

Demosim's Population Projection Model: Updates and New Developments

by Éric Caron-Malenfant^{1,2}

Abstract

This paper provides an overview of the main new features that will be added to the forthcoming version of the Demosim microsimulation projection model based on microdata from the 2011 National Household Survey. The paper first describes the additions to the base population, namely new variables, some of which are added to the National Household Survey data by means of data linkage. This is followed by a brief description of the methods being considered for the projection of language variables, citizenship and religion as examples of the new features for events simulated by the model.

Keywords: Population projections, microsimulation, demography.

1. Introduction

Demosim is a microsimulation projection model developed and maintained by Statistics Canada. Its main purpose is to produce projections by various population characteristics for small geographies, including census metropolitan areas (CMAs). It starts with microdata from Canadian population censuses and projects the individuals in the population one at a time by having them “experience” various events over time until they die, emigrate or reach the end of the simulation. During the simulation, this initial population is augmented by immigrants and births. On the basis of assumptions concerning the probabilities of experiencing the simulated events, the model produces population tables that cross-tabulate the variables considered for future reference dates.³

Built essentially on a cost-recovery basis, Demosim has permitted the development of various projection products. These products include official population projections (such as *Projections of the Diversity of the Canadian Population, 2006 to 2031* (Caron-Malenfant et al. 2010) and *Population Projections by Aboriginal Identity in Canada, 2006 to 2031* (Caron-Malenfant and Morency 2011)),⁴ methodological and analytical papers,⁵ custom tabulations, and short-term projections used by Statistics Canada surveys. To date, these products have been developed using versions of Demosim based on data from the complete 2001 and 2006 censuses.

¹ Statistics Canada, Main Building 1708, 150 Tunney's Pasture Driveway, Ottawa, Canada, K1A 0T6, eric.caron-malenfant@statcan.gc.ca.

² The author would like to thank the rest of the Demosim team, on whose behalf he spoke at the 2014 International Methodology Symposium. In addition to the author, the following people worked on updating Demosim: Melanie Abeysondera, Jonathan Chagnon, Simon Coulombe, Karla Fox, Harry François, Nora Galbraith, Dominic Grenier, Chantal Grondin, René Houle, Mark Knarr, Stéphanie Langlois, Laurent Martel, Jean-Dominique Morency, Soumaya Moussa and Martin Spielauer. Thanks also to Marc Lachance for reviewing a preliminary version of this paper.

³ For a general description of the microsimulation methods used in population projections, see Van Imhoff and Post (1997) and Willekens (2011).

⁴ Please see these two publications for more information on the methodology of the previous version of Demosim (based on 2006), which forms the basis of the new version outlined here. The first version of the model (based on 2001) is documented in Bélanger and Caron-Malenfant (2005) and Bélanger et al. (2008).

⁵ For example, see Caron-Malenfant et al. (2011) and Spielauer (2014).

Since then, funding has been allocated to update Demosim with more recent data, in particular data from the 2011 National Household Survey (NHS).⁶ This paper provides an overview of some of the new features that will accompany this update. Since Demosim is being completely redesigned, only some of the new features being added are described here. Developments affecting Demosim's base population are covered first. Then, examples of changes in the methods and data sources used in some aspects of the modelling process itself are presented.

2. New developments related to the base population

The first developments in Demosim relate to its base population. The base population that will serve as the starting point for the new version of the model will be generated from 2011 NHS microdata. These microdata will undergo certain adjustments, including adjustments for census net undercoverage and incompletely enumerated reserves.

The new base population will include all the variables of the previous versions, the most important of which are: age, sex, immigrant status and period of immigration, generation status, detailed place of birth, visible minority group, Aboriginal identity and Registered Indian status, religion, mother tongue, marital status, labour force participation, and place of residence.

New variables will be added, however, to enhance the model's analytic potential. The first two are intended to support the production of detailed language projections: language spoken most often at home (English, French, English and French, other)⁷ and knowledge of official languages (French only, English only, English and French, neither English nor French). The base population will also include household headship status and census family reference person status, so that derived household and family projections by Aboriginal status can be produced. Two other new variables relate more specifically to immigrant populations. The first variable is citizenship, which will make it possible to identify whether people are Canadian citizens, and / or citizens of one or more other countries. The second variable is immigrant admission category. This will make it possible to identify whether immigrants admitted since 1980 entered Canada under the immigration program's economic component, under the family reunification program, as protected persons (refugees), or as other immigrants. In addition, the base population will include a new variable for Registered Indians: type of registration in the Indian Register (registration under subsection 6(1) or 6(2) of the federal *Indian Act* of 1985). This variable defines the rules for transmission of Registered Indian status from parents to children and is therefore a key factor in projecting the Registered Indian population.⁸

Note that the last two variables are not in the 2011 NHS database. They were added through data linkage. Immigrant admission category was added on the basis of a pre-existing linkage between the 2011 NHS and Citizenship and Immigration Canada administrative files for 1980 to 2011. Linkage was successful for 82% of the individuals who reported in the 2011 NHS that they were admitted to Canada as immigrants in 1980 or subsequently. For the 18% of records for which linkage failed, the information was imputed. The type of registration in the Indian Register was obtained from a pre-existing linkage between the 2011 NHS and the Indian Register, an administrative database maintained by Aboriginal Affairs and Northern Development Canada. The linkage was successful for 66% of the persons who reported in the 2011 NHS that they had Registered Indian status. The information was imputed for the remaining 34%.

⁶ Funding for the update is being provided by Aboriginal Affairs and Northern Development Canada and by Citizenship and Immigration Canada.

⁷ Multiple languages consisting of a non-official language and one or both official languages are assigned to one of the official language categories.

⁸ Type 6(1) Registered Indians can transmit Registered Indian status to their children without restriction, whereas Type 6(2) Registered Indians can do so only when the other parent also has Registered Indian status. This variable was imputed in the previous version of Demosim for a small portion of the Registered Indian population. The new development is that the information is now available for the entire Registered Indian population within Demosim's base population and that it is obtained mostly through record linkage.

It is worth noting that some of the variables projected in the past will be projected in greater detail with the new version of Demosim. One such variable is geography. The model already included Canada's 33 CMAs,⁹ the provinces and territories, and specific geographies for Aboriginal projections.¹⁰ For the purposes of projections of the language situation in Canada, other geographies will be added. First, the Montréal CMA will be divided into two regions that differ notably in their language profile: the island of Montréal and the rest of the CMA. Second, for the same reason, regions with a concentration of francophones will be created in the non-CMA part of Ontario and the non-CMA part of New Brunswick; they will consist of the census divisions in which at least 20% of the population in 2011 had French as their first official language spoken. The maps in the appendix show, for illustrative purposes, the resulting geographic infrastructure in Ontario and New Brunswick.

3. New developments related to certain projected events

To update the projected variables at the individual level, Demosim uses various methods and data sources to dynamically simulate events that together form the longitudinal histories of the individuals being projected. In addition to dynamically simulating events, the model imputes characteristics at fixed dates, again for the purpose of updating individual attributes.¹¹ The version of Demosim currently being developed will continue to model all the events modelled by previous versions, primarily fertility, mortality, immigration, emigration, internal migration, change of education level, religious mobility,¹² ethnic mobility of Aboriginal peoples,¹³ registration in the Indian Register, and departure of children from the parental home. It will also continue to impute marital status and labour force participation status to individuals annually.

However, additional components must be developed in order to properly simulate the newly added characteristics that are likely to change during a lifetime. For two of these components, those related to projecting the two new language variables and to citizenship, a brief description of the methods we plan to test is provided below.¹⁴ A third characteristic relates to the projection of households, but it is not presented here, as the methods are described in Jean-Dominique Morency's paper in this same publication.

Linguistic mobility has two distinct but complementary dimensions: intragenerational linguistic mobility (over an individual's lifetime) and intergenerational linguistic mobility (from parents to children). The probabilities associated with intragenerational language shifts¹⁵ will be determined through the linkages between the 2001 and 2006 censuses (for changes in home language and mother tongue) and between the 2006 and 2011 censuses (for knowledge of official languages).¹⁶ Using these sources, logistic regression models specific to the languages of origin and destination will estimate the probabilities of experiencing a transfer over a five-year period according to

⁹ In fact, it contained 34 CMAs, since the Ottawa-Gatineau CMA was divided into its Ontario portion and its Quebec portion.

¹⁰ These geographies include all Indian reserves and the Inuit Nunangat regions, where most Inuit live.

¹¹ The events and the other components of Demosim are managed by Modgen, a language used to program Demosim. For more information about Modgen, see the following page on Statistics Canada's website: www.statcan.gc.ca/microsimulation.

¹² Religious mobility refers to shifts in religion that may occur from generation to generation or over the course of an individual's life.

¹³ Ethnic mobility of Aboriginal peoples refers to shifts in reported Aboriginal identity that may occur from one generation to the next or over the course of an individual's life.

¹⁴ The methods presented were either under consideration or in development at the time of the 2014 Symposium.

¹⁵ The possible shifts are learning French and/or English, losing one's knowledge of French and/or English, changing one's language spoken most often at home and, more rarely, changing one's mother tongue. A change of mother tongue is possible because, under the census definition, the first language learned in childhood must still be understood by the respondent. For more on this subject, see Lepage (2011).

¹⁶ There are comparability issues between the 2011 Census and previous censuses, particularly the 2006 Census, with regard to mother tongue and language spoken most often at home (Houle and Corbeil 2013). For this reason, the estimates of the transition parameters for these two variables are based on the linkage between 2001 and 2006 rather than on the linkage between 2006 and 2011. The data on knowledge of official languages in the 2006 and 2011 censuses are comparable.

various characteristics projected by Demosim. The modelling of intergenerational linguistic mobility will involve probabilistically assigning to children born during the simulation a language profile based on their mother's profile.¹⁷ The probabilities will be estimated on the basis of 2011 NHS data, within which the youngest children will have been linked to the women most likely to be their mothers. Once this link is established, it will be possible to calculate probabilities of linguistic mobility between mothers and children via logistic regression models that will take into account the relevant variables projected by Demosim.

The citizenship module will essentially enable immigrants who are added to the population during the projection to acquire Canadian citizenship after their arrival.¹⁸ Two methods are currently under consideration. One method would take advantage of the linkages used in the development of the intragenerational linguistic mobility module. It would estimate the probability that immigrants who did not have Canadian citizenship at the beginning of the period would have acquired it five years later, according to certain key characteristics. The other method would be based solely on data from the 2011 NHS. It would derive Canadian citizenship acquisition rates from an estimate of the proportion of immigrants who had become citizens after three years, four years, five years, six years and so on, by place of birth, age and sex.¹⁹ There is a possibility that these two methods will be tested and compared.

Aside from the addition of new components, the Demosim update is also an opportunity to redesign the methods related to certain events already being modelled. One such event is religious mobility, which was simulated on the basis of data from the 2002 Ethnic Diversity Survey, which contained a question on the respondent's religion and another on his or her mother's religion up until the respondent was 14 years old. However, since the survey has not been conducted since 2002, an alternative method will be implemented in the new version of Demosim; as in the case of linguistic mobility, this method will have an intragenerational dimension and an intergenerational dimension. Intragenerational religious mobility will be modelled by applying a residual method to data from the 2001 Census and the 2011 NHS²⁰ adjusted for net undercoverage. This method will involve estimating the portion of the increase or decrease in the populations of the various religious denominations that remains unexplained after controlling for the effects of births, deaths, immigration, and emigration; this remaining unexplained portion is deemed to be the result of religious mobility. For denominations that have lost members through religious mobility, the estimated net losses will be divided by the initial population to generate 10-year net "exit" rates. The "leavers" will be distributed among the "gaining" denominations in proportion to the latter's net gains. Estimates based on previous censuses may be used to develop alternative hypotheses. Intergenerational religious mobility will be modelled by a method similar to the one used for intergenerational linguistic mobility, whereby the religion of mothers will be compared with that of their children in the 2011 NHS.

4. Conclusion

This paper aimed to provide an overview of the changes currently underway in the update to Demosim. While the paper helped illustrate the new potential that the availability of data sources based on linkages holds for microsimulation, space constraints made it possible to present only a brief description of some of the changes in the model. This update is an opportunity to re-estimate all of Demosim's parameters and to introduce other developments, some of which are fairly complex, particularly with regard to the many dimensions of projecting Registered Indian status. For these reasons, this paper should be regarded as only a preliminary introduction to some aspects of the methodology used in the new version of the model currently in development. The methodology will be described at greater length in forthcoming publications.

¹⁷ This strategy is employed because in Demosim, only the link between the child and his or her mother (not father) is created.

¹⁸ Citizenship at the time of admission will be assigned to immigrants by imputation.

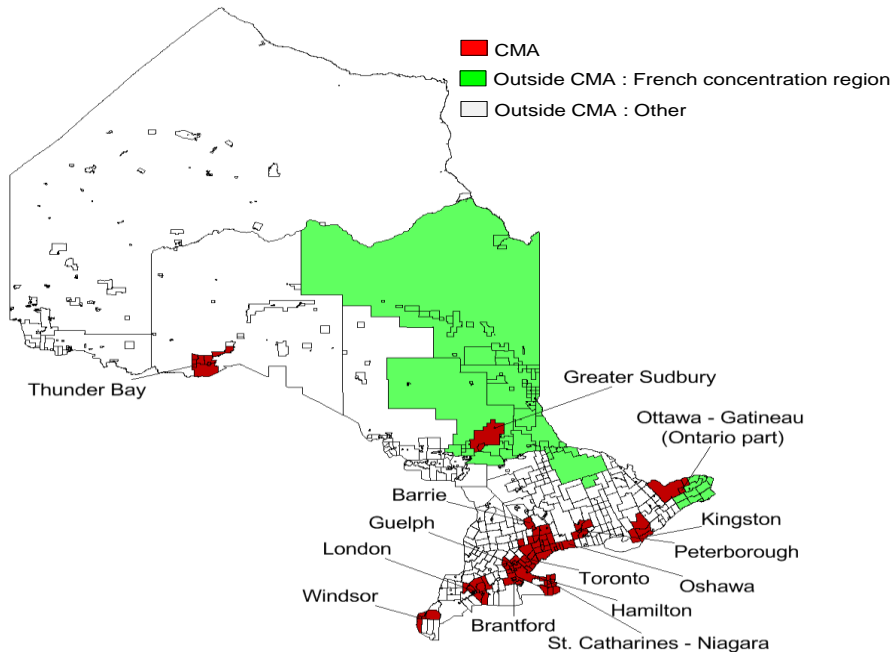
¹⁹ Calculating such rates would likely require certain adjustments to account for emigration and mortality of immigrants after their arrival.

²⁰ Note that the religion question is asked only every 10 years in the Canadian census.

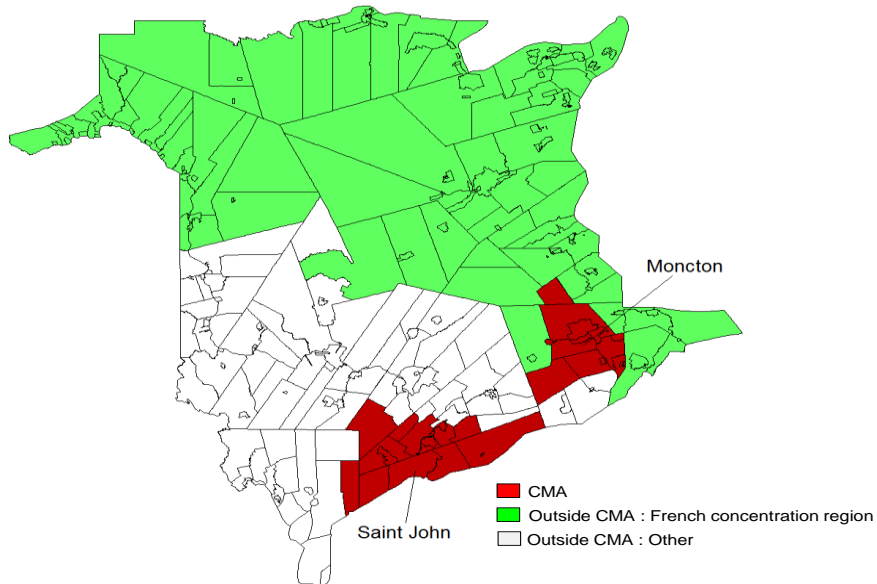
References

- Bélangier, A., and É. Caron-Malenfant. (2005), *Population Projections of Visible Minority Groups, Canada, Provinces and Regions, 2001-2017*, Statistics Canada Catalogue no. 91-541.
- Bélangier, A., É. Caron-Malenfant, L. Martel, and R. Gélinas. (2008), "Projecting ethno-cultural diversity of the Canadian population using a microsimulation approach", in *Proceedings of the Joint Eurostat/UNECE Work Session on Demographic Projections in Bucharest*.
- Caron-Malenfant, É., P. Dion, A. Lebel, and D. Grenier. (2011), "Immigration et structure par âge de la population du Canada : quelles relations?", in *Cahiers québécois de démographie* 40(2).
- Caron-Malenfant, É., A. Lebel, and L. Martel. (2010), *Projections of the Diversity of the Canadian Population, 2006-2031*, Statistics Canada Catalogue no. 91-551.
- Caron-Malenfant, É., and J.-D. Morency. (2011), *Population Projections by Aboriginal Identity in Canada, 2006-2031*, Statistics Canada Catalogue no. 91-552.
- Houle, R., and J.-P. Corbeil. (2013), *Methodological Document on the 2011 Census Language Data*, Statistics Canada Catalogue no. 98-314-X2011051.
- Lepage, J.F. (2011), "L'oubli des langues maternelles : les données du recensement sous-estiment-elles les transferts linguistiques?", in *Cahiers québécois de démographie* 40(2).
- Spielauer, M. (2014), "The relation between education and labour force participation of Aboriginal peoples: A simulation analysis using the Demosim population projection model", in *Canadian Studies in Population* 41(1-2).
- Van Imhoff, E. and W. Post (1997), "Méthodes de micro-simulation pour des projections de population", in *Population* 52(4).
- Willekens, F. (2011), "La microsimulation dans les projections de population", in *Cahiers québécois de démographie* 40(2).

Map 1. Demosim geographic infrastructure: Ontario



Map 2. Demosim geographic infrastructure: New Brunswick



Note: The CMAs (in red) are all projected separately. The portion of the non-CMA part of each of the two provinces that has a high concentration of francophones (in green) forms a single region for projection purposes, as does the non-francophone portion (in white). Note that some regions of Ontario and New Brunswick also have an on-reserve portion, not shown here.