic coding (e.g., Machteuiatsh instead of Mashteuiatsh).

In some cases, respondents enter the name of the Indian band/First Nations instead of the municipality associated with the Indian reserve, which can lead to inaccurate geographic coding. For example:

- Atikameksheng instead of Whitefish River (Part) 4
- Nipissing First Nation instead of Nipissing 10
- Peigan instead of Pikani 147
- Norway house Cree Nations instead of Norway house 17
- Mississauga First Nation instead of Mississagi River 8.

Sometimes, the name of the Indian band/First Nations provided by the respondent corresponds exactly to a CSD name (which is often an off-reserve location), which complicates the geocoding. For example:

- Big River (Indian Band) and Big River (CSD off-reserve)
- Cumberland House (First Nations) and Cumberland House (CSD off-reserve)
- Kingsclear (First Nations) and Kingsclear (CSD off-reserve)
- Lytton (Indian Band) and Lytton (CSD off-reserve)
- Moosomin (Indian Band) and Moosomin (CSD off-reserve)
- Oromocto (First Nations) and Oromocto (CSD off-reserve)

The analysis of responses showed that a relatively large proportion of the discrepancies found occurred among individuals who reported living in a different place 5 years earlier, but who had the same place of residence in the two consecutive censuses. The majority of these cases involved individuals who were living on an Indian reserve in 2011 (according to the 2011 Census), but who retrospectively reported in the 2016 Census that they were living off reserve five years earlier (in 2011), or who were geocoded as such. This type of discrepancy would lead to an overestimation of the migration from an off-reserve area to an Indian reserve in a non-linked census.³⁵

³⁵ In some remote areas, data collection is done a few months before the official census date. As a result, the reference date for the place of residence of some respondents in the 2011 Census was not May 10, 2011, but a few months before. Therefore, it is possible that some individuals moved between the collection date and the official reference date of the census. This could impact the comparability of the place of residence obtained from the retrospective question (which stated a precise reference date, such as May 10, 2011, in the case of the 2016 Census long-form questionnaire) and the one where individuals were enumerated in the previous census. Reference date inconsistencies can cause some discrepancies between sources, but this discrepancy is assumed to be only marginal.

This analysis of the linked files also showed that imputing the place of residence 5 years prior to the census variable during census processing (when the retrospective response is missing) also has an effect, although this effect is probably only minor. In most of these cases, an off-reserve location was imputed as a response for individuals who had been enumerated on an Indian reserve in 2011. This contributed to the positive net migration on Indian reserves that was observed in the non-linked censuses.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s42650-021-00042-2>.

Acknowledgements The authors would like to thank Indigenous Services Canada, the department that funded this study. They also thank the people who reviewed preliminary versions of this document and provided comments: John Clement, Eric Guimond, Stéphanie Langlois, Laurent Martel, Erin O’Sullivan, Jean-Pierre Seguin, Jennie Thompson, and Samuel Vézina. They are also grateful to Coryse Ciceri and Stéphanie Tudorovsky for their help at different stages of this project and to Dominic Grenier who prepared the supplementary document explaining the methodology used to micro-match the 2011 and 2016 censuses.

Funding This research was funded by Indigenous Services Canada

Data Availability Not available

Code Availability (software application or custom code) Not available

Declarations

Conflicts of Interest/Competing Interests No

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

- Amorevieta-Gentil, M., Bourbeau, R. & Robitaille, N. (2015). Migration among the First Nations: Reflections of inequalities. *Population Change and Lifecourse Strategic Knowledge Cluster Discussion Paper Series*, 3(1), Article 10. Available at <https://ir.lib.uwo.ca/pclsc/vol3/iss1/10>.
- Clatworthy, S., & Norris, M. J. (2007). Aboriginal mobility and migration: Trends, recent patterns, and implications: 1971–2001. *Aboriginal Policy Research Volume IV Setting the Agenda for Change*. Available at <https://ir.lib.uwo.ca/aprci/103/>.
- Clatworthy, S., & Norris, M. J. (2014). Aboriginal mobility and migration in Canada: Trends, patterns and implications, 1971 to 2006. In F. Trovato & A. Romaniuk (Eds.), *Aboriginal populations: Social, demographic and epidemiological perspectives* (pp. 119–160). The University of Alberta Press.
- Cooke, M. (1999). *On leaving home: Return and circular migration between first nations and prairie cities*. Graduate thesis. Department of Sociology, University of Western Ontario.

- Cooke, M., & Bélanger, D. (2006). Migration theories and First Nations mobility: Towards a systems perspective. *Canadian Review of Sociology and Anthropology*, 43(2), 141–164.
- Cooke, M., & Penney, C. (2019). Indigenous migration in Canada, 2006–2011. *Canada Studies in Population*, 46(2), 121–143.
- Dion, P., & Coulombe, S., (2008). Portrait of the mobility of Canadians in 2006: Trajectories and characteristics of migrants. *Report on the Demographic Situation in Canada, 2005 and 2006*. Statistics Canada (cat. no. 91–209-X).
- Falconer, P. (1985). *Urban Indian needs: Federal policy responsibility and options in the context of the talks on Aboriginal self government*. Unpublished discussion paper. Institute of Urban Studies.
- Morency, J.-D., Caron-Malenfant, É., & Daignault, D. (2018). Fertility of Aboriginal people in Canada: An overview of trends at the turn of the 21st century. *Aboriginal Policy Studies*, 7(1), 34–61.
- Norris, M. J., & Clatworthy, S. (2011). Urbanization and migration patterns of Aboriginal populations in Canada: A half century in review (1951 to 2006). *Aboriginal Policy Studies*, 1(1), 13–77.
- Norris, M. J., Beavon, D., Guimond, E., & Cooke, M. (2004). *Registered Indian mobility and migration: An analysis of 1996 census data*. Strategic Research and Analysis Directorate, Indian and Northern Affairs Canada, Catalogue no. R2–167/1996E.
- Siggner, A. J. (1977). *Preliminary results from a study of 1966–1971 Migration patterns among status Indians in Canada*. Department of Indian Affairs and Northern Development.
- Statistics Canada. (2017a). *Mobility and migration reference guide, Census of Population, 2016*. Statistics Canada (cat no. 98–500-X).
- Statistics Canada. (2017b). Chapter 7 – Field operations. In *Guide to the Census of Population, 2016*. Statistics Canada (cat no. 98–304-X).
- Statistics Canada. (2018). *Dictionary, Census of Population, 2016*. Statistics Canada (catalogue no. 98–301-X).
- Statistics Canada. (2019). *Aboriginal Peoples Technical Report, Census of Population, 2016*. Statistics Canada (cat no. 98–307-X).
- Statistics Canada. (2020). *Annual Demographic Estimates: Subprovincial Areas*. Statistics Canada (cat no. 91–214-X).
- Trovato, F., Romaniuc, A., Addai, I. (1994). *On- and Off-reserve migration of Aboriginal peoples in Canada: A review of the literature*. Indian and Northern Affairs Canada.
- Wright, A., Thurber, K. A., Yap, M., et al. (2020). Who responds? An examination of response rates to a national postal survey of Aboriginal and Torres Strait Islander adults, 2018–2019. *BMC Medical Research Methodology*, 20, 149. <https://doi.org/10.1186/s12874-020-00970-8>.

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