CIMI

METHODOLOGY OVERVIEW

MARCH 2020

Canadian Index for Measuring Integration

TABLE OF CONTENTS

For more information on this study, please contact **Paul Holley:**
paul.holley@acs-aec.ca
514 925-3099

INTRODUCTION: OUR APPROACH

The Canadian Index for Measuring Integration (CIMI) is an evidence-based assessment tool used to evaluate the ongoing state of immigrant integration in Canada. The CIMI identifies factors that underline successful immigrant integration, assesses changes and trends over time (currently from 1991 to 2020), enables detailed examination of four dimensions of integration (economic, social, civic and democratic participation and health) and provides rankings based on empirical evidence for Canadian geographies (provinces and Census Metropolitan Areas or CMAs).

The CIMI compares outcomes for immigrants relative to the Canadian-born population using a two-step process:

1. First, through the use of descriptive or unadjusted data to demonstrate differences or "gaps" between immigrants and non-immigrants per indicator (without controlling for socio-demographics), which offers snapshots of integration trends for Canadian geographies at a specific point in time;

2. Second, by examining several integration-related outcomes while adjusting for socio-demographic differences between immigrants and Canadian-born population, allowing for equal comparisons between geographies and across time.

The second phase of the CIMI is conducted using multiple regression analyses, including both Linear Regression for continuous dependent variables (e.g., wages, # friends living in the same city/community) and Logistic Regression for binary (0,1) dependent variables (e.g., unemployed, voted in the last election).

The primary advantage of using multivariate regression analysis is that it allows us to estimate the main effects of key independent variables (i.e., immigrant status, geography, and immigrant status x geography) on our dependent outcomes while holding constant several demographic factors (e.g., age, sex, ethnicity, language) and socioeconomic factors (income, occupation, education). Furthermore, multiple regression allows for robust estimation and the reduction of error/inaccuracies through confidence interval estimation and hypothesis testing procedures.

For example, the regression models allow us to estimate the gap in voting patterns between immigrants and non-immigrants in Ontario (or any other province/CMA) relative to the gap found in the rest of Canada, while holding constant a number of sociodemographic characteristics of the sample. As another illustration, we can estimate the wage gap between immigrants and the Canadian-born in Ottawa and compare this to the wage gap between immigrants and non-immigrants in all other CMAs in Canada.

CIMI GEOGRAPHIES

The CIMI examines integration outcomes for Canadian provinces and Census Metropolitan Areas (CMAs) across six time-periods (1991-1995, 1996-2000, 2001-2005, 2006-2010, 2011-2015, 2016-2020) corresponding to the five-year cycles of the Census.

Table 1 below presents the list of CIMI geographies (provinces and CMAs). In 2016, a total of 10 provinces and 35 CMAs were included in the analysis. If a geography is not represented in the list below, this could be due to one of two reasons: (1) the city does not meet the minimum population criteria to be considered a Census Metropolitan Area (CMA) by Statistics Canada; and/or (2) the geography has too small of an immigrant population to produce reliable estimates in the regression analysis.

Table 1. CIMI Geographies (Provinces and CMAs)

PROVINCES	CITIES (CMAS)
Newfoundland	St. John's
Prince Edward Island	—
Nova Scotia	Halifax
New Brunswick	Moncton, Saint John
Québec	Montréal, Québec, Saguenay, Sherbrooke, Trois-Rivières
Ontario	Barrie, Brantford, Guelph, Hamilton, Kingston, Belleville, Kitchener-Cambridge-Waterloo, London, Oshawa, Ottawa-Gatineau, Peterborough, St. Catharines-Niagara, Greater Sudbury, Thunder Bay, Toronto, Windsor
Manitoba	Winnipeg
Saskatchewan	Regina, Saskatoon
Alberta	Edmonton, Calgary, Lethbridge
British Columbia	Abbotsford, Kelowna, Vancouver, Victoria

CIMI DATA SOURCES

The CIMI uses three primary data sources in the analysis of economic, social, civic and democratic and health outcomes for immigrants and Canadian-born populations, which include the Census, Canadian Community Health Survey (CCHS), and General Social Survey (GSS).

Each CIMI dimension relies on the following data sets:

- **Economic:** Census cycles 1991, 1996, 2001, 2006, 2011 National Household Survey (NHS), and Census 2016 - Public use Micro Data File (PUMF) data and relevant Master data files;

- **Social:** General Social Survey (GSS) Social Identity cycles 17 (2003), 22 (2008) and 27 (2013); PUMF data and relevant Master data files;

- **Civic and Democratic Participation:** General Social Survey (GSS) Social Identity cycles 17 (2003), 22 (2008) and 27 (2013); PUMF data and relevant Master data files;

- **Health:** Canadian Community Health Survey (CCHS) cycles 2000-01, 2005, 2010, 2014, 2018; PUMF data and relevant Master data files.

The Canadian Census is the primary data source for the economic dimension and provides the largest counts of immigrants and most reliable data source at the sub-provincial level. Other Statistics Canada surveys used by this study (e.g., the General Social Survey and Canadian Community Health Survey) do not produce as reliable of estimates for all CMAs in all years due to lower counts of immigrants surveyed in smaller geographies (e.g., Saguenay in Québec or Lethbridge in Alberta). Table 2 below describes the CIMI data sources used for each time-period across the four dimensions of integration.

Table 2. CIMI Data Sources by Dimension of Integration

CIMI PERIODS	ECONOMIC	SOCIAL	CIVIC & DEMOCRATIC PARTICIPATION	HEALTH
2016 to 2020	Census 2016	—	—	CCHS 2018
2011 to 2015	Census 2011	GSS 2013	GSS 2013	CCHS 2014
2006 to 2010	Census 2006	GSS 2008	GSS 2008	CCHS 2010
2001 to 2005	Census 2001	GSS 2003	GSS 2003	CCHS 2005
1996 to 2000	Census 1996	—	—	CCHS 2000/01
1991 to 1995	Census 1991	—	—	—

CIMI INDICATORS

Table 3 presents the complete list of CIMI indicators across four dimensions of integration: economic, social, civic and democratic participation and health. Although dozens of indicators were considered, this final list of 22 indicators was selected based on conceptual and methodological considerations from our Expert Advisory Committee (EAC). For more details on the CIMI indicators, please refer to the Codebook on the Resources page of the website.

Table 3. CIMI Indicators by Dimension

CIMI DIMENSIONS AND INDICATORS		VARIABLE	SOURCE
ECONOMIC DIMENSION	*Wages*[1] – refers to gross wages and salaries before deductions for such items as income taxes, pension plan contributions and employment insurance premiums during the reference period	Continuous	Census (1991-2016)
	Low-Income Cut-offs (LICOs) – refers to the proportion of individuals who have lived under Statistics Canada's low-income cut-offs, before tax	Binary	Census (1991-2016)
	Labour Force Participation – refers to the percentage of individuals who are active in the labour force, either employed or unemployed - but looking for work	Binary	Census (1991-2016)
	Unemployment Rate – refers to the percentage of individuals who, during reference week, were available for work and were either on temporary layoff, had looked for work in the past four weeks or had a job to start within the next four weeks	Binary	Census (1991-2016)
	Employment Rate – refers to the number of persons employed in the reference week expressed as a percentage of the total population aged 18-64.	Binary	Census (1991-2016)
	Full-time Employment Rate – refers to the percentage of individuals who are working full-time, divided by the total population working full-time and part-time	Binary	Census (1991-2016)
	Non-official Language at Work – refers to the percentage of individuals who use non-official languages most often at work	Binary	Census (2001-2016)
	Subsidized Housing – refers to the percentage of renters who live in subsidized housing (i.e., rent applied to income, social housing, public housing, government-assisted housing, or non-profit housing)	Binary	Census (2011-2016)
SOCIAL DIMENSION	*# Close Friends* – refers to an individual's number of close friends who are not relatives, but who they feel at ease with, can talk to about what is on their mind, or call on for help[2]	Continuous (count)	GSS 2003 GSS 2008 GSS 2013
	# Close Friends Living in Same Community/City – refers to the number of close friends that live in the same local community or city who the individual feels at ease with, can talk to about what is on their mind, or call on for help	Continuous (count)	GSS 2008 GSS 2013
	Sense of Belonging to Local Community – refers to the extent to which individuals report a strong or very strong sense of belonging to their local community	Continuous (Likert-scale)	GSS 2003 GSS 2008 GSS 2013
	Sense of Belonging to Province – refers to the extent to which individuals report a strong or very strong sense of belonging to their province of residence	Continuous (Likert-scale)	GSS 2003 GSS 2008 GSS 2013
	Sense of Belonging to Canada – refers to the extent to which individuals report a strong or very strong sense of belonging to Canada	Continuous (Likert-scale)	GSS 2003 GSS 2008 GSS 2013
	Victim of Discrimination in the Past 5 Years – refers to the extent to which an individual has said they experienced discrimination over the past 5 years	Binary	GSS 2013

1 In the regression models, we used the square root of "wages" in lieu of the original "wages" variable, which was highly skewed to the left (because of few individuals with very high salaries).

2 In 2003, the value for this indicator is the average of the following categories of close friends: (1) None, (2) "1-2", (3) "3-5", (4) "6-10", (5) "11-20", (6) "More than 20". In 2008 and 2013, the value is the average of the actual number of close friends.

CIVIC AND DEMOCRATIC PARTICIPATION	*Unpaid Volunteer Work in the Past 12 Months* – refers to the percentage of individuals who have done unpaid volunteer work in the past 12 months	Binary	GSS 2003 GSS 2008 GSS 2013
	Involvement in Organizations in the Past 12 Months – refers to the percentage of individuals who reported membership, participation or involvement in groups or organizations in the past year	Binary	GSS 2003 GSS 2008 GSS 2013
	Voted in the Last Provincial Election – refers to the proportion of individuals who voted in the last provincial election	Binary	GSS 2003 GSS 2008 GSS 2013
	Voted in the Last Federal Election – refers to the proportion of individuals who voted in the last federal election	Binary	GSS 2003 GSS 2008 GSS 2013
HEALTH DIMENSION	*Has a Regular Medical Doctor* – refers to the proportion of individuals who have a regular medical doctor	Binary	CCHS 2000-2018
	Self-perceived Unmet Health Care Needs – refers to the proportion of individuals who perceive that their health care needs are not being met[3]	Binary	CCHS 2000-2018
	Self-perceived Life Stress – refers to the proportion of individuals who report being quite a bit or extremely stressed	Continuous (Likert-scale)	CCHS 2000-2018
	Satisfaction with Life – refers to the proportion of individuals who say they are very satisfied with their lives	Continuous (Likert-scale)	CCHS 2005 - 2018

CIMI CONTROLS

CIMI control variables include relevant socio-demographics such as the respondent's sex, age, knowledge of official languages, visible minority status, education level, occupation, income and mobility status. Other dimension-specific/thematic controls were incorporated into the models as needed. The consistent use of these controls across all CIMI models ensures an "apples to apples" comparison of immigrant and non-immigrant outcomes across Canadian geographies (10 provinces and 35 CMAs/cities) and various time-periods (1991-1995, 1996-2000... 2016-2020).

Table 4 summarizes all CIMI control variables used for each dimension of integration. These control measures were selected using conceptual and methodological considerations and guided by the recommendations of the Expert Advisory Committee.

For more details on the CIMI controls, please refer to the Codebook on the Resources page of the website.

3 In 2018, this optional variable was only administered in the provinces/CMAs of Nova Scotia, New Brunswick, Ontario, Manitoba, Alberta.

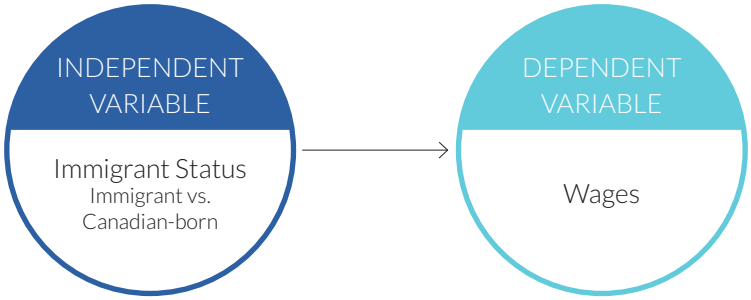
Table 4. Control Variables Used in the Analysis

CONTROL VARIABLE	DEFINITION	ECONOMIC	SOCIAL	CIVIC	HEALTH
Sex	Refers to whether the person is male or female	X	X	X	X
Age	Refers to the age (in years) at last birthday before the reference date	X	X	X	X
Visible Minority Status	Includes persons who are non-Caucasian in race or non-white in colour and who do not report being Aboriginal	X	X	X	X
Knowledge of Official Languages	Refers to the ability to conduct a conversation in English only, in French only, in both English and French, or in neither English nor French	X			X
Mother Tongue	Refers to the first language learned at home in childhood and still understood by the person		X	X	
Education	Information indicating the person's most advanced certificate, diploma or degree	X	X	X	X
Occupation	Refers to the kind of work performed by employed persons based on the National Occupational Classification System	X	X	X	X
Income	Refers to best estimated total personal income from all sources, before taxes and deductions, from all sources in the past 12 months		X	X	X
Full-time Employment status	Refers to the percentage of individuals who are working full-time	X	X	X	X
Mobility status (5-year)	Refers to the person's usual province or city of residence five years prior to the reference day	X			
Self-perceived Health (physical)	Refers to the percentage of people who reported that their (physical) health is very good or excellent.				X
Self-perceived Health (mental)	Refers to the percentage of people who reported that their mental health is very good or excellent.				X

ADJUSTED VS. UNADJUSTED DATA

Unadjusted or descriptive data looks at the relationship between two variables – an independent variable or predictor (e.g., immigrant status) and a dependent variable or outcome (e.g., wages). Unadjusted data can answer the question: What is the raw difference or 'gap' in wages between immigrants and non-immigrants? Descriptive data does not control for pre-existing differences between immigrants and non-immigrants such as their age, sex or gender, ethnicity, education levels and occupation. It simply measures the raw difference (in wages) between immigrants and the Canadian-born (see Figure A).

Figure A. What the unadjusted CIMI models measure



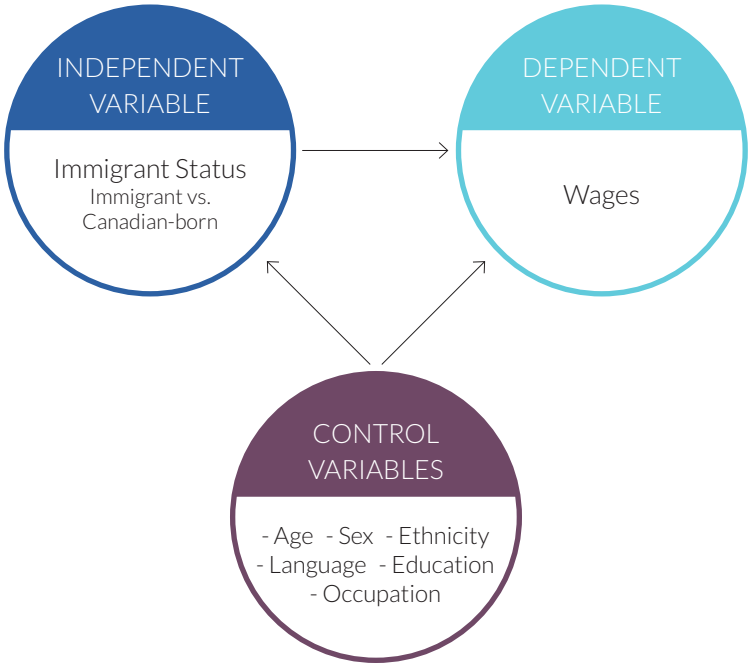
As an illustration, let us look at the difference in wages for immigrants and non-immigrants in Ontario. In 2016, Ontario's immigrants who worked full-time earned an average annual salary of $56,270 compared to $59,851 for non-immigrants – a wage gap of $3,581.

ONTARIO - 2016	IMMIGRANTS	NON-IMMIGRANTS	WAGE GAP
Wages for FT employees	$56,270	$59,851	$3,581

Adjusted or controlled data looks at the relationship between two variables while taking into consideration some third variable (or set of variables) that influences both the independent AND the dependent variable. For example, adjusted data can answer the question: What is the difference or 'gap' in wages between immigrants and non-immigrants, after controlling for their level of education?

Without 'adjusting' wages by education level, we are assuming that both immigrants and non-immigrants have the same educational attainment level or that education does not have an impact on wages. If we test the first part of this hypothesis, we find this is a false assumption: 41.9% of immigrants in Ontario obtained a university degree compared to 27.4% of non-immigrants in Ontario. Since higher education is associated with higher income levels, we might also expect that immigrants would earn more than the Canadian-born, but the data show the opposite trend.

Figure B. What the adjusted CIMI models measure



The adjusted models show us that having a university education is beneficial for both immigrants (+$19,238) and non-immigrants (+$21,350), but that the benefit is not constant (i.e., the same for both groups). In other words, there is something unique about immigrant status that correlates with lower wages that education does not account for. Otherwise, the wage gap difference in the table below would be equal to $0. Hence, adding control variables to the models adjust the original estimate or 'gap' in wages between immigrants and the Canadian-born population.

ONTARIO - 2016	IMMIGRANTS	NON-IMMIGRANTS	WAGE GAP
Average wages for FT employees without a university degree	$47,873	$53,659	$5,786
Average wages for FT employees with a university degree	$67,111	$75,009	$7,898
Added benefit of having degree	+$19,238	+$21,350	Difference = +$2,112

Note: Sample includes labour force population age 18 to 64 in the labour market, currently employed and earned wages are between $1 and $200,000.

In this example, only one control variable is added (education level) for demonstration purposes. In our full CIMI models, we add several controls to adjust the estimated wage gap between immigrants and non-immigrants. The descriptive data provides as with a snapshot of the raw differences while the controlled models (using regression analysis) adjust for differences between groups (i.e., their education level and their occupation as well as age, sex, ethnicity, language and other sociodemographic characteristics).

CIMI SAMPLE AND FILTERS

The CIMI sample consists of the adult population (aged 18-64) living within Canada. Non-permanent residents and northern Canadian sub-populations were excluded from analysis due to sampling issues and data availability.

Each CIMI indicator is analyzed using a uniquely pre-defined sub-population/sample that is filtered by various socio-demographic control variables or specific categories within these controls/indicators.

For more details on the CIMI filters, please refer to the Codebook on the Resources page of the website.

Economic Filters:

INDICATOR	FILTER
Wages	Age: 18 to 64; population of interest is limited to currently employed, paid workers working for wage, salary, tips or commission
Low Income Cut-Off (Lico)	Age: 18 to 64
Labour Force Participation	Age: 18 to 64
Employment Rate	Age: 18 to 64
Unemployment Rate	Age: 18 to 64, in the labor force
Full-Time Employment Rate	Age: 18 to 64
Non-Official Language At Work	Age: 18 to 64
Subsidized Housing	Age: 18 to 64

Social Filters:

INDICATOR	FILTER
Number of Close Friends	Age: 18 to 64
Number of Close Friends Living In The Same City/Community	Age: 18 to 64
Sense of Belonging to Local Community	Age: 18 to 64
Sense of Belonging to Province	Age: 18 to 64
Sense of Belonging to Canada	Age: 18 to 64, exclude Quebec[4]
Victim of Discrimination in the Past 5 Years	Age: 18 to 64

4 The Quebec province is excluded from this indicator because of its uniqueness in the Canadian context.

Civic & Democratic Participation Filters:

INDICATOR	FILTER
Unpaid Volunteer Work in the Past 12 Months	Age: 18 to 64
Involvement in Organizations In the 12 Months	Age: 18 to 64
Voted in Last Provincial Election	Age: 18 to 64
Voted in Last Federal Election	Age: 18 to 64

Health Filters:

INDICATOR	FILTER
Have a Medical Doctor	Age: 18 to 64
Self-Perceived Life Stress	Age: 18 to 64
Self-Perceived Unmet Health Care Needs	Age: 18 to 64
Satisfaction with Life	Age: 18 to 64

SAMPLE ATTRITION

The economic dimension of integration is based on the Census, which has the highest and most reliable counts of immigrants in Census Metropolitan Areas (CMAs), however, the other dimensions do not yield as large a sample as they are based on the GSS (social and civic and democratic participation dimensions) and CCHS (health dimension) datasets.

Larger CMAs like Toronto, Vancouver and Montreal have immigrant sample sizes in the hundreds or even thousands in any given survey year and are likely to produce reliable regression estimates. Even midsize CMAs such as Winnipeg or Ottawa-Gatineau still have sufficient samples across survey years (typically in the hundreds). However, immigrant sample sizes in the GSS and CHHS for several smaller CMAs (e.g., Peterborough in Ontario or Lethbridge in Alberta) can be very minimal. After conducting a series of reliability tests, we established a baseline of n=15 immigrants minimum (for the unweighted sample) per geography and across the majority of survey years (i.e., for at least two out of three GSS survey periods and three out of five CCHS survey periods). It was determined that this was the minimum threshold for a CMA to be included in the final rankings.

The final analysis includes rankings for 20 CMAs across all four dimensions of integration (economic, social, civic and democratic participation, and health). A list of these CMAs is provided in the Appendix along with immigrant sample sizes across years for both surveys (GSS and CCHS). By taking this precautionary measure and only ranking the top 20 CMAs in terms of immigrant population sizes, we are able to produce more reliable regression estimates and avoid making inferences about smaller CMAs with a limited representation of immigrant diversity.

It should be noted that within each dimension of integration, adjusted and unadjusted data is presented for all CMAs with available and reliable data: economic (n=35), social (n=20), civic and democratic participation (n=20) and health (n=24). The overall index presents information for n=20 CMAs as this is the lowest common denominator across the four dimensions and survey periods.

ANALYTIC MODEL

Immigrant status and geography are the key independent variables. An immigrant * geography interaction term is used to measure the gap between immigrants and the Canadian-born population across Canadian geographies.

Regression Equation: $Y = b_0$ constant + b_1 immigrant + b_2 geography + b_3 immigrant x geography + ... controls

Where:

A = Immigrant Status	**B = Geography**	**A * B = Immigrant Status * Geography**
0 = Canadian Born (ref)	0 = Rest of Canada (ref)	
1 = Immigrant	1 = Geography of Interest (e.g., Quebec)	

$b_1 = M_{ROC, Immigrant} - M_{ROC, Canadian\text{-}born}$	Interpretation: The difference between the means for immigrants and the Canadian-born population in the ROC (i.e., when b2 GEO = 0).
$b_2 = M_{QC, Canadian\text{-}born} - M_{ROC, Canadian\text{-}born}$	Interpretation: The difference between the means for Canadian-born Quebecers and Canadian-born in the ROC (i.e., when b1 immigrant = 0).
$b_3 = M_{QC, Immigrant\text{-}Canadian\text{-}born} - M_{ROC, Immigrant\text{-}Canadian\text{-}born}$	Interpretation: The difference between immigrants and the Canadian-born population in Quebec vs. immigrants and non-immigrants in the ROC.

The interaction term coefficient (b_3) is used as the basis for the provincial/CMA's rankings, which will be explained further in the following section. Please refer to **Appendix A** for an explanation/illustration of two regression models under the Economic Dimension.

HOW ARE GEOGRAPHIES RANKED IN THE CIMI?

The analytic procedure described above (regression analysis) is replicated for each province and CMA with available data, across survey years and for all 22 indicators of the Canadian Index for Measuring Integration (CIMI). The results of the regression analysis produce a series of standardized coefficients for the interaction between immigrant status and geography. Canadian geographies (provinces and CMAs) are then ranked based on these regression coefficients (*standardized beta* for the linear regression models and odd ratios for the logistic regressions models).

Table 5 presents a snapshot of what this analysis and ranking procedure looks like. In the table, the value for "score" represents the standardized regression coefficient for the interaction between *immigrant status * geography*. The rankings column simply orders these scores from highest to lowest (or from lowest to highest) in terms of influence on outcomes. Dimension rankings (in this case, the Economic Dimension) are based on the average scores across all indicators (wages, labour force participation, unemployment, employment, low-income cut-off, non-official language at work, etc.). A normalization procedure (min-max) is used to rescale the standardized regression scores to have values between 0 and 1 prior to averaging them across indicators (see below for a more detailed explanation). This strategy helps to reduce any instances of "ties" in the rankings.

Table 5. Ranks and Scores for the Economic Dimension (2016)

2016	ECONOMIC DIMENSION		WAGES		LABOUR FORCE PARTICIPATION		UNEMPLOYMENT RATE		...	
Province	Overall Score	Overall Rank	Score	Rank	Score	Rank	Score	Rank	Score	Rank
NL	.891	1	1.000	1	1.000	1	1.000	1	...	...
PE	.288	10	.508	6	.000	10	.095	9	...	...
NS	.549	6	.797	2	.384	8	.408	7	...	...
NB	.463	8	.605	4	.355	9	.524	5	...	...
QC	.378	9	.000	10	.433	7	.000	10	...	...
ON	.569	5	.474	7	.584	6	.420	6	...	...
MB	.807	2	.556	5	.814	2	.752	2	...	...
SK	.743	3	.342	8	.810	3	.644	3	...	...
AB	.520	7	.1710	9	.615	5	.361	8	...	...
BC	.670	4	.6315	3	.671	4	.561	4	...	...

Note: "Score" represents rescaled regression coefficients (0,1) for the interaction between immigrant status and geography. "Overall Rank" (Dimension Rank) is based on the average across all indicator scores within the dimension.

Data source: Census 2016.

The same process demonstrated in Table 5 is used when analyzing the other dimensions of the integration (social, civic and democratic participation, and health dimensions).

It is important to note that **CIMI rankings should be interpreted with caution** when establishing their overall significance. Differences in rankings are not equal and can be minimal in some cases. The statistical significance of the results has been denoted (at $p < 0.05$ or $p < 0.1$) in the exportable spreadsheets.

DATA TRANSFORMATIONS AND WEIGHTING

Various transformations and recoding of the data for indicators and control variables have been conducted (for details, please refer to the Codebook). Diagnostic tests and procedures (including normality, linearity, heteroscedasticity, multicollinearity and curve estimation procedures) as well as multiple tests of correlational and statistical significance have been deployed in order to ensure that the data and models have produced results of trusted quality. Population and bootstraps weights provided by Statistics Canada and calculated normalized weights were applied as needed.

DIMENSION WEIGHTS

Overall rankings for the CIMI are based on the following weights:

- Economic Dimension: 0.4 (40%)
- Social Dimension: 0.3 (30%)
- Civic & Democratic Participation Dimension: 0.2 (20%)
- Health Dimension: 0.1 (10%)

The weighting system was developed by CIMI researchers and approved by the EAC. A unique aspect of the CIMI's website is that it allows users to change these weights to reflect their personal research interests (www.integrationindex.ca/search-rank-compare/rank/). For instance, a health service provider organization may be more interested in the Health Dimension and could increase the .1 weight to .5, which would change the standardized scores and ranks produced by the Index.

A NORMALIZATION PROCEDURE (MIN-MAX)

In CIMI 1.0, we ran separate regression models for each indicator and **rank-ordered** the regression coefficients across geographies (standardized betas for linear regression and odds ratios for non-linear regression). Provinces were ranked from 1 to 10 and CMAs were ranked from 1 to 35 based on the regression coefficients for the interaction between immigration status * geography. **These rankings were then used as scores to calculate the overall provincial/CMA rankings** within each of the four CIMI dimensions (economic, social, civic and democratic participation, health) and across all indicators of immigrant integration.

CIMI 1.0 ranking methods led to several instances of "ties" in the overall rankings. In other words, when calculating scores based on the ranks for each indictor, some provinces or CMAs received the same overall ranking. E.g., 1+2+3+4+5 = 15/5 = 5+3+4+1+2 = 15/5.

To correct for this issue in CIMI 2.0, a **min-max normalization procedure** is used to rescale each indicator's regression scores to have values between 0 (min) to 1 (max) prior to averaging. Other indicator values in the middle are mapped to have a value ranging between 0 and 1.'[5]

Rescaled score = (regression coefficient - min) / (max - min)

For more details about the pros and cons of the summative ranking used in CIMI 1.0 and the min-max procedure used in CIMI 2.0, please refer to the **Appendix C**.

HOW TO HANDLE THE OUTLIERS?

While using min-max procedure for the CIMI's ranking helps avoid "ties" in the overall rankings, this method does not handle the outliers properly. Outliers or extreme scores can be identified using the standard inter-quartile range formula in Excel (see below). In order to ensure a more normal distribution of regression scores, scores that are identified to fall outside the normal range are transformed to the upper or lower-bound estimates. Note that this procedure does not change the order of scores prior to ranking, but it does "reel in" extreme values for Provinces/ CMAs on each indicator so that the overall Index score is not overly dependent on single indicators.

- QUARTILE (array, quart): formula identifies the 1st and 3rd quartiles.
- Lower range limit = Q1 – (1.5* IQR). This is 1.5 times the inner quartile range subtracting from your 1st quartile.
- Higher range limit = Q3 + (1.5*IQR). This is 1.5 times IQR + quartile 3.

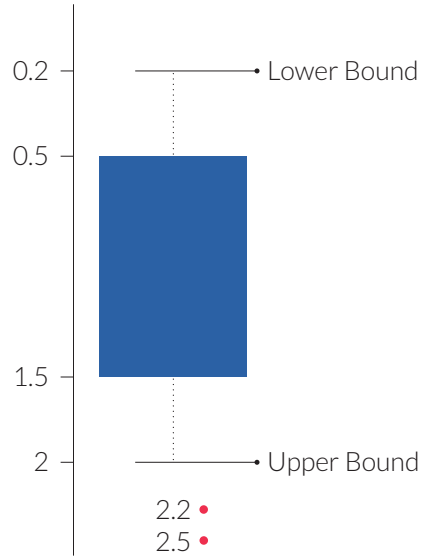
5 For several indicators such as low-income cut-off, subsidized housing, victim of discrimination, etc., the following formula is used: Rescaled score = (regression coefficient - max) / (min - max).

In the case that there were multiple outliers, we assigned values to Provinces/CMAs that fell within the normal distribution of scores. The most extreme value would receive the upper or lower bound limit while the next extreme score would be assigned an equidistant value between the upper/lower limit and the next observed score that falls within the normal distribution.

For example: Consider the scenario where the lower range limit for a specific indicator is 0.2 and the upper bound range is 2 (odds ratios), normal values fall between .5 and 1.5, but we had two extreme values of 2.2 and 2.5 in the distribution.

In this example, the extreme scores would be assigned new values that fall within the lower and upper bound ranges: '2.5' would be assigned the value of '2' and '2.2' would be assigned the value of '1.75'.

Figure 1 – Example of outlier box plot



SIMULATION MODEL

The Simulation Model is a new feature in CIMI 2.0, which allows users to view their predicted wages relative to their immigrant or Canadian-born counterparts.

Users can use this interactive tool by inputting *their demographic information* such as age group, sex, city of current residence, education level, knowledge of official languages, immigrant status, etc. and the tool will generate their corresponding estimated outcomes (please refer to Appendix D for the detailed model).

However, due to time and budget limitations, CIMI 2.0 only simulate "wages", which means that users input their demographic information and the model will generate *their estimated wages* compared to their immigrant or Canadian-born counterparts. The model is based on the linear regressions from the most recent Census (2016). The sample applied to this model is comprised of the adult population (aged 18-64), Canadian-born or landed immigrants (non-permanent resident is not included), currently employed, working for wage, salary, tips or commission, and wages from $1-$200,000.

Regression Equation: Y (estimated wages) = b_0 constant + b_1 established immigrant + b_2 recent immigrant + … controls

The dependent variable in the models is "wages". We did not use the square root transformation to correct for skewed data, however, we did remove outliers (i.e., annual wages that exceeded $200,000) to ensure a normal distribution of the dependent variable.

The key independent variable is immigrant status ("Canadian-born" is a reference groups; "established immigrants" and "recent immigrants" are inserted in the model).

The control variables include: visible minority status (binary variable), sex (binary variable), full-time status (binary variable), highest level of education (binary variable: "less than bachelor" and "bachelor and above"), age (5 groups), knowledge of official languages ("English" is a reference group; "French", "Both English and French", "Neither English or French" are inserted in the model), mobility status over 5 years (binary variable).

Each city (CMA) has its own regression equation. Therefore, a city filter will be applied along with the other filters mentioned above. For example, to have a regression equation for Montreal, the following filter will be applied: [CMA=462 (Montreal's code), age 18-64, employed, and working for wages, wages from $1-$200,000]. The regression equation generated from this model will be used for the simulation of people currently residing in Montreal. A similar procedure is replicated for each city (CMA), which finally results in 34 equations for 34 cities/CMAs), except for Guelph due to the sampling reason.

APPENDIX A: CONTROL VARIABLE DEFINITIONS

All CIMI indicator models allow for the simulation of immigrant integration outcomes compared to the Canadian-born population when taking into consideration (controlling for) their socio-economic/demographic characteristics as well as their geography (provinces and CMAs) over time. The following examples illustrate the Linear Regression as well as the Logistic Regression method that are deployed in order to estimate/model all indicator variables within each of the five dimensions (economic, social, civic and democratic participation, and health).

Linear Regression Example:

To illustrate, the following is a linear regression model predicting earned wages, using Census 2016, constructed in the following mathematical denotation:

$$Y = b_0 \text{ constant} + b_1 \text{ immigrant} + b_2 \text{ geography} + b_3 \text{ immigrant x geography} + \ldots \text{controls}$$

Predicted Wages (sqrt WAGES)* = 118.077 (y-intercept) – 14.488 (immigrant) – 11.369 (Quebec) – 6.579 (immigrant*Quebec) – 27.262 (female) + 1.408 (age) – 14.110 (visible minority) – 7.802 (French) + 2.405 (both English and French) – 31.444 (neither English or French) + 6.087 (education) – 7.697 (mobility status 5 years) + 77.003 (FT worker) – 40.233 (NOC1) – 16.736 (NOC2) – 26.640 (NOC3) – 41.511 (NOC4) – 70.765 (NOC5) – 65.770 (NOC6) – 45.985 (NOC7) – 63.070 (NOC8) – 44.378 (NOC9)

Notes: NOC = National Occupational Code 2015 Version 1.3. Omitted category is NOC0 = Management occupations. NOC1 = Business, finance and administration occupations; NOC2 = Natural and applied sciences occupations; NOC3 = Health occupations; NOC4 = Occupations in education, law and social, community and government services, NOC5 = Occupations in art, culture, recreation and sport; NOC6 = Sales and service occupations; NOC7 = Trades, transport and equipment operators and related occupations; NOC8 = Natural resources, agriculture and related production occupations; NOC9 = Occupations in manufacturing and utilities.

After controlling for socio-demographic characteristics to include sex, age, visible minority status, language, education, mobility status and occupation in the above model:

- Immigrants in the rest of Canada earned an average of 14.488 "units" less than the Canadian-born. Results are statically significant at p<.001. As previously mentioned, *wages* was transformed using the square root function to account for a non-normal distribution. When the outcome variable is transformed using the square root, it is possible to interpret the squared regression coefficients (14.448^2 = $208.74), which implies that being an immigrant reduces one's wages by $208.74 after adjusting for other sociodemographic characteristics. In this instance, the wage gap may seem minimal, but that is due to the control variables accounting for residual differences;

- The Canadian-born population in Quebec province earned an average of 11.369 units less than the Canadian-born population in the rest of Canada ($129.25). Results are statically significant at p< .001;

- The wage gap between immigrants and the Canadian-born population in Quebec was 6.579 units less than (-$43.28) the immigrant wage gap in the rest of Canada. Results are statically significant at p< .001.

* In our regression model, a square root of wages is used instead of the original "wages" variable.

MODEL	UNSTANDARDIZED COEFFICIENTS		STANDARDIZED COEFFICIENTS	T	SIG.
	B	Std. Error	Beta		
(Constant)	118.077	.132		893.456	.000
Immigrant status	-14.488	.078	-.060	-185.043	.000
QB	-11.369	.097	-.047	-117.342	.000
QBxIMM	-6.579	.146	-.012	-45.128	.000
Female	-27.262	.052	-.134	-524.891	.000
Age	1.408	.002	.175	735.373	.000
Education: Highest certificate, diploma or degree	6.087	.008	.204	778.683	.000
Visible minority	-14.110	.077	-.056	-184.015	.000
French	-7.802	.121	-.022	-64.652	.000
Both EF	2.405	.081	.010	29.870	.000
Neither EF	-31.444	.297	-.024	-105.882	.000
Mobility over 5 years	-7.697	.059	-.029	-129.377	.000
FT worker	77.003	.064	.286	1201.015	.000
NOC1 (Business)	-40.233	.091	-.148	-441.815	.000
NOC2 (Natural Sciences)	-16.736	.109	-.043	-152.854	.000
NOC3 (Health)	-26.640	.112	-.068	-237.006	.000
NOC4 (Education)	-41.511	.097	-.136	-428.084	.000
NOC5 (Arts/Culture)	-70.765	.170	-.101	-416.762	.000
NOC6 (Sales/Service)	-65.770	.090	-.271	-731.343	.000
NOC7 (Trades)	-45.985	.098	-.157	-471.107	.000
NOC8 (Agriculture)	-63.070	.187	-.082	-337.906	.000
NOC9 (Manufacturing)	-44.378	.127	-.095	-350.469	.000

This model demonstrates data findings of trusted quality based on its "Best Linear Unbiased Estimates" (BLUE) results for the model with conformity to normality, linearity and homoscedasticity. The model is also showing a non-multicolleanarity conformity based on tolerance and VIF statistics analysis and has an overall healthy indicator explanatory/predictive power R2 value of about 32.4%.

Logistic Regression Example:

To further illustrate, the following is a logistic multivariate regression model estimating the dependent dichotomous variable Non-Official Language (NOL) at work – based on Census 2016 – constructed in the following mathematical denotation:

$$\text{Probability of use of NOL at work} = e^{a+b_1x_1+b_2x_2+\ldots b_nx_n} / 1 + e^{a+b_1x_1+b_2x_2+\ldots b_nx_n}$$

Where $b_1, b_2, \ldots b_n$ are the logistic (probability) model's coefficients that correspond to independent variables (predictors/controls) $X_1, X_2, X_3 \ldots X_n$.

STEP 1[a]	B	S.E.	WALD	df	SIG.	Exp(B)
IMM	1.957	.007	81445.588	1	.000	7.081
ON	-.483	.009	2713.900	1	.000	.617
ONxIMM	.253	.010	633.608	1	.000	1.288
Vismin	.950	.005	33452.359	1	.000	2.586
Female	-.011	.004	7.825	1	.005	.989
Age	.008	.000	2549.538	1	.000	1.008
Education Level	-.048	.001	7090.442	1	.000	.953
French	-.700	.010	4726.592	1	.000	.496
Both E & F	-.521	.006	6631.062	1	.000	.594
NeitherEF	3.245	.007	217173.105	1	.000	25.656
Business	-.445	.007	3562.528	1	.000	.641
Natural	-1.105	.011	10284.084	1	.000	.331
Health	-.703	.010	4895.509	1	.000	.495
Education & Law	.017	.008	4.768	1	.029	1.017
Art & Culture	.391	.011	1317.190	1	.000	1.478
Sales & Services	-.160	.006	620.240	1	.000	.852
Trade Transport	-.293	.007	1564.108	1	.000	.746
Natural & Agriculture	.517	.012	1776.473	1	.000	1.677
Manufacturing	-.630	.009	4563.098	1	.000	.533
Mobility	.255	.004	3422.743	1	.000	1.290
Constant	-5.128	.011	234987.581	1	.000	.006

Similar to the linear regression illustration above, this logistic regression model estimates the likelihood of speaking a non-official language at work for immigrants and non-immigrants in Ontario relative to the likelihood found in the rest of Canada (controlling for socio-demographics). According to the model above, immigrants in Ontario have a 28.8% [(1.288 – 1.000) x 100] higher likelihood of using a non-official language at work compared to immigrants in the rest of Canada. Results are statically significant at $p<.001$.

This logistic regression model has reliable findings and demonstrates linear conformity as per the statistically significant results of the Hosmer and Lemeshow test. The model's tolerance and VIF statistics are within normal range, indicating that the assumption of multicollinearity has not been violated; the model also shows decent explanatory/predictive power with a Nagelkerke R^2 value of 31.5%.

APPENDIX B: DESCRIPTIVE DATA EXAMPLES

Descriptive analyses were conducted to show the raw difference or gap across all CIMI indicators for immigrants and the Canadian-born population. The descriptive data aims to:

- complement the multivariate inferential techniques;
- provide the ability to examine both snapshots and patterns/trends of economic immigrant integration outcomes over time;
- assist in further situating and interpreting ranking outcomes.

The following is a summary descriptive data of average wages and proportion of people speaking non-official language most often at work at the national and provincial levels from the Census 2016. Please note results are based on the same corresponding CIMI filters outlined above, but without controlling for socio-demographic differences.

PROVINCE	WAGES (full-time workers)			WAGES (part-time workers)			NON-OFFICAL LANGUAGE AT WORK		
	Mean			Mean					
	Total CIMI population	Canadian born	Immigrants	Total CIMI population	Canadian born	Immigrants	Total CIMI population	Canadian born	Immigrants
Newfoundland and Labrador	$ 61,167	$ 60,771	$ 76,061	$ 16,242	$ 16,216	$ 17,395	0.2%	0.2%	0.7%
Prince Edward Island	$ 45,125	$ 45,134	$ 44,974	$ 16,193	$ 16,291	$ 14,705	0.8%	0.1%	18.1%
Nova Scotia	$ 52,051	$ 51,705	$ 57,748	$ 15,772	$ 15,846	$ 14,634	0.3%	0.2%	2.8%
New Brunswick	$ 48,820	$ 48,730	$ 50,810	$ 16,058	$ 16,149	$ 14,275	0.4%	0.1%	6.9%
Quebec	$ 52,657	$ 53,703	$ 46,513	$ 17,009	$ 17,230	$ 15,688	1.2%	0.5%	5.1%
Ontario	$ 62,579	$ 64,895	$ 57,494	$ 16,566	$ 16,548	$ 16,613	2.4%	0.3%	7.0%
Manitoba	$ 54,268	$ 56,794	$ 45,145	$ 18,390	$ 18,683	$ 17,360	1.8%	0.8%	5.5%
Saskatchewan	$ 61,799	$ 63,373	$ 50,985	$ 19,638	$ 19,982	$ 17,612	0.9%	0.4%	5.3%
Alberta	$ 75,631	$ 78,957	$ 65,248	$ 22,145	$ 22,332	$ 21,560	1.4%	0.3%	4.8%
British Columbia	$ 58,396	$ 60,544	$ 53,481	$ 17,586	$ 17,717	$ 17,232	4.7%	0.3%	14.6%
Canada	**$ 60,405**	**$ 61,849**	**$ 55,752**	**$ 17,579**	**$ 17,676**	**$ 17,232**	**2.1%**	**0.4%**	**7.7%**

APPENDIX C: COMPARISON OF CIMI 1.0 & 2.0 RANKING METHODS

METHOD USED	CIMI 1.0	CIMI 2.0
	Summative ranking	Min Max transformation
Explanation	Rescaled the regression scores from 1 to 10 (for provinces) and from 1 to 35 (for CMAs)	Rescaled the regression scores between 0 to 1 for both provinces and CMAs.
Formula	Giving the ranking '1' to the biggest number, '2' to the second biggest value and so on.	Rescaled score = (regression coefficient - min) / (max - min)
Pros & Cons	Pros: - All ranks are equally far apart from each other, so the impact of outliers is mitigated. Cons: - Distorts correlations and original distances (variations) across geographies - Some geographies may have same rankings	Pros: - Avoid "ties " in the overall rankings - Keep the original distances across geographies Cons: - Do not handle the outliers well

APPENDIX D: SIMULATION MODEL

This page gives you a glimpse into the wage gap between immigrants and the Canadian-born population in the city/CMA of your choice. Once you input your demographic information below, your predicted annual wage and the estimated wage of your immigrant or Canadian-born counterpart (whichever applies) will be generated based on Census 2016 data.

Please note that this simulation:

- does not represent actual wages;
- only applies to adults aged 18-64 who are Canadian-born or immigrants (non-permanent resident is not included), currently employed, work for wages, salary, tips or commission, and earn annual wages from $1-$200,000.

To obtain an estimate of your wage, please answer the following questions:

1. Were you born in Canada? [select between two: Yes/ No];
2. How long have you been a permanent resident? [Select between two: <= 5 years / >5 years];
3. What city (CMA) do you currently reside in? [select from list of 35 cities];
4. How old are you? [Select among five age groups: 18-24/ 25-34/ 35-44/ 45-54/ 55-64];
5. What is your sex at birth? [select between two: Female/ Male];
6. What is the highest level of education that you have completed? [select between two: University degree or above/ Less than a University degree];
7. Which official languages can you conduct a conversation in? [select among four: English/ French/ Both English and French/ Neither English or French];
8. Do you identify as a visible minority person? [select between two: Yes/ No];
9. Are you a…? [select between two: Full-time employee/Part-time employee];
10. Have you changed neighbourhoods in the past 5 years? [Select between two: Yes/ No].

Your expected wage	$ X
Your immigrant (or Canadian-born) counterpart's wage	$ X'
Wage gap	$ a

Example of "wages" simulation model for the City of Montreal:

- **Dependent variable:** Wages (original variable);
- **Independent variables:** Immigrant status, vismin status, age groups, education level, knowledge of official languages, mobility status, full-time status;
- **Filter:** (AGE >= 18) & (AGE <= 64) & (EMPL = 1) & (COWfilter = 1) & (CMACIMI = 462) & (Wages >= 1) & (Wages <= 200000).

MODEL	UNSTANDARDIZED COEFFICIENTS		STANDARDIZED COEFFICIENTS	T	SIG.
	B	Std. Error	Beta		
2 (Constant)	949.765	128.183	—	7.409	.000
RecIMMsimu	-14179.415	119.766	-.085	-118.393	.000
EstIMMsimu	-6079.243	68.175	-.074	-89.171	.000
Visible minority	-5477.547	69.953	-.066	-78.303	.000
CIMI age groups	6703.772	18.084	.259	370.692	.000
SEXre	-9708.533	42.692	-.148	-227.411	.000
University	20013.443	47.618	.281	420.289	.000
Frenchre	1443.244	109.588	.019	13.170	.000
BothEF	8481.591	105.147	.120	80.664	.000
NeitherEF	-11777.349	401.848	-.019	-29.308	.000
Mobility over 5 years	-1176.211	56.862	-.014	-20.685	.000
FT Worker	25361.507	58.415	.295	434.162	.000

This document was produced as part of a larger project funded by Immigration, Refugees and Citizenship Canada (IRCC).

PUBLISHED 2020

Institut canadien pour les identités et les migrations | The Canadian Institute for Identities and Migration

ACS • AEC
Association for Canadian Studies • Association d'études canadiennes