IUSSP Working Paper

Prepared for the *Seminar on Registering and Counting the Population: The production and exploration of census information from early modern times to the end of 20th century*, organized by the IUSSP Scientific Panel on Historical Demography and the Department of Demography, NEPO-UNICAMP, in Campinas, Sao Paulo, Brazil, 19-21 May 2016.

Comparing information from vital events to the 1891 census data in the Saguenay region of Quebec: a critical appraisal of the two sources

Jean-Sébastien Bournival
Centre interuniversitaire d'études québécoises (CIEQ), Université Laval

Hélène Vézina
Projet BALSAC - Université du Québec à Chicoutimi

Marc St-Hilaire
Centre interuniversitaire d'études québécoises (CIEQ), Université Laval

Date of last revision: (May 2016)

Abstract

The recent linkage of Canadian censuses from 1852 to 1911 to the vital records found in the BALSAC database in the context of the creation of the Integrated Infrastructure of the Quebec Population Historical Microdata (IMPQ) offers a rare opportunity to evaluate the reliability of the two types of data for research. In fact, few studies have focused on this methodological aspect. Our goal is to assess the quality and completeness of the two documentary sources by comparing one to another. Our approach will enable us: 1) to comprehensively assess the reliability of the two sources; 2) to provide a portrait of unlinked people and their household/family to better understand who they are and why we could not link them; 3) to carry a critical evaluation of the linkage program used at BALSAC. In this presentation, we will first briefly review the steps of the matching process and expose a method used to optimize individual success rates. Then, we will present our approach to confront the two sets of data in order to identify missing individuals in both sources. Lastly, we will highlight opportunities offered by the linked datasets for research on demographic behavior and more especially on spatial mobility.

Introduction

The use of historical microdata has boomed since the past 20 years. Whether it is in demography, in geography, in history or in the social sciences in general, the growing interest for the analysis of individual or family behavior at the micro level is observed and supported by a sustained intellectual production. At the same time, datasets on various populations are flourishing everywhere and form a spatiotemporal mosaic to which are regularly added new perspectives.

Several sources such as assessment rolls, land registry, medical records, and censuses can be used in conjunction to improve the information on individuals, families or cohorts. For example, the censuses allow to follow individuals and families on a more or less long period and to extract indices or conclusions on trends in terms of migration, fertility, nuptiality, social mobility, etc. In the same spirit, family reconstitution based on civil (or parish) registers opens the way to the observation of life courses, of intergenerational variations of various demographic outcomes and of genealogical and kinship links. Computational possibilities at our disposal make it increasingly easy to link these sources together and to add depth to analyses.

Research is however confronted with the inherent limitations of each set of data. Demographic analyses sometimes suffer of a lack of depth or representativeness due to incomplete sources. Moreover, the assessment of the accuracy and completeness of microdata can be complicated by the absence of information on the objectives and rules used to parameterize the data collection. Sometimes assimilated to each other, these two features lead us to observe the data from different angles. The accuracy, or quality, refers here to the content of the registers in relation to the requirements established by the responsible authorities. It answers the question whether the personal information which should appear in the historical source really does. In terms of completeness, which is the object of our concern, we rather seek to investigate problems directly related to the registration process measured in terms of presence or absence of individuals who should have been recorded.

Historically, some sources are known to have been better kept than others. This is particularly true of civil registers. The legal aspect of the registration of vital events (births,

marriages and deaths) by the religious and civilian authorities makes it one of the most rigorous sources (Vézina, St-Hilaire, & Bellavance, 2014). The period and the area of observation are also factors to consider in the characterization of the sources. As a general rule, the most recent registers tend to be of better quality than the oldest (Willigan & Lynch, 1982). For censuses, the quality of the information is dependent on the objectives underlying the enumeration of the population (Curtis, 2002) as well as on the constancy of enumerators.

But beyond the intrinsic quality of each source (the conditions of their registration), the best way to measure and to assess the completeness of a source is to compare it to another source also aimed at collecting exhaustive data on the population. Few studies have focused on this methodological aspect, most likely because of the lack of means (time, money, and technology) or for reasons pertaining to the datasets. Indeed, the concurring sources may be rare or non-existent, and their relative quality, or even their scope, may constitute obstacles to the achievement of this objective. The linkage is also a heavy and costly process which may restrict the initiatives in this direction. Also, depending on the research objectives, the completeness of the sources is not always fundamental to undertake relevant analyses on a given population. But for those who are interested in these issues, the linkage of two different sources is a way to certify the reliability of both sources and allows as a corollary to broaden the field of possible analyses.

In this sense, the recent linkage of nominative censuses from 1852 to 1911 in the context of the creation of the Integrated Infrastructure of the Quebec Population Historical Microdata (IMPQ) (Vézina, St-Hilaire, & Bournival, 2015) to the civil registers found in the BALSAC database (BALSAC 2015) offers a great opportunity to perform a systematic comparison between the two datasets and allows for a thorough evaluation of the completeness of the sources. Registering and counting the population are clearly motivated by different objectives though they both aim to be exhaustive. Very few studies have proceeded to a comparison of the two sources.

In this presentation, we compare the 1891 census for the Chicoutimi census division (eastern half of the Saguenay-Lac-Saint-Jean region) located in Quebec, Canada with the civil registers. Our goal is to assess the quality and completeness of the two documentary sources by comparing one to another. Our approach will enable us: 1) to comprehensively

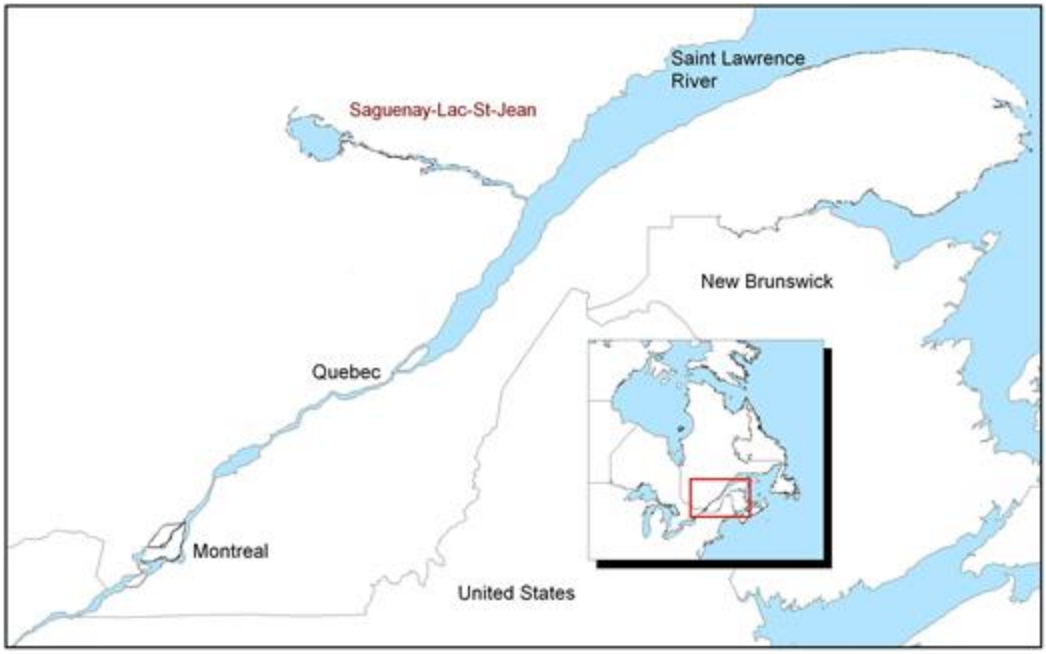
assess the reliability of the two sources; 2) to provide a portrait of unlinked people and their household/family to better understand who they are and why we could not link them; 3) to carry a critical evaluation of the linkage program used at BALSAC. In this presentation, we will first briefly review the steps of the matching process and expose a method used to optimize individual success rates. Then, we will present our approach to confront the two sets of data in order to identify missing individuals in both sources. Lastly, we will highlight opportunities offered by the linked datasets for research on demographic behavior and more especially on spatial mobility.

Data

The Saguenay region

Located 200 km north of Quebec City the Saguenay-Lac-Saint-Jean (SLSJ thereafter) region is a natural enclave along the Saguenay River and its reservoir, bordered by the Laurentians to the South and by unsettled territories to the North (Fig. 1).

Fig. 1. Location of the Saguenay-Lac-Saint-Jean region (Quebec, Canada).



French Canadian settlers arrived in the 1830s, but the first vital event officially recorded was a marriage celebrated in 1842. At the beginning the area consisted almost exclusively of farmers who came mostly from the nearby area of Charlevoix (Bouchard & Larouche, 1988). By 1852 the regional population was around 6,000 individuals and reached 37,000 50 years later. Most of the population remained rural up to the 1930s. Until the end of the 19th century, the majority of farmers alternately worked on the family farm during the summer and logged the forests during the winter to maintain incomes flow. In 1901, Chicoutimi barely exceeded the threshold population of a city set to 5,000 inhabitants but managed to evolve as an urban pole as it benefited from diverse regional functions (administration, hospital, courthouse, etc.). It is not before the Second World War that the population becomes predominantly urban.

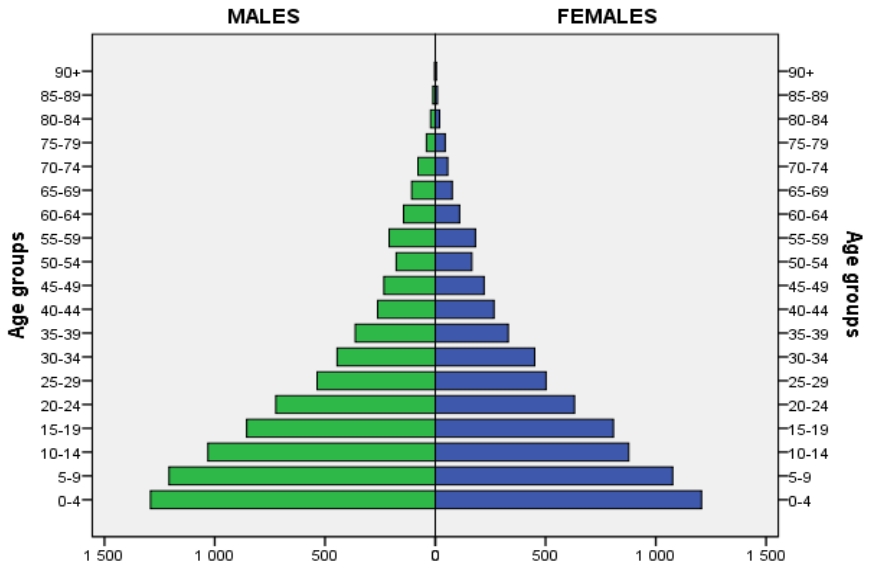
In terms of demographics, the region is characterized by a 99.7% French-Canadian population whose fertility levels as high as 9.79 children in complete families in the second half of the 19th century remained high until the 1960's (Roy, Bouchard, & Declos, 1988). Mortality levels were generally lower than the rest of Quebec for the same period as the age structure, the characteristics of immigrants (selection effect) and the rural sanitary conditions acted as favorable factors (Pouyez et al., 1983). Net migration rates were positive up until 1871 in the surge of settlement, but the region suffered from emigration from 1880 to 1891 as part of a general movement where families migrated to United States or great cities like Montreal in search for better opportunities (Gauvreau & Bourque, 1988). After that, net migration remained almost null until 1961 (Pouyez et al., 1983).

Sources of data

As mentioned above, our project aims at comparing two sources of data. First, we rely on vital events from the BALSAC database maintained at the Universite du Quebec a Chicoutimi. Created more than 40 years ago, the database contains all the Catholic marriages since the early days of French settlement in the 17th century to 1965. Births and deaths are available for the SLSJ region only but ongoing work aiming at the creation of the Integrated Infrastructure of the Quebec Population Historical Microdata (IMPQ) will make available births and deaths records for the whole province of Quebec up until 1850. Currently, BALSAC contains 2.9M records which were linked according to the family

reconstitution method inspired by Louis Henry (Fleury & Henry, 1956) thus providing information on 2.6M families and 5.3M individuals. For this study, we work with the 645,000 events recorded between 1842 and 1971 in the SLSJ region relating to 125,000 families and 500,000 individuals.

Fig. 2. Age structure in SLSJ, Quebec, Canada, 1891



The second set of data comes from the 1891 Canadian historical census. We use full count data (100%) for the Chicoutimi census district where 2300 households and 14,776 individuals were enumerated that year. The distribution by age groups (Fig. 2) portrays a young population where people aged less than 25 represent more than 61% of the total. Our choice of census year was based on practical considerations. Among the Saguenay censuses (1852-1911) linked to the BALSAC database, we chose the one which had the highest linkage rate, and for which the number of unlinked individuals was proportionally the lowest in order to alleviate the processing of data (see Table 1 below). Also, for 1891, the relationship to the household head was reported, an information which greatly assists the linkage process.

Before undertaking the comparison of the sources, we must perform a qualitative assessment of the data in order to verify if the characteristics related to the quality of entries could possibly restrain our objective. It must however be emphasized from the outset, that

we expect no big surprises in this attempt to prove their quality, at least for the civil registers.

Assessing the quality of Quebec civil registration

The precision and rigor related to the keeping of the registers are based on at least three factors: the existence and application of rules on the registration of events, the level of competence of the registrars and the historical and geographical circumstances which may have had a negative impact on the number of recordings (Willigan & Lynch, 1982, p. 61). It is now well known that the Quebec province has the advantage of an exceptional and remarkably well-preserved documentary resource with the wealth of birth, marriage and death certificates recorded by the Catholic Church. As of 1679, the regulation on the registration of vital events carried a double legislation, one clerical and the other civilian. A single registrar, the parish priest, was mandated to collect the information with the state acknowledging the official status of events recorded by the officiating priests.

Only a handful of scholars have looked at the regulation of the content of acts (Bouchard & LaRose, 1976), while others, more practical, have verified the quality of the information found in the registers. Using a population sample from Nouvelle-France in 1681, René Jetté (1980) attempted to reconstruct the census population from the registers reaching interesting results. The complex methodology used to reconstitute the population deemed present makes it possible to certify, even indirectly, the reliability of the records, at least at an aggregated level. With more recent data, Bouchard and Bergeron (1975) have identified for Laterrière, a village in the Saguenay region the problems pertaining to the registration of births, marriages and deaths for the period 1855-1911. Despite some shortcomings, results show interesting and reassuring findings. In terms of content, the rigor with which parish priests executed their duty is almost flawless as only the declaration of the occupation at marriage, especially for women for whom it is virtually non-existent, can be regarded as incomplete. According to standard procedures, information on most variables is found in more than 97% of records. With respect to occupation, the least consistently registered variable, its declaration varies around 50% (Bourque, Markowski, & Roy, 1984).

Regarding the coverage of civil registers, we know from the judicial district administrations responsible for their conservation that none is missing for the SLSJ region (Bourque et al.,

1984). However, nothing indicates that the scribes have recorded all of the events that occurred on the territory nor that the information contained in the records is comprehensive. The existence of "loose sheets" has been reported as early as the 17th century in Nouvelle-France (Charbonneau, 1975). Moreover, under-registration of births remains possible. Priests sometimes had to travel long distances in order to fulfill their task of registrar, especially on the pioneering fringe where small settlements are served by an established parish from around. These trips sometimes occurred after the death of a child born a few days or even a few weeks earlier. By recording only deaths, the parish priests went against the civil prescription to record births and may have caused a surplus of deaths over births. Nevertheless, despite the frontier nature of the region until the turn of the 20th century, no trace of such a practice was reported for Laterrière and its missions between 1855 and 1911 (Bouchard & Bergeron, 1975).

In the course of various studies, the vital events contained in the BALSAC database have been scrutinized in many ways. Internal validity tests like the ratio of births on marriages, sex ratio and the geographical variations across administrative units have all been examined at one time or another. Sex ratios of ten parishes during the settlement period (1838-1902) show high values that are explained by an excess of young men aged between 10 and 19 years. The impossibility of establishing sons on the surrounding lands played an important role in the migration of families from the nearby Charlevoix region to the new terroirs in Saguenay (Bouchard & Larouche, 1988; Gauvreau & Bourque, 1988). Also, Raymond Roy concluded that the potential record undercounts were all explained by changes in boundaries across time (cited in Bouchard, Roy, & Casgrain, 1985). The absence of births records for infants whose deaths are reported in the 1891 census could have been a way to highlight potentially missing records. Unfortunately, this information was not recorded on that year.

In a whole, from a qualitative point of view, civil registers in SLSJ appear to be very reliable as for the coverage and as for the conformity of the information registered.

Assessing the quality of Canadian censuses (1851-1911)

To fully assess the 1891 census, it is worthy to look at the modern censuses series computerized and available for research: the 1851-1911 censuses. Censuses are not oriented

toward the same purposes and they do not respond to legal objectives as civil registration. The first modern decennial census of Canada occurred in 1851, only a few years after the opening of the SLSJ region to settlement. First and foremost a tool for political representation and redistribution among Canadian provinces, the census also allows to follow the main socio-economic trends at various scales and to guide state policies. Some scholars have however argued that policies were not always supported by noble intentions, making the census a political tool of more or less relevance for social or economic objectives. Bruce Curtis revealed that the political ambitions and objectives behind the 1851 and 1861 censuses included the assimilation of the French Canadian population. According to him, the first census of real scientific value was the first Dominion of Canada's census which took place in 1871, four years after the Canadian Confederation (Curtis, 2002). Despite the relative independence gained from the British Crown after the creation of the Dominion, the imperial imprint over the census taking lasted for a few decades, the 1891 census being the last one undertaken simultaneously with the rest of the British Empire (Dunae, 1998).

The content of censuses varied greatly between 1851 and 1911. The 1891 census was the last of a series where the questions remained mostly constant. However, it included for the first time the relationship to head of the household as well as the place of birth of parents, variables which, combined with the names, forenames, marital status, and age, are extremely relevant to trace individuals or families in the vital events registers. From 1901 onward, more specific questions on origin, year of immigration, native language, color (race) and nationality were added.

Assessing the quality of the censuses regarding their level of accuracy has to be done census by census. In a whole, their quality may show from the manuscripts themselves, the documentation surrounding their taking is clearly lacking (Inwood & Reid, 1995; Ruppert, 2015). In a rare article specifically on the taking of the 1891 census, Patrick A. Dunae evokes the under-coverage of the city of Victoria, British Columbia. Displeased with its position in the face of its rival, Vancouver, and with the under-representation of its "white" population, the City Council of Victoria ordered not one but two civic censuses to correct the figures issued by the Canadian authorities. Such political agendas are not the only factor that dictates or orientates census taking. The Victoria case is one of the main issues

encountered for this census, but "the tyranny of terrain and the challenge of distance, regionalism and sectional rivalry, alienation from Eastern Canada, anti-Asian feeling, and ambivalent attitudes toward Aboriginal peoples (1998, p. 225)" were also challenging the success of the operation. Although specific to this particular census, these issues clearly highlight the different challenges and pitfalls that sometimes arise during such an undertaking. Despite all the good will of enumerators and commissioners, census taking is inseparable from the context in which it is taken and external factors that exerted pressure at different degrees had an influence on the accuracy of the enumeration. The stronger these influences the more the census is a representation of the reality rather than the reality itself. Assuming this, we support Curtis when he says that the censuses are not "taken" but rather "made" (Curtis, 2002; Gaffield, 2007), whether the influences follow a top-down or a bottom-up kind of approach. Dunae (1998) nevertheless concluded that the degree of accuracy of the 1891 census is "reasonable". Although valuable, this information is not sufficient to draw conclusions on the SLSJ population as the comparative basis between the two provinces is weak. British Columbia is the westernmost province of Canada located almost 5000 km from Quebec. Thus, the conflicts that have marked British Columbian history around 1891 have but little echoes in the SLSJ's reality[1].

A concern that could have impacted on the regional population relates to the criteria for inclusion of the population enumerated in the census. The census identifies the *de jure* population, i.e. individuals deemed residents on the territory the day the census is taken. This includes persons present during the passage of the enumerator as well as the individuals temporarily absent. In 1891, the rule on absence is redefined and allows the inclusion of workers or students who have temporarily left for their activities. Individuals whose absence extends beyond a year are excluded (Pouyez et al., 1983, p. 25). However, significant discrepancies from the rules have been observed. Numbers from the 1891 census aroused the suspicion of a Canadian politician, Sir Richard Cartwright, who audited twenty districts of Quebec (on 45) and soon identified a problem of over-enumeration totalling 40,000 individuals[2]. This was explained by the inclusion of individuals and families who had migrated several years before. Although partial, these counts suggested inflated numbers for the whole province of Quebec (Cartwright, 1912, pp. 325–326 et pp. 402-405; Pouyez et al., 1983, pp. 28–29). Brought to the attention of the government at the

turn of the century, the situation remained unheeded (Cartwright, 1912, p. 325). Confusion between *de jure* and *de facto* populations probably aroused in the flexible application of the rules of inclusion for families who left to work outside but always maintained links or dwellings in Quebec and came back every year. From a political viewpoint, perhaps it was for the sake of political representation that the number of French Canadians had thus been inflated. Regardless, the districts being not named, it remains to identify and measure, where applicable, the impact of this on the Saguenay census data.

On a statistical level, Lavoie and Pouyez (1983, pp. 32–33) have measured the gap between deaths declared in the census and the death records in civil registers between 1890 and 1891 in Saguenay. The gross rates calculated highlight the under-estimation of the mortality in the censuses by a proportion close to 30% for 1891 only.

Overall, the little we know about the 1891 census suggests inverse trends: on one hand, Quebec's population may have been overestimated; on the other hand, undercount of deaths in the census casts doubts on the diligence with which the data was recorded. It is difficult at this stage to draw conclusions about the impact of these issues on the census. In the end no source is flawless. We could have chosen another census for the comparison, and therefore another observation period and the same pitfalls would have marked our course. Despite some discrepancies with the reality, nothing seems to indicate that the sources are incomplete as no missing documents were reported. We consider for the purpose of this analysis that 1) civil registration is complete and that all vital events were recorded and 2) all living individuals residing in the census districts were enumerated (some more than once). This encourages us to confront the two sources.

Linkage methods

The linkage program

We have already provided elsewhere a detailed description of the program developed for linkage between civil records and censuses (Vézina et al., 2015). We summarize here the different steps carried out and present the additional work done to optimize the linkage results.

The linkage process is divided into two major steps, or modules, independent and not synchronous. First, the linkage between censuses and civil records, and second the linkage across censuses. Each module is ruled by the same principle: the comparison of the nominative elements of census household heads. The modules use comparison and decision algorithms to generate scores that classify potential candidates (either from BALSAC or from the following census depending on the module). In order to be selected in the process, a household must contain at least two members forming a nuclear-type familial unit (husband and wife or widow-er with a child) since a minimum of three nominative elements is needed for the linkage. Therefore, lone-headed households, institutions, working places such as lumber camps as well as parentless sibships are excluded from the process since they do not fit the selection criterion. These households as well as the individuals in it are classified as unprocessed. We will get back to them later on.

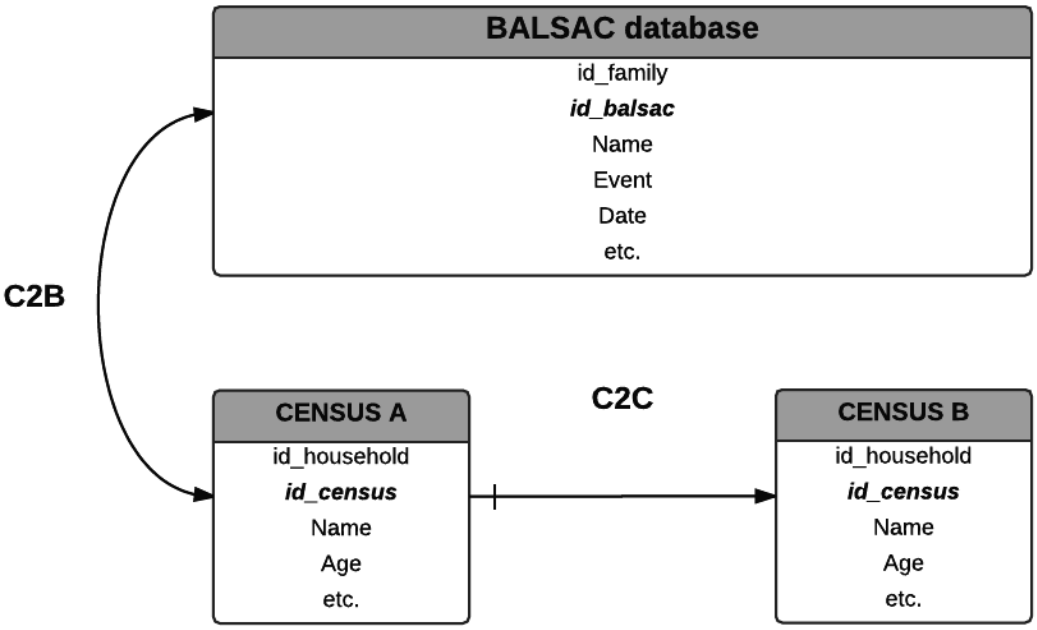
For the linkage of census records to vital events (hereafter called "C2B" for Census to BALSAC), the method is essentially borrowed from the family reconstitution system developed at BALSAC. Husband's and wife's names and surnames constitute the basic unit of information for linkage, to which children's first names are added in the comparison routines. But since the linkage process relies on highly variable data (i.e. names), dates and ages are also taken into account as criteria to select candidate families, reflecting the stages of the family cycle. In the BALSAC database, each individual has a unique number, but can be part of multiple families across time. In fact, every individual is first a child in a nuclear family ant then, at marriage, starts his own family. Places and time are therefore of high importance to link an individual coherently with his life course.

Linkage between two censuses (hereafter called "C2C" for Census to Census) is done after the two censuses have been linked to BALSAC. The same "heads" are used for comparison of households and here, too, nominative and contextual data is used to assign score for potential candidates. Individuals inside a household, apart from headship, are linked or not according to their presence in the subsequent census. For instance, children who marry and form their own household between two censuses will obviously not be found in their parents' household in the second census.

At this point what is important to keep in mind is that both modules generate a distinct type of link between individuals. Each individual has a unique number in a given census

(*id_census*) and each time an individual is linked to another census it creates a "sequence", which means that this particular individual can be followed in more than one census. This sequence is extended by each type of link (two consecutive C2B links or a C2C link between two censuses), or a combination of both (as represented in Figure 3). Figure 3 below shows one possible case of linkage between two censuses that uses the two types of link.

Fig. 3. Schematic view of the sequence formed by linkage modules



This schematic view of the result (as opposed to the process) will greatly help the comprehension of how we can optimize linkage by the process itself. In our example, the C2B link is represented by an upward arrow from the census A toward the family file contained in the BALSAC database. It is mainly a "biographical link" as individuals are matched with themselves (*id_B*) in their personal history told by the chronological series of events in civil registers. They are simultaneously linked to the family (*id_fam*) they belong to at the time of the census. If the same individual is linked to BALSAC in two consecutive censuses, it creates, by default, a sequence of *id_census* that are related to the same BALSAC individual. A C2C link is created when the comparison between two households between consecutive censuses allows keeping track of individuals within a similar structure and composition of household. With this "structural link", it is possible to take into account

the elements of the household and to better identify the individuals in it. This link is represented by the horizontal arrow. Here, the C2C link indicates that our individual in census B is the same as the one that was linked to BALSAC in census A. In our example it appears that the same individual was not linked to BALSAC in census B. There are multiple reasons that could explain this kind of situation, with age and name variations being the most common. With the two households side by side, it is possible to assess whether some declared information seems wrong or if the match between households is incorrect. Our missing C2B link in census B is therefore logical (age/name were not matching with family file) but the C2C confirmed that it was the same individual. Other combinations of links are possible as well as the absence of link when individuals do not match at all.

This is, in a nutshell, how to understand the processes of linkage and the nature of results it generates. For the 1891 Chicoutimi census, results for households and individuals after the two linkage steps are shown in Table 1. Since our aim is to analyze and improve individual linkage, household results are presented here only to illustrate the potential of the linkage program.

Table 1. Linkage status of 1891 census households and individuals to civil registers using the linkage program

	Households	**%**	**Individuals**	**%**
Linked to BALSAC (C2B)	2205	95.9	13,854	93.8
Unlinked	42	1.8	822	5.5
Unprocessed	53	2.3	100	0.7
Total	**2300**	**100**	**14776**	**100**

On one hand, we can say that the linkage method is very effective. The high linkage rates suggest that almost all families (couple and children) in the census were found in the BALSAC database and that almost all individuals had at least one record in the registers.

The presence of births and deaths records for the SLSJ region had a real impact on the likeliness to be linked (Vézina et al., 2015). Census children, and to some extent singles, could definitely more easily find a connection with the registers. On the other hand, these results do not necessarily mean that every individual who could be linked actually was. By definition, no attempt was made to link unprocessed individuals as they were part of households headed by non-nuclear-type families. Furthermore, at this stage unlinked individuals should only be considered as missed links and not as missing individuals. Children and unmarried people born outside SLSJ, couples married outside Quebec, or non-Catholic couples are expected cases of missed links as they concern events and people not contained in the BALSAC database. But this group also contains individuals that were not linked to civil registers and not to the previous or next census but could have been born, married or dead in the region. Servants and lodgers are the most common examples of people moving through different households between censuses but we lack information on which family they belong to and we are therefore unable to link them. Birth place retrieved from the census is of no help here since the geographic scale is at the level of the province for the 1891 census.

Manual optimization of linkage results

We decided to investigate a little bit further using a simple methodology to see what kind of improvement we could get. In order to compare census with civil records we must ensure to maximize C2B links, otherwise missed links will be interpreted as missing individuals and biased results.

We found two ways to improve linkage results: one based on the sequence and one that relies on the individual and familial persistence between censuses. To really put it in context, we change “census A” for “1881” and “census B” for “1891”. First, we use the sequence to highlight how we could benefit from the linkages of other census years. Our example (Fig. 3) illustrates a quite common case of sub-optimality that suggests an underestimation of links to civil records in 1891. The same individual was identified in two consecutive censuses by a C2C link. The linkage to BALSAC in 1891 is not optimal as compared to 1881 because if it is the same individual it should have been linked. In this scenario we must manually verify if the association is correct or not[3]. There are only three

possibilities: the C2C link is good/wrong, the C2B link is erroneous. Wrong links must be undone to restore coherence in the sequence of individuals. If the links are all good, we can deduce the *id_B* of our individual in 1891 from his *id_B* in 1881 since we acknowledged the C2C link between them. The inverse situation of our schematic view is also of great interest for us in this presentation: no C2B in 1891 but one in 1901, linked by a valid C2C. This is precisely what we did.

We verified 96 cases corresponding to these situations. No more than 3 of them were in fact erroneous links. In fact, we extended the method to see if we could recover more links by searching in more distant censuses. Only 5 links were found. This might seem surprising but the explanation lies in the high linkage rates. Once individuals are found, they are most of the time linked directly to BALSAC with the first module. Individuals linked over 2 or more consecutive censuses by C2C links only are very rare, especially for the SLSJ region. Probability of being linked is slightly different for places outside SLSJ because of the absence of birth and death records in the database (Vézina et al., 2015). It also means that the comparison between pair of consecutive censuses rather than comparing each census with all the others is more efficient.

The sequencing of individuals across censuses is an important step that allows the cross-validation of links between the two modules (C2B and C2C). It highlights the limits of the linkage process itself and allows the measurement of its own validity. Each census can also take advantage of links made in the others in order to maximize connections between the two sources.

As a second strategy to improve linkage results, we verified to what extent we could recover links from the remaining unlinked and unprocessed individuals by trying to find them in the previous and the next census. The rationale we estimated that the persistence of individuals and families, i.e. the percentage of people enumerated in at least two consecutive censuses, for this particular geographical area between 1881 and 1891 is around 60%[4]. At this point that means that an unmeasured part of the unlinked individuals in 1891 could have been linked to BALSAC in the previous census.

The linkage process is designed mostly to match families, i.e. individuals within close kinship ties who gravitate around the same common core. Unlinked individuals are

sometimes members of the extended family such as cousins, or simply unrelated individuals like maids or lodgers in households that share no evident ties, if any, with their family of origin (parents) or their conjugal family. They are harder to match due to this distance with the core members of the household. Relationship to head is often of great help, especially for close or extended family members, but not always. In fact, among the 822 unlinked individuals, 368 (44.7%) are said to be related to the head as son or daughter. We also wanted to fill the gap for those who remained in the unprocessed category. It is not uncommon to find a widow as the head of a lone-headed household or a married man in a lumber camp or in a small shop. The selection criterion rejected these households for their non-conformity that particular year, but it is reasonable to think that the same individuals could have appeared in previous censuses in a nuclear-type household as a child or as a married/widowed person. For these individuals, the sequence of appearances shows an interruption in 1891 caused by their exclusion from the linkage process. We need to take these situations into account in order to maximize linkage results.

But looking manually for these individuals on a nominative basis in the BALSAC database is an almost impossible task: it would be a massive time-consuming process not only to search for candidates but to sort the huge amount of them due principally to homonymy. Also, since outside Saguenay the database only contains marriages, we would have to focus mainly on married or widowed individuals, which is not optimal as only 12% of unlinked and unprocessed individuals are married or widowed. An easier way to find them is to search in another census. This is based on the simple assumption that people enumerated in another family's household could have appeared in their own household (family or conjugal) at one time or another if they had not left the area. But also, we can locate the more mobile individuals, those who never appear in a household sheltering their family of origin or the one they might have founded.

To perform this manual linkage operation, we used the same principle underlying the linkage program, namely the nominative information comparison, for lone individuals instead of couples. All individuals (unlinked and unprocessed) were compared to linked individuals in the 1881 and 1901 censuses. According to Table 1, it represents 922 individuals that were compared to the 12,762 and 14,965 linked ones for 1881 and 1901 respectively[5]. A score was assigned to the first and last names according to their degree of

similarity. Age difference between censuses was also calculated to sort results and set a tolerance interval between 8 and 12 years[6]. Matches were done manually to attest the plausibility that the two individuals would be the same.

What this method does is to find almost perfect matches on names for people who were enumerated in others' households. Although only three household duplicates were already found in the linkage process (before our attempt to maximize the results), comparing names highlighted 36 single individuals that had remained unaccounted for until then[7]. It does not take into account name substitutions, the number of cases being obviously difficult to estimate since we could not find them. However, homonymy remains the main issue and certainly prevents us from matching all potential individuals. For example, in 1891 the last name "Tremblay" is carried by 13% of all individuals, and the 17 most frequent last names cover half of the census population. Common names like "Joseph Tremblay" and "Marie Tremblay" each represents 2% of all men and women the same year, an unexceptional proportion compared to other censuses. Out of their familial context, these individuals are nearly impossible to link.

Table 2. Linkage status of 1891 census individuals to civil records before and after optimization

	Before	**%**	**After**	**%**
Linked to BALSAC (C2B)	13,854	93.8	14,081	95.3
Unlinked	822	5.5	630	4.3
Unprocessed	100	0.7	65	0.4
Total	**14776**	**100**	**14776**	**100**

Table 2 shows the results after manual optimization of linkage with civil records. Since the population was already linked to civil records in a very high proportion, our gain here is relatively modest compared to the total. But if we consider only the unlinked and unprocessed individuals (without the 36 duplicates) the improvement is noticeable. The last method added around 130 C2B links for a total of 227. The final portrait shows that 1 out

of 4 unlinked and unprocessed individuals have been added to the sequence of linked individuals. These results take also into account some minor corrections that occurred during the comparison process. For instance, we found a few links that were erroneous and in a few cases matching an individual led us to link a kin living with him that had not been found by the previous methods. Altogether, correcting links or adding new ones improved the global coherence of the sequence by bridging some gaps in census biographies.

As we took care along the process not to create erroneous links, the final picture is still probably a suboptimal result. At this point, though, only the comparison between the two sources can improve our overall understanding and help provide a general explanation for the 695 remaining unlinked individuals.

Comparing sources

Recovering families in the BALSAC database

A thorough comparison of sources would require that we attempt to reconstruct the entire population of Saguenay in 1891 from the civil records. As we mentioned earlier, the reconstruction of a population from the civil records is a colossal task that has already been performed by René Jetté (1980) on a sample of the Quebec population of the late 17th century. In his work, individuals were classified according to the level of certainty they actually inhabited the place observed at the census date by retrieving information on place of residence, whether they were present or not at the event, and whether they signed or not the record. In addition to the time required to reconstitute this population from thousands of records, the main problem encountered is related to the nature of the source. Within a family, births, marriages and deaths are recorded when they occur and although they are constrained by biological factors they do not follow a strictly defined calendar. The family life cycle has indeed some influence on the frequency of events as well as the intervals between each of them. Long intervals always raise the question of the actual presence in the census while the open intervals (censoring) do not allow choosing between presence, migration or death (in the absence of a death record).

These issues are important enough to bring us to choose for another comparison strategy. To avoid the pitfalls related to censoring and to very long intervals, we used a fraction of the population as a reference sample by selecting families with events recorded before and after a certain time interval around the time of census. We will then extrapolate general conclusions on the quality of the records from the results of this sample.

Unlike the previous section where we have tried to maximize individual results, we use the family as a unit of analysis. As the recorded events are compiled by family, it is easier to follow the fate of the family unit than a particular individual because the frequency of events is much higher. In a family file, the interval between the birth of a child and his marriage is often greater than 20 years, while between the birth and death of an unmarried child it may take more than 80 years. If we were to consider individuals, we should deal with these multiple scenarios scattered in time, as René Jetté did, to recompose if only a small percentage of the total population. The family represents then a more effective proxy for assessing the presence of the individuals enumerated in the census. Thus, the selection of families in BALSAC is based on the assumption that a family that declared a place of residence in Saguenay before and after the census is likely to be present during the census enumeration. We were inclusive insofar as we have not limited our selection to the residence of a single family member (e.g. spouse). By considering all residences reported in a record we want instead to highlight cases where the spouses/parents declared different places. A longer interval before and after the census will yield more families to investigate, but the risk to include individuals and families that were in fact absent on the census day will increase.

We tested various intervals to provide an overview of the progression of results. The official date of the census, April 6, 1891, is the center of the interval and we include in our selection families for which the only recorded event occurs precisely on that date. Table 3 shows the results of our investigation for different intervals. For example, for a one-year interval (6 months before and after the census), we found 69 related families, or 2.4% of census families linked to BALSAC (N = 2869). Of these, only 9 families present in BALSAC with an event were not located in the census. When the interval increases only a fraction of the families found was not recorded. Moreover, as the intervals overlap, the same unrecorded families are found in each of the larger intervals. The 9 families found

with the one-year interval are also recorded in the 2-year interval, and so on. However, the first two intervals seemed too restrictive as they are inferior to a birth interval (considering relatively long breastfeeding) and using that basis could eliminate many families for whom two consecutive births overlap our interval. Beyond the two-year interval the interpretation becomes more sensitive because of possible mobility, like temporary workers who left for a period of over a year[8]. The intervals of 10 and 20 years are certainly more at risk to the extent that families could have left the region before the census and come back after without registering any event during the period of exile. Interestingly, with the 20-year interval we only found 3 families out of 4. This is mainly explained because families whose interval remains open are not included, and because families in post-parental stage for which no event occurred during the interval are missing.

For the purposes of this presentation, the interval of two years is appropriate. First, it brings together more than a third of all linked census families, making it a very acceptable sample for our comparison. Second, even if the interval allows the inclusion of families or individuals absent during the census, the low proportion of families not listed suggests that the risk is rather low. Indeed, only 44 families possibly present on the territory do not appear in the census. Their number is small but should be enough to raise the possible scenarios and explain the reasons for their absence in the census.

Table 3. BALSAC family with a recorded event before and after the 1891 census according to interval length

	Total interval length (years)				
	1	**2**	**4**	**10**	**20**
BALSAC families (N)	69	349	1060	1770	2186
Families found in census (N)	60	326	1016	1653	2003
% of linked census families	*2.1*	*11.4*	*35.4*	*57.6*	*69.8*
Families not found in census (N)	9	23	44	117	183

First, we detected a geographical discrepancy between the two sources. Inhabited since the onset of settlement and founded only in 1890, the parish of l'Anse St-Etienne (Petit-Saguenay) is now a part of the Saguenay region. In 1891, the administrative boundaries had not yet been established at the provincial level, and under the federal jurisdiction the sector

belonged to the Côte-Nord, the neighboring region. As we also linked censuses for this region (1852-1911), we were able to trace the 17 families involved in the 1891 census to confirm this fact. It is not unusual to find variations in boundaries over time, but to find at least one case reasserts the importance of keeping track of these changes, especially for those interested in tracing mobility using both sources.

That left us with 27 families to investigate. Of them, 3 had erroneous links. Also, we recovered two missed links: a widow in an unrelated household and an individual married early in 1891 but declared in the census as single and living with his parents. Of the 22 remaining families, 5 were excluded because their selection was not appropriate. Two situations were observed to explain this: the family was dissolved before the census and the event taking place after the census was a child's marriage (which is listed in the family file of his parents); the couple lived outside the region but one spouse declared a residence in Saguenay.

According to the information contained in the records, 17 of the missing families should have been enumerated in the 1891 census. These are mostly young families as 15 of them were married (first marriage) between 1886 and 1891. Eight of them were found in 1901 and/or 1911 censuses. These families are interesting insofar as they left a trace in Saguenay at one time and their omission may suggests a potential undercount of the census. The absence of the last 7 families is harder to explain. They did not appear in any of the 7 censuses from 1852 to 1911, strengthening the idea of families that had left the area. But their BALSAC file suggests that they lived in Saguenay during the interval and some of them even have events recorded very close to the census date. In order to find our missing couples in the census we also looked for the husband's parents to see if they were present and if the spouses were not listed in their household. When parents were also missing and/or they declared a place of residence outside Saguenay, the possibility that related families migrated together made sense.

This supposition is reinforced by the analysis of the temporal sequence of our 17 families (successive appearances in the census). This sequence would allow for the detection of suspicious omissions or gaps in the series of appearances suggesting that maybe families were still in the region in 1891. We observed only 1 family that was enumerated in both the 1881 and 1901 censuses, but where the head of household could not be found in the 1891

census even if he remarried in 1893 in Saguenay. We found 8 families that appeared in censuses after 1891. But there are 8 families with no appearance at all in the censuses, suggesting that they have never inhabited the region. In most cases, family records support the migration of parents and children because the marriage of the latter took place outside of Saguenay and the declared residence of the parents was also outside the region.

The declared places of residence of families in the interval do not seem to have any particular influence on the presence/absence in the census. Different sectors are represented showing no obvious trend. We found 3 missing census families who come from the same place, St-Ambroise, a frontier area in 1891 whose official foundation occurs only in 1903. Although the case seems similar to that of St-Etienne mentioned earlier, we found in the same period 33 linked families in the census coming from the same place, which confirms that the sector has not been forgotten by enumerators.

Describing unlinked individuals in the 1891 census

From the census standpoint, the remaining 647 individuals (without St-Etienne's families) who could not be linked to BALSAC are distributed in Table 4 according to sex, age, marital status, religious and occupation. Total numbers are quite similar but there is a slight imbalance in favor of men at a young age while women are more present at a later age, a result that is consistent with the fact that there are 4 times more widows than widowers in this subgroup. Men are mostly single, reinforcing the idea of a mobile sub-population composed of young males. Permanent celibacy is difficult to disentangle from migration (unless we have the death record), and it is risky to extrapolate on exact proportions given that three quarters of people are under the age of 30, age at which it is premature to talk about permanent celibacy. Singles are more likely to be mobile workers hard to link because of a lack of contextual information to properly locate individuals within a coherent family network. Lodgers are a good example as they constitute 13% of unlinked individuals (data on relationship to head not shown). At this point of the analysis, the presence of married and widowed individuals suggests more probably that marriages occurred outside of Quebec than missed potential links. But the difficulty we are facing here is to know for sure that the marriage was or was not registered in Quebec. Being unable to find the couple in BALSAC does not mean that it's not registered. Error in names, ages, and relationship to

head can lead to miss a link. But this uncertainty is largely lifted when it concerns non-Catholic marriages, only partly included in BALSAC. Among the 9 non-Catholic households, 5 are said to be linked but in fact, most of the time only the parents were connected to the BALSAC database. Children remain for the most part unlinked. Given the small number of families involved, religion does not represent an additional explanation to our investigation.

Table 4. Distribution of unlinked census individuals

		MALE	FEMALE	TOTAL
Unlinked individuals (N)		309	338	647
Age groups				
	0-14	44.7	30.2	37.1
	15-29	34.6	36.1	35.4
	30-64	17.8	26.0	22.1
	65+	2.6	7.7	5.3
Marital status				
	Single	90.6	82.5	86.4
	Married	7.1	8.3	7.7
	Widowed	2.3	9.2	5.9
Religion				
	Catholic	90.6	91.4	91.0
	Protestant	6.8	5.9	6.3
Occupation		(107)	(62)	(169)
	Non manuals	26.2	0.0	16.6
	Skilled workers	12.1	8.1	10.7
	Farmers	36.4	3.2	24.3
	Lower skilled workers	20.6	74.2	40.2
	Unskilled workers	4.7	14.5	8.3

Despite low numbers for occupations, the distribution shows that the vast majority of unlinked individuals are manual workers. Farmers and mill workers are amongst the most

represented in the lower socioeconomic scale. High rank professionals and religious people as well as professors compose for the most part the non-manual category. As expected, women are underrepresented both in terms of number and distribution. In this sample, they worked for the most part as domestic.

Consistent with age groups is the relationship to head that is dominated (63%) by 299 sons and daughters, excluding 31 non-Catholic children. Those are removed because we know for sure that they are not in the database. By removing these cases we focus mainly on the French-Canadian population for which it is more certain that it left a trace at one time or another in the civil records. It is not surprising to find so many sons and daughters as children represent 66% of the whole census. What is striking is that 294 of them declared being born in the province of Quebec (see Table 5 below). Moreover, they are enumerated in 204 distinct households of which 193 are linked to BALSAC. Complete families are linked but households contain almost 1.5 unlinked individual that is said to be a son or a daughter. This becomes particularly interesting when we add that 62% of headships were married in Saguenay and 26% in Charlevoix.

Table 5. Distribution of unlinked census Catholic children

		MALE	FEMALE	TOTAL
Nb of unlinked children		161	138	299
Surname				
	Same	94.4	84.1	89.6
	Different	5.0	15.2	9.7
Birth place				
	Quebec	97.5	99.3	98.3
	Outside Quebec	2.5	0.7	1.7
Age difference with head				
	0-15	7.5	8.0	7.7
	16-54	85.1	78.3	81.9
	55+	6.8	11.6	9.0

Significant disparities exist among unlinked children when considering age and surnames. Regardless of age, we found 29 cases where the child did not carry the same surname as the family head. These cases are difficult to explain without a thorough analysis of the extended family to detect identical surnames and explore the trail of distant relatives. Otherwise, we have little evidence to guide our investigation[9].

We also measured the gap between age of head of household and age of the unlinked child. Some children are older than the head of household and others show an age difference of less than 15 years suggesting that they were probably family members mistakenly counted as children.

Given the quality of the sources, it is odd to find so many children absent from the civil records. They share the same characteristics: they are either born outside Saguenay, or the event is really missing in the database and they married outside the region, or they remained single. They also died outside the region or the event is missing. It seems really unlikely that this apply for all the Saguenay families represented in this sample.

A more fruitful explanation would be to think that this is the first expression of a difficult phenomenon to measure: the children transfers or, in other words, the circulation of orphans. In critical times, for example in case of death of a parent, these children may have been "given" to another couple usually related, who kept custody. Unable to bear the family burden, the living parent, usually the father, tried to place one or more children in his close family (Collard, 1991; Garneau, 1988). Civil records are silent on such transfers and suggests, by the nature of their content, that families remain united even after the death or remarriage of a parent. As emphasized by Chantal Collard (1991, p. 139), the "gift" of children was not a practice framed by state institutions as adoption or placement in orphanage. Furthermore, for 1891, declared cases of adoption are very rare: only 4 for the whole census. In a sample of Charlevoix, the neighboring region, Collard has confronted civil records between 1900 and 1960 to oral genealogies collected in the field and was able to trace 253 orphans (one parent is deceased, or both) aged 14 years and whose fate is known. It portrays a rather closed community where family dynamics played a major role in the inclusion of orphans and where they fit into "their new relationship following the dominant model of biological filiation" (Collard, 1991, p. 146. Our translation). In other

words, when the children were transferred they took the surname of the head, if not already the same, and were considered as children in their new family.

The advantage of Collard's analysis is that she knew the origin and the fate of the transferred children. But from the census, by definition unlinked individuals have no known biography, which leaves room for a number of competing interpretations. Yet, the similarity between surnames of unlinked children and the household head seems to echo his comments. However, we shall keep in mind that the patronymic variations are not impossible, for nothing indicates that the family name carried by the child is the one he received at his baptism. Also, some relationship to head may have been distorted and sons and daughters may actually only be visiting cousins the day the census was taken. Only a methodical investigation of family files and kinship network could further isolate real cases of orphans from erroneous relationship to head within the unlinked children.

Discussion

The final results showed a linkage rate of 95% for census individuals which clearly demonstrate how complementary the two sources are. Despite the local nature of our analysis, this complementarity allows us two basic conclusions: only high quality sources can allow that level of concordance between them; and the linkage program developed at BALSAC adequately fulfill the task entrusted to it. The work done here proved the reliability of both civil records and census. Comparing the sources did not reveal any major issues suggesting that one source is incomplete or unreliable. As each census is different from the others, this paper will add to the understanding of the 1891 census in Quebec. This is a modest contribution which however proves important given the rudimentary documentation accompanying this source. We also performed additional manual linkage on individuals in order to maximize the success rate. This allowed to expose the limits of the linkage program and helped quantify the possible gains. In addition to identifying duplicates, the most interesting gain for research is probably related to the linkage of individuals in unrelated households, who are also the more mobile individuals.

Some unlinked individuals raised more questions than anticipated. Nearly 300 children or just over 3% of census children have an unknown destiny among individuals and families whose biographies are known. The hypothesis of child transfer seems quite plausible although difficult to observe because of the informality of this practice and the uncertainty relatively to the reliability of the information on the relationship to head of household. Knowledge of godparents, among other things, would certainly have helped contextualize the origin and fate of some unlinked individuals. Unfortunately, this information is not available in the database at this moment. This is a limit of the BALSAC database that surely restrain analyzes on family, kinship and support networks.

In this sense, even if we insist on the complementarity of sources, we must also take into account the disparities between them. This is particularly true for the geography of each source. Their respective boundaries are under jurisdiction of different government levels, the geographical scales used in each one are also different, and boundaries vary over time. We are currently working on family network and residential itineraries within Saguenay and the level of concordance between the geographic variables will be part of our analysis. Another current project aims to compare the declared origins in the censuses with the genealogical origins of census individuals. This work will add to our understanding on the joint use of sources for research.

Finally, sequencing the individual links proved to be a powerful tool both to validate the linkage process and to improve overall results. In continuation of the IMPQ project, we are working on the complete sequence of the 7 Saguenay censuses to draw clearer conclusions about its potential for research. Ultimately, censuses for every region involved in the project will be sequenced, adding scope and depth in the analysis of the Quebec longitudinal data.

Notes

[1] The lack of documentation or the absence of important conflict may explain this different reality. In fact, Aboriginals were occupying the territory long before French Canadian settlement. Mostly in the Lac-St-Jean part of the region, a sector for which we do not have the census data yet, it would have been interesting to compare the issues related to their census enumeration with those related in the Dunae article.

[2] The total population of the province was just under 1.5M individuals in 1891.

[3] Technicians link one census at a time with the first module. The second model, census to census, keeps them blind from previous linkage with the first module (unless they go back into it). This ensures impartial linkage and prevents that the same technicians always link the same households.

[4] We calculated the percentage of linked individuals in 1881 that also appeared in the 1891 census. This was calculated before any attempt to improve the linkage results and it does not include 1881 unlinked individuals. That percentage is therefore slightly underestimated. This rate is consistent with migration rates of the same period (Pouyez et al., 1983) and the persistence rates for families in SLSJ highlighted by St-Hilaire (St-Hilaire, 1996).

[5] In theory, children less than 10 years old should not appear in the 1881 census. Since age is not always accurate, we compared all individuals to maximize results and to highlight potential discrepancies between the two censuses.

[6] We found three matches outside this interval. They had highly discriminant names with only one potential match in the other census.

[7] The most frequent case is where a married individual is enumerated in both his family (and/or family by marriage) and conjugal households.

[8] As mentioned earlier, the enumerators had to record people that had left for a period of less than one year. If the family left for a longer period but still less than our interval, we would then consider people that should not have been enumerated and therefore overestimate our sample.

[9] In the previous section we compared unlinked individuals with 1891 matched 1881 and 1901 and these cases have not been found. However, it is possible that they were present in these censuses without being linked.

Bibliography

BALSAC. (2016). Retrieved from http://balsac.uqac.ca/

Bouchard, G., & Bergeron, M. (1975). Les registres de l'état civil de Notre-Dame de Laterrière (1855-1911). *Archives 75.3*, *3*(3), 164–173.

Bouchard, G., & LaRose, A. (1976). La réglementation du contenu des actes de baptême, mariage, sépulture, au Québec, des origines à nos jours. *Revue d'histoire de l'Amérique française*, *30*(1), 67–84. doi:10.7202/303510ar

Bouchard, G., & Larouche, J. (1988). Dynamique des populations locales: la formation des paroisses rurales au Saguenay (1840-1911). *Revue D'histoire de l'Amérique Française*, *41*(3), 363–388. doi:10.7202/304583ar

Bouchard, G., Roy, R., & Casgrain, B. (1985). *Reconstitution automatique des familles: le systeme SOREP*. Chicoutimi: SOREP.

Bourque, M., Markowski, F., & Roy, R. (1984). Évaluation du contenu des registres de l'état civil saguenayen, 1842-1951. *Archives*, *16*(3), 16–39.

Cartwright, S. R. (1912). *Reminiscences*. Toronto: W. Briggs.

Charbonneau, H. (1975). *Vie et mort de nos ancêtres: étude demographique*. Montréal: Presses de l'Université de Montréal.

Collard, C. (1991). Les orphelins "propres" et les autres... Carence parentale et circulation des orphelins au Québec (1900-1960). *Culture*, *XI*(1-2), 135–149.

Curtis, B. (2002). *The Politics of Population: State Formation, Statistics, and the Census of Canada, 1840-1875*. University of Toronto Press.

Dunae, P. A. (1998). Making the 1891 Census in British Columbia. *Histoire Sociale / Social History*. Retrieved from http://hssh.journals.yorku.ca/index.php/hssh/article/view/4666

Fleury, M., & Henry, L. (1956). *Des registres paroissiaux à l'histoire de la population: manuel de dépouillement et d'exploitation de l'état civil ancien*. Paris: I.N.E.D. Retrieved from http://ariane.ulaval.ca/cgi-bin/recherche.cgi?qu=a2188992

Gaffield, C. (2007). Conceptualizing and Constructing the Canadian Century Research

Infrastructure. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *40*(2), 54–64. doi:10.3200/HMTS.40.2.54-64

Garneau, B. (1988). La circulation des orphelins au Saguenay entre 1900 et 1970. *Anthropologie et Sociétés*, *12*(2), 73–95. doi:10.7202/015024ar

Gauvreau, D., & Bourque, M. (1988). Mouvements migratoires et familles: le peuplement du Saguenay avant 1911. *Revue d'histoire de l'Amérique française*, *42*(2), 167–192. doi:10.7202/304677ar

Inwood, K., & Reid, R. (1995). Introduction: The Use of Census Manuscript Data for Historical Research. *Histoire Sociale / Social History*, *28*(56), 301–311. Retrieved from http://hssh.journals.yorku.ca/index.php/hssh/article/view/16638

Jetté, R. (1980). *Reconstitution de recensements à partir de registres paroissiaux: analyse méthodologique sur échantillon*. Université de Montréal.

Pouyez, C., Lavoie, Y., Bouchard, G., Roy, R., Simard, J.-P., & St-Hilaire, M. (1983). *Les Saguenayens: introduction à l'histoire des populations du Saguenay XVIe-XXe siècles*. Sillery: Presses de l'Université du Québec.

Roy, R., Bouchard, G., & Declos, M. (1988). La première génération de Saguenayens: provenance, apparentement, enracinement. *Cahiers Québécois de Démographie*, *17*(1), 113 – 134. doi:10.7202/600632ar

Ruppert, E. (2015). Infrastructures of census taking. In G. Darroch (Ed.), *The Dawn of Canada's Century: Hidden histories* (pp. 51–70). McGill-Queen's University Press.

St-Hilaire, M. (1996). *Peuplement et dynamique migratoire au Saguenay, 1840-1960*. Québec: Presses de l'Université Laval.

Vézina, H., St-Hilaire, M., & Bellavance, C. (2014). The linkage of micro census data to vital records: New perspectives for longitudinal studies. In *Using micro census data: recent developments and new perspectives*. First Conference of the European Society of Historical Demography (ESHD). Alghero, Sardinia, Italy. September 25th-28th.

Vézina, H., St-Hilaire, M., & Bournival, J.-S. (2015). The linkage of micro census data and vital records: an assessment of results on Quebec historical censuses (1852-1911). Paper presented at the 40th Annual meeting of the Social Science History Association, Baltimore, MA.

Willigan, J. D., & Lynch, K. A. (1982). *Sources and methods of historical demography*. New York: Academic Press.