Issues regarding the use of the Canadian census sample of 1852 for data linkage

Catalina Torres

Presented at the workshop ‘Population Reconstruction’
International Institute of Social History (IISH), Amsterdam
February 21st 2014
Introduction

20% sample of the Canadian census of 1852

- Despite the critiques, this census offers unique opportunities to study the Canadian population of the mid-19th century

- Information at the individual level: suitable for data linkage (for example with the complete census of 1881)

- Variables:
  - Demographic: sex, age, marital status
  - Socioeconomic: occupation, type of house (e.g. stone, log house, brick)
  - Other characteristics: religion, place of birth, place of residence

- Issues: missing data, interpretation of the manuscripts and homogeneity of the French names
1) The missing urban data

- Absence of 1/3 of the records (mostly of the urban population)
  - Population living in the biggest cities (Toronto and Montreal) particularly under-represented

- Some urban data available (Hamilton and Québec city): the urban population can be analyzed if weights are used

- Weights are not always a solution
  - Underrepresentation of the individuals living in the biggest cities
  - Be careful with the methods and the interpretations

*Population by origin, four Canadian cities, aggregated statistics of the Canadian census of 1852*
1) The missing urban data

Distribution of some occupations in four Canadian cities: % of the population aged 15 + having reported an occupation on the census of 1881

- ...Only 10% of the population was urban in 1852
2) B for “Bas” or for “British”? 

- Mentions like “Canada b” coded as Québec (“Bas Canada”)
- Does “B” always mean “Bas”?
  - e.g. what does it mean when the enumerator was Anglophone?
- Most of the individuals (≈95%) whose place of birth is accompanied by a “b” have an English last name
- Sometimes the enumerators indicated the cultural origin in the column *place of birth* (e.g. “Canadien f”, “irlandais”, etc.)
- Previous work: “Canada f” coded as “Canada French (province not specified)”. Includes mentions relative to the cultural origin, e.g. “french canadian”

**Solution:**

1) Variable “B”
2) Birthplace code: “Canada B (province not specified)”:
   - Place of birth with “b” but lacks precision (e.g. “Canada b”) and
   - English origin of the individual

<table>
<thead>
<tr>
<th>Place of birth</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Canada”</td>
<td>54215</td>
<td>21.82</td>
</tr>
<tr>
<td>“Canada F (province not specified)”</td>
<td>25643</td>
<td>10.32</td>
</tr>
<tr>
<td>“Canada B (province not specified)”</td>
<td>2408</td>
<td>0.97</td>
</tr>
</tbody>
</table>

*20% sample of 1852*

Better codification of the information provided by the manuscripts: cases where a specific place of birth cannot be attributed and where “f” or “b” suggest the cultural origin rather than the place of birth

Regarding data linking, these variables may be useful in the stage of manual verification
2) B for “Bas” or for “British”? 

1852
Charles Gitty
1 year old
Born in “Canada b”
Religion: “Church of England”
Son of Robert Gitty, farmer

1881
Charles Gitty
29 years old
Born in “Quebec”
Origin: “English”
Religion: “Church of England”
Occupation: farmer
3) The homogeneity of French last names

- 10 000 French immigrants before 1760 vs. regular immigration of people from the British Isles (Charbonneau et al., 2000)

- Ratio of individuals to a single last name in the Canadian census of 1881 (Dillon, 2002):

  \[
  \begin{align*}
  &46:1 \text{ (French)} \\
  &22:1 \text{ (English)}
  \end{align*}
  \]

  More French-Canadians shared the same surname than English-Canadians did

- Focusing on uncommon names is inappropriate as linking strategy for the French-Canadian population. Additional variables should be considered.
  - However, homogeneity in other aspects as well (e.g. religion and place of birth)
  - Advantage: availability of the marriage registers of the catholic population living in Québec

- Under-representation of common names does not necessarily introduce bias into a linked sample
Example of verification using the marriage registers: linked sons
1852-1881

<table>
<thead>
<tr>
<th>Verification variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>X</strong></td>
<td>The linked individual (identified by the presence of his parents) does not have the same spouse in 1881</td>
</tr>
<tr>
<td><strong>X</strong></td>
<td>Someone else with the same name (different parents) is married with the woman that appears in 1881</td>
</tr>
<tr>
<td><strong>OK</strong></td>
<td>The linked individual (identified by the presence of his parents) has the same spouse in 1881</td>
</tr>
<tr>
<td><strong>OK</strong></td>
<td>The linked individual has the same spouse in 1881 but the information about 1 parent only is equal (remarriage of one of the parents)</td>
</tr>
<tr>
<td><strong>OK</strong></td>
<td>The linked individual did not marry/widowed during the interval and other members of the family present in 1852 appear also in 1881 to confirm the link</td>
</tr>
<tr>
<td>?</td>
<td>The linked individual was not living with his parents in the first census (e.g. apprentices)</td>
</tr>
<tr>
<td>?</td>
<td>More than 1 possible register found (rare...this implies that the son and parents have common names)</td>
</tr>
<tr>
<td>?</td>
<td>Register not found and no other members of the family are present to confirm the link</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1852 (Census)</th>
<th>1864</th>
<th>1879</th>
<th>1881 (Census)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linked son: François Patoine</td>
<td>Parents of the husband: Etienne Patoine Marguerite Dion</td>
<td>Previous wife: Philomène Beaulieu (d)</td>
<td></td>
</tr>
</tbody>
</table>
Example of verification using the marriage registers

1844
Couple: Narcisse Poitras & Thérèse Pelletier

Previous wife: Domithilde Vaillant (d)

Father remarries!

1852 (Census)
Couple: Narcisse Poitras & Thérèse Pelletier

Linked sons:
- Alfred Poitras
- Isidore Poitras

1864
Couple: Alfred Poitras & Marie Philomène Hetu

Parents of the husband:
- Narcisse Poitras
- Domithilde Vaillant (d)

1877
Couple: Isidore Poitras & Virginie Jalbert

Parents of the husband:
- Narcisse Poitras
- Thérèse Pelletier

1881 (Census)
Couple: Alfred Poitras & Philomène Etu
Conclusion

- Despite the limitations of the Canadian census of 1852, the 20% sample is a rich source of data about the population of the mid 19th-century. Combined with other censuses or with other sources of data it can provide valuable information about the Canadian population in an era of socioeconomic transformations.

- Being aware of the issues presented -the missing urban data, the meaning of the mentions in the column *place of birth* and the homogeneity of the French last names - is important to avoid bias in studies willing to use the sample of the Canadian census of 1852 for linkage purposes.

- Regarding data linking: it is possible to verify the validity of the links using the marriage registers (BALSAC), which are available for the catholic population living in Québec until the early 20th century.

References

