THE CANADIAN MICROSIMULATION MODEL (LSD-C)

Content, Modules, and Some Preliminary Results

Alain Bélanger, Patrick Sabourin, Samuel Vézina, Guillaume Marois,
Kevin D'Ovidio, David Pelletier and Olivier Lafontaine

Notes de recherche / *Working paper*, n° 2018-01

CENTRE
URBANISATION CULTURE SOCIÉTÉ

INRS
UNIVERSITÉ DE RECHERCHE
A RESEARCH UNIVERSITY

www.ucs.inrs.ca

THE CANADIAN MICROSIMULATION MODEL (LSD-C)

Content, Modules, and Some Preliminary Results

Alain Bélanger, Patrick Sabourin, Samuel Vézina, Guillaume Marois,
Kevin D'ovidio, David Pelletier and Olivier Lafontaine

Institut national de la recherche scientifique
Centre - Urbanisation Culture Société
Montréal

February 2018

Alain Bélanger, Patrick Sabourin, Samuel Vézina, Guillaume Marois, Kevin D'Ovidio, David Pelletier and Olivier Lafontaine

Centre Urbanisation Culture Société

Institut national de la recherche scientifique (Montreal, Canada)

alain.belanger@ucs.inrs.ca

Centre - Urbanisation Culture Société
Institut national de la recherche scientifique
385, rue Sherbrooke Est
Montréal (Québec) H2X 1E3

Téléphone : (514) 499-4000
Télécopieur : (514) 499-4065

www.ucs.inrs.ca

This document can be downloaded without cost at:
http://www.ucs.inrs.ca/ucs/publications/collections/

TABLE OF CONTENT

Abstract

This document provides an overview of how the LSD-C projection model works. LSD-C is a microsimulation demographic projection model designed to project the Canadian population according to various characteristics. LSD-C stands for *Laboratoire de Simulations Démographiques-Canada*. The model has been developed by the authors at the *Institut national de la recherche scientifique* (INRS) in Montreal, Canada. More precisely, the LSD-C model simultaneously projects demographic (age, sex, place of residence, place of birth, generation and immigrant status), ethnocultural (mother tongue, language spoken at home, knowledge of official languages, visible minority status, religion) and socioeconomic (education, labour force participation) characteristics of the Canadian population. It allows for changes in individual characteristics over the life course as well as for intergenerational transfers of some characteristics from the mother to her child. This document describes the base population, data sources and methods for every projection modules contained in the model. The conceptualisation of the model events and the derivation of the corresponding parameters are also described. This document provides is a technical supplement to analytical reports and papers published in peer-reviewed journals presenting results generated by LSD-C.

Keywords: Microsimulation; Population projections; Canada; Ethnocultural diversity; Immigration; Labour force; Human capital.

Résumé

Le présent document donne un aperçu du fonctionnement du modèle de projections LSD-C. Il s'agit d'un modèle de projections démographiques par microsimulation conçu pour projeter la population canadienne selon plusieurs caractéristiques. LSD-C signifie *Laboratoire de Simulations Démographiques-Canada*. Le modèle est développé par les auteurs à l'*Institut national de la recherche scientifique* (INRS) à Montréal, Canada Précisément, le modèle LSD-C projette la population canadienne selon des variables démographiques (âge, sexe, lieu de résidence, lieu de naissance, statut de génération et statut d'immigration), ethnoculturelles (langue maternelle, langue le plus souvent parlée à la maison, connaissance des langues officielles, groupes de minorités visibles, religion) et socio-économiques (éducation, statut d'activité sur le marché du travail). En cours de simulation, le modèle permet des faire évoluer les différentes caractéristiques des cas simulés et permet les transferts intergénérationnels de certaines caractéristiques de la mère à ses enfants. Ce document décrit la population de base du modèle, les sources de données et la méthodologie utilisée pour chacun des modules de projection contenus dans le modèle. La conceptualisation des événements, des états dérivés et les paramètres utilisés pour la modélisation sont également décrits. Ce document constitue un document technique aux rapports d'analyses et aux articles publiés dans les revues scientifiques qui présentent des résultats de projections générés par LSD-C.

Mots-clés : Microsimulation; Projections démographiques; Canada; Diversité ethnoculturelle; Immigration; Main-d'œuvre; Capital humain.

Introduction

Most developed countries are facing similar demographic challenges: population aging, possible labour shortages and reduced population growth, if not population decline (Mahroum, 2001; Martin & Ruhs, 2011; Ruhs & Anderson, 2010; United Nations Population Division, 2012). Consequences of population aging and labour force population decline are posing a serious challenge to Western countries' policy makers regarding long-term sustainability of social security programs, healthcare and retirement plans (Anderson & Hussey, 2000; Auerbach & Lee, 2011; Esping-Andersen, 2000; OECD, 2000, 2013; Pammolli, Riccaboni, & Magazzini, 2012). In response to these challenges, most developed countries have increased immigration levels such that current immigration has reached historical levels (Australian Department of Immigration and Citizenship, 2012; Citizenship and Immigration Canada, 2012; Coleman, 2009; Jennissen, Van Der Gaag, & Van Wissen, 2006; Zaiceva & Zimmermann, 2008). International net migration is actually accounting for two thirds of the total population growth of developed nations (Population Reference Bureau, 2012). In consequence, the ethnocultural makeup, especially in large urban areas, is changing rapidly (Vertovec, 2007), forcing political decision-makers to deal with a number of challenges in the areas of urban development, labor market integration, health and social services, and public institutions (Meissner & Vertovec, 2014). In addition, with the replacement of older, less educated cohorts by younger, more educated cohorts, developed countries are also facing rapid changes in the educational composition of their labor force, generating potential mismatch between labour force demand and supply in terms of skills (Lutz, Butz, & KC 2014).

Microsimulation models constitute innovative tools to better inform "the public, but also to provide the means for policy makers and advisors to assess the impacts of their policies and programmes and to increase their effectiveness and cost-efficiency" (Wolfson, 2011). These models can account for the growing diversity of the population, whereas standard cohort-component projection models are not agile enough to handle multi-dimensional and dynamic forms of diversity since they are based on aggregated population estimates and rates. Indeed, the cohort-component projection method tends to ignore heterogeneity within groups and have practical limits in the number of events, states, and groups that can be modeled (van Imhoff & Post, 1998). Microsimulation enables the modeling of complex behaviours and the simultaneous projection of several characteristics and thus becomes the necessary technique to produce demographic and socio-economic projections in the context of super-diversity. This is important because immigration alters the population demographically, socio-economically, spatially, and culturally (Briggs, 2003; Vertovec, 2007). According to some authors, this process, described as a “Third Demographic Transition” can be a source of growing political conflict, cultural disunity and loss of community or cohesion (Alba & Foner, 2015; Coleman, 2006; Lee & Bean, 2010; Lichter, 2013). It poses challenges that should not be ignored or dismissed, as *laisser-faire* may lead to more fractionalisation and ultimately to more inequalities (Alesina, Devleeschauwer, Easterly, Kurlat, & Wacziarg, 2003; Alesina, Harnoss, & Rapoport, 2016; Patsiurko, Campbell, & Hall, 2012). The ultimate objective the research program of our team is to project population characteristics beyond age and sex so that generated results may guide policy makers in their decisions regarding immigration policy, social cohesion, labour market needs and changes, poverty and inequalities, and education and language skill formation needs.

This working paper presents a high-level technical overview of a microsimulation model that projects the Canadian population by several demographic, ethnocultural, and socioeconomic dimensions. This working paper describes the variables included in the microsimulation model. It also provides simple descriptive statistics for these variables. The *Modules* section describes in detail the modules of the model in terms of their events, their parameters and their outputs. Interdependence between the different modules is also described.

General Description of the Model

The presented model is named LSD-C, which stands for *Laboratoire de Simulations Démographiques – Canada*. The model has been developed by the authors at the *Institut national de la recherche scientifique* (INRS) in Montreal, Canada. This Canadian microsimulation projection model is theoretically grounded on the idea that forecasting socioeconomic changes can be realized through population forecasts following the *Demographic Metabolism* theory developed by Ryder (1965) and more recently by Lutz (2013). It puts forward the importance of cohort succession to explain social changes, building on the fact that the process of population replacement provides opportunity for social change through the constant flow of new people entering the social process and continuous withdrawals of older individuals through death.

LSD-C is a dynamic microsimulation projection model of the Canadian population coded using the Modgen programming language. Its point of departure is 2011 and its starting population is based on the 2011 National Household Survey public-use microdata file (NHS-PUMF). The use of a publicly available database as a starting population distinguishes this model from other models produced by Statistics Canada (such as Demosim) which are based on confidential microdata and whose code is not available to the general public. With LSD-C, the goal is to create an accessible and transparent model, so that other similar models may be built for other countries, thus making international comparisons possible.

The LSD-C model simultaneously projects demographic (age, sex, place of residence, place of birth, generation and immigrant status), ethnocultural (mother tongue, language spoken at home, knowledge of official languages, visible minority status, religion) and socioeconomic (education, labour force participation) characteristics of the Canadian population. It allows for changes in individual characteristics over the life course as well as for intergenerational transfers of some characteristics from the mother to her child.

The model is case-based, meaning that each individual is simulated separately from other individuals and that no interactions between individuals are allowed (with the exception of interactions between mother and children). The model is also dynamic and in continuous time, meaning that characteristics of individuals are modified continuously (in “real time”) through scheduled events whose timing is determined by the values of their specific input parameters at any given time during the projection period. That is in contrast to discrete-time models where characteristics are only changed within predefined time units (typically one year). The parameters were derived from a variety of sources and methods: censuses, surveys, vital statistics.

Each event depends on a selection of characteristics. Interregional mobility, for instance, varies with age, sex, language knowledge and use at home, religion, visible minority status, immigrant status, education and labour force participation. All events are thus highly interdependent.

The conceptualisation of the model events and the derivation of the corresponding parameters are further described below in the *Modules* section.

Modgen

Modgen is a meta language of C++ developed and maintained by Statistics Canada. Modgen and its documentation may be downloaded for free on the Statistics Canada website (Statistics Canada, 2017).

Modgen models are coded and implemented in the Microsoft Visual Studio software suite. Once compiled, a Modgen model takes the form of a stand-alone executable file (.exe) that allows manipulation of the model through a graphical user interface. From this interface the user is able to modify simulation parameters and create customized scenarios. More details on Modgen can be found on Statistics Canada's website[1]. For a full didactical presentation of Modgen, you may consult the book "Microsimulation and population dynamics: an introduction to Modgen 12" written by Bélanger and Sabourin (2017) published by Springer.

[1] http://www.statcan.gc.ca/eng/microsimulation/modgen/modgen

The Base Population

A base population is the database that is used as the starting point of a simulation: it is a synthetic representation of a population at a given time.

The base population in LSD-C is constructed by recoding and extracting relevant variables from the 2011 Canadian National Household Survey public-use microdata file (NHS-PUMF). The extracted information is saved in a flat file (CSV) readable by Modgen.

The NHS PUMF contains 887,012 records, which corresponds to 2.7% of the total Canadian population in 2011. When weighted, these records amount to a population of 32,852,323. In the NHS, weight values range from 32.4 to 460.6, for an average of 37.0.

From those 887,012 records, 4,725 had to be removed because of missing age information. The definitive base population in LSD-C then contains **882,287** records.

Each record includes values for 15 variables, all of which are further described in the next section. Those variables are age, sex, mother tongue, language spoken at home, knowledge of official languages, education, religion, age at immigration, place of residence (region), visible minority status, place of birth, weight, generation status and location of study.

Variables

All 15 variables are presented below in the order in which they appear in the CSV file (this order is further specified in parenthesis as the position number).

Age (position 0)

In the NHS-PUMF, age is provided in groups of two to five years in the variable *AGEGRP*. There is also an open age group at age 85+.

As a first step, age groups are broken down by randomly assigning a year of age based on the population distribution by age, sex and province found in Statistics Canada's demographic estimation for the 1st of July 2011[2]. A random number between zero and one is then added to the reassigned age in order to get a continuous age.

[2] This data can be found in Table 051-0001 on Statistics Canada's website. The estimations contained in this table include the effect of the census' net undercount as well as collective dwellings, which were not surveyed by the NHS. It should be noted that Statistics Canada updates from time to time the data found in this table. For the purpose of LSD-C, the table data was extracted at the end of 2015.

Code	Description	n	%
[0,100]	Continuous age	882,287	100

Sex (position 1)

The sex variable is directly recoded from the *SEX* variable in the NHS-PUMF.

Code	Description	n	%
0	Female	449,254	50.92
1	Male	433,033	49.08

Mother Tongue (position 2)

Mother tongue is derived from the variables *MTNEN*, *MTNFR* and *MTNNO* in the NHS-PUMF. English and French are the first two listed categories of the LSD-C mother tongue variable[3]. Languages other than the two official languages are grouped in two additional categories – *Others-Anglotrope* and *Others-Francotrope* – whose construction is further explained below.

In Canadian provinces other than Québec, most Allophones (people having a mother tongue that is neither French nor English) linguistically integrate to the English majority: most of them know English and, if they change the language they use in their home, it will usually be English. In Québec, however, both French and English act as integration languages and Allophones will lean towards one or the other based on their ethnocultural backgrounds. Allophones leaning towards English are labeled *Others-Anglotrope,* while Allophones leaning towards French are labeled *Others-Francotrope. Others-Francotrope* in Québec are usually Allophones with a Latin or Francophone background, either speaking a latin language or coming from a country part of the Francophonie.

To determine if Allophones of a given language group are Francotrope or Anglotrope, we use a binomial test to check if members of that group, when speaking an official language at home, mostly use French or English. When the binomial test was not significant, the language group was randomly assigned to the Francotrope or the Anglotrope category.

All Allophones outside Québec are considered to be in the category *Others-Anglotrope.*

[3] NHS question: “What is the language that this person first learned at home in childhood and still understands? If this person no longer understands the first language learned, indicate the second language learned.”

Finally, individuals having declared more than one mother tongue are randomly assigned one of the provided languages.

Code	Description	n	%
0	English	502,139	56.91
1	French	191,491	21.70
2	Others - Anglotrope	172,593	19.56
3	Others - Francotrope	16,064	1.82

Language Spoken Most Often at Home (position 3)

Language **most often** spoken at home[4] is derived from the variables *HLAEN*, *HLAFR* and *HLANO* in the NHS-PUMF. English and French are the first two categories of this variable. Languages other than the two official languages are grouped in a third category *Others*. Individuals having declared more than one language at home (29,979) are randomly assigned to one of the declared languages.

It should be noted that the NHS-PUMF also contains a variable indicating languages spoken **regularly** at home. This variable is not considered in the base population.

Code	Description	n	%
0	English	578,445	65.56
1	French	187,415	21.25
2	Others	116,427	13.20

Knowledge of Official Languages (position 4)

The knowledge of official languages[5] variable is directly recoded from the variable *KOL* in the NHS-PUMF.

Code	Description	n	%
0	English only	594,751	67.41
1	French only	111,192	12.60
2	English and French	160,631	18.21
3	Neither English nor French	15,713	1.78

[4] NHS question: "What language does this person speak most often at home?"
[5] NHS question: "Can this person speak English or French well enough to conduct a conversation?"

Immigrant Status (position 5)

The immigrant status[6] variable is directly recoded from the *IMMSTAT* variable in the NHS-PUMF.

Code	Description	n	%
0	Non-Immigrant	685,145	77.66
1	Immigrant	187,637	21.27
2	Non-Permanent Resident	9,505	1.08

Education (position 6)

The education variable describes the highest level of educational attainment and is derived from the variable *HDGREE* in the NHS-PUMF. It contains four categories:

1. *No diploma*: includes individuals without at least a high school diploma as well as individuals aged 0 to 14.
2. *High school diploma*: includes individuals with a high school degree or equivalency certificate.
3. *College diploma*: includes individuals with an apprenticeship or trade certificate or diploma; college, CEGEP or other non-university certificate or diploma as well as university certificate or diploma below bachelor level.
4. *University diploma*: Includes all individuals with a university diploma or certificate at or above bachelor level.

Missing values for this variable (3,596) were randomly reassigned based on the weighted distribution of non-missing educational attainment.

Code	Description	n	%
0	No diploma	294,303	33.36
1	High School	185,280	21.00
2	College	244,293	27.69
3	University	158,411	17.95

Religion (position 7)

The religion[7] variable is derived from the variable RELIGDER in the NHS-PUMF.

[6] NHS question: “Is this person now, or has this person ever been, a landed immigrant? A ‘landed immigrant’ (permanent resident) is a person who has been granted the right to live in Canada permanently by immigration authorities.”

Missing values for this variable (4,806) were randomly reassigned based on the weighted distribution of individuals with non-missing religion.

Code	Description	n	%
0	No religion	212,387	24.07
1	Christian	591,341	67.02
2	Hindu	13,598	1.54
3	Sikh	12,607	1.43
4	Muslim	29,098	3.30
5	Jewish	8,793	1.00
6	Buddhist	9,969	1.13
7	Other Religions	4,494	0.51

Age at immigration (position 8)

The age at immigration variable is derived from the *YRIMM* (year at immigration) and *AGEIMM* (age at immigration) variables in the NHS-PUMF. Both variables are used in order to extract a maximum amount of information as the values of both variables do not always overlap (*AGEIMM* comes in groups of five years of age plus one open age group, while *YRIMM* sometimes has a one year precision).

The continuous age at immigration is randomly assigned in a range that is coherent with the assigned continuous age, as well as the value of *AGEIMM* and *YRIMM*.

Missing values for this variable (12,929) were randomly reassigned based on the weighted distribution of individuals with non-missing age at immigration (*AGEIMM*). Canadian-born individuals are assigned an age at migration of zero (the value does not really matter as these individuals are distinguished with the immigration status variable).

[7] NHS question: "What is this person's religion? Indicate a specific denomination or religion even if this person is not currently a practising member of that group. For example, Roman Catholic, United Church, Anglican, Baptist, Lutheran, Muslim, Presbyterian, Pentecostal, Jewish, Buddhist, Hindu, Sikh, Greek Orthodox, etc."

Code	Description	n	%
[0,100]	Continuous age	882,287	100

Place of residence (position 9)

The geography of the model is divided in 24 areas derived from the province (*PR*) and the census metropolitan area (*CMA*) of residence.

Categories “Rest of [province]” include individuals residing in the province outside of the predefined CMAs. Hence, “Rest of” categories include residents of rural regions as well as residents of census agglomerations and small CMAs. As an example, “Rest of Québec” includes individuals from rural zones (ex. Gaspésie), census agglomerations (ex. Drummondville) and small CMAs (ex. Saguenay).

The original CMA variable in the NHS-PUMF contains 35 categories. Small CMA with less than 500,000 inhabitants and less than 15% visible minorities were merged in the “Rest of” categories. Small provinces with few visible minorities still constituted a region of their own (Newfoundland and Labrador for instance). However, the three federal territories (Yukon, Northwest territories, and Nunavut) were combined into a single region because of their small population size.

Code	Description	n	%
0	Newfoundland and Labrador	12,642	1.43
1	Prince Edward Island	3,333	0.38
2	Nova Scotia	23,119	2.62
3	New Brunswick	18,726	2.12
4	Québec (CMA)	20,450	2.32
5	Montréal	105,373	11.94
6	Gatineau	8,220	0.93
7	Rest of Québec	78,810	8.93
8	Ottawa	25,049	2.84
9	Oshawa	9,614	1.09
10	Toronto	151,621	17.18
11	Hamilton	18,864	2.14
12	Kitchener-Cambridge-Waterloo	12,769	1.45
13	Windsor	8,271	0.94
14	Rest of Ontario	114,026	12.92
15	Winnipeg	19,502	2.21
16	Rest of Manitoba	11,720	1.33
17	Saskatchewan	25,619	2.90
18	Calgary	32,579	3.69
19	Edmonton	30,792	3.49
20	Rest of Alberta	31,407	3.56
21	Vancouver	63,043	7.15
22	Rest of British Columbia	53,915	6.11
23	Territories	2,823	0.32

Visible Minority Status (position 10)

The Visible Minority Status variable combines information from the *ABOID* (aboriginal identity groups) and *VISMIN* (visible minority groups) variables in the NHS-PUMF. In Canadian official statistics, visible minority status does not apply to Aboriginals who are then always identified as “not a visible minority”. We combined all aboriginal subgroups (First Nations, Métis, Inuit and/or Registered Indians) into a single group and treated it as other, official visible minority groups. All other individuals remain in the group identified by *VISMIN*.

Individuals having declared to belong to more than one visible minority (4,514 cases) were randomly assigned to one of the declared visible minorities (except to the category “Not a visible minority”).

Records with missing information (3,082) were randomly reassigned based on the weighted distribution of individuals with non-missing visible minority status.

Code	Description	n	%
0	South Asian	44,897	5.09
1	Chinese	39,053	4.43
2	Black	25,549	2.90
3	Filipino	17,784	2.02
4	Latin American	10,623	1.20
5	Arab	10,348	1.17
6	Southeast Asian	8,598	0.97
7	West Asian	5,565	0.63
8	Korean	4,482	0.51
9	Japanese	2,249	0.25
10	Visible Minority (not included elsewhere)	2,698	0.31
11	Aboriginal	36,658	4.15
12	Not a visible minority	673,783	76.37

Place of Birth (position 11)

The Place of Birth is directly recoded from the *POB* variable in the NHS-PUMF.

Records with missing information (7,949) were randomly reassigned based on the weighted distribution of individuals with non-missing place of birth information. Individuals with missing place of birth information but who had indicated they were non-immigrants were assumed to have been born in Canada.

Code	Description	n	%
0	Canada	683,080	77.42
1	United States of America	8,196	0.93
2	Central America	4,509	0.51
3	Jamaica	3,180	0.36
4	Other Caribbean and Bermuda	6,383	0.72
5	South America	8,605	0.98
6	United Kingdom	15,438	1.75
7	Germany	4,371	0.50
8	Other Northern and Western Europe	9,437	1.07
9	Poland	4,143	0.47
10	Other Eastern Europe	10,053	1.14
11	Italy	6,834	0.77
12	Portugal	3,755	0.43
13	Other Southern Europe	6,319	0.72
14	Eastern Africa	4,321	0.49
15	Northern Africa	5,318	0.60
16	Other Africa	4,436	0.50
17	West Central Asia and the Middle East	13,525	1.53
18	China	17,140	1.94
19	Hong-Kong, Special Administrative Region	6,190	0.70
20	Other Eastern Asia	6,858	0.78
21	Philippines	14,416	1.63
22	Other Southeast Asia	8,030	0.91
23	India	16,372	1.86
24	Pakistan	4,511	0.51
25	Other Southern Asia	5,396	0.61
26	Oceania and Others	1,471	0.17

Weight (position 12)

Weights were taken from the *WEIGHT* variable in the NHS-PUMF and then adjusted according to age, sex and province of residence to account for census net undercoverage and for residents living in collective dwellings.

Data in the NHS-PUMF is derived from the “raw” confidential census data, which is not corrected for net undercoverage and only apply to private households. Statistics Canada does postcensal estimates of net undercoverage and regularly publishes updated estimates of the entire population, including both private and collective dwellings.

The NHS-PUMF data was calibrated to estimates of the Canadian population on the 1st of July 2011. Data used for weight calibration can be found on Statistics Canada’s website[8].

Range	Description	n	%
[32.39361116, 460.61]	Weight (uncorrected)	882,287	100
[4.591702, 491.1391]	Weight (corrected)	882,287	100

Generation Status (position 13)

The Generation Status variable is derived from the age at immigration (see description above) and the *GENSTAT* variable in the NHS-PUMF. The age at immigration is used to distinguish between generations 1 and 1.5. Generation Status includes five categories:

- Generation 1: Born outside Canada from two parents born outside Canada. Arrived in Canada at or after age 15.
- Generation 1.5: Born outside Canada from two parents born outside Canada. Arrived in Canada at age 14 or before.
- Generation 2: Born in Canada from two parents born outside Canada.
- Generation 2.5: Born in Canada from one parent born in Canada and one parent born outside Canada.
- Generation 3+: Born in Canada from parents born in Canada.

Non-permanent residents are all considered to be Generation 1. Records with missing information (2,642) were randomly reassigned based on the weighted distribution of individuals with non-missing Generation Status information and according to Immigrant status information: immigrants were reassigned to Generation 1 or 1.5 whereas non-immigrants were reassigned to Generation 2, 2.5 or 3.

[8] Table 051-0001. See footnote 2 in the description of the age variable.

Code	Description	n	%
0	Generation 3	530,380	60.11
1	Generation 2.5	69,619	7.89
2	Generation 2	85,146	9.65
3	Generation 1.5	54,088	6.13
4	Generation 1	143,054	16.21

Location of Study (position 14)

Location of Study is a dichotomous variable derived from the *LOCSTUD* in the NHS PUMF indicating the geographical location of highest educational attainment. Since this variable is only useful for describing immigrants' location of study, all individuals born in Canada are defined as having Canada as their Location of Study.

Records with missing information were reassigned based on the following rule: immigrants arrived before the age of 25 were reassigned to the "Canada" category whereas those arrived at age 25 or after were reassigned to the "Outside Canada" category.

Code	Description	n	%
0	Canada	788,864	89.41
1	Outside Canada	93,423	10.59

Example of a Record in the Base Population

Below is an example of a typical record found in the base population file:

55.3196,1,2,2,0,1,2,1,34.40294,10,1,18,35.46220050130,4,1

The record is from a 55 years old man residing in Toronto. He has earned his college degree outside of Canada and immigrated to Canada at the age of 34 (he is therefore Generation 1). He was born in China and is of Chinese visible minority. He is Christian. His mother tongue and home language are neither French nor English, but he is able to speak English (but not French). His sampling weight is 35.46.

Modules

Typical Modgen models are based on two main elements. The first element is the main model file, which is the driver of the simulation. This file includes control flow structures looping through all cases in the simulation, creating and starting the simulation of all actors (a case is build from a record in the base population). The second main element of the model is the *Person* class (essentially a C++ class) which contains all of the elements required for the simulation of a case: event functions (rules for changing state variables and creating new actors), state variables (age, sex, etc.) and output tables. The description of the *Person* class is usually divided in several files or "modules" according to functional criteria (mortality, fertility, migration, etc.).

The main module is usually called the *PersonCore* module, as it is the only one automatically created by the Modgen model Wizard. Technically, all the code pertaining to the description of the actor (the *Person* class) could be contained in this single file, but this would prove impractical for big models. The extension of Modgen files (main file or module files) is .mpp.

Each of the modules also contains a declaration of Modgen-specific data structures (range, classification, etc)[9] and parameters. Parameters declared in the modules are containers whose values are externally defined in a separate data file. Parameter values are defined in a separate file so that users may build alternative scenarios through the user interface without having to alter the source code.

LSD-C mainly comprises the following files:

- LSD-C.mpp: This is the main file. It has been modified to include the simulation of future immigration (this will be discussed later on).
- PersonCore.mpp: This is the main module. In LSD-C, this file is mostly used for basic simulation parameters and time keeping (age and calendar time).
- ModgenInputMicroDataModule.mpp: This module only serves the purpose of reading data from the base population file.
- Tables.mpp: This module contains all of the output tables generated by the model.
- Modules (14): Education.mpp, Emigration.mpp, Fertility.mpp, HomeLanguage.mpp, Immigration.mpp, KnowledgeOfficialLanguage.mpp, Labour.mpp, Language.mpp, Literacy.mpp, Mobility.mpp, Mortality.mpp, MotherTongue.mpp, Religion.mpp, VisibleMinority.mpp

[9] See Modgen documentation for details

Each of these modules will now be described in more details.

The Main Simulation File (LSD-C.mpp)

The Main Simulation file is not a module *per se*, as it does not contain any information on the *Person* class. Rather, it is the main file and driver of the model: it creates the actors and simulates each one of them in succession, one at a time. The main file contains two functions: *Simulation* and *CaseSimulation*.

The *Simulation* function contains a simulation loop whose number of iterations is specified in the scenario. Each iteration of this loop reads a record from the base population file and creates an actor (in C++ parlance, an actor object is instantiated from the *Person* class). Simulation is then launched by calling the *CaseSimulation* function, which in turn calls the *start*() function which will initialize all of the actors state variables. After an actor dies or leaves the simulation (through emigration for instance), the loop iterates again and move on to the next record, and so on until the end of the base population file.

The Main Module (PersonCore.mpp)

The PersonCore module contains basic characteristics (state variables) of the actor such as age and sex. It also keeps track of calendar time.

The module also contains the definition of a very important function, the *Start* function, which is called at the creation of a new actor. This function initializes all the states of the actor.

Ranges and classifications	
AGE	Age range. Can take integer values between 0 and 100.
YEAR	Simulation year range. Can take values between 2011 and 2060.
SEX	Sex of actor. Male or female.

In addition to these data structures, other classifications and partitions are defined for table production or parameter definition. These can be found in the program code.

Parameters are listed below.

Parameters	
yearENM	Year of the NHS (2011)
SimEnd	Last year of simulation
nPopBase	Number of cases to read in the population file
sexratio	Sex ratio at birth (105/205)
CohortOnly	Allows the simulation of a specific cohort (birth cohort or immigration cohort)

The preceding Modgen elements are declared outside the *Person* class. State variables and event function declarations are made inside de *Person* Class.

Main state variables	
alive	LOGICAL. It is turned to FALSE after a death event.
sex	SEX. Sex of the actor.
age_int	AGE. Age of the actor in completed years.
BasePop	LOGICAL. Indicates if an actor was taken from the base population or was created during the simulation through birth or immigration.
year	INTEGER. Calendar year during the simulation.
cohort	INTEGER. Year of arrival for immigrants and year of birth for native.

The *Person* class also includes declaration of event functions. Event functions are used to change the value of state variables.

Events and functions	
Birthday	Increments the age of the actor by one year
New year	Increments the calendar year by one unit
Start	This function is very important. It is called to create and start a new actor in the simulation. It initializes all of the actor's state variables according to the type of actor that is being simulated (an actor taken from the base population, an actor born during the simulation or an immigrant actor arriving during the simulation).
Finish	This function is called to terminate the simulation of an actor, typically after a mortality or an emigration event.

The Education Module (Education.mpp)

Highest educational attainment is taken directly from the base population file for immigrants 25 years and older and Canadian-born 30 years and older. Immigrants aged less than 25 and Canadian-born aged less than 30 years old are assumed to have an incomplete education and are thus reassigned a new educational pathway based on the results of a multinomial logit regression model and distributions of age at graduation. Different age thresholds are set according to immigration status to maximize the use of observed data on education. A significant share of Canadian-born individuals earn university degrees beyond the age of 25 years old. Hence, it is pertinent to model the schooling pathways up to 30 years old. As for immigrants, in keeping the threshold at 30 years old, more than half of observed data on educational attainment would we lost. Indeed, the age structure of the foreign-born population comprises mainly individuals aged 25 to 34 years. Therefore, the threshold is set at 25 years old to reduce the proportion of imputed data.

The multinomial analysis is stratified by region and sex and includes cohort, visible minority, mother tongue and generation status as independent variables. The cohort variable is included in the analysis so that hypotheses on future educational expansion may be formulated.

Analyses were performed on the 2011 NHS confidential microdata, which provided a larger sample size than the NHS-PUMF and yielded more robust results in smaller regions.

Newborns and actors with incomplete educational history are assigned a highest educational attainment upon the start of their simulation. Educational history is reconstructed from distributions of graduation ages derived from the Canadian National Graduates Survey. Specific graduation ages are assigned for all diplomas (high school, college and university). Additional rules are included so that graduation ages stay coherent: graduation from high school can only occur at least two years before graduation from college, and graduation from college must occur at least three years before graduation from university.

Ranges and classifications	
EDUCATION	There are four possible levels of educational attainment: no diploma, high school, college or university (see the description of the education variable in the base population for more detail)

Parameters included are regression coefficients from the multinomial logit as well as age at graduation distributions.

Parameters	
Coefficients	Coefficients by visible minority status, mother tongue, generation status and cohorts are provided for all strata (sex and region) and education levels.
Age distributions	Distribution for age at diplomation are included for every level of education (high school, college and university)

The education module has a single relevant state variable.

Main state variables	
education	EDUCATION. May take any of the four values listed in the EDUCATION classification.

Events and functions	
Highest degree event	Determines the highest education attainment of the actor and the age at graduation for all degrees. This function is run at birth for actors born in the model, and upon the start of the simulation for immigrant under 25 years old and Canadian-born under 30 years old in the base population.
Secondary	Changes the value of the education state variable to high school
College	Changes the value of the education state variable to college
University	Changes the value of the education state variable to university

The Emigration Module (Emigration.mpp)

Emigration is modeled as an annual risk of out-migration. Rates of out-migration are estimated by Statistics Canada and vary according to age, sex and province of residence.

Relative risks with respect to country of birth are also incorporated into the model. It is well known that a large share of emigration is due to recent immigrants moving back to their home country. To estimate relative risks according to region of birth, we gather the observed number of immigrants from India, the Philippines, the rest of Asia, Africa, Latin America and the Caribbean's and the rest of the world in the 2001 Canadian census. These numbers are then compared to the number of immigrants received five years prior

(1996) according to Citizenship and Immigration Canada. The resulting probability of being in Canada five years after immigration is then used in LSD-C to calibrate relative risks according to region of birth.

The emigration module also manages emigration time of non-permanent residents, whose duration of stay in Canada is set to exactly one year. The inclusion of non-permanent residents is further described in the immigration section.

An emigration event terminates the simulation of the out-migrating actor.

Ranges and classifications
The emigration module does not contain any classification of its own.

Parameters	
Emigration hazards	Emigration hazards vary according to sex, age and province of residence.
Relative risks	Risks of out-migration are modulated according to country or region of birth: Canada, India, the Philippines, the rest of Asia, Africa, Latin America and the Caribbean's and the rest of the world.

Main state variables	
emigration	LOGICAL. A flag that identifies emigrants. Mainly used for tabulation.

Events and functions	
emigration	Calculates time before emigration. If emigration occurs, turns emigration state variable to TRUE and removes actor from simulation.

The Fertility Module (Fertility.mpp)

Fertility in LSD-C is modeled in three steps.

In the first step, age-specific fertility rates by province of residence are taken from vital statistics and standardized so that the TFR is equal to 1 in any given province. These curves provide the tempo information of fertility.

Relative risks were also calculated using the 2011 NHS confidential microdata file. A logistic regression on the probability of having a child aged 0 was computed with respect to generation status (1, 1.5 or 2+), visible minority status, religion and education as independent variables and age as a control variable. Risks relative to the population as a whole were computed using the *contrast* function in Stata.

Finally, the model was calibrated so that the number of births in the model corresponded to the observed number of births in each province as given by vital statistics. The calibration year was 2012.

After the occurrence of a fertility event, a new actor is created and simulated. This new actor is linked to its mother through a Modgen link, a programming element akin to a C++ pointer. This link allows for the intergenerational transfer of characteristics between mother and child, such as place of residence, mother tongue, religion, visible minority status, etc.

In Modgen, waiting time before the next event is recalculated as soon as the event has occurred, so that a female actor would be at risk of giving birth immediately after a fertility event. This was prevented in the model by allowing female actors to give birth only once per year of age.

Ranges, classifications and links	
AGEFERTILE	This range determines the ages at which a female actor is able to give birth, between 15 and 49.
Links	Modgen links are created between the mother and her children (one link goes from the mother to her child, and a vector of links goes from the children to their mother). See the developer's guide for more information on links

Parameters	
Fertility hazards	Baseline fertility hazards vary according to age and province of residence.
Relative risks	Fertility risks are modulated according to generation status (1, 1.5 or 2+), visible minority status, religion and highest educational attainment.

Main state variables	
fertileyear	LOGICAL. Flag that prevents actor from giving birth twice in any given year of age.

Events and functions	
fertility	Calculates time before a fertility event. If a birth occurs, a new actor is created and simulated (its Start function is called). This new actor is linked to its mother actor through a Modgen link. This link allows the transmission of characteristics between mother and child (this is done in the Start function of the child).

The Mother Tongue Module (MotherTongue.mpp)

The only purpose of the mother tongue (MT) module is to simulate intergenerational language shifts. Intergenerational language shifts are defined to occur when a child acquires a MT that is different from his mother's MT.

At birth, an actor is first assigned the MT of its mother. The actor is then instantaneously and immediately at risk of changing its MT: based on a Monte Carlo experiment, a MT state value is then reassigned. This operation is performed by using origin-destination matrices (mother's MT to actor's MT) that vary according to the mother's home language, generation status and region of residence. These origin-destination matrices were extracted from the 2011 NHS confidential microdata file using the own-children method. Note that the region of residence used in all language modules is a 7-category aggregation of the place of residence variable, *regionlang*, defined in the mobility module.

Following the intergenerational language shift event, home language and knowledge of official languages are modified to insure linguistic coherence: for instance, an actor having French as its starting MT but who shifted to English will consequently have its home language set to English and his knowledge of official languages set to English only.

After this one-time event, the mother tongue of an actor is fixed for the rest of the simulation.

Ranges, classifications and links	
MOTHERTONGUE	This classification has four possible values: English, French, Other-Anglotrope and Other-Francotrope. See section on base population for further details.

Parameters	
Mother tongue origin-destination matrix	Origin (mother's MT) – destination (actor's MT) matrices that vary according to the mother's region of residence, home language and generation status.

Main state variables	
mothertongue	MOTHERTONGUE. Contains the information on the actor's mother tongue. May only change at birth.

Events and functions	
Intergenerational language shift	Determines if an actor has a MT that is different from its mother's MT. Subsequently modifies all language characteristics accordingly.

The Acquisition of Official Languages Module (KnowledgeOfficialLanguages.mpp)

This module simulates the acquisition of French and English in Canada. The acquisition of each of the two languages is modeled separately.

Acquisition rates are based on survival curves derived from the 1991 to 2006 censuses[10]. Survivors are defined as actors having not yet learned one or both of the official languages. The survival curves are based on age for the Canadian-born or on time since migration for immigrants. They vary according to mother tongue, region of residence and, additionally in the case of immigrants, age at immigration. Rates of acquisition are calculated each year based on the simple survivorship formula 1-($Survivors_{t+1}$/$Survivors_t$).

In addition to official language acquisition during the lifecourse, the module also takes two additional phenomena into account: the instantaneous acquisition of languages "at birth" and the forgetting of French after high school in provinces other than the province of Québec.

The first phenomenon is an artifact of the census instructions provided to respondents, which ask parents to indicate the official languages that their children are currently learning if they haven't acquired speech yet. As a consequence, children may be considered to "know" French and English "at birth". This "instantaneous" acquisition of French and English is taken care of by an event occurring at birth and only once, immediately following the intergenerational language shift event.

The second phenomenon occurs in provinces other than Québec, where a significant proportion of Anglophones and Allophones having learned French in school "forget" this language after high school. The module explicitly takes this forgetting into account so

[10] For more details on the derivation of these survival curves, see Patrick Sabourin's PhD thesis: *Projections démolinguistiques des populations francophones, anglophones et allophones au Canada: Une analyse par microsimulation.*

that Non-Francophone Canadians outside Québec show a net loss of French knowledge after age 15.

Ranges, classifications and links	
KOL	Knowledge of official languages classification. It has four possible values: English only, French only, French and English and Neither French nor English. See section on base population for further details.

Parameters	
Acquisition survival curves: Canadian-born	Survivors are actors who don't know one or both of the official languages at age *t*. Survivorship at time 0 is not equal to 1, as some actors learn French or English "at birth" (see explanation above). Survival curves vary according to acquired language (French or English), mother tongue and region of residence.
Acquisition survival curves: Immigrants	Survivors are actors who don't know one or both of the official languages at a duration *t* following migration. Survival curves vary according to acquired language (French or English), mother tongue, age at migration and region of residence.

Main state variables	
kol	KOL. Contains the information about the actor's knowledge of official languages.
kolfr and kolen	LOGICAL. Indicates if an actor knows French (kolfr) or English (kolen). Actors knowing an official language include actors knowing only that official language as well as bilingual actors.

Events and functions	
Kolen	This event simulates the acquisition of knowledge of English during the lifecourse.
Kolfr	This event simulates the acquisition and the loss of knowledge of French during the lifecourse.
Kolbirth	This event simulates the acquisition of French and English at birth. Occurs only once immediately after the intergenerational language shift event.

The Language Most Often Spoken at Home Module (HomeLanguage.mpp)

The module for the language most often spoken at home (thereafter the *home language* module) simulates changes in the language used at home occurring during the lifecourse. Changes in the language used at home are particularly common for children of immigrants and linguistic minorities, as many of them gradually stop using their mother tongue at home, for instance through linguistic assimilation in school or through exogamy after leaving the family home. When an actor starts predominantly using at home a language that is different from its mother tongue, it is said that the actor has performed a language shift.

Language shift rates are based on survival curves extracted from the 1991 to 2006 censuses[11]. Survivors are defined as actors still predominantly speaking their mother tongue at home. The survival curves are based on age for the Canadian-born and on time since migration for the immigrants. They vary according to mother tongue, region of residence and, in the case of immigrants, age at immigration. Rates of language shifts are derived each year based on the simple survivorship formula $1-(Survivors_{t+1}/Survivors_{t})$.

Like interregional migrations, language shifts are modeled in two separate steps. First, a time to language shift is probabilistically set based on the provided survival curves. If a language shift actually occurs, a destination language is chosen according to an origin-destination matrix linking the mother tongue of the actor to its new language used at home. Whereas most shifts are made towards English in Canada, they are generally divided between English and French in Québec, as both languages attract a significant number of Allophones. Origin-destination matrices vary according to mother tongue, region of residence and immigrant status.

To be at risk of performing a shift toward English or French, an actor must first be able to speak the language towards which it is shifting. Since rates of language shift are calculated for the population as a whole and independently of the destination language, they have to be adjusted to take the population at risk into account. For instance, actors who don't know English are not at risk of making a language shift towards English. Since the population at risk is smaller than the whole population on which the rates are based, rates of language shift have to be increased accordingly[12].

[11] For more details on the derivation of these survival curves, see Patrick Sabourin's PhD thesis: *Projections démolinguistiques des populations francophones, anglophones et allophones au Canada: Une analyse par microsimulation.*

[12] Details for the adjustments of language shift rates are available in Patrick Sabourin's PhD thesis: *Projections démolinguistiques des populations francophones, anglophones et allophones au Canada: Une analyse par microsimulation.*

The census questions for mother tongue and language most often spoken at home are identical for children who are still unable to speak (they should both correspond to the language most often spoken to the child). Nevertheless, many parents still provide different answers to the two questions: they may for instance indicate Arabic as the mother tongue and French as the language most often spoken at home, even though both should be the same. As a consequence, some children may be considered to have performed a language shift "at birth", a phenomenon that is for the sake of the projection similar to the acquisition of French or English «at birth» that was noted in the knowledge of official languages module. Although the sociological meaning of this is unclear, language shifts «at birth» still have to be modeled. A special "Instantaneous" language shift at birth event is therefore added and occurs only once, immediately following the intergenerational language shift and the instantaneous acquisition of official language events.

Ranges, classifications and links	
HOMELANG	Language most often spoken at home classification. Can take three values English, French or Other. See section on base population for further details.

Parameters	
Language shift survival curves: Canadian-born and non-Allophone immigrants	Survivors are actors who still use predominantly their mother tongue at home at age *t*. Survivorship at time 0 is not equal to 1, as some actors make a language shift “at birth” (see explanation above). Survival curves vary according to mother tongue and region of residence.
Language shift survival curves: Allophone immigrants	Survivors are actors who still use predominantly their mother tongue at home at a duration *t* following migration. Survival curves vary according to age at migration and region of residence.
Language shift origin-destination matrices	Origin (mother tongue) – destination (home language) matrices varying according to region of residence and immigrant status. Particularly useful in Québec where language shifts are split between English and French.

Main state variables	
homelang	HOMELANG. Indicates the actor’s home language.
homelangmother	HOMELANG. Indicates the home language of the actor’s mother. Used in the derivation of intergenerational language shifts.

Events and functions	
LanguageShift	This event simulates the occurrence of a language shift during the lifecourse. If a language shift occurs, a new home language is chosen randomly according to an origin-destination matrix.
LanguageShiftBirth	This event simulates the occurrence of a language shift at birth. If a language shift occurs, a new home language is chosen randomly according to an origin-destination matrix.

The Derived Language Variables Module (Language.mpp)

The Derived Language Variables module is not a module *per se*, as it doesn't drive any event by itself. Its sole purpose is to generate state variables derived from the other three language variables (mother tongue, home language and knowledge of official languages). The only event in this module is called externally, that is from other event functions generating a change of state in one of the core language variables.

The derived variables are used either analytically (as variables of interest for the analysis) or for modeling (as dimensions of parameters). They are further described in the table below.

Language Derived Variables	
Combination of home language and knowledge of official languages	This variable is used in the modeling of internal mobility. It is used as a determinant of out-migration in the logistic regression and as a determinant of the origin-destination matrices. It can take one of six possible values: English (does not speak French at home and knows only English), French (does not speak English at home and knows only French), English bilingual (speaks English at home and knows French), French bilingual (speaks French at home and knows English), Other bilingual (speaks neither French nor English at home but knows both languages) and Other (speaks neither French nor English and does not know either languages).
Combination of mother tongue and home language	This variable is used as a determinant of the literacy level. It can take one of four possible values: English (mother tongue English), French (mother tongue French), Other – official (mother tongue neither French nor English but speaks an official language at home), Other – non official (mother tongue and home language neither French nor English).
Linguistic proficiency index	This derived variable is used as a proxy for official language proficiency, itself a determinant of economic integration. It can take one of seven values (or levels). Level 1: Mother tongue official, knowledge of both official languages; Level 2: Mother tongue official, knowledge of only one official; Level 3: Mother tongue non official, home language official, knowledge of both official languages; Level 4: Mother tongue non official, home language official, knowledge of only one official language; Level 5: Mother tongue non official, home language non official, knowledge of both official languages; Level 6: Mother tongue non official, home language non official, knowledge of only one official language; Level 7: Mother tongue non official, home language non official, no knowledge of official languages.

The Labour Force Participation Module (Labour.mpp)

Labour force participation is a dichotomous state variable: an actor may be considered *active* (employed or unemployed looking for work) or *inactive* (unemployed and not looking for work). The value of the labour force participation state variable is derived from the actor's characteristics: there are no specific transition probabilities between the *active* and the *inactive* states.

Using the 2011 NHS data, the probability of being active at any point in time is calculated for actors aged 15 and over[13] using a logistic regression with labour force participation as the dependent variable and age (five-year age groups), region of residence, visible minority status and immigrant status as regressors. The regression

[13] In fact, the described methodology is applied to all simulated cases aged 15 to 79 years old. The activity rates for individual aged 80+ are set constant throughout the simulation at the 2011 observed level. Precisely, the activity rate of 80-84 years old is equal to 2.50%, of 85-89 years old is equal to 1.67%, 90-94 years old is equal to 0.85%, and of 95+ to 0%.

was further stratified by sex and level of education. Whenever a state variable affecting labour force participation changes, the probability of being active is recalculated and the labour force status reassigned following a Monte-Carlo trial.

The participation rates were also assumed to evolve in the course of the simulation. The evolution factor varies according to sex, age (five-year age groups) and level of education, and was derived from extrapolated trends observed between 2002 and 2013 in the Canadian Labour Force Survey. This 11-year period was chosen to emphasize recent trends without incorporating too much noise.

The evolution of participation rates is constrained using a triple mechanism. First, participation rates are not allowed to increase beyond the maximum rate observed in 2011 (97,9%). Second, actors in age groups at or above 55 years old are not allowed to have participation rates higher than the preceding age group. As participation rates of older workers have increased significantly in the reference period, this constraint prevents the generation of unlikely participation rate profiles, namely profiles with increasing rates above age 50. Finally, the third constraint prevents evolution beyond a fixed time horizon, typically 15 years (although this may vary depending on the scenario).

This method of deriving participation rates provides adequate aggregate cross-sectional descriptions, but yields incoherent individual lifecourses, as individual actors are susceptible to change their participation status every time they move from one region to another or reach another age group. Further development of the model should introduce transition probabilities between states according to characteristics of the actor as well as according to duration in a given state.

Ranges, classifications and links	
LFSTAT	Labour force participation status. Can take two possible values: active (at work or unemployed and looking for work) or inactive (unemployed and not looking for work).

Parameters	
Logistic regression coefficients	Logistic regression coefficients for age (five-year age groups), region of residence, visible minority status and immigration status. Coefficients are further stratified by sex and education.
Evolution factor	The probability of being active evolves throughout the simulation. The evolution slopes vary according to age (five-year age groups), sex and education and are based on labour force participation trends observed from 2002 to 2013 (Canadian Labour Force Surveys). Projected probabilities are bounded by a triple mechanisms (see description above).
Evolution duration	Duration beyond which labour force participation probabilities stop changing.
Maximum rate	Maximum allowable labour force participation rate.

Main state variables	
lfstat	LFSTAT. Indicates if actor is active or inactive.
ChangesStates Labour	INT. A derived variable who is incremented every time a factor affecting labour force participation is changed. A change in the value of this variable triggers the labour force participation event.

Events and functions	
Labour	This event computes the probability of being active and randomly assign a new labour force status based on a Monte-Carlo trial. Probability of being active is derived from the regression coefficients and is recalculated every time characteristics included in the regressions change. Only actors aged 15 or more can be active.

The Mortality Module (Mortality.mpp)

The modeling of mortality in LSD-C is done by attributing each actor an annual risk of death that combines a baseline rate and a relative risk. Baseline rates were established according to age, sex, province of residence and year of simulation by Statistics Canada for the more recent version of their cohort-component projection model (Bohnert et al. 2015).

Relative risks were calculated in two steps. They were first estimated using the 1991 Canadian Census Health and Environment Cohort (CanCHEC), a large longitudinal dataset that links individual census data from 1991 to 20 years of official death records and annual tax return data. Since the annual tax data allows tracking changes in province of residence throughout these 20 years, province was treated as a time-varying variable. The sample at risk of dying was slightly adjusted by removing potential emigrants from the dataset: individuals who did not file a tax return for five consecutive years and whose death was not observed were considered to have emigrated just after their last filing and were not considered at risk of dying in Canada from that point on. Cox models were used to estimate relative risks of dying according to educational attainment, generation status (1, 1.5 or 2+), place of birth (3 regions), and aboriginal status. Separate models were run for each sex and large age groups (25-44, 45-64, 65-74, 75-84, 85+). All models were stratified on province of residence (option strata in Stata), thus assuming a different baseline hazard in each province, but common relative risks.

In a second step, the estimated relative risks were adjusted in order to account for the different composition of the population in the dataset used for estimation (1991 CanCHEC) and the base population of the model (2011 NHS). The adjustment was made separately by province. Note that below age 25 only the baseline rates are applied as CanCHEC only followed people aged 25 and older at census day in 1991.

Ranges, classifications and links	
AGEMORT	Aggregation of the AGE classification into age groups at which relative risks change: 25-44, 45-64, 65-64, 75-74, and 85+.
POBMORT	Aggregation of the POB classification into three regions: 0) Canada; 1) the United States, Oceania, and Western, Northern and Eastern Europe (including all ex-USSR countries); 2) Southern Europe, Africa, Asia, and Central and Southern America. Regions were identified based on empirical mortality differentials.

Parameters	
Mortality rates	Baseline mortality rates vary according to year of age, calendar year, sex and province of residence. Projection of mortality rates were done by Statistics Canada.
Relative risks	Mortality risks are modulated according to highest educational attainment, generation status (1, 1.5 or 2+), region of birth and aboriginal identity.
Maximum age	The maximum age that can be reached in the model: 100 years.

Main state variables	
pobmort	POBMORT. May take any of the three values listed in the POBMORT classification.
isaboriginal	LOGICAL. Indicates if an actor identifies as aboriginal. Derived from the *vismin* state variable.

Events and functions	
Mortality	This event computes the waiting time before a death event, taking into account the maximum allowable age. After a death event, the *alive* state variable is set to *FALSE* and the actor is removed from the simulation.

The Visible Minority Module (VisibleMinority.mpp)

The visible minority module takes care of intergenerational transfers of visible minority status from mother to child. Transfers are established according to origin-destination matrices (from mother visible minority status to child visible minority status) extracted from the 2011 NHS confidential microdata file using the own-children method. The origin-destination matrices are computed separately by mother's generation status.

At birth, an actor is automatically given the visible minority status of its mother and is then immediately and instantaneously subjected to an intergenerational transfer of visible minority status. Visible minority is then fixed for life.

Ranges, classifications and links	
VISMIN	Visible minority status. Can take 13 possible values (see relevant section above in the description of the base population).

Parameters	
Visible minority status origin-destination matrix	Origin (mother's visible minority status) – destination (actor's visible minority status) matrices for intergenerational visible minority status transfers. Vary according to the mother's generation status.

Main state variables	
vismin	VISMIN. Visible minority status.
isvismin	LOGICAL. A derived variable who takes the value *TRUE* when visible minority status is not equal to 12 (i.e. when visible minority status is non-white).

Events and functions	
Intergenerational transfer of visible minority status	This event assigns a visible minority status based on an origin-destination matrix. This event only occurs once at birth.

The Religion Module (VisibleMinority.mpp)

The religion module takes care of intergenerational transfers of religious affiliation from mother to child. Intergenerational transfer probabilities were established according to origin-destination matrices (from mother's religious affiliation to child's religious affiliation) extracted from the 2011 NHS confidential microdata file and using the own-children method. The origin-destination matrices were computed separately according to the mother's generation status and her region of residence (living in Québec or in Canada outside Québec).

At birth, an actor is automatically given the religious affiliation of its mother and is then immediately and instantaneously subjected to an intergenerational transfer of religious affiliation. Religious affiliation is then fixed for life.

Ranges, classifications and links	
RELIGION	Religious affiliation. Can take 8 possible values (see relevant section above in the description of the base population).

Parameters	
Religious affiliation origin-destination matrix	Origin (mother's religious affiliation) – destination (actor's religious affiliation) matrices for intergenerational religious affiliation transfers. Vary according to the mother's generation status and her region of residence (Québec and Canada outside Québec).

Main state variables	
religion	RELIGION. Religious affiliation.
isjew	LOGICAL. A derived variable that takes the value *TRUE* when religious affiliation is Jewish. This variable is used in the mobility module as Jews are often Caucasian, educated, speakers of official languages, but show mobility patterns that are different from mobility patterns of other religious groups having the same characteristics.

Events and functions	
Intergenerational transfer of religious affiliation	This event assigns a religious affiliation to a newborn actor based on an origin-destination matrix. This event only occurs once at birth.

The Literacy Module (Literacy.mpp)

The literacy score in LSD-C is a continuous derived variable taking a value between 0 and 500. The source of data used for the analysis was the Programme for the International Assessment of Adult Competencies (PIAAC) database of 2012. The analysis was based on more comprehensive work done by Samuel Vézina as part of his PhD thesis.

The literacy score is derived from parameters of linear regressions stratified by immigration status in order to consider the specific characteristics of the foreign-born population. The regression models contain socio-demographic variables (age, sex, province of residence), human capital variables (education level, mother tongue and language spoken at home), and labour force participation status. For immigrants, the model also includes variables linked with characteristics that are specific to foreign-born people, such as age at immigration, length of stay in Canada, country of birth, and country of highest diploma. A simple logarithmic transformation is made on the dependant variable (literacy score) to ensure that the linear regression model will not lead to illogical predicted score (below 0 or above 500) during the simulation

Only the population aged 25 to 64 may be assigned a literacy score. Literacy scores are updated whenever relevant actor’s characteristics undergo changes.

Since literacy and labour force participation share many determinants and are both imputed variables, labour force participation status takes precedence when both states are scheduled to be updated at the same time. This is necessary as labour force participation is a determinant of literacy.

Ranges, classifications and links	
PROVLIT, EDUCLIT, GENLIT, LANGLIT, YSMLIT, POBLIT, AGELIT	Classification that are aggregates or reconstructions of existing classifications, for parameterisation of literacy. PROVLIT (7): Atlantic provinces, Quebec, Ontario, Manitoba-Saskatchewan, Alberta, British-Columbia, Territories EDUCLIT (3): no diploma, secondary or college, university. GENLIT (5): generation 3+, 2.5, 2, 1.5, 1. LANGLIT (4): mother tongue English, mother tongue French, non-official mother tongue and official language used at home, non-official mother tongue and non-official language used at home). YSMLIT (4): less than 5 years since migration, 5 to 9, 10 to 14, 15+ years) POBLIT (2): born in poor country, born in rich country. AGELIT: ten years age groups.
POD	Place of diplomation. Takes one of two values: in Canada or Outside Canada.

Parameters	
Age, sex, region of residence, education, language, labour force participation status, generation status, years since migration, place of birth and place of diplomation	Regression parameters used to impute actors' literacy scores. All parameters are available for immigrants and non-immigrants. Parameters are set to zero when they are not applicable (e.g. coefficients corresponding to "years since migration" for non-immigrants).

Main state variables	
provlit, educlit, genlit, pod and poblit	State variables of corresponding classifications used as indices in parameter matrices.
scorelit	DOUBLE. A derived variable giving the literacy score of the actor based on the result of a linear regression (see above).
ChangesStatesLiteracy	INTEGER. A derived variable that is incremented every time a factor affecting literacy changes. A change in the value of this variable triggers the literacy event.

Events and functions	
literacy	This event computes the literacy score based on the results of a linear regression.

The Mobility Module (Mobility.mpp)

Interregional mobility in the model unfolds in two separate steps. First, waiting time before an outmigration event is calculated from the parameters of a logistic regression. Second, once an outmigration event has been scheduled, a destination region is chosen according to an origin-destination matrix, whose values vary according to the actor's characteristics.

Parameters used to calculate outmigration rates are derived from the results of logistic regressions performed on the 2011 NHS confidential microdata. For individuals aged 15 years or older, the probability of living in a region that is different from one's region of residence one year ago was regressed against sex, age, language (combination of language spoken at home and knowledge of official languages), visible minority status, immigrant status, duration since migration, religion, education and labour force participation status. For individuals aged less than 15 years old, sex, labour force participation status and education were removed from the equation. Regressions were further stratified by region of residence (all 24 regions were used).

Origin-destination matrices between all 24 regions were derived according to age, language characteristics (a combination of knowledge of official languages and language used at home), visible minority status, immigration status, religion, labour force participation and education. Only a subset of all the variable categories was used: details are provided in the corresponding box below.

Since the probability of outmigration is modeled over a one-year time span, i.e. migrants are counted and not migrations, only one migration event is allowed in a single year of age.

Classifications related to the geography of the model are included in the mobility module.

Ranges, classifications and links	
REGION	All regions of residence included in the model. See the place of residence section above.
PROVINCE	Aggregation of regions into provinces and territories. PROVINCE has thus 11 categories: Newfoundland and Labrador, Prince Edward Island, Nova Scotia, New Brunswick, Quebec, Ontario, Manitoba, Saskatchewan, Alberta, British Columbia, Territories.
REGIONLANG	Regional grouping that are relevant with respect to language behavior and characteristics. It contains 7 categories: New Brunswick, Montreal, Gatineau, Rest of Quebec, Ottawa, Rest of Ontario, Rest of Canada. Those regions are used as strata for the parameters in the language modules.

Parameters	
Outmigration parameter matrix	Includes all regression coefficients necessary to calculate probability of outmigration. It contains 48 columns (24 regions for actors aged less than 15 years old and 24 regions for actors aged 15 years and above) and the lines correspond to regression parameters. These parameters include the intercept, two sex categories (male, female), age (13 categories: 0-4,5-9,10-14,15-18,19-21,22-24,25-27,28-29,30-34,35-49,50-64,65-69,70+), language (6 categories: knowledge of English only, knowledge of French only, bilingual and speaks English at home, bilingual and speaks French at home, bilingual but speaks neither French nor English at home, no knowledge of official languages), visible minority status (13 categories: see relevant section above), immigration status (3 categories: non immigrant, immigrant arrived in Canada less than five years ago, immigrant arrived in Canada five or more years ago), religion (8 categories: see relevant section above), labour force participation status (2 categories: active, inactive) and education (4 categories, see relevant section above). This storage of regression coefficients in matrix form should eventually be modified so that each regression coefficient has its own parameter variable.
Origin-destination matrices	Matrices containing the distribution of movers in destination regions (24 categories) according to region of origin one year ago (24 categories), age (2 categories to distinguish student migrations: between 18 and 24 years old, all others), language (4 categories: bilingual using English at home, bilingual using French at home, no knowledge of French and no French used at home, others), visible minority status (2 categories: visible minority, not a visible minority), immigration status (2 categories: immigrant, non-immigrant), religion (2 categories: Jewish, non Jewish; this distinction is necessary as Jews are often Caucasian, educated, speakers of official languages, but show mobility patterns that are different from mobility patterns of other religious groups having the same characteristics), labour force participation status (2 categories: active, inactive) and education (2 categories: university, others).

Main state variables	
region	REGION. A state variable giving the region of residence of the actor.
province	PROVINCE. A state variable derived from region and giving the corresponding province of residence.
regionlang	REGIONLANG. A state variable derived from region and used in the language modules.
isqc	LOGICAL. Indicates if actor is residing in Quebec or in the rest of Canada.

Events and functions	
mobility	Calculates waiting time before outmigration and picks a new region of residence based on an origin-destination matrix. Actors may only move once per year of age.

The Immigration Module (Immigration.mpp)

The immigration module includes all classifications, state variables and parameters relevant to immigrants and immigration: immigration level and composition, immigrant status, age at immigration, duration since immigration, generation status and place of birth.

The module includes a single event whose purpose is to increment the number of years since migration. Immigration itself (i.e. the creation and simulation of a new immigrant actor during the projection) is not included in this module. Why is that so? Unlike actors born during the projection, immigrant actors are not related to existing actors in the simulation. Whereas a newborn actor has a link to one of the female actors, an immigrant is not linked to any pre-existing actor in the simulation[14]. Hence, the creation of immigrants must be "external", that is defined outside the description of the *Person* class. Immigrant actors must be created in the same way as are the actors from the base population: in the main simulation file (see section on the main simulation file above).

Whereas the starting time for the actors of the base population is the year of the NHS (2011), starting times for new immigrants must be spread all along the course of the simulation.

[14] This is a specificity of this model and is not an inherent characteristic of microsimulation models. One could equally produce a microsimulation model simulating chain migration, in which actors could generate new immigrants.

The characteristics of recent immigrants in the NHS are used as a basis to determine the characteristics of future immigrants in the simulation. In the simulation, whenever the record of a recent immigrant (defined as having lived in Canada for less than five years) is read from the base population file, it is first simulated as a member of the base population, and then re-simulated once in every year of the projection (2011 to the end of the projection) as a new immigrant.

To achieve this in Modgen, a second “immigration” loop is inserted in the main simulation loop (in the *Simulation* function of the Main file). For each iteration of this second loop, the *CaseSimulation* function is called to create an immigrant actor arriving in the simulation at year t + a random number between 0 and 1, where t is the iteration number corresponding to the year of the simulation.

This method for generating immigrants implies that all immigrant cohorts have the same characteristics as recent immigrants from the base population. To allow for some variation, the individual weights are adjusted to match desired distributions. Through reweighting, it is thus possible to change the total volume of immigration during the simulation. It is also possible in each region to change the immigrant distribution according to mother tongue, age, education and visible minority status. Only one of these four variables may be used for reweighting in a single projection as joint distributions were not included in the model.

Non-permanent residents are included in the simulation as well, but only for bookkeeping purposes. They are inserted in the model just like immigrants are, but are scheduled to emigrate exactly one year after entry. Only the number (and not the composition) of non-permanent residents may be modified in the simulation.

For more information on the integration of immigration in a Modgen microsimulation model, please consult chapter 5 of the book *Microsimulation and population dynamics* written by Bélanger and Sabourin (2017) and published by Springer.

Ranges, classifications and links	
IMMSTAT	Immigrant status. Includes four categories: non-immigrants, established immigrants (in Canada for five years or more), recent immigrants (in Canada for less than five years) and non-permanent residents.
GENERATION	Generation Status. Includes five categories: generation 3+ (Canadian-born from Canadian-born parents), generation 2.5 (Canadian-born from 1 Canadian-born parent), generation 2, (Canadian-born from 2 foreign-born parents), generation 1.5 (foreign-born from foreign-born parents, arrived in Canada before the age of 15), generation 1 (foreign-born from foreign-born parents, arrived in Canada at the age of 15 or older). Other generation classifications are used in other parts of the model to accommodate specific model designs (for instance, generations 3+, 2.5 and 2 are aggregated for intergenerational transfers of religion, visible minority and mother tongue).
POB	Place of birth. See relevant section above for categories.
AGEIMM	Age at immigration. An integer number between 0 and 100.
YSM	Years since migration. An integer number between 0 and 100.

Parameters	
Number of immigrants	Number of immigrants entering the simulation each year.
Weight of recent immigrants	Weight of recent immigration in the base population (recent immigrants serve as a basis for future immigration in the simulation)
Number of non-permanent residents	Number of non-permanent residents entering the simulation each year. Non-permanent residents only stay one year in the simulation (after which time they emigrate).
Weight of non-permanent residents	Weight of non-permanent residents in the base population (serves as a basis for future cohorts of non-permanent residents)
Distribution of immigrants according to age	Distribution of recent immigration according to age in the initial population and distribution of immigration according to age during the simulation. If both distributions are equal, future immigration is assumed to have the same characteristics as recent immigrants in the base population.
Distribution of immigrants according to mother tongue	Distribution of recent immigration according to mother tongue in the initial population and distribution of immigration according to mother tongue during the simulation. If both distributions are equal, future immigration is assumed to have the same characteristics as recent immigrants in the base population.
Distribution of immigrants according to education	Distribution of recent immigration according to education in the initial population and distribution of immigration according to education during the simulation. If both distributions are equal, future immigration is assumed to have the same characteristics as recent immigrants in the base population.
Distribution of immigrants according to visible minority	Distribution of recent immigration according to visible minority in the initial population and distribution of immigration according to visible minority during the simulation. If both distributions are equal, future immigration is assumed to have the same characteristics as recent immigrants in the base population.
Intergenerational transfer of generation status	Origin-destination matrix linking the generation status of a mother to the generation status of her children. Since the characteristics of the father are unknown, generation status of newborns must be probabilistically assigned using a predetermined distribution: an immigrant mother, for instance, may give birth to a child that is either generation 2 or generation 2.5, depending on the (unknown) generation status of the father. The origin-destination matrix was obtained from the NHS confidential microdata file using the own-child method.

Main state variables	
immstat	IMMSTAT. Immigration status.
ageimm	AGEIMM. Age at immigration.
generation	GENERATION. Generation status.
ysm	YSM. Number of years since migration.
isimm	LOGICAL. A binary variable indicating if actor is an immigrant or not.
pob	POB. Place of birth.

Events and functions	
YSMevent	Increments the number of years since migration for an immigrant actor. Also updates accordingly the immigrant status of the immigrant actor (from recent to established immigrant).

Conclusion

LSD-C projection results were first presented at the 2016 Annual Meeting of Canadian Population Society in Calgary, Canada (Bélanger et al., 2016). The presentation focused on labour force projection results showing that the labour force participation rate is likely to decline in the coming decades in Canada, notwithstanding the future immigration level. Other results showed the diversification process going on within the Canadian workforce in terms of the share of visible minority and language skills.

The model LSD-C serves as a prototype for the development of other microsimulation models for the United States (LSD-USA) and for the 28 European Union member states countries (CEPAM-Mic). The core structure of these two models is based on LSD-C. These were presented at the 6th World Congress of the International Microsimulation Association (IMA 2017) in Torino, Italy (Marois, Sabourin, Bélanger, & Lutz, 2017; Sabourin, Bélanger, Vézina, Marois, & Pelletier, 2017; Van Hook, Morse, Sabourin, & Bélanger, 2017).

For additional information about the three microsimulation models, it will be possible to refer to a paper in preparation for submission and entitled “A framework for the prospective analysis of super-diversity” which presents the common analytical framework undergirding the three models as well as some projection results regarding the ethnocultural composition of the future population.

References

Alba, R., & Foner, N. (2015). Integration's challenges and opportunities in the Wealthy West. *Journal of Ethnic and Migration Studies, 42*(1), 3-22. doi:10.1080/1369183x.2015.1083770

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., & Wacziarg, R. (2003). Fractionalization. *Journal of Economic Growth, 8*(2), 155-194.

Alesina, A., Harnoss, J., & Rapoport, H. (2016). Birthplace diversity and economic prosperity. *Journal of Economic Growth, 21*(2), 101-138.

Anderson, G. F., & Hussey, P. S. (2000). Population aging: a comparison among industrialized countries. *Health Affairs, 19*(3), 191-203.

Auerbach, A. J., & Lee, R. (2011). Welfare and generational equity in sustainable unfunded pension systems. *Journal of Public Economics, 95*(1-2), 16-27.

Australian Department of Immigration and Citizenship. (2012). The Outlook for Net Overseas Migration. Retrieved from http://www.immi.gov.au/media/publications/statistics/immigration-update/nom-jun12.pdf

Bélanger, A., & Sabourin, P. (2017). *Microsimulation and population dynamics: An introduction to Modgen 12*: Springer.

Bélanger, A., Sabourin, P., Vézina, S., Marois, G., Lafontaine, O., & D'Ovidio, K. (2016). *Projecting Superdiversity and Socioeconomic Composition with a Microsimulation Model*. Paper presented at the Annual Meeting of Canadian Population Society, Calgary, Canada.

N. Bohnert, J. Chagnon, S. Coulombe, P. Dion and L. Martel (2015) Population Projections for Canada (2013 to 2063), Provinces and Territories (2013 to 2038): Technical Report on Methodology and Assumptions, Ottawa: Statistics Canada, 85 pages

Briggs, V. M. (2003). *Mass immigration and the national interest: policy directions for the new century*: ME Sharpe.

Citizenship and Immigration Canada. (2012). *Facts and Figures 2012: Immigration Overview–Permanent and Temporary Residents*. (Ci1-8/2012E-PDF). Ottawa: Government of Canada.

Coleman, D. (2006). Immigration and Ethnic Change in Low-Fertility Countries: A Third Demographic Transition. *Population and Development Review, 32*(3), 401-446. doi:10.1111/j.1728-4457.2006.00131.x

Coleman, D. (2009). The demographic effects of international migration in Europe. *Oxford Review of Economic Policy, 24*(3), 452-476. doi:10.1093/oxrep/grn027

Esping-Andersen, G. (2000). The sustainability of welfare states into the twenty-first century. *International Journal of Health Services, 30*(1), 1-12.

Jennissen, R., Van Der Gaag, N., & Van Wissen, L. E. O. (2006). Searching for similar international migration trends across countries in Europe. *Genus, 62*(2), 37-64.

Lee, J., & Bean, F. D. (2010). *The diversity paradox: Immigration and the color line in twenty-first century America*: Russell Sage Foundation.

Lichter, D. T. (2013). Integration or Fragmentation? Racial Diversity and the American Future. *Demography, 50*(2), 359-391. doi:10.1007/s13524-013-0197-1

Lutz, W. (2013). Demographic Metabolism: A Predictive Theory of Socioeconomic Change. *Population and Development Review, 38*(s1), 283-301. doi:10.1111/j.1728-4457.2013.00564.x

Lutz, W., Butz, W. P., & KC , S. (2014). *World population and human capital in the twenty-first century*. Oxford: Oxford University Press.

Mahroum, S. (2001). Europe and the immigration of highly skilled labour. *International Migration, 39*(5), 27-43.

Marois, G., Sabourin, P., Bélanger, A., & Lutz, W. (2017). *Immigration, Diversity, Human Capital and the Future Labor Force of Developed Countries: The European Model*. Paper presented at the 6th World Congress of the International Microsimulation Association, Torino, Italy.

Martin, P., & Ruhs, M. (2011). Labor Shortages and US Immigration Reform: Promises and Perils of an Independent Commission. *International Migration Review, 45*(1), 174-187.

Meissner, F., & Vertovec, S. (2014). Comparing super-diversity. *Ethnic and Racial Studies, 38*(4), 541-555. doi:10.1080/01419870.2015.980295

OECD. (2000). *Reforms for an Ageing Society*. Paris: OECD Publishing.

OECD. (2013). *Pensions at a Glance 2013: OECD and G20 Indicators*. Paris: OECD Publishing.

Pammolli, F., Riccaboni, M., & Magazzini, L. (2012). The sustainability of European health care systems: beyond income and aging. *The European Journal of Health Economics, 13*(5), 623-634.

Patsiurko, N., Campbell, J. L., & Hall, J. A. (2012). Measuring cultural diversity: ethnic, linguistic and religious fractionalization in the OECD. *Ethnic and Racial Studies, 35*(2), 195-217. doi:10.1080/01419870.2011.579136

Population Reference Bureau. (2012). *2012 World Population Data Sheet*. Washington, DC: Population Reference Bureau.

Ruhs, M., & Anderson, B. (2010). *Who needs migrant workers?: labour shortages, immigration, and public policy*: Oxford University Press.

Ryder, N. (1965). The Cohort as a Concept in the Study of Social Change. *American Sociological Review, 30*(6), 843-861.

Sabourin, P., Bélanger, A., Vézina, S., Marois, G., & Pelletier, D. (2017). *Immigration, Diversity, Human Capital and the Future Labor Force of Developed Countries: the Canadian Model.* Paper presented at the 6th World Congress of the International Microsimulation Association, Torino, Italy.

Statistics Canada. (2017). Modgen (Model generator). *Microsimulation.* Retrieved from http://www.statcan.gc.ca/eng/microsimulation/modgen/modgen

United Nations Population Division. (2012). *World Population Prospects: The 2010 Revision*. New York: United Nations.

Van Hook, J., Morse, A., Sabourin, P., & Bélanger, A. (2017). *Immigration, Diversity, Human Capital and the Future Labor Force of Developed Countries: The US* *model*. Paper presented at the 6th World Congress of the International Microsimulation Association, Torino, Italy.

van Imhoff, E., & Post, W. (1998). Microsimulation Methods for Population Projection. *Population: An English Selection, 10*(1), 97-138.

Vertovec, S. (2007). Super-diversity and its implications. *Ethnic and Racial Studies, 30*(6), 1024-1054. doi:10.1080/01419870701599465

Wolfson, M. (2011). *Linking Policies to Well-Being Outcomes Through Micro-Simulation*. (1815-2031). Paris: OECD Publishing.

Zaiceva, A., & Zimmermann, K. F. (2008). Scale, diversity, and determinants of labour migration in Europe. *Oxford Review of Economic Policy, 24*(3), 427-451.