Canadian Studies in Population, Vol. 16 (1), 1989, pp. 61-86

POPULATION STUDIES AND GENETIC EPIDEMIOLOGY IN NORTHEAST QUEBEC

Gérard Bouchard
Université du Québec à Chicoutimi, Chicoutimi, Québec, Canada

Résumé — SOREP est un centre de recherches inter-universitaire qui mène des études de population sur les regions du nord-est du Québec. Depuis 1972, les chercheurs ont travaillé sur le développement d'une base de données généalogiques informatisée couvrant la période allant du commencement de la colonisation jusqu'à nos jours. On est en train d'utiliser cette ressource dans le domaine de la génétique afin d'étudier le développement historique des populations régionales à des niveaux variés. A partir des régions du Saguenay et de Charlevoix, cette étude fournit une vue d'ensemble du travail fait en ce qui concerne le fichièr de population elle-même (données, méthodes de jumelage, système de gestion informatique) et les traveaux en cours en histoire démographique et sociale, génétique de population et épidémiologie génétique.

Les populations étudiées ici sont caractérisée par un degré relativement élevé d'homogénéité génétique (du moins en certaines parties du génôme) et par des prévalences ou incidences élevées de quelques maladies rares. L'origine de es traits se trouve dans l'histoire démographique (modèle d'immigration à effet fondateur, haute fécondité, isolement relatif...). L'étude décrit aussi les efforts faits pour développer une approche épidémiologique originale basée sur une combinaison de'études démographiques (reserche de subdivisions génétiques), d'inférence généalogique et d'analyses de marquerurs dans le but de fournire à la prévention et au conseil génétique des informations plus riches. La dimension juridique et éthique de ces travaux est également abordée.

Abstract— SOREP is an Inter-University research center dealing with population issues in northeastern Quebec. Since 1972, researchers have been working on the development of a computerized genealogical database covering the period from the beginning of the white settlements up to the present time. That resource is now being used in the field of social sciences and in the field of genetics, in order to study the historical development of regional populations at various levels. Focusing on the regions of Saguenay and Charlevoix, this paper provides an overview of the work carried out with respect to the database itself (data, computerized record linkage, database management system) and the research being done in demographic and social history, population genetics and genetic epidemiology.

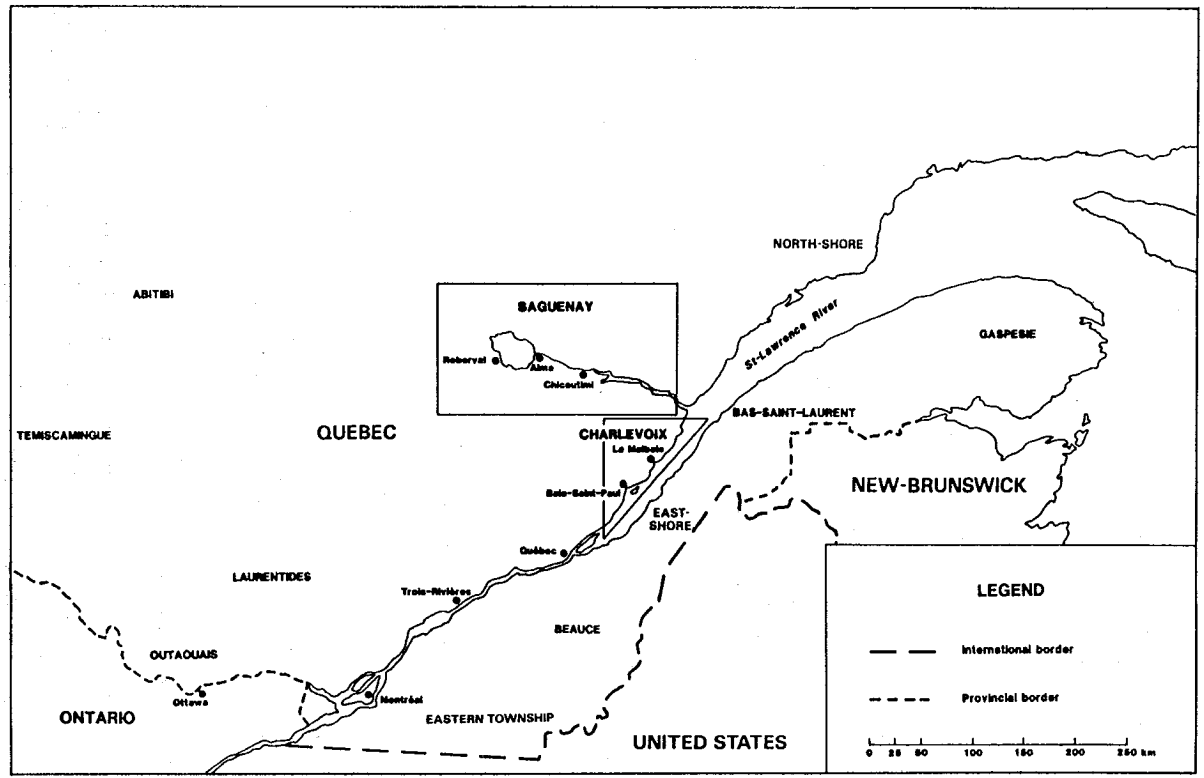
The populations studied are characterized by certain amount of suspected genetic homogeneity and a high prevalence or incidence of rare disorders. The origin of those traits is found in population history (founder-effect migration pattern, high fertility, relative isolation...). This paper also describes attempts to set up an original epidemiological approach based on a combination of population studies (search for gene subdivisions), genealogical inference, and genetic markers in order to provide genetic counselling with enriched information. The legal and ethical issue is also addressed.

Key Words — **Population register, founder effect, genetic epidemiology**

The Populations of Northeastern Quebec

The peopling of Canada by French immigrants began in the 17th century, along the shores of the St. Lawrence river. The process started from the Quebec city area and rapidly expanded upstream and downstream. By the beginning of the 19th century, rapid natural increase had resulted in a scarcity of land in some parts of the St. Lawrence valley. As a response to this, many French-Canadian families had to move to the hinterland and open new settlements. This led to the creation of new regions like Saguenay, Côte-nord, Outaouais, and Mauricie. From 1765 on, the colonization process was fed almost entirely by descendants of the oldest immigrants families, since immigration from France virtually stopped after England took over the colony. In fact, this whole population derives from around 8,000 original founders (Boleda, 1984). Still today, 85 per cent (around 5.5 million) of the Quebec population is Francophone, giving rise to a kind of "macro-isolate" among Anglophones of North America.

The ethnic homogeneity of the Quebec population has been favoured by cultural factors, including the firm Church stand against the so-called "mixed"



MAP 1. LOCATION OF SAGUENAY AND CHARLEVOIX WITHIN THE PROVINCE OF QUEBEC

marriages (outside the Catholic religion). Furthermore, parts of the province suffered from isolation, which prevented people from acculturating. This is particularly true of the northeast area, including regions like Saguenay, Charlevoix, Côte-Nord, Bas-St. Laurent, and Gaspésie. A study of regional surname frequencies illuminates the homogeneity of the northeast, as opposed to the southwest where the populations were more mobile. Also, during the last century, the former area (particularly the region of Montreal) received immigrants of various ethnic backgrounds.

The isolation of northeast populations along with a very high rate of natural increase (around three per cent annually in the 18th and 19th centuries) combined to produce and perpetuate some specific cultural, social and genetic traits. For instance, we find in this area the highest incidence or prevalence ever known for a few Mendelian disorders such as tyrosinemia, agenesis of the corpus callosum, and myotonic dystrophy.

In 1972, we launched a long-term project whose far-reaching goal was to build a computerized population register covering the area from the beginning of the white settlements up to the present time. Today, the BALSAC project (the acronym is made up of initial letters of regions and sub-regions involved) is nearly half-way toward its objectives. The register now contains 700,000 vital records covering the whole Saguenay and Lake St. John, and part of Charlevoix (total population involved: about 350,000). Data are processed through a computerized record linkage software which gives access to family histories and genealogies. Through the years, the project has developed into a pluridisciplinary and inter-university research center (SOREP) supported by University of Quebec at Chicoutimi, Laval University (Quebec City) and McGIll University (Montreal). The center contains about 60 researchers, assistants and technicians, and its works unfold along three lines: a) development and management of the database, b) studies in social dynamics at the regional level, and c) research in population and medical genetics.

This paper is aimed at summing up very briefly what has been achieved with respect to these three fields in terms of methods, issues and findings.[1]

The Population Register

The data consists, first, of a central file which contains identification of virtually every individual who ever lived in the target area (at the present time: Saguenay and Charlevoix). This main file was fed with parish records (660,000 birth, marriage and death certificates for Saguenay and 40,000 for Charlevoix). The database is also made up of auxiliary or peripheral files pertaining to various

specific topics or sub-populations such as patients suffering from genetic and multifactorial disorders, staffing of an aluminum plant, contractors, students, or members of religious orders.

Data are processed through a computer record linkage system devised by SOREP researchers (Bouchard, 1986b; Bouchard, *et al.*, 1985), the output of which is either individual or family histories and genealogies. Basically, this record linkage system is designed to process mentions of married couples, which are taken as the basic units of the linkage work. Subsequently, complementary links within and between family files give rise to individual biographies. Indeed, all information contained in a vital record always relate to a married couple (hereafter: couple mention).

The linkage process as such consists of sorting and pairing these mentions to produce couple or family records. Because of numerous name variations (Bouchard and Roy, 1982a; 1985), the process had to be divided into several steps, each of them dealing with a specific type of name variations, including a complete change of name or surname — which is called substitution. From this, the genealogies are obtained by linking the family records together, which is achieved in two ways: backward, through the marriage of the parents, and forward, through the marriages of the children.

In the first step of the family reconstitution, completeness of the data is verified and minor, strictly orthographic, name spelling variations are eliminated by the use of phonetic devices (FONEM, NETTOIE, INIT). Major name variations are processed in steps 2 (phonetic variations) and 3 (name substitutions). The construction of individual biographies from family histories is performed in step 4. Since they relate to individuals only, rather than married couples, these links are based on variables that are much less discriminant and they give rise to many ambiguities (such as when two or more records compete for the same link). The linkage system provides also for that kind of problem, deciding which records are the most likely to be linked. Finally, these operations, in order to be carried out properly, need to be sustantiated by several routines which are listed on the right side of Figure 1.

This linkage work has been submitted to various validation and consistency checks. We also have conducted numerous accuracy and efficiency measurements (Bouchard and Roy, 1982b). Overall, according to the most recent evaluations, the percentage of good links is well over 99 per cent after step 3 (dependable efficiency measures are very difficult to establish after step 4, but some are currently being carried out).

Data are accessed through a relational Data Base Management System (INGRES), running on a DEC computer (VAX-8350). By use of this software, family histories can be linked "vertically," which allow for genealogical queries.

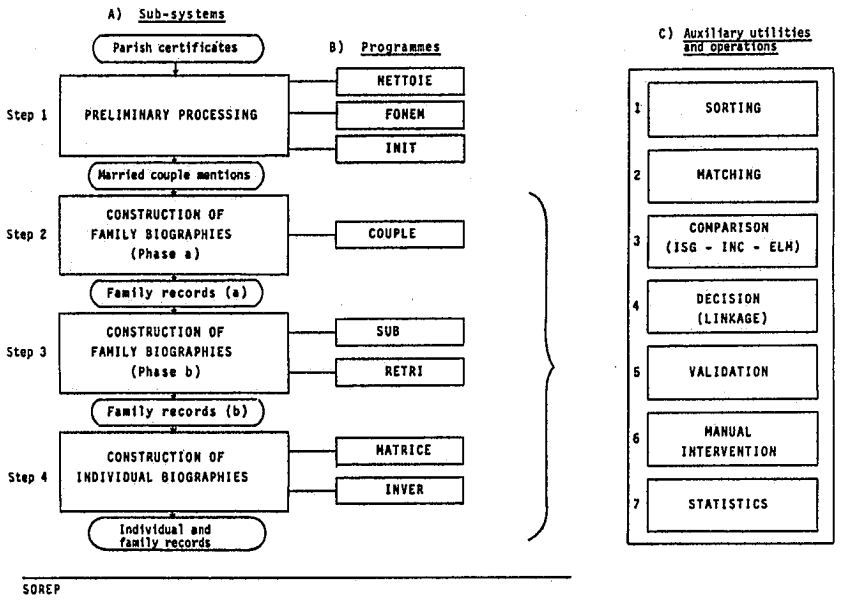


FIGURE 1. CHART OF THE SOREP SYSTEM

Several research projects are underway within the three components of the center: a) record linkage and data management, b) studies in social change at the regional level, c) research in population genetics and genetic epidemiology. Each of them takes advantage of the time-depth provided by the register and of the invaluable opportunity to study regional populations from their very beginning up to the present time. We will introduce first, very briefly, the social component and then, in more detail, the genetic aspect of the research.

Social Change

Work on this side has focussed heavily on the Saguenay region. It has been carried out in its own right by historians and sociologists but the researchers always have kept in mind the interactions between the social processes (such as economic and cultural.) and the evolution of the gene pool. In this regard, it is clear that the basic question to be addressed is the cultural homogeneity of this

society and whether it applies as well to the gene pool, as suggested by the high incidence of rare recessive disorders.

As said above, starting from Quebec City by the end of the 17th century, French settlement stretched eastward along the north shore of the St. Lawrence river and expanded into the Charlevoix region. Traditionally, people living in this remote area hardly made a decent living from various part-time occupations like fishing, farming, timbering and household industries. Other ventures like maple syrup production and fruit picking were also frequent. Overall, because of a lack of communication facilities other than the river up to the 20th century, the rural economy was poorly integrated with the extra-regional market. Unfavourable economic conditions, however, did not prevent a sharp population growth. Between 1765 and 1850, total population went from 1,000 to 13,000, experiencing during the same period a net-migration loss of nearly 3,000. This implies a natural annual growth rate of 3.2 per cent.

Charlevoix rural ecology could not absorb such an increase and, very soon, a land shortage triggered an emigration process. Most emigrants went to the Saguenay area, further north, to start new settlements and create a whole new society in the wilderness. While Charlevoix population, society, and culture changed very slowly up to the middle of the 20th century and never experienced urbanization, the Saguenay region took a rather different course. In the first decades, everything repeated the pattern of Charlevoix: isolation, high fertility (an average of 11 births in completed families, with an *Ig* index at some point higher than Hutterite's — see Chapter 6 of Pouyez *et al.*, 1983), fast population growth (current population: near 300,000), frontier economy, strong family bonds, pervasive influence of priests and religion, and no clear sign of birth control before 1920-1930 (the birth rate was still around 50/100 at this time) thanks to the wilderness which meant available land for numerous offspring.

Very soon though, the industrial and market economies were introduced in the area. Since 1840, winter timber work for Montreal and Quebec-based companies was a widespread occupation among farmers, settlers and labourers. From the end of 19th century onward, big pulp and paper factories were built in Chicoutimi, Jonquiére, Kénogami, Port-Alfred, etc. These were complemented by aluminum plants in Arvida and Isle-Maligne. This process gave rise to a dozen small cities throughout the region, typical of the "milltown" pattern.

In the meantime, starting in the 1880's, the dairy industry developed in the countryside. Small butter and cheese producers were active in every parish, selling their goods on the British market through Montreal brokers. Nevertheless, modern industrial activities and market involvement did not seem to exert a strong impact on local culture and society, which remained quite unchanged

until the Great Depression. At that time, many factors combined to break up the old order.

One such factor originated in the countryside where famillies experienced a shortage of new land for establishment of children. Progressively, a new pattern of social reproduction unfolded, based on schooling and leading to urban occupation and family limitation. In the cities, where the Second World War stimulated production and growth, people took advantage of the hydro-electric power of the Saguenay river and its affluence.

Together, these developments culminated in the radical climax of 1960s (the so-called "quiet revolution"). Across the Saguenay region and the whole province of Quebec, people turned their back on the past and shed the traditions that had for so long characterized the French-Canadian society. In a few years, the Church was stripped of most of its social and political privileges, and Quebec families became — as they still are today — the most contraceptive across Canada.

Several research projects are underway within SOREP, which aim to clarify the demographic, socioeconomic and cultural processes underlying this change.[2] A major question seems to be why they have not occurred earlier. A second one, pertaining to the demographic patterns and perhaps more to the point here, is: to what extent have they affected the gene pool itself and the risk of genetic disorders, particularly those which are specific to this population?

Population and Medical Genetics

The latter question, and others related, are dealt with through research projects carried out at different levels, some very fundamental, some more applied. Actually, with the help of several universities and medical institutions, SOREP would like to contribute to the creation of an original population approach in genetic epidemiology. This approach would be characterized by the combination of population studies, genealogical inference and genetic markers in order to provide genetic counsellors with richer information. Hence a four-step model that can be outlined as follows:

1. Through population studies (demography, history, genetics, anthropology and so on), to detect any clustering or segmentation in the gene flow and the population structure.

   Those studies focus on migration, fertility (trends and variance), mating patterns, family dynamics, patterns of settlement, mortality rates, surname distribution, etc.

2. Using genealogical inference, to calculate the relative risk for any particular individual to be a carrier of any defective gene.

   This kind of calculation has been pioneered by scholars like C. Cannings (Cannings, *et al.*, 1978) and E.A. Thompson (1986). It consists of analysis of genealogies in order to estimate the probability of the transmission of any gene to the descendants in a pedigree. Ideally, it provides a likelihood or a relative risk coefficient for any individual related to a particular set of ancestors. Such a procedure is being implemented on the SOREP database.

3. Following procedures that still have to be developed, to combine information obtained from genealogical inference and from gene markers in order to support genetic counselling.

   Since most of the markers that are now available are linked-markers, our approach would lead to a combination of two types of likelihood, strengthening each other. If the marker involved is the deleterious gene itself, then genealogical inference may prove useful in other ways, for instance by targeting subgroups at risk — without, of course, impinging on individual and family rights of freedom and privacy.

4. The last step does not pertain to research as such. It consists of all appropriate services to be set up in order to meet population's needs in this regard.

As described above, steps 3 and 4 have not been addressed directly yet, although they are next on the agenda. Step 2 is underway and will not be reported on here. The following will be devoted to the main findings of population studies carried out primarily in the last three years.

Two Regions, One Gene Pool?

Because of the very close relationships they have developed since the 19th century, the Charlevoix and Saguenay populations share many genetic traits. Between 1840 and 1880, about 80 per cent of the settlers coming to the Saguenay area originated in Charlevoix. Moreover, at least 70 per cent of those settlers were siblings or related as first and second cousins (Roy, *et al.*, 1988). Finally, we have been able to show that most of the Charlevoix surnames and parishes contributed to the emigration stream according to their relative weight or size (Bouchard, *et al.*, 1987). In a manner of speaking, the old gene pool has been reproduced in the new area, which is illustrated today by some Mendelian autosomal gene carrier frequencies in the two regions (Table 1). The conditions in which these population have developed (isolation, high fertility ratio along with little immigration, regional endogamy) largely account for these traits,

TABLE 1. MAIN MENDELIAN DISORDERS IN THE SAGUENAY AND CHARLEVOIX REGIONS

A. DISORDERS SPECIFIC TO SAGUENAY-CHARLEVOIX

1 - Recessive

a) Tyrosinemia
(Carriers: 1/15 to 1/20)

b) Vitamin D dependency rickets
(Carriers: similar to tyrosinemia)

c) Spastic ataxia
(Carriers: 1/14 to 1/27)

d) Agenesis of the corpus callosum
(Carriers: 1/20 to 1/28)

2 - Dominant

Myotonic dystrophy (Steinert disease)
(Prevalence: 1/480)

- Cases in Saguenay: 600
- Cases in Charlevoix: 63 (estimation)

B. NON-SPECIFIC DISORDERS

1 - Haemochromatosis
(Carriers: 1/40)

2 - Cystic fibrosis
(Carriers: 1/23)

3 - Multiple endocrine neoplasia, Type II-A (Sipple's syndrome). Hereafter: MEN

- Cases in Sageunay and Charlevoix: under study.

Note: Also present in these populations: cases of Tay-Sachs, multiple intestinal atresia, late onset ptosis and dysphagia.

raising significantly the probability for any deleterious gene to be transmitted and reproduced in many copies.

As a result, the Charlevoix and Saguenay populations show a fair amount of social and cultural homogeneity. In Charlevoix, the 15 most common surnames account for 42 per cent of the population, which is the highest proportion among all Quebec regions (Bouchard, *et al.*, 1985), the Saguenay ranking second. In the latter case, the situation has not changed radically since the 19th century (Table 2). In any decade between 1842 and 1971, the relative weight of the 15 most frequent surnames was never less than 35 per cent. Above all, genetic homogeneity is witnessed by the fact that, despite the high incidence of autosomal recessive disorders, consanguineous marriages are not very frequent. The coefficient of consanguinity *(F)* has been calculated for six case groups and 18 control groups. It appears that the *F* values are rather low[3] and the differentials between cases and controls are not so striking, except for haemochromatosis (Table 3). This statement also holds the Phi values, directly calculated from genealogies. Along the same lines, analysis of frequency of consanguineous marriages through dispensation records does establish that the Saguenay rates were relatively moderate over the period 1842-1971, the peak being 22.0 (*F* x 10-4) at the turn of the century (Gradie, 1986).

Founder Effect

For no Mendelian disorder has it been possible to trace Saguenay patients back to one, two or a very small group of immigrants. For every disease, there are quite a large number of founders who seem to have introduced the defective gene, again most of them coming from Charlevoix (Bouchard *et al.*, 1988). Every one of the six groups analysed (Table 5) exhibits a very low value of Phi. Actually, each "proband" in those groups was related to an average of 11 founding couples and among the latter, only 15 per cent were related to more than one patient. This finding suggests that the deleterious genes were already frequent among the first generation of settlers, a conclusion that has been validated through a very simple simulation. Knowing the parameters of the population growth in the Saguenay and the current frequency of carriers for each recessive disorder, we have been able to estimate the number of initial carriers that would account for their number in the present-day population, assuming no selective advantage or disadvantage associated with the heterozygotes (Bouchard *et al.*, 1984). For a disease like tyrosinemia (carriers frequency in Saguenay: about 1 per 15-20), the minimum number is 350-400. If we extend this estimation to the six recessive disorders appearing in Table 5, we are led to

TABLE 2. FREQUENCIES OF THE MOST FREQUENT SURNAMES, SAGUENAY, 1842-1971

Surnames (a)	Decades													Total for the whole period	
	1842-51	1852-61	1862-71	1872-81	1882-91	1892-01	1902-11	1912-21	1922-31	1932-41	1942-51	1952-61	1962-71	N. A.	%
TREMBLAY	10,42	10,66	10,82	12,20	11,72	10,83	11,08	10,34	9,34	9,96	9,05	8,76	8,18	16587	9,27
BOUCHARD	4,78	3,37	4,78	4,39	4,04	3,72	3,90	4,01	3,31	3,17	3,33	3,41	3,11	6108	3,42
GAGNON	2,45	2,76	4,01	3,95	3,43	3,35	3,72	3,09	3,22	3,28	3,06	3,08	3,17	5724	3,20
SIMARD	4,78	4,68	3,88	4,51	3,59	3,37	3,67	3,66	3,26	3,15	2,73	2,96	2,89	5567	3,11
GIRARD	3,31	2,45	3,41	2,75	2,69	2,84	2,66	2,76	2,17	2,59	2,48	2,63	2,52	4598	2,57
FORTIN	2,82	3,30	2,84	2,41	3,02	3,45	2,65	2,87	2,56	2,57	2,27	2,45	2,12	4388	2,45
LAVOIE	1,72	2,76	2,20	2,63	2,27	2,38	2,41	2,16	1,87	2,03	1,95	1,97	2,15	3731	2,09
COTÉ	1,72	1,61	1,03	1,73	2,17	2,09	1,65	2,01	1,91	2,03	1,88	1,86	1,82	3350	1,87
GAUTHIER	1,47	2,22	1,47	1,91	1,87	1,97	1,69	1,69	1,74	1,52	1,56	1,37	1,68	2875	1,61
LAROUCHE	0,00	0,31	1,03	1,05	2,01	1,55	1,68	1,74	1,54	1,65	1,62	1,54	1,49	2777	1,55
GAGNE	3,55	1,46	2,28	2,13	1,44	1,34	1,50	1,60	1,33	1,28	1,47	1,41	1,37	2575	1,44
BERGERON	0,49	0,61	1,59	1,39	1,39	2,04	1,43	1,43	1,76	1,56	1,16	1,36	1,33	2495	1,39
HARVEY	0,74	1,30	1,21	1,70	1,25	1,60	1,21	1,36	1,12	1,11	1,21	1,13	1,04	2085	1,17
SAVARD	2,21	2,38	1,85	1,48	1,44	1,31	1,09	1,18	1,17	1,32	1,08	0,93	1,17	2072	1,16
BOIVIN	1,59	1,46	1,29	1,70	1,42	1,39	1,79	1,26	0,99	1,07	1,09	0,99	1,02	2009	1,12
TOTAL	42,05	41,33	43,69	45,93	43,75	43,23	42,13	41,16	37,29	38,29	35,94	35,85	35,06	66941	37,42

(a) Order based on the frequencies of 1842-1971.

(SOREP)

TABLE 3. COEFFICIENTS OF CONSANGUINITY (F) AND KINSHIP (PHI) FOR SOME AUTOSOMAL RECESSIVE DISORDERS, CASE CONTROL GROUPS, SAGUENAY

Groups	Number of "probands"	F x 10^{-4}	Cases of common ancesters in ascendances	Phi x 10^{-4}
RICKETS (VIT. DEPEN.)				
- Cases	25	0,0	0	5,8
- Controls I	25	14,1	3	1,5
- Controls II	25	19,1	3	1,4
- Controls III	25	12,5	1	1,2
SPASTIC ATAXIA				
- Cases	65	13,2	9	5,4
- Controls I	65	2,1	2	2,6
- Controls II	65	0,3	2	1,1
- Controls III	65	2,5	2	1,6
AGENESIS (CORP. CALL.)				
- Cases	58	18,2	9	3,8
- Controls I	58	8,9	3	3,0
- Controls II	58	4,1	3	3,4
- Controls III	58	4,4	4	2,9
CYSTIC FIBROSIS				
- Cases	55	9,1	4	4,5
- Controls I	55	3,0	3	1,8
- Controls II	55	3,8	4	1,8
- Controls III	55	1,6	2	2,6
TYROSINEMIA				
- Cases	68	16,1	6	4,4
- Controls I	68	6,2	4	1,7
- Controls II	68	18,8	7	2,6
- Controls III	68	7,0	4	1,5
HAEMOCHROMATOSIS				
- Cases	18	83,0	4	10,1
- Controls I	18	0,0	0	0,1
- Controls II	18	2,1	1	4,7
- Controls III	18	0,0	0	0,6

conclude that about one-third of the immigrants between 1842 and 1971 carried one of those mutant genes.

Another research project focused on an autosomal dominant disorder (mytonic dystrophy or Steinert disease). The main findings (to be published soon) may be summed up as follows: a) over 60 founders may have introduced the gene in the Saguenay (current prevalence 1 per 480); b) most of these

immigrants came from Charlevoix; and c) evidence suggests that the affected couples might have been able to reproduce almost normally and transmit many copies of the gene to descending generations, a finding which is not in accord with most of the literature on the subject (Bouchard *et al.*, 1988b).

Whatever the disease considered, it is clear that the bulk of the carriers came from Charlevoix. Therefore, it appeared logical to address at this larger scale the founder-effect issue. To that end, we have used the Saguenay genealogies of tyrosinemia, vitamin D-dependent rickets, and a control group. Those genealogies covering the period 1840-1980 were reconstructed further back to the 17th century, up to the very first Canadian immigrants (Bouchard *et al.*, 1988a).

This research revealed that a very small cluster of ancestors were related to every one of the Saguenay cases of tyrosinemia and rickets, and to every control as well, as though a handful of founders had transmitted all sort of genotypes in the Charlevoix and Saguenay gene pool. Thus, trying to relate one recessive gene or trait back to one particular very remote ancestor is a very uncertain task.

Another result of the same study suggests a kind of regional fragmentation of the whole French-Canadian founder effect. Indeed, it seems that diseases that are currently specific to one region or macro-region in the province of Quebec

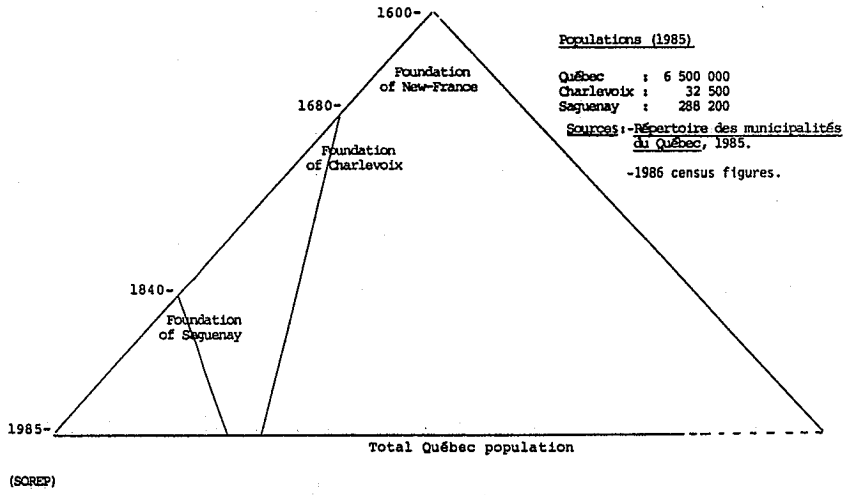


FIGURE 2. TRIPLE FOUNDER EFFECT, FRANCE/NEW FRANCE (QUEBEC) CHRLEVOIX/SAGUENAY (17th-20th CENTURIES)

are related to particular, specific sets of 17th century ancestors, thus revealing distinct subsequent migration patterns among those first immigrants.

Other studies along the same lines are underway within SOREP, focusing on the formation of Charlevoix population, its initial settlement and reproduction pattern as well as the migration dynamics that fostered the opening of the Saguenay region. The overall goal is to highlight what could be called the triple founder effect that led to the creation of the Saguenay population: first, from France to the St. Lawrence Valley; second, from there to Charlevoix; third, from Charlevoix to Saguenay.

Evolution of the Saguenay Gene Pool

Of course, the central question here is whether each of these three steps resulted in increased homogeneity in the corresponding gene pool. In this regard, as far as Saguenay is considered, a set of hypotheses may be put forward, assuming that its gene pool is made up of three components:

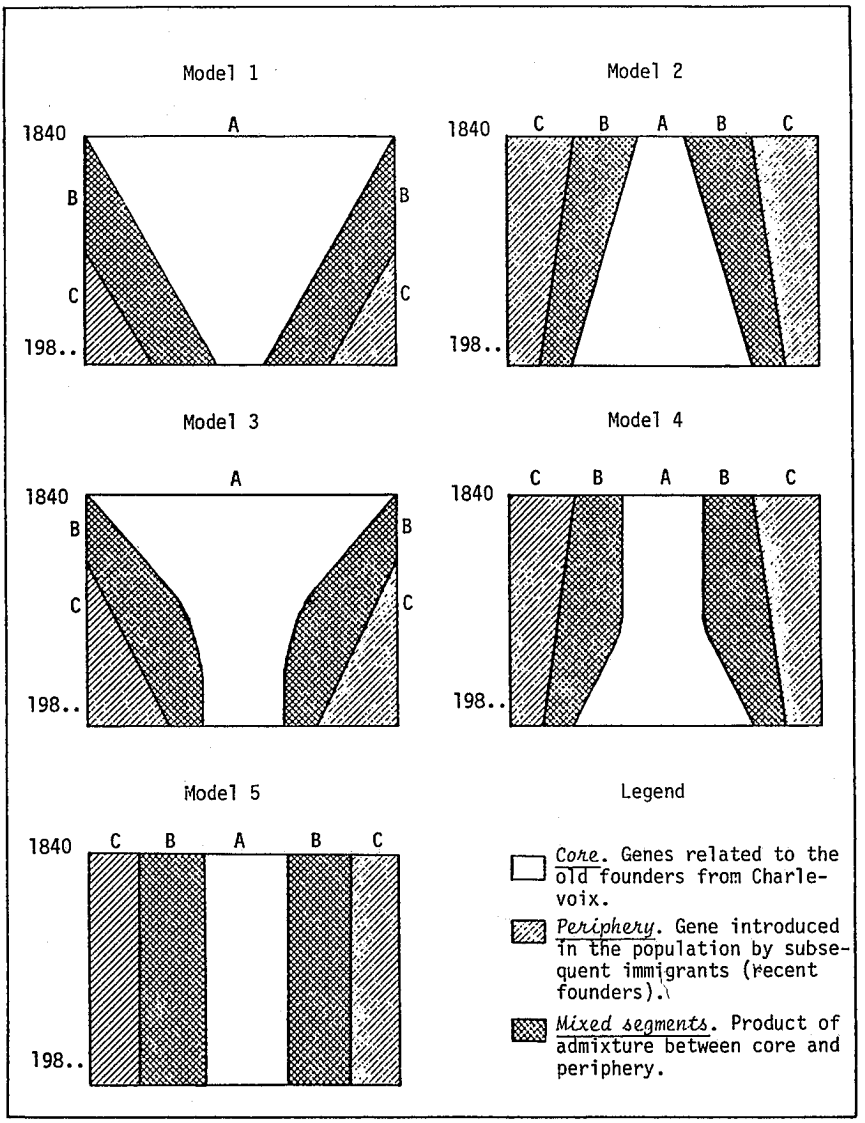
a) old genotypes originating in the Charlevoix founder effect (the "core");

b) genotypes brought in by subsequent immigration from other regions (the "periphery");

c) other genotypes that, through inter-marriage, are a mixture of a and b ("mixed segments").

These three components may have combined in various ways according to the population dynamics upon which they depended. Figure 3 illustrates in a very rough way some of the possible combinations. At this time, from what we know of the history of the Saguenay population, models 1 and 3 are the most likely to be found, while models 2 and 4 are the most pessimistic with regard to genetic epidemiology.

At the present phase of our research, evidence suggests that a fair amount of genetic homogeneity still exists, although a diversificatiton process is at work. Homogeneity, first, is witnessed by a high incidence of recessive disorders coinciding with relatively low coefficients of consanguinity (*F*). Second, as mentioned above, homogeneity is suggested by the evolution of the surname pool (see Table 4). In the same vein, it is also useful to note that during the formative years (1842-1881), all of the most frequent Charlevoix surnames contributed heavily to the emigration stream toward Saguenay. As a result, the Saguenay and Charlevoix name pools were almost identical in those years (Table 4).



(SOREP)

FIGURE 3. DYNAMIC VIEWS OF THE SAGUENAY GENE POOL (1980-198...)

TABLE 4. FREQUENCIES OF THE MOST FREQUENT SURNAMES, CHARLEVOIX AND SAGUENAY, 1842-1881 (%)

Surnames (1)	Region	Decades				1842-81	
		1842-51	1852-61	1862-71	1872-81	%	N.A.
1. TREMBLAY	Charlevoix	16,1	12,4	17,2	16,8	15,6	1548
	Saguenay	10,4	10,7	10,8	12,2	11,3	870
2. BOUCHARD	Charlevoix	4,7	6,2	7,6	6,6	6,2	615
	Saguenay	4,8	3,4	4,8	4,4	4,4	336
3. SIMARD	Charlevoix	4,4	4,6	4,0	5,1	4,6	453
	Saguenay	4,8	4,7	3,9	4,5	4,4	336
4. LAVOIE	Charlevoix	2,6	3,0	3,3	3,6	3,1	310
	Saguenay	1,7	2,8	2,2	2,6	2,4	186
5. GAGNON	Charlevoix	2,5	2,8	3,3	3,1	2,9	289
	Saguenay	2,4	2,8	4,1	4,0	3,6	277
6. FORTIN	Charlevoix	2,2	1,8	1,6	1,6	1,8	179
	Saguenay	2,8	3,3	2,8	2,4	2,7	210
7. GIRARD	Charlevoix	2,0	2,6	2,5	2,3	2,3	232
	Saguenay	3,4	2,4	3,4	2,8	3,0	227
8. PERRON	Charlevoix	2,0	2,0	1,7	1,5	1,8	179
	Saguenay	0,2	0,9	1,2	0,9	0,9	69
9. DUFOUR	Charlevoix	1,9	1,8	3,4	3,2	2,6	254
	Saguenay	1,1	0,8	1,3	1,2	1,1	87
10. DUCHESNE	Charlevoix	1,6	1,1	1,9	1,7	1,6	158
	Saguenay	0,7	0,5	0,5	0,7	0,6	43
11. MARTEL	Charlevoix	1,5	2,2	0,6	0,6	1,3	125
	Saguenay	0,6	1,2	1,6	1,0	1,2	90
12. GAUTHIER	Charlevoix	1,4	1,7	1,3	1,1	1,4	138
	Saguenay	1,5	2,2	1,5	1,9	1,8	137
13. GAUDREAULT	Charlevoix	1,4	1,2	1,6	0,8	1,2	124
	Saguenay	0,9	1,1	1,4	0,9	1,1	83
14. SAVARD	Charlevoix	1,4	2,2	1,7	2,3	1,9	189
	Saguenay	2,2	2,4	1,8	1,5	1,8	140
15. COTE	Charlevoix	1,3	1,5	1,5	1,9	1,6	154
	Saguenay	1,7	1,6	1,0	1,7	1,5	115
TOTAL	Charlevoix	47,0	47,1	53,2	52,2	49,9	4947
	Saguenay	39,2	40,8	42,3	42,7	41,8	3206

(1) The order of the surnames (1 to 15) corresponds to the frequencies of the decade 1842-1851.

(SOREP)

However, the figures in Table 4 also signal a slight diversification since the frequency of the 15 most common surnames, altogether, fell from 41.2 per cent to 35.1 per cent in the periods 1912-21 and 1962-71, a tendency that has still accelerated in more recent years. Change in the migration pattern is the key factor underlying this process. From 1930 on, immigrants were no longer settlers from Northeast Quebec looking for cheap land in a frontier area. Rather, they came from every part of Quebec — cities as well as countryside — and they moved to take up jobs in industry and service sectors. Of course, the decline in the most frequent surname percentage is not dramatic and the Saguenay picture is not to be compared with France's, for instance, where the five most frequent surnames account for no more than one per cent of the population, against 21.6 in Saguenay (Tesnieres, 1980). At the same time however, it is far from the Amish figures showing that the 15 most common surnames gather up to 95 per cent of the population (McKusick *et al.*, 1978, papers A1, C1, D9).

As the reader might have noticed, the foregoing pages focus heavily on immigration, the role of which is so prominent in the formation of a new population and a new gene pool. Clearly, emigration also took place at a high rate, but since we have no evidence yet of a selection in terms of gene carriers, this topic has not been a first priority on our research agenda.

Finally, one has to take into account the effect of contraceptive practices that have spread out massively between 1960 and 1970. The Quebec birth rate, which used to be the highest among Canadian provinces before 1960, ranked the lowest after 1970, where it remains today (for Saguenay, Québec and Canada gross birth rate trends, see Table 5).

Of course this effect has to be clarified, which will be done in a forthcoming research project. But it seems doubtless that a sharp reduction in fertility and a new approach to pregnancy are bringing major changes to some old genetic patterns.

Tracing the Defective Genes in the Population

The sketches appearing in Figure 3, however rude or approximate, help raise pertinent questions about the gene pool with regard to prevention of the disease: what means do we have to identify the "core," that is, the kindreds within which most of the Charlevoix genes have segregated? What is the size of this core? How has it evolved since the 19th century? More specifically, how have migrations and recent changes in demographic behaviors affected gene frequencies and overall risk of disease?

First, following the four-step approach outlined above, various population studies have been carried out in order to capture some clues or hints of possible

TABLE 5. GROSS BIRTH RATE PER DECADE
SAGUENAY, QUEBEC, CANADA, (0/000).

(in brackets: sources)

Decade (4)	Saguenay	Quebec	Canada
1844-1850	67,2 (1)		
1851-1861	61,4 (1)	50,0 (1)	45,6
1861-1871	54,9 (1)	47,5 (1)	45,0
1871-1881	50,9 (1)	44,5 (1)	42,2
1881-1891	48,2 (1)	43,0 (1)	37,5
1891-1901	48,7 (1)	41,0 (1)	35,9
1901-1911	55,2 (1)	41,0 (1)	35,2
1911-1921	51,7 (1)	40,0 (1)	32,5
1921-1931	53,1 (1)	33,5 (1)	25,2
1931-1941	42,5 (1)	29,5 (5)	22,1
1941-1951	43,9 (1)	29,3 (5)	24,8
1951-1961	37,2 (1)	29,0 (5)	27,8
1961-1971	22,2 (1)	20,0 (5)	20,6
1971-1981	18,2 (1-2-3)	14,9 (2-3-5)	15,6 (2-3-5)

(1) C. POUYEZ et al. (1983): Table 6-13 p. 277 and Annexes C-1, C-2, C-3.

(2) Statistiques de l'état civil (Ottawa), Catalogue 84-204 (1979-1980) et Catalogue 84-542, Tableau 1.1 (1981).

(3) Statistique Canada: Estimations annuelles intercensitaires de la population des divisions de recensement, 1976-1981, (1981), Catalogue 91-521.

(4) J. HENRIPIN (1968), Tableau B-6, p. 370 (used as a control only since the time periods are different).

(5) Statistiques de l'état civil (Ottawa), 1973, 1976: volume I, Catalogue 84-204.

genetic subdivisions. However, the evidence provided so far by these studies has not yet permitted us to pinpoint any clear gene concentration asssociated with an elevated risk of disease. The study of consanguinity through church dispensation records (Gradie, 1986) has not shown a significant spatial distribu-

tion. In another direction, the construction of parent-offspring matrices was used to study the drift from one generation to another between 1842 and 1911 (Gradie, *et al.*, 1988). The research concluded that there is very little drift, due to the high level of in-migration. This finding has been validated by isonomy studies.

Therefore, it is not surprising that for all of the recessive disorders, we observe an almost random spatial distribution of the affected families (Table 6). Actually only the micro-region of Bas-Saguenay stands out with a clearly higher score. But this area is very low-populated, accounting for only 2.5 per cent of the Saguenay population and 5.2 per cent of the overall 478 affected families. Also, its higher score is attributable mainly to one particular disease (spastic ataxia), with seven affected families in one single parish. Another attempt (reported in Bouchard, *et al.*, 1987) looked for correlation between genes and surnames: are mutant genes susceptible to association with particular surnames? Again, our initial research yielded no convincing evidence of such a relationship. A second, more sophisticated attempt is underway.

Aside from searching for indications of gene segmentation, research in population and family history has provided useful insights on the gene flow. For instance, we have been able to demonstrate that, in this frontier society, people who were related were more likely to stay in the region and reproduce. In the same way, those who reproduced were very successful in establishing their married children as farmers, thanks to strong family bonds and abundant cheap land. Actually in each peasant family between 1842 and 1911, three children out of four were established on new land (Bouchard, 1987). All this resulted in more offspring from big Charlevoix families staying in the region and reproducing what we have called the "core" or Charlevoix genotypes (see also Bouchard, 1986a).

New Challenges for Genetic Epidemiology

Along with population studies and the search for genetic subdivisions, we also try to gather individual information in order to provide people with better genetic counselling. On the one hand, development or implementation of markers is underway at McGill and Laval University for disorders like tyrosinemia and myotonic dystrophy (Steinert disease). A marker for cystic fibrosis should be available soon and carrier tests already exist for haemochromatosis and mutiple endocrine neoplasia (Type II-a).

Another type of genotype evidence may consist of the likelihoods derived from genealogical inference. Such a device is currently being developed at SOREP. If the risk values obtained are discriminating enough for any one

TABLE 6. DISTRIBUTION OF AFFECTED FAMILIES (PER 10,000) FOR SIX RECESSIVE DISORDERS, SAGUENAY (MICRO-REGIONS)

Micro-regions	Population in 1985	Rickets vitamin -D-dependent	Agenesis of corpus callosum	Cystic fibrosis	Haemochro-matosis	Tyrosinemia	Spastic Ataxia	All disorders
Bas-Saguenay	7 300h.	0,0	6,8	0,0	0,0	8,2	19,1	34,2
La Baie	25 080h.	1,1	1,1	1,9	0,0	4,7	2,7	11,9
Chicoutimi	121 580h.	0,9	3,2	1,8	1,7	4,3	2,6	14,8
St-Ambroise	16 900h.	2,3	3,5	2,3	0,5	5,9	3,5	18,3
Alma	52 590h.	2,2	6,2	3,6	1,3	3,4	5,1	22,0
Roberval	32 800h.	0,3	1,8	3,0	0,3	4,2	5,4	15,2
Dolbeau-Mistassini	31 571h.	2,2	2,5	5,0	1,2	1,2	2,2	14,5
REGION	287 821h.	1,3	3,5	2,6	1,1	4,0	3,8	16,6

Notes: Of the 290 families involved, 51 were not included in this count because their residence were unknown in the region. There is no reason to suspect that their distribution may differ from the others.

disorder, then it would be consistent to combine kinship inference with the DNA technology.

As mentioned above, two situations may be tentatively anticipated. In the first case, the marker involved may be the mutant gene itself, as happens with the thalassemia and sickle cell disease. Then, genealogical inference might be used in order to delineate subgroups at risk and to target the utilization of markers in a large population. However, this strategy could work only with reference to population-based screening, a field which is hampered with very serious legal and ethical questions.

In this regard, another approach seems much more acceptable. In most cases — at least at this time — carrier tests consist of linked markers and they provide probability evidence, not certainty. Hence the idea of putting together all sort of indications, each of them imperfect but strengthening each other, in order to make more complete information available to genetic counsellors; indications yielded by population studies should be included here as well.

Ancestral inference might prove useful to genetic epidemiology in another way. The utilization of DNA markers must closely fit the pecularities of the population involved. As a result, one has to devise a specific model, taking into account allele frequencies, number of mutations and sources of the gene, etc. In this regard, genealogical inference will help to establish and to evaluate proper strategies for genetic counselling.

The scope of this approach is widened by the fact that it addresses not only Mendelian disorders but also complex etiologies like cancers, heart disease, diabetes, mental disorders and all other so-called multifactorial disorders. But at the same time, it is faced with a major difficulty pertaining to law and ethics. Although the level of the difficulty may vary from one country to another, what is generally questioned is the right a) to accumulate personal, medical information on individuals who are not given the opportunity to agree or not, and b) to violate the privacy for screening purposes. On the first point, the development of a population database seems perfectly justified when the data are in the public domain (e.g., vital records) or have been collected through legal procedures, and when its utilization carries the possibility of being a public service (Goulet, *et al.*, 1983; SOREP, 1986). On the second point, the researcher has to cope with apparently conflicting rules:

1. A very strong concern for privacy rules undermine any kind of population-based screening. Instead, people are expected to know from public advertising about the existence of the register and its facilities, and individuals themselves may only decide to make requests in order to access personal information. Thus, the register's manager is not allowed to take any initiative.

2. The reverse side of this is that, in some instances, researchers may have the obligation to intervene without request. In Saguenay, this can be exemplified by MEN-Type II, an autosomal dominant disorder diagnosed in some families. Very often, cancer develops among the carriers. The age of onset is variable but mostly adults are affected. A quite simple carrier test is available. If performed early enough, risk of cancer can be almost entirely eliminated through routine surgery. Using genealogical inference, it would be very easy to delineate individuals at risk. Moreover, in this particular case, we have been able to identify the couple who introduced the gene in Saguenay in the 19th century.

Clearly, the foregoing illuminates the diversity of the situations to be encountered and the complexity of the decisions to be made. Lawyers and ethicists are currently working with SOREP to help establish principles and guidelines. In the meantime, as a policy, we took the rigid stand not to interfere in any way with privacy and wait for individuals to ask for accessing their record. We have also elected to focus our work on disorders for which treatment or significant help is available.

Acknowledgments

The research works reported here were made possible through financial assistance from Université du Québec à Chicoutimi, Fondation de l'Université du Québec à Chicoutimi, Fonds FCAR (Québec), Social Sciences and Humanities Research Council (Ottawa). Margaret I. Gradie and Marc De-Braekeleer have been kind enough to read the manuscript and to improve it.

Footnotes

1. For more detailed accounts, SOREP makes available to the reader a collection of annual reports, articles, working papers, etc.

2. They are reported to regularly in SOREP annual reports.

3. Comparative data available in E.K. Ginter *et al.* (1980), E. Bois *et al.* (1976), and D. Salmon *et al.* (1980).

References

Bois, E., J. Feingold, P. Frenay and M.L. Briard. 1976. Infantile cystinosis in France: Genetics, incidence, geographic distribution. Journal of Medical Genetics 13:434-438.

M. Boleda. 1984. Les migrations au Canada sous le régime français (1608-1760). Cahiers québécois de démographie 13(1):23-40.

Bouchard, G. 1986a. La dynamique communautaire et l'évolution des sociétés rurales québécoises aux 19e et 20e siècles. Construction d'un modèle. Revue d'histoire de l'Amérique française 40(1):51-71.

_____. 1986b. The Processing of Ambiguous Links in Computerized Family Reconstruction. Historical Methods 19(1):9-19.

_____. 1987. Sur la reproduction familiale en milieu rural: systèmes ouverts et systèmes clos. Recherches sociographiques XXVIII(2-3):229-251.

Bouchard, G., M.-A. Desjardins-Ouellette, F. Markowski and K. Kouladjian. 1985. La distribution des patronymes au Québec: témoins des dynamiques de population. Anthropologie et Sociétés 9(3):197-218.

Bouchard, G., C. Laberge, C.-R. Scriver, and M. Declos. 1988a. Recherche d'un effet fondateur de la tryosinémie et du rachitisme vitamina-dépendant dans la population québécoise (17e-20e siècles). Forthcoming.

Bouchard, G., C. Laberge, C.-R. Scriver, F. Glorieux, M. Declos, L. Bergeron, J. Larochelle and S. Mortezai. 1984. Etude démographique et généalogique de deux maladies héréditaires au Saguenay. Cahiers québécois de démographie 13(1):117-137.

Bouchard, G., J. Morissette and K. Kouladjian. 1987. La statistique agrégée des patronymes du Saguenay et de Charlevoix comme indicateurs de la structure de la population aux XIXe et XXe siècles. Cahiers québécoise de démographie 16(1):67-98.

Bouchard, G. and R. Roy. 1982a. Les variations nominatives dans les registres paroissiaux du Saguenay. Annales de démographique Historique, 354-368.

_____ and R. Roy. 1982b. La reconstitution automatique des familles et le problème des substitutions nominatives: mesures de rendement. Informatique et Sciences Humaines (Paris), 14e année, nos. 56-57, pp. 75-94.

_____. 1985. Registre de population et substitutions nominatives. Population (3): 473-490.

Bouchard, G., R. Roy and B. Casgrain. 1985. Reconstitution automatique des familles. Le système SOREP. Université du Québec à Chicoutimi, Dossier no. 2, septembre, 2 vol.

Bouchard, G., R. Roy, M. Declos, K. Kovaladjain and J. Mathieu, 1988b. La diffusion du gene de la dystrophic myotomique au Saguenay (Québec). Journal de génétique humaine 36(3):221-237.

Cannings, C., E.A. Thompson and M.H. Skolnick. 1978. Probability functions on complex pedigrees. Advances in Applied Probability 10:26-61.

Ginter, E.K., R.F. Garkavtzeva and A.A. Revazov. 1980. Population structure and hereditary in Uzbekistan. In A.W. Eriksson, H.F. Forsius, H.R. Nevanlinna, P.L. Workman and R.K. Norio (eds.), Population Structure and Genetic Disorders, New York: Academic Press.

Goulet, J., M. Gagne and D. Girard. 1983. Régles de droit et confidentialité. SOREP, Dossier no. 1.

Gradie, M. 1986. Spatial land Temporal Tends in Inbreeding in the Saguenay Region of Quebec. Paper presented at the annual conference of the American Association of Physical Anthropologists, Albuquerque (New-Mexico).

Gradie, M., L. Jorde and G. Bouchard. 1988. The genetic structure of the Saguenay, 1852-1911. American Journal of Physical Anthropology (forthcoming.)

Henripin, J. 1968. Tendances et facteurs de la fécondité au Canada. Ottawa: Bureau fédéral de la statistique.

McKusick, V.A. (assembled with commentary by V.A. McKusick) 1978. Medical Genetic Studies of the Amish. Baltimore: Johns Hopkins University Press, (selected papers).

Pouyez, C., Y. Lavoie, G. Bouchard, R. Roy. 1983. Les Saguenayens. Introduction à l'histoire des populations du Saguenay, 16e - 20e siècles. Québec: Presses de l'Université du Québec.

Roy, R., G. Bouchard and M. Declos 1988. La première génération de Saguenayens: provenance, apparentement, enracinement. Câhiers Québecois de Débecois de Démographie 17:113-34.

Salmon, D., M.-F. Landre, G.R. Fraser, S.K. Buehler, J. Crumky and W.H. Marshall. 1980. A familial aggregate of common variable immunodeficiency, Hodgkin disease and other malignancies in Newfoundland -II. Genealogical analysis and conclusions regarding hereditary determinants. Clinical and Investigative Medicine. 2:175-181.

SOREP. 1986. Règlement concernant la confidentialité des données contenues dans le fichier BALSAC. Document I-C-49.

Tesnieres, M. 1980. L'extinction des noms de famille. Héralique et généalogie: 22-24.

Thompson, E.A. 1986. Pedigree Analysis in Human Genetics. Baltimore: John Hopkins University Press.

Received November, 1987; revised August, 1988.