



The linkage of microcensus data and vital records: An assessment of results on Quebec historical population data (1852–1911)

Hélène Vézina^{a,b}, Marc St-Hilaire^c, Jean-Sébastien Bournival^{a,c}, and Claude Bellavance^b

^aProjet BALSAC, Université du Québec à Chicoutimi, Chicoutimi, Québec, Canada; ^bCentre interuniversitaire d'études québécoises, Université du Québec à Trois-Rivières, Trois-Rivières, Québec, Canada; ^cCentre interuniversitaire d'études québécoises, Université Laval, Québec, Québec, Canada

ABSTRACT

The interconnection between civil records and census data, along with linkage across censuses, can substantially broaden and enrich the avenues of research in both the social and the biological sciences. Here we present a linkage program developed to match nominative microdata from the Canadian censuses to those from Quebec civil records. We provide a critical assessment of linkage results obtained on two geographical settings by investigating household and individual characteristics that could impact on linkage rates and raise bias issues. The linkage program is at the heart of the construction of the Integrated Infrastructure of the Quebec Population Historical Microdata which will integrate available historical microdata (vital records and census data) on the Quebec population dating back to the beginning of European settlement.

KEYWORDS

Canadian censuses;
historical population
database; longitudinal data;
Quebec vital registration;
record linkage



Introduction

While research on contemporary populations relies on censuses, large surveys, or various substantial longitudinal data sets provided by governmental agencies, studies on historical populations can count on very few sets of consistent data and need inventive ways to develop and exploit this data. This represents “a real challenge,” as stated by Bloothoof et al. (2015) in their recent collection of essays.¹ Except for some European countries, such as Belgium and the Netherlands where local population registers exist since the 19th century, researchers must use modern censuses (that is from the mid-19th century) and civil or parish records to reconstruct past populations.

These sources of data are fundamental for the study of two major categories of social facts. By nature cross-sectional, nominal censuses make it possible to study household characteristics, residential patterns, occupational structures and income as well as micro-scale economic strategies. In turn, information from vital events, once linked to create longitudinal family files, fosters research on family and

kinship, such as reproduction, family formation, migration and intergenerational transmission of various traits. Each source also has its limitations: censuses do not contain information on the role of extended kinship in economic and residential strategies; vital registration provides no or very little indication on the socioeconomic conditions of individuals and families and none on living arrangements and neighborhoods.

In the North American context, census microdata have been digitized for many years now and large transversal databases have been created from US or Canadian census manuscripts, such as those developed by the Minnesota Population Centre (IPUMS) or the Canadian Century Research Infrastructure.² On the other end, due to the variability of their consistency and quality, few parish or civil registers have been digitized and linked for longitudinal research. The need for such biographical records has thus led scholars to use census microdata to create longitudinal data sets by linking nominative information, that is, the name and surname of individuals, from one census to a previous or a next one.³

CONTACT Hélène Vézina  hvezina@uqac.ca  Projet BALSAC, Université du Québec à Chicoutimi, 555 boul. de l'Université, Chicoutimi, Québec G7H 2B1, Canada.

Preliminary versions of this paper were presented in the session “Using microcensus data: recent developments and new perspectives” at the First Conference of the European Society of Historical Demography (ESHD) in September 2014 and in the session “Challenges, Methods and Solutions in Record Linkage of Micro-data” at the 40th Annual meeting of the Social Science History Association in November 2015.

Color versions of one or more figures in this article are available online at www.tandfonline.com/vhim.

The Quebec province in Canada 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 since the early days of French settlement in the 17th century. More than 40 years ago, two major projects, the Programme de recherche en démographie historique (PRDH) and the BALSAC Project, initiated the digital transcription and linkage of these parish records. More recently, a good share of Quebec data from the Canadian historical censuses prior to 1921 has also been digitized and formatted in databases by the Centre interuniversitaire d'études québécoises (CIEQ) and by the PRDH.

In order to overcome census and parish records intrinsic limitations, and build on the strengths of the two types of data, a computer program has been developed at BALSAC to create enhanced biographical data sets. This program relies on systematic and automated tools and procedures to match microdata from the Canadian censuses to those from Quebec civil records and to link census data together. In doing so, we are creating a research infrastructure that will take advantage both of the power of family and genealogical files based on civil registration and of the wealth of census data thus opening up new opportunities of research for the study of historical populations.

In this paper, we first provide a brief overview of past and ongoing population reconstruction projects in Quebec. Then we describe the two components of the linkage program and analyze the results of linkage operations performed in two geographical settings: the Saguenay region for which the BALSAC population database contains birth, marriage and death certificates linked into family files, and the city of

Trois-Rivières for which BALSAC contains only marriages⁴. In order to provide a critical assessment of these results, we investigate household and individual characteristics that could impact on linkage rates and affect the representativeness of the linked population. We also illustrate the potential of the linkage program by comparing matching results among men and women. Lastly, we discuss the possibilities and limits involved in our approach.

Population reconstruction projects in Quebec

The Registre de la population du Québec ancien (RPQA) was created in 1966 by the PRDH at the Université de Montréal. Its construction was inspired by the techniques of family reconstitution developed by the French demographer Louis Henry. It contains the longitudinal linkage of the entire Catholic population of Quebec from the beginning of French settlement in Canada in 1608 (first record in 1621) to 1799, comprising 700,000 birth, marriage and death records.

In 1972, the BALSAC Project was initiated at the Université du Québec à Chicoutimi. In the first phase, family reconstitution was performed on the population of the Saguenay-Lac-St-Jean, a region located 200 km north of Quebec City (see Figure 1), from the beginning of settlement in 1838 to 1971 and rely on 660,000 birth, marriage and death records. Since 1989, a second phase has been ongoing to perform the digital transcription and linkage of marriage records for the whole province of Quebec. The work is now completed up to 1965 and the database contains 2.2 million linked marriage records allowing for the automatic construction of genealogies.

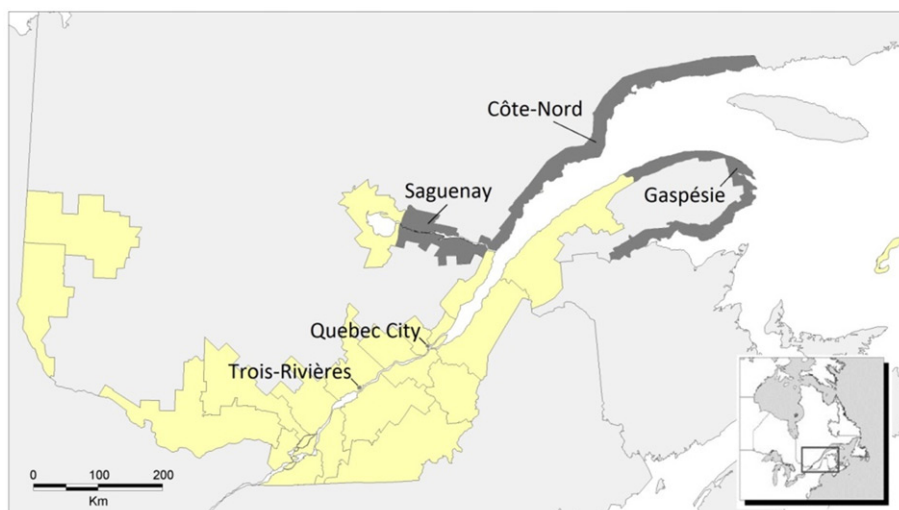


Figure 1. Census data included in the IMPOQ.

The complete census of Canada for 1881 as well as samples for all other Canadian historical censuses (from 1852 up to 1911 with the exception of 1861) are formatted in digital databases and publicly available. In Quebec, more specifically, full count censuses for the cities of Quebec and Trois-Rivières and some Quebec regions have been digitized for seven decennial censuses (1852–1911) in the course of previous projects, along with a few samples for other parts of the province.⁵

A few attempts to link Quebec civil and census records were previously carried out relying on manual work performed on small data sets (Charbonneau et al. 1970; Gauvreau 1991; Gossage 1999). In some instances marriage records were used to facilitate linkage across censuses, especially to link single individuals living with their family of origin in one census to married individuals in the next census (when the marriage occurred in the interval) (Olson and Thornton 2011; St-Hilaire 2009; St-Hilaire, Richard, and Marcoux 2014). In 2010, members of our group initiated a pilot project to evaluate the feasibility of designing a program to link census data to marriage records found in the BALSAC database (Gauvreau, Thornton, and Vézina 2010; St-Hilaire and Vézina 2010; Vézina and St-Hilaire 2011). The development of this linkage program is now at the heart of the construction of the Integrated Infrastructure of the Quebec Population Historical Microdata (IMPQ).

The IMPQ integrates available historical microdata on the Quebec population dating back to the beginning of European settlement, into a set of relational databases. The objective is to preserve, highlight and develop this major historical and scientific heritage. The creation of the infrastructure will also facilitate integration of new data and development of new tools for linkage and analyses as well as promote training in relevant disciplines and collaborations both at the national and international levels (comparative research).

The construction of the infrastructure includes three components. First we plan a full fusion of the BALSAC and RPQA databases which both contain data from vital registration. This will enable continuous updating of the longitudinal data as well as integration and linkage of new vital records. The second part focuses on the harmonization of existing census data series and expansion of the geographical coverage to include two urban environments (Quebec City and Trois-Rivières) and three regions mixing rural and urban environments (Gaspésie, Côte-Nord and Saguenay) (see Figure 1). In the third component of

the project, households and individuals from these areas are gradually linked to the BALSAC database and across censuses. The linkage program we present here is thus at the heart of the construction of the IMPQ.

Developing and implementing the program for the linkage of civil records and census data

The BALSAC linkage system of vital records as a starting point

From the start, we made the decision to use an approach based on the linkage procedures developed for the construction of the BALSAC database (Bouchard, Roy, and Casgrain 1985). The challenge was to adapt them to take into account the differences in the structures of the two types of data. Before describing these modifications, we will first present briefly the BALSAC linkage system.

The BALSAC family reconstitution system which processes vital data records, aims at grouping in the same file all mentions appearing in the records and referring to the same couple. Thus, the basic unit of information for linkage is the “couple mention” which contains four nominative elements namely the husband’s name and surname and the wife’s name and surname. All nominative information is processed to eliminate superficial name variations using FONEM, an automatic phonetization program (Bouchard, Brard, et al. 1981; Bouchard and Bouchard 1981). As shown in Figure 2, FONEM can significantly simplify the name variations to be linked.

Candidate couple mentions in BALSAC are created on the basis of at least two common nominative elements with the couple to be linked. These mentions are compared using three programs designed to detect and measure degrees and forms of similarity between two last names or two first names. ISG (*Similarity index*) calculates a score based on the degree of similarity between nominative pairs of elements based on the position of same letters in names. INC (*Inclusion*) deals with truncated names by detecting suffixes and prefixes in names and deciding if one can be treated as being included in the other. ELM (*Multiple elements*) processes situation of multiple names and surnames and decide if two entities can be treated as equivalent or not. The decision-making process relies almost exclusively on nominative information contained in the records but the coherence of dates in the family structure is also taken into account (Bouchard, Roy, and Casgrain 1985).

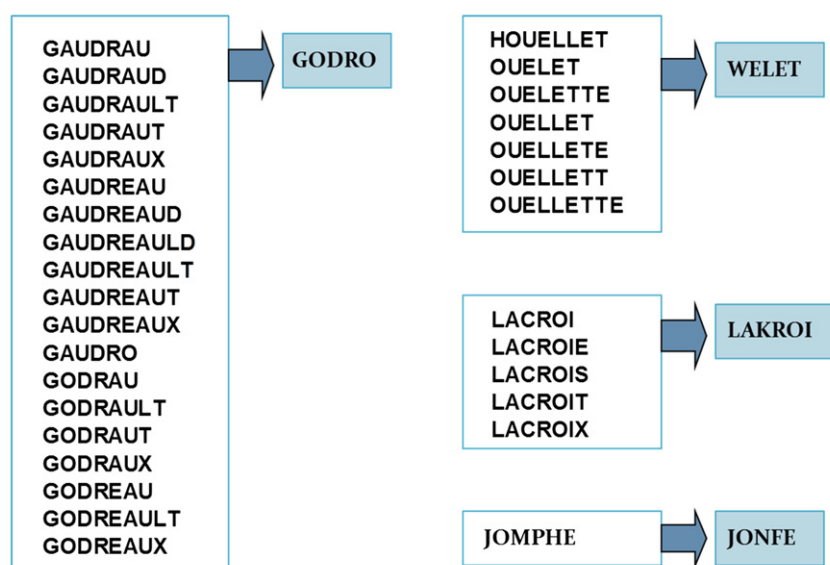


Figure 2. Examples of automatic phonetization with the FONEM program.

The linkage operation consists of four different phases processing first the easiest links and moving progressively to the most difficult. The first two phases are performed immediately after the entry of a marriage record. The first one is entirely automatic: a couple is linked to an already existing family file only if all nominative information is perfectly identical and if there is only one potential candidate for the linkage. In the second phase, when the program has been unable to make the appropriate linkage within the established safety range, it provides a list of potential candidates for linkage and the entry clerk makes a selection if an unequivocal choice can be made. The third phase is performed at a later stage. It is also based on a list of candidates provided by the program and is operated by experienced staff who might use complementary sources, such as genealogical repertoires or websites to support their decision. Lastly, after these three stages, there remains a small share of unsolved linkages which could eventually be subjected to further investigation. Since this is a costly and time-consuming operation, it will take place depending on research needs and available resources.

Linkage operations lead to the construction of a family file that includes all the records pertaining to a unique couple (their own marriage, the remarriage of a surviving spouse when existing, the marriage of their children⁶) (St-Hilaire 1990). All links submitted to automatic linkage and those carried out at the stage of the computer-assisted manual linkage are immediately validated by automatic routines of coherence (for example: acceptable timespan between two events, chronological sequence of events, reported or

calculated age, duplicate events or names) (Bouchard, Casgrain, et al. 1981).

Adapting the BALSAC method for census and civil records linkage

The program for the linkage of civil records and census data rests on the principles used at BALSAC for the linkage of civil records. Essentially, the method remains the same and linkage is done using matching programs as well as comparison and decision algorithms. However, several adjustments were made to take into account differences in structure and content between the two types of data. Hence, before describing the method of linkage itself, the main issues raised by the differences in the two data sources and an overview of modifications brought to the BALSAC system are presented.

First, the structure of the basic grouping units (family vs. household) is different. The BALSAC family structure corresponds to a nuclear family model built on information found in civil registration events. Census household composition is determined by a survey at a specific point in time and is based on a residential criterion (housing). Several scenarios of household structure are possible: one or many single individuals, nuclear family (partial or complete), blended family, intergenerational family, etc. That being said, the similarities between the two types of data are real, since the household as a grouping unit often corresponds to a structure, if not identical, at least similar to that of a nuclear family.

Second, the quality of nominative information is not the same in both sets of data. Civil records are legal documents and priests, pastors, and other civil officers usually kept them with great care. They are very consistent through time since there was little change in prescribed rules to record information (Bouchard and La Rose 1976). The context of census data collection is very different. It is a massive and complex operation aimed primarily at ensuring a fair representation at the Parliament but also used to follow the development of the nation. It is conducted in a short period of time, at 10-year intervals by people who were not always well trained and sensitive to the cultural context of the people they enumerated (Bellavance, Normand, and Ruppert 2007). These factors, combined with the fact that the census nominative information has no legal value, impact directly on the quality of the nominative data (usual instead of legal first and last names, nicknames, abbreviations, phonetic alterations, etc.).

Last but not least, whereas in the BALSAC database, first and last name are found for both spouses, in census data the maiden name of the wife is almost always missing. As the whole linkage system relies on couple mentions comprising four distinct nominative elements, this represented an important issue for the development of our program.

Given these differences, some modifications were brought to the linkage program and new elements were introduced. One of the main changes in the method consists of a prior shaping of census data in order to extract nuclear families comparable to BALSAC families. To be considered for linkage, a household must contain at least two members forming a familial unit (husband and wife or widow-er with a child) since a minimum of three nominative elements is needed.

To overcome the absence of the wife's maiden name in the households, it was decided to generate a new nominative unit of comparison (NUC) including the three nominative elements available (father's last name, father's first name and first name of the mother) to which is added the first name of a child. It is then possible to compose as many NUCs as there are children mentioned. For each set of data, the program creates a table of nominative mentions whose number per household or per family varies according to the number of children.

Because of the changes brought in the composition of NUCs, it was necessary to relax sorting and matching criteria to bring more potential candidates in the phase of nominative pairing (see below). The adapted

method of matching requires only one nominative element out of four to be perfectly identical between a NUC derived from the census and the one obtained from BALSAC. The relaxation of these rules is, however, offset by the addition of a rating system of candidates which will be reported in the next section.

The implementation of this rating system and the matching criteria used to define it represent another major change to the BALSAC linkage approach. Because of the variability in nominative information between the two sources, it seemed important to involve other attributes in the comparison process to handle situations where the nominative data would be limited or ambiguous. Thus, as described below, ages, event dates and residence are taken into account to improve the capacity of the program to quantify the degree of similarity between two families and to generate scores that truly reflect the plausibility of a family as a candidate for linkage. It must be emphasized, however, that only nominative information is used in the selection and pairing of candidates. The other attributes are considered afterwards in the comparison process and score assignment.

Lastly, while developing the linkage program, we performed various tests using a sample of 1400 households selected in the 1871 Quebec City census which was manually linked in the course of a previous project. Using this sample as a ground truth, we were able to calibrate the various parameters of the program in order to obtain the best possible matches (Bourque et al. 2013). Based on these tests, we also made the decision to exclude automatic linkage and to rely exclusively on manual computer-assisted linkage. This means that all linkage decisions are made by research assistants.

Linking census data to BALSAC families

The first module of the program is used for linking census data to BALSAC and involves five steps described in Figure 3. In the first two steps data from the census and from BALSAC are transformed into a common structure: the nominative unit of comparison (NUC). In step 3, the NUCs are sorted and paired to generate a pool of potential candidate families in BALSAC. In step 4, each candidate receives a score based on selected matching criteria. Lastly, in step 5, a final decision is made to proceed with linkage or not. When a household is linked, the program allows for linkage of individuals belonging to this household.

Step 1. Preparing the census data: Nuclear families must first be identified and extracted from census

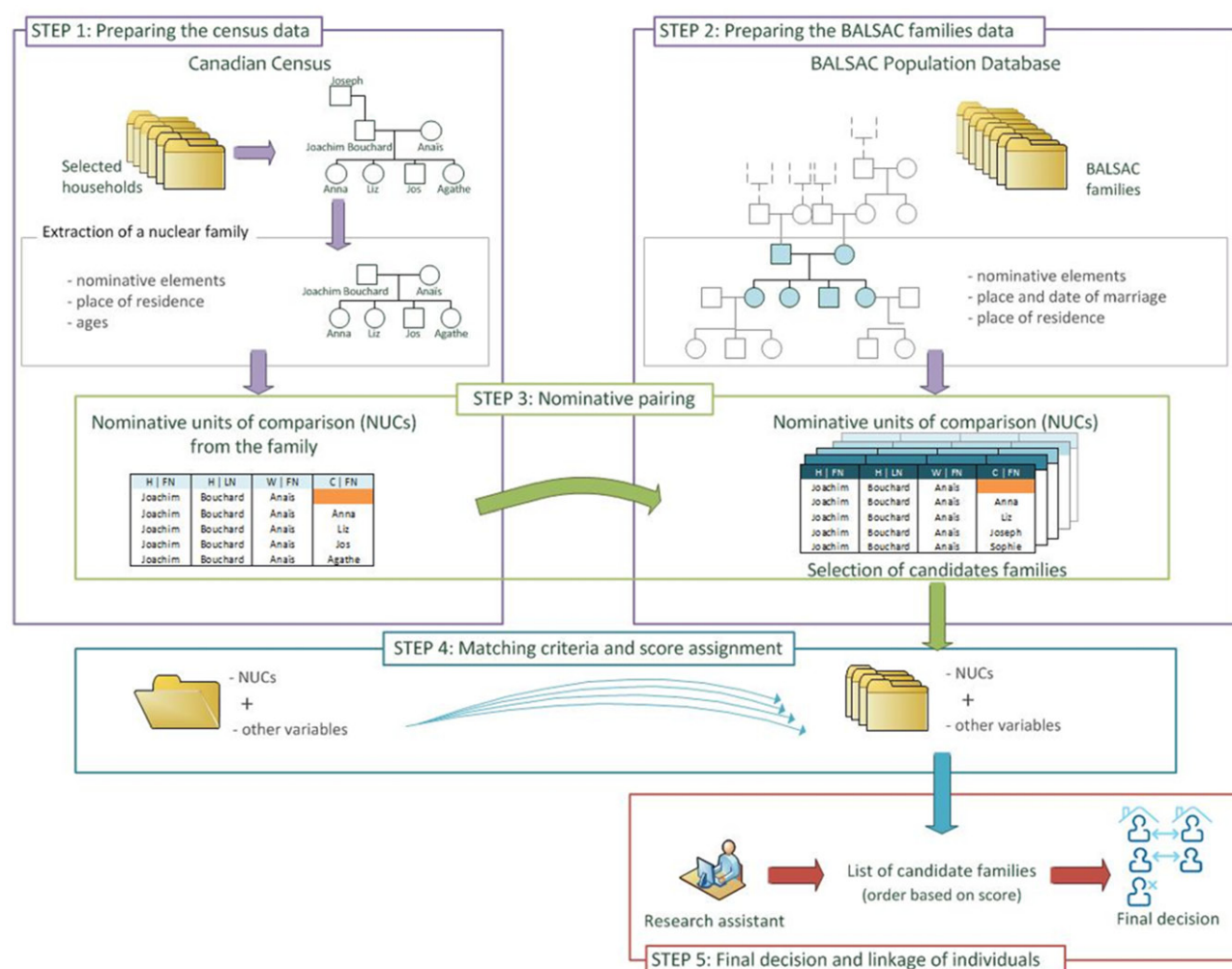


Figure 3. The linkage of census data to BALSAC families.

households in order to build the NUCs. More than one family can be extracted from a household. To compensate for the absence of the wife's maiden name in census data, the first name of each child is added to the NUC. Hence, there will be as many NUCs for a census family as there are children. The three-element combination containing the first and last names of the husband and the first name of the wife is also used and it is the only combination possible when a couple has no children recorded. Other members of the household are regarded as children of this couple if a) their surname is the same as the presumed father; b) their age is compatible with the age of the mother (minimum age difference of 15 years); and c) they are single. In addition to the nominative data, other variables that will be used to link records and to calculate a score are kept (place of residence, ages).

Step 2. Preparing the data on BALSAC families: Potential candidate families from BALSAC are also organized in NUCs to provide a pool of candidates

for each census family. First, a subset of families is extracted from the BALSAC database. These families are selected within a specified time interval based on the likelihood of having been enumerated at a given census. First and last name of husband and wife, name of children as well as other variables that will be used for linkage (place of residence, ages and date of marriage) are retrieved.

Step 3. Nominative pairing: Based on the programs of comparison of nominative similarity developed at BALSAC and described above, the application performs various sorting operations aiming at pairing each census family to candidate BALSAC families. The pairing criterion is that both NUCs (one from the census and the other from BALSAC) must contain at least one identical nominative element. A list of potential candidates for linkage with a specific census family is thus selected from the BALSAC families.

Step 4. Matching criteria and score assignment: Each of the BALSAC families selected in the pairing process is submitted to the comparison component of

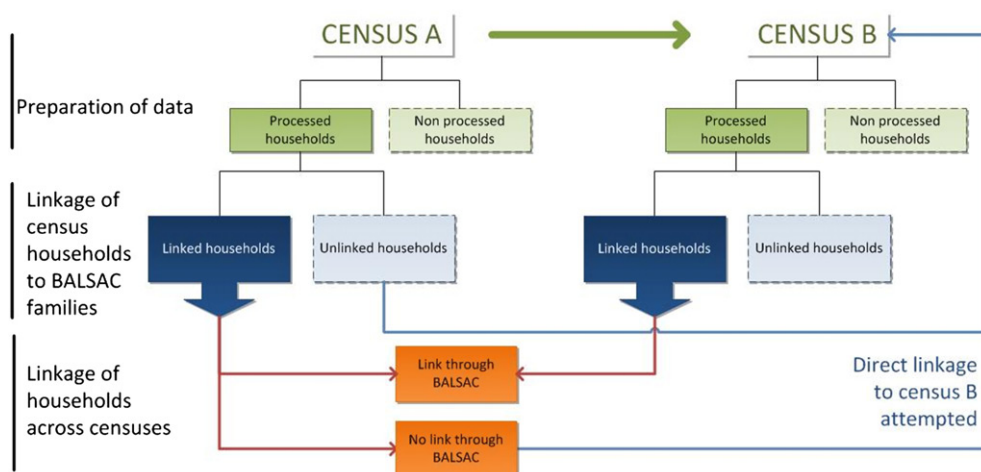


Figure 4. Linkage of households across censuses.

the program and a score is attributed to each potential match based on the degree of similarity with the census family being processed. While nominative data found in the NUCs still provides the most important linkage criterion, concordance of dates and ages as well as place of residence are also used in the comparison process. The goal is to take into account and make optimal use of the information available in each data set to sort the candidate families for linkage. Every element of the comparison receives a score weighted according to its importance. For instance, concordance of children's ages is more important than similarity of their first names. The weights are modifiable according to the characteristics of the data to be linked. Candidate families are sorted based on their total score and are presented by decreasing order to the linkage operator for the final step (Bourque et al. 2013).

Step 5. Final decision and linkage of individuals:

At this stage, manual work is involved to proceed with the linkage decision. Based on the linkage results presented in this paper, we have calculated that depending on the region and on the census year, between 72% and 92% of the created links involve the candidate family with the highest score. A final decision is made to accept a candidate or abort the linkage attempt within two minutes on average.

Linking families and individuals across censuses

The second module of the program allows for linkage across censuses.⁷ In the work, we are currently conducting for the IMPQ project, linkage between two censuses takes place after both censuses have been linked to BALSAC using the first module. The whole linkage process is similar to the one described in

Figure 3. In each of the two censuses, nuclear families are extracted from households and NUCs are produced. Nominative pairing based on the NUCs lead to the selection of candidates from Census B for linkage to a specific family in Census A. Nominative and contextual information are used for score assignment and the final decision is taken by the research assistant.

There is, however, a prior selection in Census A of households depending on if they have been linked to a BALSAC family or not. This selection process is displayed in Figure 4. It shows that in the preparation of data, some households will be processed and others will not, based on the criteria described previously (at least two people in the household and the head being married or widowed). Among processed households, some of them will then have been linked to BALSAC while for others linkage will not have been successful. Direct linkage to Census B will be attempted on all households from Census A which have not been linked to a BALSAC family. Moreover, among linked households from Census A, some will be connected to a household in Census B through their linkage to the same BALSAC family but others will not. Direct linkage to Census B will also be initiated on this latter set of households. Linkage is also performed on all individuals found in linked families.

An illustration of the approach: Linkage results in Saguenay and Trois-Rivières

In this section, we present the results obtained on the region of Saguenay and on the city of Trois-Rivières (see Figure 1) to illustrate our linkage approach. We also carry out a critical evaluation of these results by analyzing their variability depending on characteristics of households and individuals.

The Saguenay region is located about 200 km north of Quebec City. Sedentary settlement, almost uniquely by French Canadians, began in the second quarter of the 19th century and the population grew rapidly mostly through natural increase. At the beginning, the region consisted almost exclusively of farmers and industrialization began slowly with the implantation of pulp industries at the turn of the 20th century. Cities of modest size were in place from the late 19th century. Trois-Rivières, located on the north shore of the St. Lawrence River about halfway between Montreal and Quebec City, is a middle-sized city founded in 1634 and also populated by a majority of French Canadians. It was the third city in size (after Quebec and Montreal) up until the middle of the 19th century when it lost its rank due to a slower rate of industrialization than other cities of similar size. Growth resumed in the early 20th century with the development of hydroelectricity and the pulp and paper industry in the surrounding Mauricie region.

Census data for the two areas were digitized for the 1852–1911 Canadian censuses and Table 1 shows the number of households and individuals as well as the mean number of individuals per household for each census. The population was slightly bigger in Saguenay already in 1852 and as the rate of increase in this region was also higher the difference in population size became more marked with each census. The mean size of households was also bigger throughout the whole period in Saguenay where households contain on average one more member than in Trois-Rivières.

As described above, linkage between census data and BALSAC was carried out first. BALSAC contains all marriages celebrated in Quebec up to 1965 as well as birth and death records for the Saguenay region.⁸ Thus for Trois-Rivières, candidate families from BALSAC are composed of couples and their children married in Quebec whereas for Saguenay they also contain all children born and/or dead in the region whether they married or not. In a second step, linkage

to the next census was performed on households that were not already connected through BALSAC.

Based on our selection criteria, linkage was attempted on 97% of households in Saguenay and on 96% in Trois-Rivières. Non processed households are households without conjugal units (counting one member or single members only) and collective households (such as logging camps or institutions like schools, religious congregations and hospitals). After performing linkage to BALSAC and to the next census, we had 2.7% of processed households in Saguenay and 10.5% in Trois-Rivières that could not be linked at all. This means that overall, 94.5% of households in Saguenay and 85.8% in Trois-Rivières were linked to BALSAC or to the next census or in most instances to both sources.

Each time a household is linked either to BALSAC or to the next census, the linkage program offers the possibility to link household members individually. Altogether, 91.4% of individuals present in linked households in Saguenay and 61.5% in Trois-Rivières were linked to BALSAC. Among individuals in the 1852–1901 censuses, 60.4% in Saguenay and 38.1% in Trois-Rivières were linked to the next census.⁹

The linkage of households and individuals to BALSAC

The left panel of Figure 5 shows, for each of the seven censuses, the proportion of households that were linked to a BALSAC family. Linkage rates are above 90% (between 93 and 97%) for all censuses in Saguenay. They are lower and slightly more variable in Trois-Rivières: between 81 and 89% except for the 1871 census reaching only 75%. Differences between the two regions are largely explained by the availability of birth and death records in Saguenay which facilitates the linkage (for instance in the case of very common names or surnames which would make it difficult to select a candidate family) since information will be available for a higher number of children (including those who remained single or married

Table 1. Distribution of households and individuals in the 1852–1911 Canadian censuses for Saguenay and Trois-Rivières.

	Saguenay			Trois-Rivières		
	Households	Individuals	Mean household size	Households	Individuals	Mean household size
1852	681	5,031	7.4	787	4,873	6.2
1861	1,075	8,825	8.2	992	6,927	7.0
1871	1,994	11,814	5.9	1,714	8,375	4.9
1881	2,074	13,810	6.7	1,760	9,296	5.3
1891	2,300	14,776	6.4	1,700	9,119	5.4
1901	2,669	16,365	6.1	2,223	12,195	5.5
1911	3,786	23,769	6.3	2,993	16,524	5.5
Total	14,579	94,390	6.5	12,169	67,309	5.5

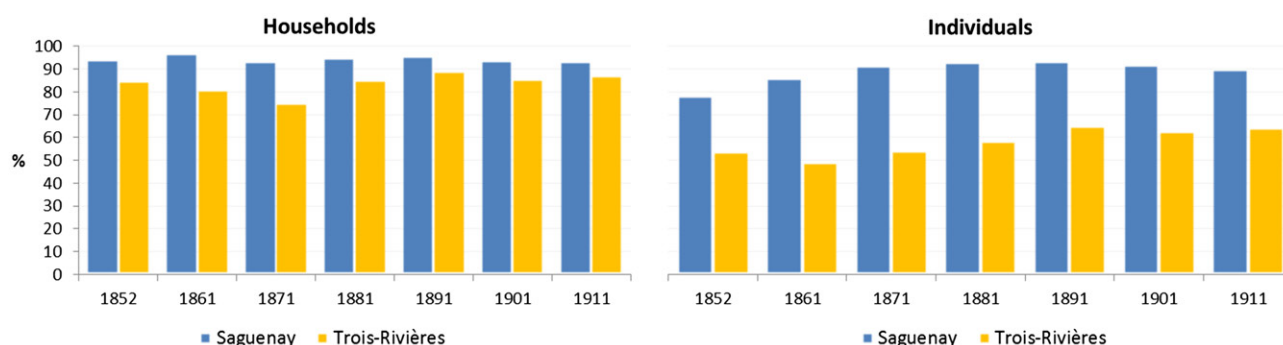


Figure 5. Proportion of households and individuals linked to BALSAC at each census.

outside Quebec). Another possible explanation is that although both areas were mostly French Canadian and Catholic, there was a small non-Catholic population in Trois-Rivières (3–8% throughout the period) while in Saguenay the population was 99% Catholic and French Canadian during the whole study period.

The right panel of [Figure 5](#) displays, for each census, the proportion of individuals that were located in BALSAC. In Saguenay, figures are slightly lower than for households (between 86 and 93%) with the exception of the 1852 where “only” 78% of individuals were linked. This could be due to the fact that as the first recorded births in Saguenay were in 1842, there were fewer individuals than for the other censuses that could be traced in BALSAC through their birth records. The linkage rates in Trois-Rivières are much lower with between 49 and 65% of individuals linked to BALSAC at each census. Again, this is probably due to the fact that when data is available on married children only, children who never married or who married outside Quebec cannot be linked.

Next, we were interested in verifying to what extent the population linked to BALSAC was different from the complete census population both at the household and individual levels.¹⁰ In order to see what factors could impact on the probability of a household to be linked to BALSAC, we looked at linkage rates by household types, household size and occupation of heads. Results are shown in [Table 2](#). Regarding household types, substantial differences can be seen among categories in both regions with linkage rates for widows and widowers (with at least one child) 16–25% lower than couples. As mentioned above, the basic unit of nominative information in BALSAC comprises the couple’s first and last names; so when we attempt linkage on households headed by a widower, we have the full name of the man but we do not have any information on his late wife’s name. When we work with families headed by a widow, the situation gets more complicated because we know only the last

Table 2. Proportion of households linked to BALSAC by household type and size and by occupation of head.

	Saguenay		Trois-Rivières	
	Linked (%)	N	Linked (%)	n
Household type				
Couple	98.1	13,101	90.7	10,194
Widower + child	91.7	411	79.3	397
Widow + child	83.0	448	69.7	864
Other	82.8	198	54.0	213
Household size				
2	90.7	1,090	77.9	1,493
3	94.5	1,431	83.4	1,637
4	95.7	1,550	87.5	1,656
5	97.2	1,639	89.2	1,639
6–9	98.7	6,075	92.3	4,192
10–19	99.1	2,345	93.9	1,022
Occupation of head				
Higher occupations	90.9	894	84.7	2,300
Skilled workers	97.1	1,253	89.8	2,285
Farmers	98.9	7,029	92.5	949
Lower skilled workers	93.8	513	91.2	1,163
Unskilled workers	97.4	3,120	92.1	2,965
Total	97.2	14,158	88.1	11,668

name of the late husband and the first name of the woman which means that we miss complete names for both spouses. The section on household types in [Table 2](#) has a category for “Other” household types which comprises mostly households where the head is said to be married but no other married individuals are found. Some of these household heads are most likely widows and widowers while others could be married individuals whose spouse was enumerated elsewhere or not at all.

Household size has a clear effect on linkage rates as the proportion of linked households increases with the number of people in the household. This increase is more marked among households of smaller size (2–5 members) but it is also apparent among larger households. This size effect is more important in Trois-Rivières (rates going from 78% in households with two members to 94% in households with 10–19 members) than in Saguenay (rates going from 91% in households with two members to 99% in households with 10–19 members). Thus the difference in linkage rates between the two regions is more important in

smaller households (13 percentage points in households with 2 members) than in larger ones (5 percentage points in households with 10–19 members). As the average household size is higher in Saguenay (6.5 individuals; 40% of households with 2–5 members) than in Trois-Rivières (5.5 individuals; 55.3% of households with 2–5 members), this has an effect on the difference in the global linkage rate between the two regions.

Lastly, we examined the linkage rates among occupational class of the household head.¹¹ As shown on the bottom part of Table 2, occupation of the household head does not seem to impact the linkage rate except for the higher occupation group where we observe slightly lower rates in both areas. Factors that could explain this are either a higher proportion of non-Catholic within that group, or a higher proportion of immigrants (married outside Quebec) or a smaller mean household size.

Once a household is linked to BALSAC, the linkage of its individual members is also attempted. To assess the linkage results at this level, we have looked at linkage rates by sex, age and marital status (Table 3). Both sexes show the same linkage rate. This is very encouraging, since one might expect female's first and last names to be more loosely registered than males', which is clearly not the case here. There are, however, more differences between age groups. They are less marked in the Saguenay region, thanks again to the availability of birth and death certificates, but they are sizeable in Trois-Rivières. There, linkage rates increase with age and this is directly related to the fact that BALSAC families outside the Saguenay region contain only married children. All other things being equal, the older a person, the more chances he or she has to have married and thus be present in BALSAC. After 50, the rate decreases a little due to widowed

individuals. In the Saguenay region, the minor variations in linkage rates are most likely explained by migration. We see a slight decrease in the rates among individuals aged 10–29 which reveals the higher mobility of young adults who as a result have more chances to marry outside Quebec. If they were born outside the region (and thus without a birth certificate in the regional vital records), we would then miss them.

Lastly, looking at linkage rates by marital status, we can again apprehend the direct effect of the availability of birth and death records in Saguenay: in that region, the highest rate is found among married individuals followed by single individuals (who are mostly children and young adults still living with their parents since we do not link single-person households or households with no married individual) and then by widowed individuals. However, the variation in rates is much less important than in Trois-Rivières where the rate is twice as high among married people as among singles.

The linkage of households and individuals to the next census (1852–1901)

Linking to the next census poses some conditions. First of all, the household or the individual must be enumerated within the same area being compared from one census to the other. Local or regional socio-economic context will thus play a significant role in the probability to find someone on the same territory over a decade. Second, despite its plasticity (some members being gone or added during the 10-year gap between the two enumerations), a household must present a minimal consistency to be linked (at least one common conjugal unit must be found in both censuses). Third, it has to be enumerated in such a way that the nominative information can be equally legible in both censuses. This involves a good knowledge of all household members (names, ages, marital status, and so on) by informants and an appropriate transcription of the information by census takers (continuity of personal information: last and first name spelling, ages, place of birth). If the information is too different, we might miss a true linkage because the household in the second census will not be selected by the program as a potential candidate for linkage.

Notwithstanding these difficulties, linkage to the next census was attempted on all processed households in the 1852–1901 censuses that were not connected to the next census through linkage with

Table 3. Proportion of individuals linked to BALSAC by sex, age group and marital status.

	Saguenay		Trois-Rivières	
	Linked (%)	<i>n</i>	Linked (%)	<i>n</i>
Sex				
Male	91.4	48,341	61.6	31,726
Female	91.5	44,776	61.3	33,559
Age group				
0–9	93.3	31,648	50.0	17,345
10–19	89.4	22,129	53.1	14,907
20–29	88.2	14,976	61.4	11,080
30–49	93.6	16,308	78.3	13,409
50 and over	91.0	8,224	74.3	8,385
Marital status				
Single	89.3	61,310	44.1	40,918
Married	96.5	29,715	88.1	21,765
Widow(er)	80.9	2,251	60.1	2,589
Total	91.4	93,295	61.5	65,309

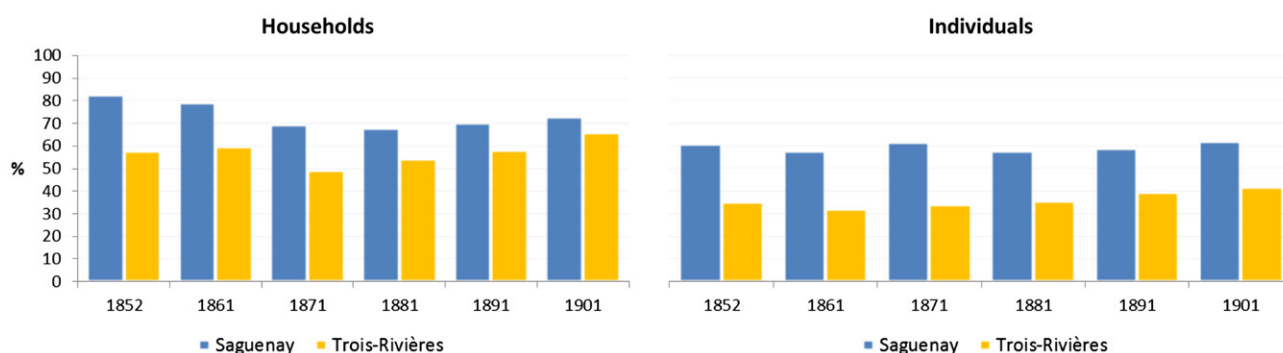


Figure 6. Proportion of households and individuals linked to the next census for the 1852–1901 censuses.

BALSAC. The results presented in the left panel of Figure 6 include all households linked to a household in the next census either through BALSAC or through direct linkage. In Saguenay, more than 99.5% of linkages were done through BALSAC indicating that a tiny proportion of households appearing in two successive censuses did not have a head married in a Catholic parish of Quebec. In Trois-Rivières, the proportion not matched through BALSAC is also very low, between 1 and 5% depending on the pair of censuses. Again this could be explained by the small but not negligible proportion of non-Catholics in this city. As shown in Figure 6, match rates are lower in Trois-Rivières (between 49 and 66% of linked households) than in Saguenay (between 67 and 83%) but interestingly both series display a similar distribution with rates decreasing and reaching the lowest values in 1871 and 1881 and then increasing again. The 1870s and the 1880s are the decades where the emigration out of the province, mostly to the United States, reached its peak, lowering the probability of finding a household in the next census.

As observed in the linkage to BALSAC, in both regions the linkage rates to the next census are lower among individuals than among households. They are also more uniform across censuses (from 57 to 62% in Saguenay and from 32 to 42% in Trois-Rivières). Lower rates can, of course, be explained by death or migration of individuals between censuses. In the case of migration, the effect could also explain in part the differences between the two areas because Saguenay represents a larger territorial unit than Trois-Rivières (region vs. single town) and thus we are most likely to miss short distance moves in the latter. Differences in the composition of the population, such as sex ratio, age distribution, size and types of households could also contribute to the observed differences. And of course, the possibility of relying on birth and death records in Saguenay would also have an effect.

Like previously done for the linkage to BALSAC, matching rates were examined by household type, household size and occupation to see what factors might affect the likelihood of a household being linked to the next census. Results are shown in Table 4 and present the proportions of households linked to the next census.

More importantly than for the linkage to BALSAC but likely for the same reasons, the household type heavily influences the probability of linkage to the next census. Compared to couples, the linkage rate for widowers decreases by 17% in Saguenay and 24% in Trois-Rivières, and by 30% and 40%, respectively, for the widows. These differences are also partly explained by mortality among widowers and widows who as a group are older than married individuals.

Household size plays a major role when attempting linkage of a household in two consecutive censuses. This is largely due to the fact that the loss of household members (by migration, marriage or death) greatly reduces the chances to find the household

Table 4. Proportion of households linked to the next census by household type and size and by occupation of head.

	Saguenay		Trois-Rivières	
	Linked (%)	<i>n</i>	Linked (%)	<i>n</i>
Household type				
Couple	75.6	9,635	63.3	7,597
Widower + child	62.7	327	47.7	310
Widow + child	53.4	335	37.8	686
Other	62.9	151	50.6	156
Household size				
2	58.0	793	40.1	1,111
3	66.3	1,041	48.7	1,188
4	69.9	1,081	60.1	1,232
5	71.2	1,168	62.4	1,222
6–9	78.0	4,527	68.7	3,185
10–19	81.2	1,813	72.3	787
Occupation of head				
Higher occupation	67.6	547	62.1	1,746
Skilled workers	71.4	839	60.8	1,777
Farmers	78.5	5,858	70.0	766
Lower skilled workers	70.1	278	66.5	813
Unskilled workers	68.4	2,008	61.2	2,179
Total	74.3	10,448	60.5	8,749

10 years after. Table 4 shows very clearly that the decrease in success rates is important especially for households counting two or three members. Among larger households, it is likely that migration explains most of the observed variation in linkage rates.

Lastly, linkage rates also vary according to the head's occupation.¹² Farmers, known to be more sedentary, show the highest score. Other groups in Saguenay have somewhat lower and very similar linkage rates. In Trois-Rivières, lower-skilled workers have slightly lower rates than farmers while the three other groups show the lowest rates. The difference between the two areas in the rates for skilled workers may be due to the progressive industrialization process in the city of Trois-Rivières which could have prompted the migration of many craftsmen. However, as a whole, the socioeconomic category displays a modest effect on success rates.

At the individual level (Table 5), the overall linkage rate between censuses corresponds to approximately two-thirds of the linkage rate to BALSAC (66.1% in the Saguenay, 62.0% in Trois-Rivières). As noted for the linkage to BALSAC, the rates are very similar for men and women while age and marital status are important factors of differentiation.

Similar to what we observed in the linkage to BALSAC, the 30-39 age group has the highest linkage results with the difference between Saguenay and Trois-Rivières probably explained by the size of the enumerated territory. The linkage rate among people aged 50 and over is much lower, however, than their linkage rate to BALSAC and this is evidently due to mortality. The decrease in the linkage rate for the 0-9 age group is smaller in Trois-Rivières than in Saguenay (25% compared to 37%). Since in Saguenay many children are linked using their birth certificate, we have a higher probability of losing them because

of death or migration before the next census than in Trois-Rivières where we can only link children who have survived to adulthood and married in Quebec. For the next two age groups, the losses are similar in both areas (34-38%) and likely reflect the effect of migrations, a part of them taking place outside the province. Finally, the general pattern for the age groups is also perceptible when considering the marital status: a better linkage rate for the married (especially in Trois-Rivières), and a strong effect of mortality for widowed individuals.

A special case: The linkage of women across censuses

Linking women across censuses is a major challenge due to the loss of maiden names after marriage. In our data, we see that a substantial share of women are enumerated with their maiden names in the 1852 and 1861 censuses. However in the following censuses, this proportion becomes negligible. This phenomenon needs to be more thoroughly documented but to our knowledge this could be explained in part by the implementation of a new version of the Civil Code of Lower Canada in 1866. Due to the connection with BALSAC, women, just like men, can be located in their family of origin and in their conjugal family allowing us to follow them throughout their life course in all censuses where they appear, even if they adopt the name of their husband after marriage.

Figure 7 shows the distribution of the 31,351 females and 31,389 males who were linked to BALSAC according to the number of censuses where they appear. In both regions and for both sexes, around 50% of individuals are found in only one census. However, half of these are individuals enumerated for the first time in the 1911 census and thus could not have been linked to another census. The proportion of individuals linked to more than two censuses is slightly higher in Saguenay, in line with the higher global rate of linkage. But the most remarkable feature of these results is the similarity between men and women in both regions. In Saguenay, among individuals linked to more than one census, we were able to find 2,595 men and 3,141 women whose marital status changes from single to married. This is also the case for 916 men and 1,056 women in Trois-Rivières. The capacity to trace both women and men points to the fact that our linkage approach overcomes some difficulties related to the nature of data and shows how complementary the two data sources are.

Table 5. Proportion of individuals linked to the next census by sex, age group and marital status.

	Saguenay		Trois-Rivières	
	Linked (%)	<i>n</i>	Linked (%)	<i>n</i>
Sex				
Male	61.3	36,320	38.7	23,922
Female	59.5	33,454	37.6	25,387
Age group				
0-9	64.6	23,643	37.2	13,229
10-19	56.3	16,935	32.4	11,498
20-29	58.6	11,257	37.6	8,218
30-49	66.1	12,109	50.0	10,020
50 and over	47.5	5,973	33.0	6,197
Marital status				
Single	58.7	46,317	32.0	31,263
Married	66.0	21,945	52.2	16,149
Widow(er)	34.2	1,649	19.7	1,903
Total	60.4	69,922	38.1	49,332

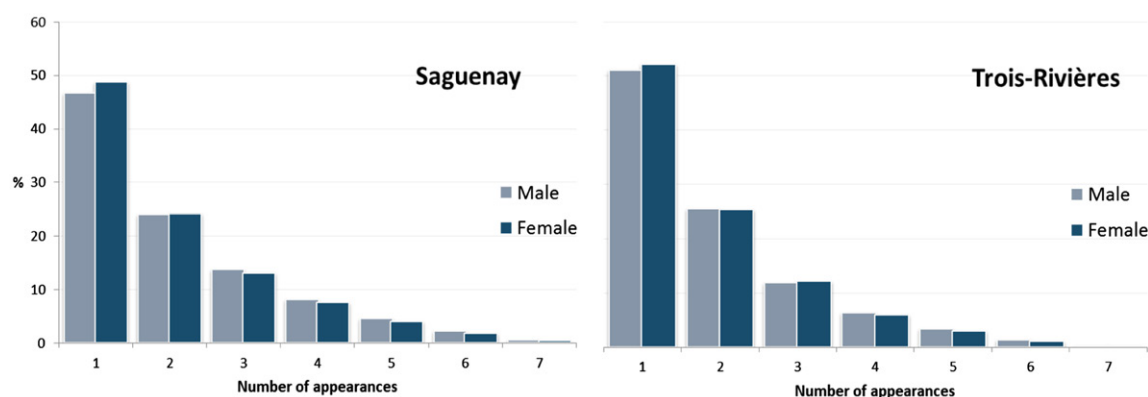


Figure 7. Distribution of census individuals linked to BALSAC according to the number of censuses where they appear.

Discussion

Possibilities and limits of the approach used for linkage

Our linkage approach is based on the comparison of two sets of nuclear families: the first one reconstituted from vital events in the BALSAC database and the second one extracted from households in Canadian historical censuses. It relies mostly on nominative information but contextual information on family members' characteristics is also used. Our objective is to link as many families and individuals from the census population as possible with the most efficient method in terms of reliability and efficiency using a program developed by our group.

Because of the differences in the two data sources, the program does not allow for a perfect match, but rather a match with various degrees of similarity between BALSAC families and census households. Family structures from BALSAC may be incomplete (in fact except for Saguenay they contain only couples and their married children). Census households can contain various types of families (nuclear with one or two parents, recomposed, multigenerational) and include members outside the family. The program can deal with situations where the nuclear family cannot be fully restored as long as the structure is not completely fragmented. The success rate is obviously dependent on the degree of alteration of the structure.

From a census perspective, the linkage with civil records, especially marriages, is of tremendous help to link censuses and build census-based longitudinal data sets. Through the BALSAC family file, it is possible to connect a single male child within his parents' household and the married man appearing as a household head 10 years later. Marriage records do not only confirm links based on individual nominative information, but are also necessary to solve a large number of homonymy problems. In addition, due to the loss of

maiden names after marriage, they are the only source which could allow the linkage of a girl within her parents' household with the woman being a member of a couple in a subsequent census. Results on Saguenay and Trois-Rivières show that our linkage method allows for reconstruction of the life course of both men and women from childhood to adulthood, which is quite unique in a North American context.

Considering couple (or couple and child) mentions as the unit of comparison, preferably to a unit based on individual information, minimizes the negative impact of nominative variations on linkage performance and increases success rates. Using this approach, linkage is attempted on all households with at least two members except those for which it is impossible to compose at least two nominative mentions, and those whose members are all single (such as religious communities, hospitals, etc.). Thus, our linkage procedures yield longitudinal data biased in favor of couples and families. However we have seen that unprocessed households in Saguenay and Trois-Rivières represent less than 5% in the seven censuses 1852–1911.

We use the BALSAC population database as the reference data for vital events. The advantage of using BALSAC is that it covers the whole territory of Quebec for the entire period for which nominal censuses are available and it is fully family-structured. Except for the Saguenay region, it contains at the moment only marriages so families are composed of parents and their married children in most instances. This also plays in favor of more linkage for couples and families. At the individual level, married people will have an advantage over single and widowed individuals. The differences between the three types of marital status are more marked for Trois-Rivières showing the impact of having access to birth and death records, especially for linking children.

Some possible biases could also be brought about by our linkage method. There are as many nominative units of comparison in a family as there are children, therefore the probability of linkage increases with the number of children. Results show that household size has a strong effect as the probability of linkage increases with the number of household members. Moreover, if we do not rely on place of residence for the selection of candidate families, we use it in the calculation of the score to rank the candidates same residence in both sources improves the total score). Thus stable families might have an advantage over those who move.

When a household is linked, whether to BALSAC or to another census, the program enables the linkage of its individual members. However, the correspondence of individuals is more difficult to establish and following individual fates is complex. From one census to the next, for example, a child can either remain in the parental household, or become part of another household as a single or married adult (in the latter case, a girl will then lose her maiden name to take the name of her husband), or be lost to observation because of death or migration. Moreover, because our linkage approach focuses on the selection and extraction of nuclear families, household members who are not part of this type of family unit are less easily included in the process.

Notwithstanding these limits and potential biases, we have obtained very high linkage rates to BALSAC both for households and for individuals. This is the case particularly in Saguenay and to a lesser extent in Trois-Rivières. Of course, since in Saguenay we rely on the complete set of vital events, we expected these higher success rates and they clearly demonstrate that availability of birth and death records allows the program to match more people. However, we have seen that other factors can explain the differences between the two areas, such as the size of the territory and its composition (modest size cities and rural areas in Saguenay vs. mostly urban setting in Trois-Rivières) or some characteristics of the population like a mean household size higher in Saguenay or the presence of a small non-Catholic population in Trois-Rivières. Occupational structure also plays a role since farmers have higher rates of linkage and are far more numerous in Saguenay than in Trois-Rivières while the higher occupation group has lower rates and is more represented in Trois-Rivières.

Emerging research opportunities

The IMPQ will make the following data available for research:

1. all marriage certificates since the implementation of parish registers in Quebec in the early 17th century to 1965 enabling automatic reconstruction of the genealogy of the Quebec population over a period of three-and-a-half centuries;
2. family reconstitution based on birth, marriage and death certificates from the beginning of the 17th century to 1849 allowing researchers to explore the historical demography of families in Quebec over a 250-year period;
3. complete-count census microdata covering two urban settings (Quebec City and Trois-Rivières) and three regions mixing rural and urban environments (Gaspésie, Côte-Nord and Saguenay) linked across seven modern nominal censuses (1852–1911) and to corresponding vital event data.

This vast array of biographical information will permit studies based on individual trajectories situated within families, households and communities and examined from a multigenerational perspective. In the North-American context, such an infrastructure is going to be unpaired at this scope and will substantially broaden and enrich the avenues of research in both the social and the biological sciences.¹³ Moreover, since it is based upon nominative data, it will easily support any linkage operation of other serial sources, such as assessment rolls, city directories, or notarial archives.

More specifically, the IMPQ will make possible detailed studies on a crucial period in the history of the Quebec population (mid-19th century to the first decade of the 20th century) focusing on the evolution and long-term consequences of phenomena, such as cultural diversification, social mobility and intercommunity relationships. During this period Quebec's inhabited space expanded as a result of new agricultural settlements and broadened maritime and forestry activity; at the same time, Quebec was transitioning to an industrial economy, and urbanization was accelerating.¹⁴ One of the most spectacular results of the linkage process will be the capacity to update the too often overshadowed life histories of one half of the population: women.

From a population genetics and bio-demographic perspective, the infrastructure will considerably enrich research on the transmission of biological and sociocultural characteristics, on the genetic diversity in Quebec regional populations and on the factors that have shaped this diversity. These will translate into original studies that can contribute to

a better understanding of the genetic determinants of health and disease.

In conclusion, we have achieved our goal of developing tools and procedures for systematic and automated linkage involving vital records and census data. Thanks to the IMPQ, the research community will have access to original data sets on the Quebec population combining information on the two types of data which is unprecedented. In future projects, we intend to pursue the analysis of the linkage results to situate them in the specific context of the studied populations. We will also look more closely at unlinked households to get a better understanding of their characteristics. Lastly, the software could also be improved. It would be desirable to come to greater automation of decision-making and to further refine the process of individual linkage. However, already in its current state, the linkage program provides an innovative tool with the ability to combine longitudinal information from BALSAC to census data and to create census-based longitudinal data sets.

Notes

1. This collection gives an excellent overview of the challenges surrounding the development of effective data sets to study past populations. Regarding national population databases built from parish population registers and censuses, see for instance (Thorvaldsen et al. 2015).
2. For the MPC databases, see <https://usa.ipums.org/usa/>; for the CCRI: <https://ccri.library.ualberta.ca/enindex.html>.
3. For a recent plea for longitudinal micro-data sets, see (Baskerville and Inwood 2015).
4. For a full description and additional information on the BALSAC database, the reader can refer to <http://balsac.uqac.ca/english/>.
5. The 1881 census is available in the North Atlantic Population Project (NAPP) website (<https://www.nappdata.org/napp/>). Data on Québec City is available in the Population et histoire sociale de la ville de Québec project website (<http://www.phsvq.cieq.ulaval.ca/>). The other data sets were constructed in the course of research projects and are not publicly available at this moment.
6. For Saguenay, family files also include spouses' death records, children's birth as well single children death records.
7. Following the same approach than for linkage to BALSAC, our census-to-census linkage process can hardly be compared to the one most commonly used by other research groups to build longitudinal data from censuses. We think for instance of IPUMS (Goeken et al. 2011) and the group from Guelph University (Antonie et al. 2014).
8. BALSAC contains all Catholic marriages up to 1965 and approximately 20% of non-Catholic marriages.
9. As data are available on seven censuses (1852–1911), we could perform linkage to next census for the 1852–1901 censuses.
10. Again we exclude here single and collective households that were not considered for linkage. They represent about 3.5% of households and about 2% of individuals in all censuses.
11. Occupations from the census were coded in HISCO (van Leeuwen et al. 2002) and classified according to HISCLASS (van Leeuwen and Maas 2011). The five occupational classes presented here are a grouping of the 12 HISCLASS categories.
12. Note that unclassified occupations have been excluded from the calculation, explaining why the global value is slightly different from the mean of the categories.
13. In Canada, the only comparable data set (but not publicly accessible) is the sample of Montreal dwellers gathered by Sherry Olson and Patricia Thornton for the period 1861–1901 and largely supporting their analyses in Olson and Thornton (2011).
14. We mean the territory occupied by sedentary populations of mostly European descent.

Acknowledgments

This project is funded by the Canada Foundation for Innovation, the Québec ministère de l'Enseignement supérieur, de la Recherche, de la Science et de la Technologie as well as the Université du Québec à Chicoutimi, Université du Québec à Trois-Rivières and Université de Montréal. Other members of the IMPQ project include: Danielle Gauvreau (Concordia University); Lisa Dillon, Alain Gagnon, Damian Labuda (Université de Montréal); Richard Marcoux (Université Laval); France Normand (Université du Québec à Trois Rivières); Marc Tremblay (Université du Québec à Chicoutimi). We also acknowledge the support of research professionals, clerks and technical staff especially Mario Bourque, Jocelyn Côté, Dominic Desbiens, Tomy Grenier, Ève-Marie Lavoie, and Jacinthe Sabourin.

References

- Antonie, L., K. Inwood, D. J. Lizotte, and J. Andrew Ross. 2014. Tracking people over time in 19th century Canada for longitudinal analysis. *Machine Learning* 95 (1): 129–46. <http://doi.org/10.1007/s10994-013-5421-0>
- Baskerville, P., and K. Inwood (Eds.). 2015. *Lives in transition: Longitudinal analysis from historical sources*. Kingston, Ontario: McGill-Queen's University Press.
- Bellavance, C., F. Normand, and E. S. Ruppert. 2007. Census in context: Documenting and understanding the making of early-Twentieth-Century Canadian censuses. *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 40 (2):92–104. <http://doi.org/10.3200/HMTS.40.2.92-104>.
- Bloothoof, G., P. Christen, K. Mandemakers, and M. Schraagen (Eds.). 2015. *Population reconstruction*. Cham: Springer. <http://dx.doi.org/10.1007/978-3-319-19884-2>.
- Bouchard, G., and M. Bouchard. 1981. *Ajouts au code de transcription phonétique FONEM*. Chicoutimi: Document de travail BALSAC II-C-66.

- Bouchard, G., P. Brard, Y. Lavoie, and G. Bouchard. 1981. FONEM: Un code de transcription phonétique pour la reconstitution automatique des familles saguenayennes. *Population (French Edition)* 36 (6):1085–103. <http://doi.org/10.2307/1532326>.
- Bouchard, G., B. Casgrain, and R. Roy. 1981. *Tests de validation des fiches de couple par ordinateur*. Document de travail BALSAC II-C-67.
- Bouchard, G., and A. LaRose. 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. <http://doi.org/10.7202/303510ar>.
- Bouchard, G., R. Roy, and B. Casgrain. 1985. *Reconstitution automatique des familles: Le système SOREP*. Chicoutimi: SOREP.
- Bourque, M., B. Casgrain, D. Desbiens, M. St-Hilaire, and H. Vézina. 2013. *Compte-rendu du développement du logiciel de jumelage Recensement-État civil et présentation de l'interface de travail*. Chicoutimi: Document de travail BALSAC I-C-253.
- Charbonneau, H., Y. Lavoie, J. Légaré, and J. Legare. 1970. Recensements et registres paroissiaux du Canada durant la période 1665–1668. Étude critique. *Population (French Edition)* 25 (1):97–124. <http://doi.org/10.2307/1529525>.
- Gauvreau, D. 1991. *Québec. Une ville et sa population au temps de la nouvelle-France*. Sillery: Presses de l'Université du Québec.
- Gauvreau, D., P. Thornton, and H. Vézina. 2010. Le jumelage des recensements aux mariages du fichier BALSAC: Présentation de l'approche et étude exploratoire des enfants de couples mixtes à la fin du xix siècle. *Cahiers Québécois De Démographie* 39 (2):357. <http://doi.org/10.7202/1003590ar>.
- Goeken, R., L. Huynh, T. A. Lynch, and R. Vick. 2011. New methods of census record linking. *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 44 (1):7–14. <http://doi.org/10.1080/01615440.2010.517152>.
- Gossage, P. 1999. *Families in transition: Industry and population in Nineteenth-Century Saint-Hyacinthe*. Montreal: McGill-Queen's University Press.
- Olson, S., and P. Thornton. 2011. *Peopling the North American City: Montreal, 1840–1900*. Montreal: McGill-Queen's University Press.
- St-Hilaire, M. 1990. *Description du format d'impression d'une fiche de couple*. Document de travail BALSAC I-C-96.
- St-Hilaire, M. 2009. Filling the gap: The use of marriage records to help with inter-censuses linkage of young adults in Quebec City (1851–1911). Paper presented at the 2009 RECORDLINK workshop, University of Guelph.
- St-Hilaire, M., L. Richard, and R. Marcoux. 2014. Individual and familial lifecourses in Quebec City, 1871–1911: Some considerations on two biographical data-sets. In *The dawn of Canada's century: Hidden histories*, ed. G. Darroch, 322–58. Montréal/Kingston: McGill-Queen's University Press.
- St-Hilaire, M., and H. Vézina. 2010. Between household and family: the use of marriage records to link census data (Quebec City, 1852–1911). Paper presented at the 2010 RECORDLINK workshop, University of Guelph.
- Thorvaldsen, G., T. Andersen, and H. L. Sommerseth. 2015. Record linkage in the historical population register for Norway. In *Population reconstruction*, ed. G. Bloothoof, P. Christen, K. Mandemakers, and M. Schraagen, 155–71. Cham: Springer International Publishing. http://doi.org/10.1007/978-3-319-19884-2_8.
- van Leeuwen, M. H. D., and I. Maas. 2011. *Hisclass: A historical international social class scheme*. Leuven: Leuven University Press.
- van Leeuwen, M. H. D., I. Maas, and A. Miles. 2002. *HISCO: Historical international standard classification of occupations*. Leuven: Leuven University Press.
- Vézina, H., and M. St-Hilaire. 2011. Entre famille et ménage: Le jumelage des données d'état civil et de recensement dans la population québécoise. In *La fin des recensements*. 79e congrès de l'ACFAS, Sherbrooke, 9–13 mai.