

Lisa Dillon
Département de Démographie, Université de Montréal
April, 2005

The Historical Demography Research Infrastructure: Creating the 1852 Canadian Census Database

English version of : DILLON, LISA, auteur. « La création de la base de données du recensement de Canada 1852 », Université de Montréal, » *Cahiers québécois de démographie*, numéro à thème sur la démographie historique, 34 (2), **automne 2005** : 297-313.

Abstract:

This research note discusses one of the latest additions to Canada's series of historical census microdata, a machine-readable sample of the 1852 Census of Canada East and Canada West. While similar to subsequent censuses in form and content, the 1852 Canadian census poses particular challenges in terms of national representativity and identifying heads of households. The 20% sample of the 1852 census of Canada East and Canada West will feature a total of 259,000 persons. Data entry procedures, the problem of missing data for urban areas and the identification of household heads using building type information are discussed. A preliminary analysis of the 1852 census frequencies and comparison of these frequencies to those derived from the 1871 Canadian census of rural-dwellers suggests that this machine-readable sample is representative of the rural Canadian population in 1852.

Lisa Dillon
Département de Démographie, Université de Montréal
April, 2005

**The Historical Demography Research Infrastructure:
Creating the 1852 Canadian Census Database**

For more than thirty years, Canadian researchers have sustained an important tradition in the creation and use of historical population microdata. This research note discusses one of the latest additions to Canada's series of historical census microdata, a machine-readable sample of the 1852 Census of Canada East and Canada West. While similar to subsequent censuses in form and content, the 1852 Canadian census poses particular challenges in terms of national representativity and identifying heads of households. The 1852 Canadian census sample is being developed under the auspices of the Programme de recherches en démographie historique (PRDH) at the Département de Démographie, Université de Montréal. The PRDH is responsible for the Registre de la population du Québec ancien (RPQA), a longitudinal linked database of baptismal, marriage and burial registers of the French population resident in the St. Lawrence valley from 1621 to 1799. The PRDH is now expanding its scope to include nineteenth-century historical census data. The 1852 Canadian census project is funded by a four-year research infrastructure grant from the Canadian Foundation for Innovation which will establish the physical and virtual facilities and networks necessary for historical population microdata projects based on grant-funded, in-house and paid data entry as well as outsourced volunteer-based genealogical data entry.¹

Nineteenth-century population data

The first phase of the 1852 Canadian census project is the preparation of a 20% sample of the 1852 Census of Canada East and Canada West. This machine-readable sample will serve as the first in the series of Canadian individual-level nineteenth-century census microdata sets which includes public-use data for the census years 1871 to 1951 and 1971 to 2001.² To situate Québec and Canadian patterns in international context, researchers will eventually be able to compare the 1852 Canadian census microdata to similar samples of the 1850 Census of the United States and the 1851 Census of England and Wales; these three mid-century national censuses will be integrated in a future

¹ Canadian Foundation for Innovation, « Historical Demography Research Infrastructure/Infrastructure de recherches en démographie historique (HDRI/IRDH) » Project #7549.

² Gordon Darroch and Michael Ornstein, 1871 Canadian Census Data (Canadian Historical Mobility Project), Institute of Social Research, York University, 1979; Lisa Dillon, The 1881 Canadian Public-Use Microdata Sample (The 1881 Canadian Census Project and the North Atlantic Population Project), Université de Montréal, 2004; Kris Inwood et. al., The 1891 Canadian Census, University of Guelph; Eric Sager and Peter Baskerville, The 1901 Canadian Census Public File (The Canadian Families Project), University of Victoria, 2001; Chad Gaffield et. al., The Canadian Century Research Infrastructure; the Data Liberation Initiative, Statistics Canada.

initiative of the North Atlantic Population Project.³ Eventually, we also hope to undertake collaborations necessary to link the 1852 census data to Québec parish register data for the early nineteenth century. Images of the nineteenth-century Québec parish registers have been digitized by l'Institut Généalogique Drouin.⁴ The PRDH has purchased a copy of these digitized images to facilitate future data entry and record linkage initiatives. At the provincial level, the period between 1800 and 1852 is an under-researched time period in Québec historical demography, largely due to a lack of province-wide historical data. Researchers interested in family and social structure have tended to focus on the late nineteenth-century and early twentieth-century periods, for which in-depth population data are available, or on particular regions such as the Saguenay and Charlevoix regions, for which earlier nineteenth-century data is available at the regional level. In the absence of systematic, provincial- and national-level microdata, general Canadian social and demographic structures during the early nineteenth-century period remain a mystery.

Nevertheless, historical demography projects are currently underway to fill this gap in the population data record. Researchers at the Université du Québec à Chicoutimi are currently working to extend the BALSAC database of parish register data for nineteenth-century Saguenay and Charlevoix regions across Québec, transcribing Catholic marriage acts for the whole of Québec.⁵ In a similar initiative, Bertrand Desjardins is adding to the RPQA burial acts for Québec Catholic persons who died in old age between 1800 and 1852. The 20% sample of the 1852 Census of Canada East and Canada West will join with these initiatives in creating a record of historical population data which sheds light on demographic behaviour of the early nineteenth-century period.

Sampling and Data Entry Procedures

The existence of Canadian national-level historical census microdata samples date back to the late 1970s in Canada, when the Canadian Historical Social Mobility Project at York University's Institute of Social Research created the 1% sample of the 1871 Canadian Census.⁶ The advent of this and similar projects, notably the Integrated Public-use Microdata Series at the University of Minnesota and the Canadian Families Project, demonstrated a consensus among database creators that these data should consist of nationally-representative, geographically stratified random samples which offer information on individuals but which are clustered at the level of dwelling and

³ Evan Roberts, Steven Ruggles, Lisa Y. Dillon, Olof Gardarsdottir, Jan Oldervoll, Gunnar Thorvaldsen and Matthew Woollard, "The North Atlantic Population Project: An Overview," *Historical Methods*, Vol. 36, No. 2 (Spring 2003): 80-88.

⁴ See Institut Généalogique Drouin, <http://institutdrouin.com/>.

⁵ Gérard Bouchard, "Current Issues and New Prospects for Computerized Record Linkage in the Province of Québec," *Historical Methods*, Vol. 25, No. 2 (Spring 1992): 67-73.

⁶ Gordon Darroch and Michael D. Ornstein, "Coding and Data Processing for the Feasibility Study: Canadian Historical Mobility Project," Paper written for the Institute for Behavioural Research, York University, Canada (revised, 1994).

household.⁷ The original manuscript census returns of the 1852 Census of Canada West and Canada East are nationally representative insofar as they represent a count of the two former colonies which had been joined together by the 1840 Act of Union.⁸ Two censuses of Nova Scotia and New Brunswick were also conducted at this time under separate British colonial administrators.⁹ For reasons of time and resources, neither Maritime province is included in the 1852 Canadian census project. In the case of New Brunswick, the number of persons enumerated on each census page varies greatly, creating difficulties for page-based sampling. The Nova Scotia census is count of household heads rather than all household inhabitants.

When considering the sampling procedures to be adopted for the 1852 Canadian census project, we examined sampling procedures used by our sister census projects. With its most recent historical census projects, the Minnesota Population Center decided to outsource data entry. To ease the process of remote data entry and improve data entry quality, the MPC changed its sampling procedure from dwelling-based sample point selection to page-based sampling. It continued to produce nationally-representative, geographically stratified random clustered samples, but these samples now consist of a 25-line window at regular page intervals rather than every tenth dwelling.

The 1852 Canadian Census Project adopted a similar strategy in order to cope with the particular exigencies of the 1852 Canadian census manuscript format. The 1852 Canadian census poses many special challenges. First, unlike subsequent censuses, this census does not include columns to number dwellings and households. The census respondents are clearly grouped into dwelling and household units: the arrangement of family names, marital status, sex and ages indicate family groupings, while a building type column indicates dwelling divisions. However, these groups were not systematically numbered, as occurred in subsequent censuses. Without such numbering, it would be difficult to undertake a dwelling-based sample point procedure, as was followed in the creation of the 1901 Canadian census microdata. Each census page in the 1852 Canadian census contains 50 lines for 50 individuals, with information recorded in 41 columns. These 41 columns stretch across four page sides, which can be viewed on three images scanned from the original microfilm. The horizontal lines which separate the information for the 50 individuals are not always clear in the scanned images, and line numbering for the 50 lines only appears on page one. In addition, the only variable which suggests household breaks, the building type column, appears far to the right on page two. Had we decided to undertake a dwelling-based sample point procedure, the data entry operator would have had to manipulate the images a great deal to understand which selected dwelling corresponded to which first and last names. As a result, to simplify our data entry procedure, we decided to adopt a page-based sampling method as well, taking every fifth group of 50 people. To date, we have sampled 240,836 persons

⁷ Steven Ruggles, "Sample Designs and Sampling Errors," *Historical Methods*, Vol. 28, No. 1 (Winter 1995): 40-42; Michael Ornstein, "Analysis of Household Samples: The 1901 Census of Canada," *Historical Methods*, Vol. 33, No. 4 (Fall 2000): 195.

⁸ Margaret Conrad, Alvin Finkel and Cornelius Jaenen, *History of the Canadian Peoples : Beginnings to 1867*, (Toronto : Copp Clark Pitman Ltd., 1993), 425; Bruce Curtis, *The Politics of Population: State Formation, Statistics and the Census of Canada, 1840-1875*, (University of Toronto Press, 2001), 98-100.

⁹ Curtis, 256.

from the 1852 Census of Canada East and Canada West. Once data entry is complete, including the addition of 20% of the 100% sample of 1852 Québec City, prepared by Marc St.-Hilaire and Richard Marcoux of the Centre interuniversitaire d'études québécoises (CIEQ), Université Laval, our sample will contain about 259,000 persons.¹⁰

Data entry of the 1852 Census of Canada East and Canada West has been greatly eased by the availability of digital images of these censuses produced by the Online Services Division of the National Archives of Canada. Working in partnership with the 1852 Canadian Census Project, the Online Services division scanned these images directly from master microfilm copies of the 1852 census, creating 86,706 images identified sequentially within each microfilm reel number.¹¹ Our project team devoted considerable hours at the outset to organizing these images into provinces, districts and sub-districts. This work was undertaken in part because we decided to stratify our sample geographically, sampling census pages from within each sub-district. Once the work of image sorting was done, we were able to generate a separate database which identified every image. Our computer programmer created a page selection system through which pre-selected images could be sampled. Each data entry operator then downloaded a series of five images at a time: three images containing the four census page sides to be sampled, and two subsequent images containing information on the following group of 50 individuals. If the sampled group of 50 people finished halfway through a dwelling, the data entry operator continued to the next set of 50 people to finish off the dwelling. For our final microdata sample, we will then eliminate dwelling fragments at the top of each sampled set of pages. In some instances, the number of persons in a sub-district for whom census manuscript pages existed was less than 250; in such cases, the number of sampled persons in those sub-districts actually exceeds 20%.

To view census images, we have used Adobe Acrobat 6.0 software, which allows data entry operators to magnify first and last names to aid in interpretation. Since much information, such as birthplace and religion, is repetitive, and the content of some other columns, such as stores, public buildings and places of worship, are often blank, the data entry operators found it quicker to enter information from the census images column by column rather than line by line. To help compensate for the absence of line numbering on pages 2a and 2b of each set of census images, our data entry program features a series of extra grey-coloured static, reference columns to the left-hand side which retain information on name, marital status, sex and age. These grey columns were populated automatically as the data entry operator entered the name, marital status, sex and age of each person. Then, as the data entry operator gradually filled in additional columns to the right, the grey columns remained still on the left side of the screen, allowing the data entry operator to keep track of which information was being entered for which person. In addition, certain lines became colour-coded as certain information was filled in: dark blue if the person was a widower, lighter blue for widows, dark pink if the person was a

¹⁰ See Richard Marcoux, Marc St-Hilaire et Charles Fleury, "Ville et population en changement: transformations urbaines et ajustements familiaux à Québec au XIXe siècle et au début du Xxe," *L'Ancêtre*, Vol. 29 (printemps 2003) : 227-230.

¹¹ This set of images includes the 1852 censuses of Nova Scotia and New Brunswick, as well as the censuses of Canada East and Canada West.

single adult male and lighter pink for single adult females, dark orange for a married male and lighter orange for married females, and plum for children. This colour-coding helped the data entry operators keep track of information as it was typed in. The organization of census images into district and sub-district folders aided the supervision and quality control of data entry work, allowing us to locate and view particular census pages quickly when a problem of interpretation arises. It also facilitated quick manual checking of the census manuscript once the phase of consistency checking and cleaning began.

Missing Data

Unlike the 1871 and later Canadian censuses from which census microdata have been created, large parts of the 1852 Canadian census manuscript had been destroyed before the National Archives undertook microfilming of the census in the 1950s. An exhaustive index prepared jointly by the National Archives of Canada and the 1852 Census Project team indicates that 349 sub-districts and divisions are entirely missing out of the 1,274 sub-districts and divisions listed in Upper and Lower Canada in 1852. According to the aggregate census volumes for 1852 Canada, these 348 sub-districts and divisions contain 508,892 persons. In addition, while many sub-districts survived in whole, other sub-districts survived just in part. There are approximately 120,558 other persons missing from the manuscript because a portion of their sub-district is missing. The total number of persons missing from the surviving census manuscripts is 629,450. The aggregate statistics list a total population of 1,842,265 persons in Upper and Lower Canada. Thus, at the individual level, 34% of persons in the 1852 census are missing from the manuscript record, and only 66%, or 1,222,447 persons are available for data transcription.¹²

If we consider the total population of Canada East and Canada West in 1852, our 20% sample represents only about 14% of the total population (and even less once we remove the dwelling fragments from the top of each page). The sub-districts and divisions which are present in the manuscript record are not representative of the whole population, resulting in a biased sample. For instance, most large Canadian cities, including most of Montréal and Toronto, and some of the smaller cities, such as Kingston, London and St. Catherines, are missing. We cannot compensate for the lack of large city dwellers in our sample, except to add to the database the CIEQ 100% sample of Québec City (weighted accordingly), and perhaps to oversample the St. Louis Ward of Montréal, all of the city of Hamilton, and the Ottawa East and West wards (if funds and time permit).

We should, however, find a way to compensate for the absence of the other rural sub-districts and divisions, creating a microdata sample useful for the study of rural Canada East and Canada West in 1852. Our initial idea was to oversample sub-districts which were geographically proximate to those with many missing persons and which bore similar socio-demographic characteristics, a procedure whose logic would be similar to the hot-decking programs used with the IPUMS data and contemporary statistical data

¹² Note that this paragraph includes revised statistics produced in July, 2011, which differ from the original statistics presented when this article was published in CQD.

to impute missing data for particular variables.¹³ In lengthy discussions with colleagues at Institut national de la recherche scientifique--Urbanisation, Culture et Société (INRS), York University and the University of Toronto, we decided against this option. First, it would be difficult to settle on a set of criteria for choosing a set of donor sub-districts, since every researcher who will use this database will approach it with a distinct set of research interests. Investing time and funds in the data entry of a particular set of donor sub-districts based on particular criteria would only solve the database bias problem for those researchers whose research questions conform to those criteria. In addition, we suspect that the aggregate 1852 census statistics, which would form the basis of choosing donor sub-districts, are beset with certain problems. In the process of creating a series of Excel spreadsheets which reproduce the aggregate 1852 census statistics, we discovered that the sub-totals for certain variables for certain sub-districts did not equal the total population for that sub-district. In other cases, totals for certain variables were grouped across several sub-districts. In doubt of the reliability of the aggregate data for all the sub-districts, we chose not to try to create oversamples by sampling from neighbouring sub-districts.

Instead, we decided that the best way to deal with missing data in the 1852 Canadian census would be first, to carefully document exactly which type of data is missing, second, to compare our microdata frequencies to the aggregate data frequencies, and third, to devise a set of weights to calibrate the microdata to the aggregate data for all of Canada East and Canada West. While weighting will not resolve the bias introduced by the absence of Canada's large cities, it could help us to compensate for missing data in rural areas. The advantage of using weights lie in their flexibility: we could create a series of different weights, depending on which research questions may be addressed. With these weights, we would still be limited to those variables for which aggregate data seems reliable.

Future work on the 1852 Canadian Census

The next phase of work on the 1852 Canadian census will consist of consistency checking, cleaning and data coding. One main challenge will be to decide on dwelling breaks among the 50 people recorded on each census page. To divide each group of 50 individuals into dwellings, we have used the building type information recorded by enumerators in column 31. Examples of responses to this question are: log house, frame house, shanty, log shanty, maison en pieces, maison en bois, stone, pierre or en charpente. When planning the 1852 sampling procedure, we observed that these responses seemed to be given by potential household heads, married men of adult age. During the data entry process, the assigning of building type to a particular line was at times the result of an interpretation made by the data entry operator. Sometimes the faintness of ruled lines and the absence of line numbers on page 3 made it necessary for the data entry typist to guess the line on which to enter the building type information. On these occasions, given the choice of four lines or so, the data entry operators would tend

¹³ Lisa Y. Dillon, "Integrating Canadian and U.S. Historical Census Microdata: Canada (1871 and 1901) and the United States (1870 and 1900)," *Historical Methods*, Vol. 33, No. 4 (Fall 2000): 189-192.

to guess that the information belonged to someone who appeared to be the head of a household.

We have now accumulated enough transcribed cases to conduct an analysis of the types of persons who reported building type information on the 1852 census. Those persons on whose line building type information was written are identified as “apparent dwelling heads”. The characteristics of these persons is compared not only to the general sampled population in 1852, but also to actual dwelling heads sampled in the 1871 Canadian census microdata, whose status was made clearer by the presence of a dwelling number column (see Table 1). Since the 1852 census data is by necessity a largely rural sample drawn from the Canada East and Canada West enumerations, Table 1 compares 1852 apparent dwelling heads who lived in the rural areas of Canada East and Canada West to 1871 household heads from Ontario and Québec who lived in rural areas.¹⁴ A comparison of the demographic characteristics of these persons suggests immediately that most apparent heads in the 1852 Canadian census were indeed the heads of households. For example, the sex ratio of all sampled Canadians in 1852 was 52% male to 48% female, while the sex ratio of apparent heads, at 94% male to 6% female, more closely resembled that of real dwelling heads in 1871, which was 93% male to 7% female (Table 1). One would expect most household heads to be male, with the exception of some widowed females. The age profile of apparent household heads in the 1852 census also closely resembled that of household heads in 1871 (see Figure 1); 85% of apparent dwelling heads in 1852 were aged 20 to 59, compared to 82% of dwelling heads in 1871. The apparent dwelling heads in 1852 were also largely married, like their counterparts in 1871. While 31% of all rural Canadians in 1852 described themselves as married, 87% of apparent dwelling heads in 1852 and 85% of real dwelling heads in 1871 did so.

The 1852 census also features socio-economic information which suggests that persons who stated building type information were genuine dwelling heads. One would expect most dwelling heads to be employed. In fact, while two-thirds of all rural Canadians in 1852 gave no occupational response, only 9% of the apparent dwelling heads in 1852 and 8% of rural dwelling heads in 1871 failed to report an occupation (Table 1). Occupational information in the 1852 census has not yet been coded, which prevents further economic analysis at this time. Nevertheless, the top 25 occupational responses of the 1852 general population include six non-occupation responses, such as none, wife, spinster or daughter. In contrast, the top 25 occupational responses of apparent household heads are all bona fide occupations such as farmer, labourer, carpenter, shoemaker, blacksmith, merchant, tailor, cooper and weaver. One would expect a larger proportion of household heads to be foreign-born, since many non-household heads would be children who would have been native-born. Accordingly, nearly half of apparent dwelling heads were born outside Canada East, Canada West and the neighbouring colonies of New Brunswick and Nova Scotia compared to only 26% of Canadians in general in 1852 (Table 1).

¹⁴ The definition of “rural” used in Table 1 differs somewhat between the 1852 and the 1871 data. The 1852 data displayed in Table 1 includes all persons who responded on a rural census schedule. The 1871 rural data excludes persons living in cities or towns with a population of 3,000 or more residents.

Since our data entry operators were occasionally forced to assign building type to an individual, we will examine potential differences in the guesses made by different data entry operators. To do so, we will compare the characteristics of dwelling heads by data entry operator, exploring whether particular data entry operators made particular assumptions when interpreting the placement of the building type information. In some cases, dwellings include multiple households or families. Once we have concluded our identification of dwelling breaks and dwelling heads, we will devise a program with a set of rules to identify probable second household heads within dwellings.

The 1852 Census of Canada East and Canada West did in fact ask a question about household membership which might potentially shed light on family and household breaks within dwellings. The 1852 Census enumerator instructions defined non-members of the family as “Those who stopped in the house on that night, but who are not members of the family, such as travellers, lodgers, clerks, servants, etc.” Unfortunately, these instructions did not specify how multiple families and households residing together in one dwelling were to be classified. Our preliminary analysis of responses to the member of the family/not member of the family variable reveals that some enumerators selectively classified certain persons as non-members of the family, while others classified all persons present as members of the family. To infer second household heads within multiple-household dwellings, we will probably have to rely on primary variables such as name, age, marital status, occupation and sequence within the dwelling. This inference work will be undertaken in the context of work to infer and impute relationship to household heads, as well as family interrelationship variables.

Preliminary analysis of different variable frequencies in the 1852 Canadian census shows some expected results, and a few unexpected ones. Some variables will require a fair amount of time for coding. For example, the occupation question received 3,897 unique responses, the birthplace question received 4,163 responses, while the religion question received 1,745 responses (Table 2). As expected, the ‘colour’ census question was a failure, as only 63 persons indicated black status; in contrast, about 1,432 persons reported themselves as aboriginal. In addition, there were few responses to the deaf (144) and blind (94) questions; somewhat more Canadians, 289, were described as lunatic, idiotic or insane. A smaller proportion of Canadians in 1852 reported school attendance, 12%, than Canadians in 1871, 19%. In the light of modestly falling fertility, one might have expected the inverse or at least similar proportions, suggesting that this variable may not have been well-recorded in 1852.

We will need to undertake careful analysis of the four variables “Residence if out of limits / Member of the Family / Not Member of the Family / Absent Member of the Family” in relation to each other to understand their exact meaning. Bruce Curtis has noted the inconsistent criteria for assigning individuals to households in the 1852 census.¹⁵ The later nineteenth-century Canadian censuses were de jure censuses; in other words, persons were to be enumerated in their usual place of residence, whether or not they were there the day the enumerator visited. In a large country with a population frequently engaged in seasonal occupations which took them into the bush or out to sea,

¹⁵ Curtis, 115.

census officials believed that the de jure census practice was the most suitable approach to keeping track of Canadians. The 1852 Census of Canada East and Canada West, however, adopted a mixed de facto/de jure approach. The census notes, “In the 1st column you will take name of every person who sojourned in the house on the night of Sunday, the 11th of January, as well strangers as members of the family, and also those members of the family who are temporarily absent, but whose usual residence it is.”¹⁶ In effect, this instruction requested the enumeration of persons in their usual place of residence, but also the recording of strangers. For those strangers, their usual place of residence was also requested: “Col. 5 – When a person chanced to stay in a house on the night of the 11th of January, you are to mark if possible the place where his usual residence is : in very many cases this will be impossible, and you are then to write the word “UNKNOWN” in the column.” Curtis has argued that that these instructions risked the double-counting of certain Canadians: persons who stayed at friend’s house the day and night of the census may have been both enumerated at home and enumerated at their friend’s house.¹⁷ A full resolution of this question would require analysis of a 100% database of the 1852 Canadian census. One of our ultimate project goals is to obtain the collaboration of genealogists to extend data entry of the 1852 Canadian census from 20% to 100% in a volunteer transcription effort similar to that undertaken for the 1901 Canadian census.¹⁸ Within three years of the online posting of images of the 1901 Canadian census online by the National Archives of Canada, data entry volunteers working online with the Automatic Genealogy.com 1901 Canadian Census Indexing Project had transcribed 99.99% of the 1901 Canadian census, a total of 5,642,088 lines on 112,797 pages. Provision of the 1852 census images online by the National Archives will help steer a similar initiative. In the meantime, we plan to analyze the characteristics of persons enumerated as strangers and their usual place of residence; case studies of those who listed specific places of residence may allow us to determine the extent to which actual double-counting occurred.

Among our unexpected results, we see many non-occupational responses in the occupation column, some of which indicate family relationships which might be usefully exploited. More importantly, about a third of the sampled Canadians gave a specific place of birth which not only noted their province of birth but also their city, town or community of birth. If these persons are similar in other respects to persons who only stated their province of birth, we could undertake more extensive analysis of their specific migration patterns, from community to community. The 1852 Canadian census asked Canadians to record the names of persons who had died the past year; accordingly, about 1% of the sampled Canadians were reported as having died during the census year. The 1852 census mortality statistics are undoubtedly underreported. For example, Marvin McInnis notes that the general death rate of Canada during the first half of the nineteenth century would have been about 20 per thousand deaths. He goes on to write

¹⁶ David P. Gagan, “Enumerator’s Instructions for the Census of Canada, 1852 and 1861,” *Histoire sociale/Social History*, Vol. VII, No. 14 (November 1974): 355-365.

¹⁷ Curtis, 116, 126-127

¹⁸ The 1901 Canadian Census Indexing Project (Automated Genealogy): <http://automatedgenealogy.com/census/cache/NationalSummary.jsp>. These images were posted online in May, 2002.

“The Canadian censuses had collected mortality data from 1851 onward, but those statistics were widely regarded as incomplete and inadequate....One of the most serious problems with census reports of deaths was the undercount of deaths of older people who left no one behind to report their passing. Infant mortality was more fully and accurately recorded.”¹⁹ McInnis’ conclusions could be tested by analyzing the characteristics of the 2,270 Canadians whose age and cause of death was detailed in the 1852 census sample.

We anticipate more extensive research possibilities which make use of the economic characteristics of Canadians in 1852, examining those who lived in shanties, log houses, frame houses and stone houses and accounting for the number of floors in their houses. In other instances, we can examine the 1,755 persons who described an attached place of business. The most frequent responses to this question were: a tavern, a store, a blacksmith shop, a grainery, a carpenter shop, a shoe shop, a tailor shop, a tannery, an auberge, a saddler shop, a waggon shop and a carriage shop. In other columns, 793 persons reported an attached factory, while 454 persons stated the number of employees they employed. The 1852 Canadian census sample will thus afford many opportunities to study the contours of economic life and small workshop and factory production in rural contexts.

A general comparison of the characteristics of all Canadians in the 1852 Census of Canada East and Canada West, and the 1871 of rural Québec and Ontario suggests that the 1852 sample is generally representative of the total rural population. The ratio of men to women in the 1852 sample is slightly higher than that of the rural 1871 sample (Table 1). This small difference may be accounted for by considering the period of high immigration which preceded the 1852 Canadian census. McInnis describes the period between 1815 and 1861 as a period of high immigration, with British immigrants seeking refuge from unemployment and famine in their home countries.²⁰ High immigration can offset sex ratios, introducing a larger proportion of men. Between 1852 and 1871, immigration fell off sharply; in consequence, one would expect to observe a more balanced sex ratio. The proportion of single and married individuals was more similar during the two census years: almost a third of Canadians were married in 1852 and 1871, while two-thirds were single. The 1852 population was very slightly younger than the 1871 population: 55% of individuals were aged 1 to 19 years in 1852 compared to 54% of their 1871 counterparts. The youthfulness of the 1852 population could be related to the higher proportion of young immigrants in the 1852 population and to higher fertility. Further tests are needed to determine to what extent the 1852 census sample is biased as a result of the omission of large cities and towns. However, these preliminary results suggest that the sample is suitably representative of the rural population of Canada East and Canada West in 1852.

Conclusion

¹⁹ Marvin McInnis, “The Population of Canada in the Nineteenth Century,” *A Population History of North America*, 379, 404.

²⁰ McInnis, 378-384.

This preliminary analysis of responses to the 1852 Census of Canada East and Canada West demonstrates how this census sample will serve as a valuable source of information on the family experiences and socio-economic conditions of rural Canadians at mid-century. Admittedly, the 1852 Canadian census sample will be characterized by some limitations, notably the absence of a systematic urban sample and a resulting bias toward rural dwellers. Some oversampling of particular urban communities for which census manuscripts remain will help to compensate for this loss. Further study of possible double-counting is also necessary to understand the results of the mixed de facto/de jure instructions sent to enumerators in 1852. Nevertheless, as the Canadian Families Project has established, it is necessary to approach every historical census database with a careful understanding of the primary source upon which it is based. “It is no longer possible, if it ever was, to treat routinely generated information in historical sources as a transparent window into the social reality of the past. The census itself must be problematized, its provenance displayed, and its internal logic...unraveled.”²¹ As the earliest national-level source of nominal population data for the modern statistical era, the 1852 Canadian census microdata sample offers an important research resource for scholars interested in the social, economic and demographic research on the country at mid-century. More importantly, this census sample will offer the first point of observation for analyses of Canadian social behaviour across a broad span of time.

²¹ Eric Sager, “The Canadian Families Project and the 1901 Census,” *Historical Methods*, Vol. 33, No. 4 (Fall 2000): 180.

Bibliography

- BOUCHARD, G. 1992. "Current Issues and New Prospects for Computerized Record Linkage in the Province of Québec", *Historical Methods*, 25, 2: 67-73.
- CURTIS, B. 2001. *The Politics of Population: State Formation, Statistics and the Census of Canada, 1840-1875*. Toronto: University of Toronto Press, 385.
- DARROCH, G., and ORNSTEIN, M. 1994. "Coding and Data Processing for the Feasibility Study: Canadian Historical Mobility Project", Paper written for the Institute for Behavioural Research, York University, Canada (revised, 1994).
- DILLON, Lisa Y. 2000. "Integrating Canadian and U.S. Historical Census Microdata: Canada (1871 and 1901) and the United States (1870 and 1900)", *Historical Methods*, 33, 4: 189-192.
- GAGAN, D.P. 1974. "Enumerator's Instructions for the Census of Canada, 1852 and 1861", *Histoire sociale/Social History*, 7, 14: 355-365.
- MARCOUX, R., ST.-HILAIRE, M., et FLEURY, C. 2003. "Ville et population en changement: transformations urbaines et ajustements familiaux à Québec au XIXe siècle et au début du Xxe", *L'Ancêtre*, 29: 227-230.
- MCINNIS, R. M. 2000. « The population of Canada in the nineteenth century », dans M. HAINES et R. STECKEL, éd. *Population History of North America*. Cambridge, Cambridge University Press : 371-432
- ORNSTEIN, M. 2000. "Analysis of Household Samples: The 1901 Census of Canada", *Historical Methods*, 33, 4: 195-198.
- ROBERTS, E., RUGGLES, S., DILLON, L., GARDARSDOTTIR, O., OLDERVOLL, J., THORVALDSEN, G., and WOOLLARD, M. 2003. "The North Atlantic Population Project: An Overview", *Historical Methods*, 36, 2: 80-88.
- RUGGLES, S. 2000. "Sample Designs and Sampling Errors", *Historical Methods*, 28, 1: 40-46.
- SAGER, Eric. 2000. "The Canadian Families Project and the 1901 Census", *Historical Methods*, 33, 4: 179-184.

**Table 1: Characteristics of apparent dwelling heads and all Canadians
1852 Rural Census of Canada East and Canada West; 1871 Census of Ontario and Québec***

	<i>1852 Census</i>		<i>1871 Census - rural areas</i>	
	<i>all persons</i>	<i>apparent heads</i>	<i>all persons</i>	<i>dwelling heads</i>
% men: % women	52:48	94:6	51:49	93:7
% married	31	87	32	85
% single	67	7	66	6
Age:				
% aged 1-19	55	2	54	0
% aged 20-59	41	85	41	82
% no occupation listed	66	9	74	8
% born outside Canada	26	45	17	40
N	236,753	36,380	16,872	2,836

*The rural census of 1852 Canada East and Canada West is compared to rural areas enumerated in the 1871 Census of Ontario and Québec

**Table 2: Number of unique responses to variables
1852 Census of Canada East and Canada West**

Occupation	3897
Birthplace	4163
Religion	1745
Residence if out of limits	1309
Type of house	1230

**Table 3: Number of persons who responded to variables,
1852 Census of Canada East and Canada West**

	<i>Number</i>	<i>% of population</i>
Colour (negro)	63	0.0
Indian	1432	0.7
Deaf	144	0.1
Blind	94	0.0
Idiotic	289	0.1
male member of the family	80998	41
female member of the family	78759	40
male not member of the family	9176	5
female not member of the family	6991	4
at school	22723	12

Source: Lisa Dillon, 1852 Canadian Public-Use Microdata Sample, Historical Demography Research Infrastructure, 2005;

Gordon Darroch and Michael Ornstein, 1871 Canadian Census Data (Canadian Historical Mobility Project), Institute of Social Research, York University, 1979