Between household and family : The use of marriage records to link census data

Marc St-Hilaire, Géographie/CIEQ, Université Laval
Hélène Vézina, Sciences humaines/BALSAC, Université du Québec à Chicoutimi

Former exploratory works on census data linkage showed the utility of marriage records to help with filling the gap between the 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's 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. The latter attribute of marriage records and the perspective of being able to elaborate longitudinal data sets from historical Canadian censuses for both male and female individuals led us to conduct further tests on the use of marriage records for the linking of 1852-1911 Quebec City's census data. Ultimately, those tests will serve as a working bench for a project just funded by the SSRHC to develop a methodology to link Quebec City's census data to the Balsac genealogical register. This paper thus aims 1) to present the methodology and the results of the census data linkage, with the help of marriage records, for two cohorts of male and female children taken from 1852 and 1871 censuses, 2) to give a quick overview of what those results show on gendered destinies, and 3) to outline the methodology we expect to develop to link census data to the Balsac register.

1. The linkage: methodology and results

The data used for this test are the 100% sample of Quebec City's Lower-Canadian and Canadian census micro-data, 1852 to 1911 (total population from 45900 to 78700) entered for the aims of the *Population et histoire sociale de la ville de Québec* project (PHSVQ), and the marriages records for the city, 1850-1910 (all denominations; 33700 unlinked certificates, provided by the BALSAC project). The test was conducted for four cohorts based upon age and sex: 5 year-old girls and 10 year-old boys in 1852 (582 girls, 479 boys[1]) and 1871 (726 girls, 822 boys). Each child was searched for, manually but computer assisted, in the four subsequent censuses following a four steps procedure aiming to get information on the destiny of the individual and his/her family:

1. First searched for within a reduced basin of candidates defined by age (expected age ± 4 years), and sorted last name/first name, paying attention to light names variations and common French aliases (Hudon/Beaulieu; Audet/Lapointe, Portugais/Dassylva...);
2. If not found, searched for within the whole city's population, sorted last name/first name, still taking into consideration spelling variations, but loosening the age tolerance.
3. If not found, searched for using the rest of the family, that is looking for one or another family member to locate the entire household and, sometimes, the individual if there had

---

[1] A problem in the application of the linking procedures for the 1852 cohorts makes impossible, at this moment, to report for the boys in that year and only partly for the girls.

been an important name variation and/or age discrepancy. At this stage, the search also used a first name/last name sort.

4.   Finally, found or not, the individual was looked for within the marriage records.

This order was followed for the boys, but steps 3 and 4 were reversed for the girls when expected age was 25 and over since the absence of a woman in a following census is likely due to a loss of her maiden name[2].

Notwithstanding the gender differentiated death rates, and considering that names variations are more likely to affect women (for instance, 4% of spouse's name is even not given for the girls' mothers of the 1852 cohort; about gendered differences regarding nominative information in the civil registers, see BOUCHARD, ROY, 1982), a better linkage rate was expected for boys. But the full results for the 1871 cohort are very similar for both sexes (Table 1): roughly half of the individuals are lost at each ten-year interval prior to reach the age of 25-30, the diminution quickly lowering after[3]. Rates by matrimonial status also show that the linked sample for this cohort is quite representative of the total population by ages, with variations widening with the diminution of the sample size. What is rather striking here is the large proportion of single women among the population, which reflects the very low sex ratio for the city in this period (between 858 and 863‰ for the censuses from 1881 to 1911).

The linkage rate is higher than what we got in previous tests for male cohorts only. It is mostly because, on one hand, the range of ages used to determine the basin of candidates was widened from ± 2 years to virtually the entire population (steps 1 and 2) and, on another hand, the search itself had been widened to the household at step 3 when the individual is not found at previous steps, overcoming the major name changes. And, of course, the linkage rate for the married or widowed women is a net gain on what we had before thanks to the use of marriage certificates. Regarding this latter aspect, marriage certificates were found for 169 men and 166 women of the 1871 cohort, that is respectively 22 and 23%[4]. The proportion for women was expected to be much higher than that for men due to the general practice of celebrating the marriage in the bride's parish. This somehow modest proportion can be explained by 1) the residual proportion of marriages celebrated out of the city limits (which apparently shows by the proportion of single women among the female cohort increasingly higher than the proportion of single women in the entire population in Table 1) and 2) the high level of singlehood among the women. Regarding the age range used to delimit the candidates basin, finally, four years still seems a bit short, especially for the women (Table 2): nearly 6% of them declared an age more than four years from the expected age, a proportion increasing with age.

What precedes worth for the 1871 cohorts. For the 1852 female one, 292 girls were found in 1861, giving a slightly higher rate of 50.3%. The fact that the local economy in the 1850s was running better than in the 1870s is likely the main reason for that score, especially if we take into account that the 1861 census provides with nominative data of a lower quality than the 1881 one.

---

[2] It is worth noting that maiden names were collected for a third of the 1852 couples (mostly for the French Canadians), a practice practically disappeared in 1871 (less than 1% of couples).
[3] If we took into account general mortality rates for the Quebec population at this period (BOURBEAU, LÉGARÉ, 1982), linkage rate for both sexes would be 50% in 1881, 28% in 1891, 22% in 1901 and 16% in 1911.
[4] In addition, 43 men declared married at a census or another could be linked without the help of their marriage certificates.

In a whole, the adopted linkage procedures yield good if not quite high rates, taking into account the entire family information largely helping to improve those scores (Table 3 gives examples of matches hardly done at the individual level). Besides, the use of the household information to locate an individual (steps 2 and 3) or to confirm a link (if there was only one candidate for a link, it was nevertheless checked with the remaining of the family) creates a bias in the linked sample, those living on their own being under-represented. Only two 1871 boys, out of 12 living without siblings with another couple or within an institution, were found in 1881; for the 1871 girls, two out of 9 were found, and three out of 11 for the 1852 girls (total 7/32 or 22%).

2. The use of longitudinal data

Once build, longitudinal data may be used in a wide array of studies, intra- and inter-generational. Follow a few basic results on Quebec City migration dynamics using both 1871 cohorts, comparing household heads attributes (marital status, cultural background, and occupation) according to the sex of the child and the head's migration status (migrant/persistent) from 1871 to 1881, a period during which the child is less likely to control his/her destiny (especially the girls, who get 15 at the end of the decade), but rather shares the parents' one. Given the fact that the persistence is calculated on household's head, the numbers differ from those for the linkage: 46 girls and 53 boys were absent in 1881, but the households in which they were living in 1871 were still present. Results are not discussed here; they will be in another paper to be presented at the Congress of the Humanities and the Social Sciences next week.

2.1 Marital status (Table 4)

Two tendencies are noticeable. The first one, hard to explain without further analysis, is that households headed by women are more likely to move than those headed by men (54% of female-headed households are gone, against 46% of male-headed ones). That may be related to a socioeconomic factor (the widow leaves for another location where she can find better conditions, thanks to family relations or to the local economic or institutional context) or to methodological bias (the widow marries again, but the new couple is not identified). The second one is that there are more women (widows, in fact) heading households with a boy than with a girl. It is partly because household heads with a boy are in average three years older than the heads of households with a girl (44.3 to 41.1), which is normal due to the cohorts ages. It is probably too because those households are more likely to count on male teenagers earnings, lowering the widow's need to remarry.

2.2 Cultural background (Table 5)

In a whole, the cohorts are quite representative of the city's population, Irish Catholics being very slightly under-represented in favour of the Anglo-Protestants. Migration by cohort sex and cultural group shows a few specificities. All together, migrants significantly outnumber non-migrants among the Anglo-Protestants, while it is the opposite (but less markedly) for the French Canadians; Irish Catholics counting for the same proportion as among the migrants, as among the non-migrants. These overall figures roughly reflect the out-migration pattern for the period, each cultural group taking part into it, but at levels moderately differentiated (French Canadians left the demographically stagnating city in slightly less numerous numbers than the other cultural groups; ST-HILAIRE, MARCOUX, 2001). According to the cohort sex, other differences appear. First, French Canadians households with a boy leave at a significantly higher rate than those with a girl while it is the opposite for the Irish Catholics. These results are likely related to the different

ethnic labour markets of the city, the French Canadians being more numerous in the labour-intensive clothing and leather industries (where unmarried women occupy an important role) while the Irish Catholics dominated the harbour-related unskilled jobs (by far mostly male). Further analyses combining ethnicity and occupation would confirm such hypotheses.

2.3 Occupation (Table 6)

As expected, occupational structure differs significantly depending on the migration status, migrants counting a few less skilled workers (manual on not), and many more unskilled manuals workers, than the persistent heads of household[5]. The differences are very similar for both male and female cohorts, showing the independence of the variables occupation/young children's sex. Besides, the high rate of migration among military people is related to the departure of the British garrison from the Citadel in the fall of 1871.

Such rough results reporting only for one decade are indicative, if needed, of the potentiality of census-based longitudinal data, encompassing both male and female lives thanks to marriage certificates. The use of marriage records also open wider perspectives for research by allowing genealogical reconstitution, which gives way to control for kin relations. That is the aim of a recently SSHRC-funded project associating PHSVQ and BALSAC projects. Since we are at the eve of the works, we can hardly go farther than outlining what we expect to do.

3. The perspectives regarding the BALSAC genealogical register

Since the beginning of the 1970s, a team lead by Gérard Bouchard at the Université du Québec à Chicoutimi has been developing a population register based upon the civil certificates (birth, marriage, and death). Previously covering the Saguenay and Charlevoix regions, the scope of the project was extended to the whole province of Quebec in 1989, but since then using only the marriage certificates. The register aims to cover the entire Quebec's population since the beginning of the European settlement to the present (4.5 M certificates, of which 3.7 M marriages). Up to date, 2.0 M marriages and 0.85 M births and deaths were entered, relating to 4.5 M individuals. It currently covers the entire catholic population from 1621 to 1955, and a part of the non catholic population (14% of the certificates are non catholic; 33% of them were entered up to 1900, 10% up to 1950, covering a few regions). Those records were linked together, allowing genealogical reconstitution[6]. The BALSAC register is mainly used for research in human genetics, but also for social sciences (especially demography, history, and geography). Its use in social sciences would of course gain from the information found in census manuscripts, as the census data would gain from the information the BALSAC register contains, which is the core-idea of our project.[7]

What we intend to do is to take advantage of the tools developed at BALSAC and to adapt them to perform linkage with census data. The linkage system developed by BALSAC is founded on

---

[5] There is no comparison between the household heads and the whole city due to the difficulty, at this moment, to build a valuable comparable set of data (based upon age and role within the household).

[6] For a more complete presentation of the BALSAC register and its current development, see BOUCHARD, VÉZINA (2009).

[7] Linking civil certificates and census data is of course not a new idea. It has been done at some rather small scales, whether geographical or temporal (for instance, see ROSENWAIKE et alii, 1998).

the couple mention, made of four elements: man's first and last names, woman's first and last names. Nominative information is processed to overcome light spelling variations (FONEM phonetic code), and use a series of sortings to link from the surest match (four identical elements) to matches involving nominative data affected by more or less important names variations (up to names substitutions; BOUCHARD, 1986). The result is a series of couple records gathering all the certificates pertaining to a unique couple (their own marriage, the remarriage of a surviving spouse when existing, the marriage of their children). The whole linkage system counts on a nominative information of the highest quality (civil registers are legal documents and priests, pastors, and other civil officers usually kept them with great care) and very consistent trough time (very little change in the legal prescriptions on the way to keep the civil records; see BOUCHARD, LAROSE, 1976).

The nature and structure of census nominative data is very different. It has no legal or administrative value, was collected in a short period of time, at ten years intervals by different people (sometimes not fully skilled, and most of them did it only once), and relates to individuals sometimes belonging to other cultural worlds than the enumerator's one, not to mention the way it was provided (was the person who answered the questions well aware of what was asked for). Information is gathered by dwelling, and additional information useful for matching (role in the household, age, place of birth, race) are whether not consistent trough time, whether provided and collected as carefully as the nominative information. Despite that apparently pessimistic portrait, by far the largest part of the collected information remains of good quality and can be processed[8].

We plan to link censuses to BALSAC register one at a time (the smaller dataset to the larger). Prior to the linkage process, census data will be formatted accordingly to BALSAC tools requirements, that is to transform the data into couple mentions in order to "feed" the linkage system (phonetic transformation, matching routines). At the same time, further information on the census names frequencies by sex is going to be gathered to better document the linkage parameters. For the moment, linkage is going to be performed on nominative information with little limitations but for dates. The additional civil records (survival of the spouse, capacity to sign, dates and location of marriages) and census variables (marital status, denomination, literacy) will be used at the validation stage. We expect few perfect matches on the four nominative elements since wife's maiden names are practically disappearing from censuses after 1861. Thus, those are mainly the subsequent sortings dealing with three, then two nominative elements that will be most used.

Once a married couple from the census is linked to a married couple in BALSAC, it also links their unmarried children to those who will eventually marry and who are present in the BALSAC family record. It also removes the remaining census children of that couple out of the stock submitted to linkage process with BALSAC. Once done, the number of unlinked individuals within the census is substantially reduced, increasing the probability to link them through other matching iteration.

Once a census is done, that is that further iteration does not add substantial numbers of new links, the next one is going to be processed, thus linking both censuses trough the BALSAC register. We should also try to link censuses together to get comparative sets of linked data and

---

[8] In addition to the nature of the data and the way it was collected, the data entry itself of course also affect the final quality of the information to be processed. Both PHSVQ and BALSAC projects documented their data entry and validation procedures.

assess performances for both ways of linking. To this end, additional limitations to the data subsets and consistency tests will have to be implemented taking advantage of the manual operations reported here for some children cohorts.

Procedures outlined here will of course yield biased longitudinal data in favour of couples and families, so weights are going to be established to properly analyse Quebec City's society. Nevertheless, the expected data sets are going to be invaluable, enabling unprecedented studies combining a fundamental social unit (family) and the basic economic unit (household). For instance, to which extent culturally mixed couples found in censuses are the fact of certain family practices? How does the family shape the urban residential structure? Which role do kin relationships play in socioeconomic differentiation? In the specific case of Quebec City, such longitudinal data will allow the study of the urban French-Canadianisation process controlling for economic and social factors. And, not the less, it is going to be possible to follow individual biographies for both genders.

REFERENCES

Gérard BOUCHARD, André 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

Gérard BOUCHARD, Raymond ROY (1982) «Les variations nominatives dans les registres paroissiaux du Saguenay». *Annales de démographie historique*, 354-368

Gérard BOUCHARD, Hélène VÉZINA (2009) *Projet BALSAC. Rapport annuel 2008-2009*. Chicoutimi, Projet BALSAC

Robert BOURBEAU, Jacques LÉGARÉ (1982) *Évolution de la mortalité au Canada et au Québec 1831-1931. Essai de mesure par génération*. Montréal, PUM («Démographie canadienne»)

Ira ROSENWAIKE, Mark. E. HILL, Samuel H. PRESTON, Irma T. ELO (1998) «Linking death certificates to early census records». *Historical Methods*, 31-2: 65-74

Marc ST-HILAIRE, Richard MARCOUX (2001) «Le ralentissement démographique». S. Courville and R. Garon, eds., *Québec, ville et capitale*, Québec, Presses de l'Université Laval, 172-179 («Atlas historique du Québec»)

Table 1. Linkage rates by sex and marital status, 1871 cohorts

		Female			Male		
		N	%	% in city	N	%	% in city
1871	5 YO girls/10 YO boys	726	100,0		822	100,0	
1881	Single	348	100,0	99,8	372	95,4	94,6
	Married	0	0,0	0,2	18	4,6	5,2
	Widow/er	0	0,0	0,0	0	0,0	0,2
	Total	348	47,9	100,0	390	47,4	100,0
1891	Single	102	57,6	55,9	70	34,8	30,0
	Married	72	40,7	42,7	127	63,2	67,6
	Widow/er	3	1,7	1,4	4	2,0	2,4
	Total	177	24,4	100,0	201	24,5	100,0
1901	Single	44	37,0	34,3	21	14,1	14,9
	Married	72	60,5	62,1	122	81,9	81,1
	Widow/er	3	2,5	3,6	6	4,0	4,0
	Total	119	16,4	100,0	149	18,1	100,0
1911	Single	26	31,0	24,7	4	4,7	10,6
	Married	56	66,7	69,2	75	87,2	84,1
	Widow/er	2	2,4	6,1	7	8,1	5,3
	Total	84	11,6	100,0	86	10,5	100,0

% in city: based upon total population of expected age ± one year

Source: PHSVQ/CIEQ, micro-données des recensements de 1871 à 1911

Table 2. Age span of linked individuals, 1871 cohorts

	1881				1891				1901				1911				All censuses			
	Women		Men		Women		Men		Women		Men		Women		Men		Women		Men	
Expected age (years)	15		20		25		30		35		40		45		50					
Observed age	N.	%	N.	%	N.	%	N.	%	N.	%	N.	%	N.	%	N.	%	N.	%	N.	%
Exact	189	54,3	177	45,4	72	40,7	80	39,8	38	31,9	58	38,9	33	39,3	38	43,7	332	45,6	353	42,7
± 1 year	130	37,4	152	39,0	54	30,5	62	30,8	46	38,7	43	28,9	24	28,6	20	23,0	254	34,9	277	33,5
± 2 years	21	6,0	42	10,8	20	11,3	30	14,9	14	11,8	25	16,8	4	4,8	13	14,9	59	8,1	110	13,3
± 3 years	5	1,4	6	1,5	14	7,9	9	4,5	6	5,0	8	5,4	6	7,1	2	2,3	31	4,3	25	3,0
± 4 years	1	0,3	9	2,3	6	3,4	11	5,5	3	2,5	9	6,0	1	1,2	4	4,6	11	1,5	33	4,0
± 5 years +	2	0,6	4	1,0	11	6,2	9	4,5	12	10,1	6	4,0	16	19,0	10	11,5	41	5,6	29	3,5
Total	348	100,0	390	100,0	177	100,0	201	100,0	119	100,0	149	100,0	84	100,0	87	100,0	728	100,0	827	100,0
Mean age	14,8		19,8		24,3		29,7		34,0		39,7		43,5		49,5					

Source: PHSVQ/CIEQ, micro-données des recensements de 1871 à 1911

Table 3. Example of matches using the entire household information, 1871 cohorts

1871		1881	
BOYS			
Beaubien	Isaic	Beaulieu	Isaic
Blakeston	W Robert	Blackston	Robert
Bourret	George	Boune	George
Brindamour	Edmond	Brandmoore	Edward
Browning	Lawrence	Brownrigg	Larns
Burns	James	Byrne	Jas.
Chamberlin	F X	Chamberland	Francois
CinqMars	Charles	Cinge	Mars Chs.
De Foy	Louis	Defoy	Degonzajue
Dion	?	Dion	Albert
Float	Pierre	Flood	Peter
Frenette	Chrisostome	Frette	Jean
Heddy	Johnny	Healy	John
Cayley	Patrick	Kiley	Patrick
L'Abbé	Perande	L'Abbe	Alexandre
Dasylva	JJ Godias	Portugais	Gaudiose
GIRLS			
Devaresne	Josephine	Debaraunes	Josephine
Edmonds	Elizabeth	Edwards	Elizabeth
Gingras	Nathalie	Gingras	Ettle
Kirouac	Eugénie	Hironae	Eugenie
Lubbert	Ellen	Hubbard	Ellen
Guéret	Eva	Latulipe	Eva
Lemay	Josephine	Lemay	Zanaide
Marsan	Lumina	Marien	Lumina
Myles	Alice	Miller	Arless
Rome	Mary	Rowen	Mary
Vézina	Clara	Vizina	Emma

Table 4. Ten-year persistence according to household head's marital status, 1871 cohorts

		Household heads, girls						Households heads, boys					
		Women		Men		Together		Women		Men		Together	
	Marital status	N	%	N	%	N	%	N	%	N	%	N	%
Migrant	Single	1	5,6	4	1,3	5	1,5	0	0,0	7	2,1	7	1,9
	Married	0	0,0	308	97,2	308	91,9	4	11,1	314	92,9	318	85,0
	Widow/er	17	94,4	5	1,6	22	6,6	32	88,9	17	5,0	49	13,1
	Sub-total	18	100,0	317	100,0	335	100,0	36	100,0	338	100,0	374	100,0
	% migrant		54,5		45,6		46,0		52,9		45,3		46,0
Persistent	Single	2	13,3	4	1,1	6	1,5	0	0,0	5	1,2	5	1,1
	Married	0	0,0	370	97,4	370	93,7	1	3,1	389	94,9	390	88,2
	Widow/er	13	86,7	6	1,6	19	4,8	31	96,9	16	3,9	47	10,7
	Sub-total	15	100,0	380	100,0	395	100,0	32	100,0	410	100,0	442	100,0
	% persistent		45,5		54,4		54,0		47,1		54,7		54,0
Both	Single	3	9,1	8	1,1	11	1,5	0	0,0	12	1,6	12	1,5
	Married	0	0,0	678	97,3	678	92,9	5	7,4	703	94,0	708	86,8
	Widow/er	30	90,9	11	1,6	41	5,6	63	92,6	33	4,4	96	11,8
	Total	33	100,0	697	100,0	730	100,0	68	100,0	748	100,0	816	100,0

Source: PHSVQ/CIEQ, micro-données des recensements de 1871 à 1911

Table 5. Ten-year persistence according to household heads' cultural background, 1871 cohorts

	Households heads, girls				Household heads, boys				All household heads				City's population (61184)
	Migrant		Persistent		Migrant		Persistent		Migrant		Persistent		
	N	%	N	%	N	%	N	%	N	%	N	%	%
French Canadians	198	59,1	290	73,4	235	63,0	296	67,0	433	61,2	586	70,0	66,4
Irish Catholics	57	17,0	53	13,4	61	16,4	79	17,9	118	16,7	132	15,8	17,4
Anglo Protestants	64	19,1	44	11,1	65	17,4	47	10,6	129	18,2	91	10,9	12,9
Others	16	4,8	8	2,0	12	3,2	20	4,5	28	4,0	28	3,3	3,4
Total	335	100,0	395	100,0	373	100,0	442	100,0	708	100,0	837	100,0	100,0

Source: PHSVQ/CIEQ, micro-données des recensements de 1871 à 1911

Table 6. Ten-year persistence according to household heads' occupation, 1871 cohorts

| | Household heads, girls | | | | Household heads, boys | | | | All household heads | | | |
| | Migrant | | Persistent | | Migrant | | Persistent | | Migrant | | Persistent | |
	N	%	N	%	N	%	N	%	N	%	N	%
Skilled non-manuals	15	5,1	31	8,1	17	5,2	33	8,0	32	5,1	64	8,0
Small entrepreneurs	30	10,3	51	13,3	42	12,7	62	14,9	72	11,6	113	14,1
Unskilled non-manuals	24	8,2	29	7,6	10	3,0	22	5,3	34	5,5	51	6,4
Farmers	2	0,7	2	0,5	2	0,6	2	0,5	4	0,6	4	0,5
Skilled manuals	110	37,7	155	40,4	130	39,4	169	40,7	240	38,6	324	40,6
Unskilled manuals	111	38,0	116	30,2	129	39,1	127	30,6	240	38,6	243	30,4
Sub-total	292	100,0	384	100,0	330	100,0	415	100,0	622	100,0	799	100,0
Unclassified/Military	27		1		11		9		38		10	
Without occupation	15		10		31		15		46		25	
Total	334		395		372		439		706		834	

Source: PHSVQ/CIEQ, micro-données des recensements de 1871 à 1911