

# Classification of Canadian immigrants into visible minority groups using country of birth and mother tongue

---

Mohammad R. Rezai, Laura C. Maclagan, Linda R. Donovan, Jack V. Tu

## **Correspondence:**

Jack V. Tu, MD, PhD  
Institute for Clinical Evaluative Sciences,  
G-106, 2075 Bayview Ave,  
Toronto, Ontario, Canada M4N 3M5.

E-mail: [tu@ices.on.ca](mailto:tu@ices.on.ca)  
Tel: 416-480-4700  
Fax: 416-480-6048

**Word Count:** 2973

**Number of Figures and Tables:** 5

**Reprints:** Requests for reprints should be addressed to Jack V. Tu (information above)

MR Rezai MD, PhD is a postdoctoral fellow and LC Maclagan MSc, an epidemiologist in Cardiovascular and Diagnostic Imaging Research Program, Institute for Clinical Evaluative Sciences (ICES). LR Donovan BScN, MBA is a project manager with Sunnybrook Research Institute. JV Tu, MD, PhD, FRCPC, is a Tier 1 Canada Research Chair in Health Services Research, and a senior scientist at Institute for Clinical Evaluative Sciences and Clinical Epidemiology Program, Sunnybrook Research Institute. He is also staff physician, Division of Cardiology, Schulich Heart Centre, Sunnybrook Health Sciences Centre and a Professor of Medicine and Health Policy, Management and Evaluation at the University of Toronto.

**Competing interests:** The authors have no conflicts of interest to disclose.

**Contributors:** MR Rezai and LC Maclagan contributed to literature review, drafted the first version of the manuscript, and provided critical appraisal of the manuscript. MR Rezai classified countries and languages, and performed statistical and geographic analyses. LR Donovan provided critical appraisal and contributed to the final draft and revision of the manuscript. JV Tu contributed to data acquisition and critical appraisal, provided funding, and supervised the entire project. All authors contributed to conception of the review, revision of the article for intellectual content and gave final approval for publication.

**Funding:** This study was supported by operating grants from the Public Health Agency of Canada (PHAC), the Heart and Stroke Foundation of Ontario, and a Team Grant from the

Institute of Circulatory and Respiratory Health-Canadian Institutes of Health Research. ICES is supported by the Ontario Ministry of Health and Long-Term Care (MOHLTC). Dr. Tu is supported by a Canada Research Chair in Health Services Research and a Career Investigator award from the Heart and Stroke Foundation of Ontario. The results and conclusions are those of the authors, and should not be attributed to any of the funding sources. The opinions, results and conclusions reported in this paper are those of the authors and are independent from the funding sources. No endorsement by PHAC, ICES or the Ontario MOHLTC is intended or should be inferred.

## Abstract

**Background:** Citizenship and Immigration Canada's Landed Immigrant Database System (CIC-LIDS) contains socio-demographic information on immigrants, but lacks ethnic group classifications. To enhance its usability for ethnicity-related research, we categorized CIC-LIDS immigrants into one of Canada's official visible minority groups (VMGs) ethnic classification using their country of birth and mother tongue.

**Methods:** Using public data sources, each of 266 country names and 245 mother tongues in CIC-LIDS were classified into one of 10 VMGs: South Asian, Chinese, Black, Filipino, Latin American, West Asian, Arab, South-East Asian, Korean, and Japanese plus a White group. Classifications using country of birth alone (Method A) or country of birth plus mother tongue (Method B) were applied to 2.5 million CIC-LIDS immigrants to Ontario between 1983 and 2010 with a valid encrypted health-card number. The ethnic categorizations were validated using linked self-reported ethnicity data from the Canadian Community Health Survey (CCHS) (n=6,499).

**Results:** The four most frequent VMGs among CIC-LIDS immigrants classified by Method B were South Asian (n=582,812), Chinese (n=400,771), Black (n=254,189), and Filipino (n=150,898). Methods A and B agreed in 94% of the categorizations; Kappa(95%CI): 0.92(0.91-0.93). Methods A and B agreed with self-reported CCHS ethnicity in 86% of total categorizations; Kappa(95%CI): 0.83(0.82-0.84). High sensitivity and specificity were found for both Method A and B in most VMGs when validated using self-reported CCHS ethnicity (e.g. sensitivity/specificity in South Asians: 0.85/0.97, Chinese: 0.93/0.99, Blacks: 0.90/0.97).

**Interpretation:** The use of country of birth and mother tongue is a validated and practical method for the classification of immigrants to Canada into different ethnic categorizations.

**Word count:** 256

## **Introduction**

As one of the world's most ethnically diverse countries, [1] Canada is home to individuals of over 200 different ethnic origins.[2] Canada's growing diversity is due, primarily, to high levels of immigration. Since the 1990's approximately 250,000 immigrants have arrived annually.[2] The major sources of Canada's immigrants include Asia, Europe, the Caribbean, South and Central America, Africa and the United States. [1] In Ontario, Canada's largest province, 2006 census data identified that 23% of the population belonged to an ethnic minority group, with the largest groups being South Asian, Chinese, and Black. [1] From 2007 to 2011, 42% of all Canadian immigrants landed in Ontario.[3]

The increasingly multiethnic nature of Canadian society and other countries is fuelling a need for ethnicity data to better understand these diverse populations. Ethnicity classifications in health research can be used to better understand the etiology of disease, the roles of environment and genetics, and the health status of disadvantaged groups, as well as improve health care delivery and target public health interventions at high-risk populations.[4,5]

Although long recognized as an important covariate in health research, individual level ethnicity data are rarely collected in health care datasets in Canada. Although some ethnicity data is captured in Canada's census, these data cannot leave Statistics Canada and be linked to many other administrative datasets available in Canada's provinces. In an effort to address this gap, alternative sources of ethnicity classification have been used. While the concept of ethnicity is complex and its definition challenging, [5-7] various methods used include surname-based approaches, geocoding of residential address, or classification based on country of birth, language or a combination of the above have been used to assign ethnicity. [8-12]

Country of birth, in particular, has been widely collected in many administrative and government datasets and represents an objective and potentially valuable source of ethnicity information.

Given Canada's high rate of immigration, Citizenship and Immigration Canada's (CIC) Landed Immigrant Data System (LIDS) database may be a useful source of ethnicity data for health research. In the past decade, this database has been used for socioeconomic and health studies among immigrants in different Canadian provinces [13-19]. The CIC LIDS provides detailed pre-landing demographic and socioeconomic information on all Canadian immigrants including country of birth. However, this dataset lacks self-reported ethnicity and the large number of options for country of birth and mother tongue (i.e. over 200 options for each category) within this dataset can also make it challenging to use for such purposes. In an attempt to improve its practical use for ethnicity-related research projects, we describe a method for classifying Ontario CIC LIDS data records into 11 official visible minority ethnic groups in Canada using a) country of birth; and b) country of birth plus mother tongue variables and validated this method using another population-based survey with self-reported ethnicity information, the current 'gold standard' for ethnicity classification.

## **Methods**

### ***Citizenship and Immigration Canada's (CIC) Landed Immigrant Data System (LIDS)***

The CIC LIDS database provides detailed socio-demographic information on all legal immigrants to Canada including country of birth, citizenship, country of last permanent residence, and mother tongue. For this analysis we used the CIC LIDS dataset held at the Institute of Clinical Evaluative Sciences, which pertains to Ontario immigrants arriving between 1985 and 2010. This data set includes 266 options for country of birth and 245 options for mother

tongue. The Ontario CIC LIDS database has been used as a source of ethnicity data for previous health research studies [13-15].

### ***Study population***

The CIC LIDS dataset used for this study contains records for 2.9 million Ontario immigrants for the 1985-2010 time period. Around 400,000 records were excluded for immigrants where it was not possible to identify a valid health card number in the Ontario Registered Persons Database, as this data was required for record linkage to the self-reported ethnicity data used for validation. Landed immigrants become eligible for health care benefits after a 3 month waiting period. Records for the remaining 2.5 million Ontario immigrants were linkable to other administrative databases available at the Institute for Clinical Evaluative Sciences. All data were deidentified and health card numbers were encrypted to ensure privacy protection.

### ***Classifying the CIC LIDS data into ethnic groups***

The CIC LIDS database lacks ethnic or visible minority group classifications. In order to facilitate use of the CIC LIDS data for health research, we tested two methods to classify the Ontario dataset into 11 ethnic categories including 10 official visible minority groups used by Statistics Canada: South Asian, Chinese, Black, Filipino, Latin American, West Asian, Arab, South-East Asian, Korean and Japanese plus a White category. According to The Employment Equity Act, visible minorities are defined as "persons, other than Aboriginal peoples, who are non-Caucasian in race or non-white in colour". [20] First, we used a country of birth (method A) and then we used country of birth plus mother tongue (method B).

We used self-reported ethnic group information available in Statistics Canada's Canadian Community Health Survey (CCHS) as the gold standard reference to test our ethnic group classifications after linking to the same individuals' record in the Ontario CIC LIDS dataset.

### **Method A: Country of birth**

Each of the 266 country of birth names in the Ontario CIC LIDS (including previous names changed due to political reasons) were mapped to one of 12 categories: 10 visible minority groups specified by Statistics Canada (South Asian, Chinese, Black, Filipino, Latin American, South-East Asian, Arab, West Asian, Korean and Japanese [1]), a White category and an Excluded category. A combination of publicly available resources were used for this purpose, including Statistics Canada Census 2006 ethnic origin categories (our preferred source), [21] United Nations Standard Country or Area Codes for Statistical Use (also known as M49 list), [22] World Bank list of economies (July 2012),[23] and The World Factbook by the United States Central Intelligence Agency (CIA). [24] These resources consider the ethnic mix of countries and provide more information needed to appropriately assign each country to the predominant ethnic group in that country.

The 10 visible minority categories used by Statistics Canada are heterogeneous. Some categories are associated with a single country, and classification is straightforward (e.g., Japan for *Japanese*). Other categories such as *South Asian* and *Latin American* relate to relatively well-defined world geographic regions. For example, we assigned the countries in South America and most of Central America to the *Latin American* category. In contrast, categories such as *Black* and *Arab* may be considered primarily ethno-cultural classifications associated with overlapping geopolitical boundaries (see Methodological Details in the supplement).

The *White* category included European countries and those with predominantly European origin population (e.g., Australia). The *Excluded* category was created to account for the countries whose nationals are not expected to be among the 10 major visible minority groups defined by Statistics Canada, or the White category defined above, or those countries whose names in CIC LIDS data were problematic phrases (e.g., country not



stated, British Overseas Citizen). Details of countries included in each of the 12 categories can be found in the online supplement.

***Method B: Country of birth plus mother tongue***

In an effort to further refine the classification based on country of birth, we completed a second classification based on country of birth plus mother tongue. The ethnic makeup of countries is often heterogeneous, and there may be individuals whose ethnic background is different from the predominant ethno-cultural group(s) of their country of birth. For instance, a person may be born to South Asian parents in a country with a predominantly white population (e.g., United Kingdom). In such cases, the mother tongue of an individual may be more representative of their ethnic background than their country of birth.

We mapped each of the 245 languages found in the Ontario CIC LIDS dataset to one of 15 categories: 10 Statistics Canada visible minority groups, a White category, plus four additional language categories (*World Language*, *Other*, *Excluded*, and *Unknown*). Publicly available data sources such as The Ethnologue: Languages of the World [25], and The CIA World Factbook [24] were used to gather language information and assign each language to an ethnic group. For instance, Cantonese and Mandarin were categorized as *Chinese*, and Persian and Kurdish as *West Asian*.

The *World Language* category, created to account for languages spoken in multiple categories and by various visible minority groups, included English, French, Spanish, Portuguese, and Russian. Less specific language options were assigned to the closest category (e.g., “Other European Languages” to *White*) or to the *Other* category (e.g., Hebrew) (see Methodological Details in supplement). The *Excluded Language* category was created for three languages associated with the countries in the *Excluded* category. The *Unknown* category was created for languages for which a region of origin or single ethnic group could not be identified. The number of immigrants

speaking *Excluded* or *Unknown* languages was relatively small (<0.1% of total sample).

An algorithm was developed to determine a final category for each individual immigrant record using the two methods, country of birth (method A) and country of birth plus mother tongue (method B) (see Figure A in Supplement).

### ***Validation: Testing classification accuracy using CCHS data as reference***

For validation, we compared the ethnic group assigned by our two methods to self-reported ethnic group data in the Statistics Canada's Canadian Health Survey (CCHS). The CCHS, a population based cross-sectional health survey of Canadians aged 12 and older, includes self-reported ethnicity information. We linked four cycles of the CCHS (2000/2001 to 2007/2008) using encrypted health card numbers, to our Ontario CIC LIDS dataset. From the Ontario CCHS data set (n=134,567), we linked 6,585 records to our CIC LIDS dataset. After excluding 86 individuals with multiple ethnicities listed in CCHS, 6,499 remained for validation analysis.

### ***Statistical analysis***

Percent agreement and simple kappa statistics were calculated to compare Method A and Method B with the CCHS ethnic classification (reference standard). Visible minority categories were considered polytomous ratings by two raters when Methods A and B were compared to the CCHS classification. Overall percent agreement was defined as the number of similar ratings by two methods divided by the total number of ratings. Sensitivity, specificity, positive and negative predictive values (PPV and NPV) were calculated for each visible minority category when Methods A and B were compared to the CCHS classification. SAS version 9.2 (SAS Institute, North Carolina, USA) was used for all statistical analyses.

### ***Geographical visualization***

The global distribution of major ethnic groups associated with the countries of birth of Ontario immigrants as per CIC LIDS is depicted in the world map presented in Figure 1. ArcGIS Desktop software version 10 (ESRI, Redlands, California, USA) was used to create the map. World country boundaries dataset was obtained from thematicmapping.org [26].

This project received ethics approval through Research Ethics Board of Sunnybrook Health Sciences Centre.

## Results

The Ontario CIC LIDS study sample ( $n=2,500,514$ ) had a mean $\pm$ SD age of  $30\pm 17$  years at landing and 51% were female. Table 1 lists the top 20 most frequent countries of birth and mother tongues among these immigrants.

Figure 1 displays the world distribution of the ethnic groups assigned to countries of birth in our sample. A list of all countries and mother tongues found in the Ontario CIC LIDS with the assigned ethnic categories is available as a spreadsheet supplement.

The frequencies of the immigrants by ethnic categories based on Methods A and B are shown in Table 2. As indicated, the two methods used to classify immigrants (i.e., on the basis of country of birth alone versus country of birth plus mother tongue) can result in some differences. For instance, among 523,855 immigrants classified as *White* by country of birth, 8,271 and 2,502 individuals were classified as *South Asian* and *Chinese*, respectively, by country of birth plus mother tongue. Methods A and B showed agreement for 94% of the ratings with a Kappa coefficient (95%CI) of 0.92(0.91-0.93).

The validation sample of 6,499 CCHS records linked to our Ontario CIC LIDS dataset had a mean $\pm$ SD age of  $29\pm 15$  years at landing and 52% were

female. Table 3 shows the self-reported visible minority data in the CCHS versus those assigned by Method B (country of birth plus mother tongue). The results using Method A were very similar (Table A, online supplement).

The ethnic categorizations by our two methods (Methods A and B) agreed with the self-reported CCHS ethnic group in 86% of total ratings. Simple Kappa coefficient (95%CI) was 0.83(0.82-0.84) for both methods when compared to self-reported CCHS data.

Table 4 shows the classification accuracy of Method B when compared to self-reported CCHS ethnic groups. Consistently high specificity and NPVs were found for all groups. Sensitivity for South-East Asian, and PPV for Latin American and South-East Asian groups were relatively lower. For the majority of the indices, Method B (country of birth plus mother tongue) showed a slight improvement or no change compared to Method A (country of birth alone). The validation results for Method A are also reported in Table B of the online supplement.

### **Interpretation**

We used two methods (country of birth alone or in addition to mother tongue) to classify Ontario immigrants in the CIC LIDS into 11 pre-defined ethnic groups. We found a high degree of agreement between the self-reported ethnic group from CCHS data and that assigned by our two classification methods. Compared to country- or world region-specific classifications used previously,[13] our classification by visible minority groups may be more practical for researchers and health policy makers as it is comparable to other important population statistics on visible minorities produced by Statistics Canada and other international organizations. The utility of our proposed methods is not limited to the Canadian context, and may prove useful (with local customization) in other countries where databases exist on immigrants' country of birth and/or mother tongue, but

where health-related information regarding self-reported ethnicity is not available or not routinely collected. The online spreadsheet supplement to this report can facilitate classification in these other datasets. Similarly, in the settings where an organization feels uncomfortable asking about individuals' ethnic background, our proposed methods could be used to infer ethnicity using information on country of birth and mother tongue which might be more practical to collect.

Using country of birth to define ethnicity has been reported as a robust method for health care research in countries such as the Netherlands where it was closely correlated with self-reported ethnicity.[12,27] Nevertheless, this method has received criticism [12,28] as problems can arise with multi-ethnic countries (e.g., Australia, the USA), or with individuals born to a family whose ethnicity is different from the predominant ethnic group of their country of residence. Classification methods that use additional information such as language and parents' country of birth have been shown to improve classification accuracy as compared to methods based on country of birth alone.[12] In our study, adding mother tongue to country of birth resulted in only slight improvement in ethnicity classification. Methods using mother tongue alone to define ethnicity also have their limitations. Second or third generation immigrants in a country (e.g., in the USA) may not speak the mother tongue of their ancestors. Moreover, native individuals may report their mother tongue to be a world language (originated from predominantly *White* countries) accepted as their birth country's official language (e.g. French in Congo, English in India). The latter problem was controlled in our data by recording the world languages spoken as official languages specific to such countries.

Despite unanimously high specificity values, sensitivity of our methods to detect South-East Asian immigrants was low compared to self-reported ethnicity. Among South-East Asians there was considerable misclassification into South Asian category. This may be due to individuals' uncertainty about

world geographic boundaries of South and South-East Asia (e.g. a South Asian might think his country is located in South-East Asia) or self-identification by country of residence rather than country of birth (e.g. a person born in India who had lived in Malaysia for a long time before immigrating to Canada may self-identify as South-East Asian). We also found relatively low PPV for Latin American group despite the high sensitivity and specificity of our methods. Some immigrants from certain Latin American countries (e.g., Brazil and Argentina) are descendants of European immigrants, and self-identify as White. Moreover, a large proportion of the population in Guyana, the Latin American country with the largest number of immigrants to Ontario, are South Asian diaspora.

Some limitations may exist for the CCHS data we used to validate our classification methods. First, the CCHS population linked to our Ontario CIC LIDS dataset may not be a representative sample of Ontario immigrants. Second, the sample size for some visible minority groups (e.g. Japanese and Koreans) may have been too small to produce reliable estimates. Third, self-reported ethnicity, although often considered a preferred method of ethnic group classification, has some shortcomings. Self-reported ethnicity may change with time, and can be influenced by psychosocial factors like the feeling of pride someone attaches to their ethnic/national identity, uncertainty about ethnic origin, or even concern related to disclosing one's ethnicity.[5-7,12,29,30] For instance, 5% of the immigrants who self-reported as White in the CCHS validation dataset were classified as West Asian or Arab based on country of birth and mother tongue. It is likely that these CCHS participants were West Asians/Arabs who reported their ethnicity as White based on skin color.

Care should be taken when interpreting positive and negative predictive values which depend on prevalence of the ethnic groups in the CCHS linked validation dataset.[31] The distribution of ethnic groups in the validation dataset and Ontario CIC LIDS data are different (Tables 2 and 3).

In conclusion, we found a close agreement between the self-reported ethnic categories and those classified using a) country of birth alone and b) country of birth plus mother tongue, in our large dataset of Ontario immigrants. These findings suggest that our method of classification is a valid method of categorizing most immigrants to Canada into its official visible minority groups. Use of a larger validation dataset in future studies may further illuminate the external validity of this method.

### **Disclosures**

The authors have no conflicts of interest to disclose.

## Tables

**Table 1.** Top 20 most frequent countries of birth and mother tongues of immigrants in CIC LIDS who landed in Ontario from 1985-2010

<b>Top 20 Countries of Birth</b>	<b>No.</b>	<b>Top 20 Mother Tongues</b>	<b>No.</b>
<b>1</b> India	296,805	1 English	365,194
<b>2</b> China, People's Republic of	263,450	2 Mandarin	170,317
<b>3</b> Philippines	163,223	3 Cantonese	166,533
<b>4</b> Pakistan	134,967	4 Tagalog	143,603
<b>5</b> Sri Lanka	96,110	5 Arabic	135,219
<b>6</b> Hong Kong	94,038	6 Punjabi	134,238
<b>7</b> Poland	78,368	7 Urdu	129,566
<b>8</b> Iran	74,957	8 Spanish	123,156
<b>9</b> Jamaica	72,782	9 Tamil	96,376
<b>10</b> United States of America	60,155	10 Russian	81,746
<b>11</b> United Kingdom and Colonies	58,180	11 Polish	78,601
<b>12</b> Guyana	50,643	12 Gujarati	58,070
<b>13</b> Korea, Republic of	41,005	13 Chinese	48,702
<b>14</b> Vietnam, Socialist Republic of	39,666	14 Portuguese	48,365
<b>15</b> Romania	38,223	15 Hindi	45,879
<b>16</b> Trinidad & Tobago, Republic of	34,819	16 Farsi	43,637
<b>17</b> Yugoslavia	34,788	17 Korean	41,489



<b>18</b>	Russia	34,523	<b>18</b>	Romanian	37,099
<b>19</b>	Iraq	33,648	<b>19</b>	Bengali	37,024
<b>20</b>	Bangladesh	32,331	<b>20</b>	Persian	33,692

Note: The country and language labels are exactly quoted from LIDS formats, and therefore may refer to old country names in some cases.

**Table 2.** The frequency of CIC LIDS immigrants to Ontario (1985-2010) in each ethnic category classified by country of birth alone (Method A) and country of birth plus mother tongue (Method B).

		Method B: Using Country of Birth plus Mother Tongue												
	Frequency	Excluded	White	South Asian	Chinese	Black	Filipino	Latin American	West Asian	South-East Asian	Korean	Arab	Japanese	Total
Method A: Using Country of Birth	Excluded	<u>11,406</u>	<70*	<70*	<70*	<70*	<70*	<70*	<70*	<70*	<70*	<70*	<70*	11,406
	White	1,312	<u>501,235</u>	8,271	2,502	1,785	600	2,007	2,357	310	331	3,063	82	523,855
	South Asian	3,449	164	<u>547,439</u>	5,768	541	94	<70*	4,538	1,068	<70*	286	<70*	563,361
	Chinese	1,592	<70*	320	<u>372,298</u>	174	212	<70*	<70*	736	175	<70*	<70*	375,670
	Black	2,737	498	9,687	445	<u>249,547</u>	124	<70*	<70*	233	<70*	8,160	<70*	271,499
	Filipino	405	<70*	548	1,699	<70*	<u>148,013</u>	<70*	<70*	12,115	<70*	<70*	<70*	162,968
	Latin American	2,053	3,251	74	569	123	<70*	<u>177,111</u>	<70*	<70*	86	156	<70*	183,491
	West Asian	4,413	1,168	1,564	<70*	143	<70*	<70*	<u>119,945</u>	1,544	<70*	3,202	<70*	132,067
	South East Asian	1,250	102	2,517	17,008	<70*	244	<70*	<70*	<u>47,556</u>	<70*	<70*	<70*	68,817
	Korean	232	<70*	<70*	94	<70*	<70*	<70*	<70*	<70*	<u>40,680</u>	<70*	<70*	41,106
	Arab	8,559	329	12,214	<70*	1,766	1,276	<70*	12,569	<70*	<70*	<u>122,874</u>	<70*	159,689
	Japanese	174	<70*	166	329	<70*	217	<70*	<70*	<70*	98	<70*	<u>5,506</u>	6,585

	<b>Total</b>	<b>37,582</b>	<b>506,882</b>	<b>582,812</b>	<b>400,771</b>	<b>254,189</b>	<b>150,898</b>	<b>179,118</b>	<b>139,608</b>	<b>63,652</b>	<b>41,452</b>	<b>137,842</b>	<b>5,708</b>	<b>2,500,514</b>
--	--------------	---------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------	---------------	---------------	----------------	--------------	------------------

\* Cells with size < 70 including cells of size 0 were suppressed to protect privacy.

Note: The diagonal cells are underlined to highlight similar categorization by two methods.

CIC LIDS: Citizenship and Immigration Canada Landed Immigrant Database System

**Table 3.** The frequency of Ontario CIC LIDS immigrants in each self-reported ethnic category (from CCHS) and that categorized using country of birth plus mother tongue (Method B).

		Method B: Using Country of Birth plus Mother Tongue												
	<i>Frequency</i>	<b>Excluded</b>	<b>White</b>	<b>South Asian</b>	<b>Chinese</b>	<b>Black</b>	<b>Filipino</b>	<b>Latin American</b>	<b>West Asian</b>	<b>South-East</b>	<b>Korean</b>	<b>Arab</b>	<b>Japanese</b>	<b>Total</b>
<b>CCHS: Self-reported</b>	<b>White</b>	40	<u>2,021</u>	<10*	<10*	19	<10*	91	74	<10*	<10*	40	<10*	<b>2,300</b>
	<b>South Asian</b>	26	<10*	<u>1,105</u>	<10*	60	<10*	66	13	<10*	<10*	<10*	<10*	<b>1,296</b>
	<b>Chinese</b>	<10*	<10*	<10*	<u>720</u>	<10*	12	<10*	<10*	22	<10*	<10*	<10*	<b>769</b>
	<b>Black</b>	<10*	12	<10*	<10*	<u>586</u>	<10*	26	<10*	<10*	<10*	15	<10*	<b>645</b>
	<b>Filipino</b>	<10*	<10*	<10*	<10*	<10*	<u>321</u>	<10*	<10*	27	<10*	<10*	<10*	<b>355</b>
	<b>Latin American</b>	<10*	<10*	<10*	<10*	11	<10*	<u>364</u>	<10*	<10*	<10*	<10*	<10*	<b>389</b>
	<b>West Asian</b>	<10*	<10*	10	<10*	16	<10*	11	<u>155</u>	<10*	<10*	16	<10*	<b>224</b>
	<b>South East Asian</b>	<10*	<10*	96	12	<10*	12	<10*	<10*	<u>61</u>	<10*	<10*	<10*	<b>205</b>
	<b>Korean</b>	<10*	<10*	<10*	<10*	<10*	<10*	<10*	<10*	<10*	<u>80</u>	<10*	<10*	<b>83</b>
	<b>Arab</b>	<10*	<10*	<10*	<10*	<10*	<10*	<10*	12	<10*	<10*	<u>185</u>	<10*	<b>212</b>
	<b>Japanese</b>	<10*	<10*	<10*	<10*	<10*	<10*	<10*	<10*	<10*	<10*	<10*	<u>19</u>	<b>21</b>
	<b>Total</b>	<b>96</b>	<b>2,059</b>	<b>1,227</b>	<b>735</b>	<b>715</b>	<b>354</b>	<b>562</b>	<b>259</b>	<b>128</b>	<b>82</b>	<b>262</b>	<b>20</b>	<b>6,499</b>

\* Cells with size < 10 including cells of size 0 were suppressed to protect privacy.

Note: The diagonal cells are underlined to highlight similar categorization by two methods.

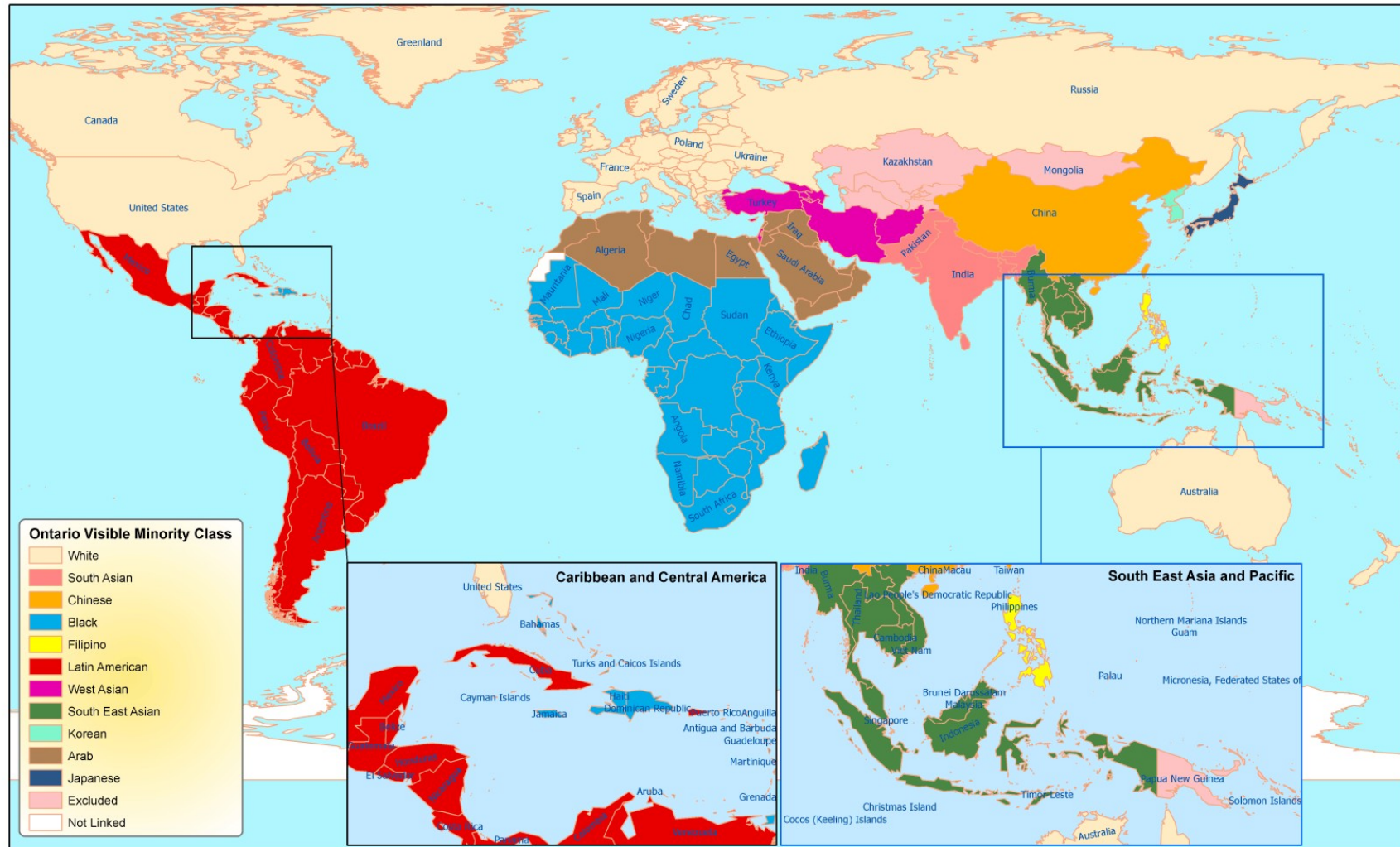
CCHS: Canadian Community Health Survey, CIC LIDS: Citizenship and Immigration Canada Landed Immigrant Database System

**Table 4.** Validation of ethnic classification method using country of birth plus mother tongue (Method B) against self-reported ethnicity (from CCHS) as reference (n=6,499).

	<b>Sensitivity (95% CI)</b>	<b>Specificity (95% CI)</b>	<b>Positive Predictive Value (95% CI)</b>	<b>Negative Predictive Value (95% CI)</b>
<b>White</b>	0.87(0.86-0.89)	0.99(0.98-0.99)	0.98(0.97-0.98)	0.93(0.92-0.94)
<b>South Asian</b>	0.85(0.83-0.87)	0.97(0.97-0.98)	0.90(0.88-0.91)	0.96(0.95-0.96)
<b>Chinese</b>	0.93(0.91-0.95)	0.99(0.99-0.99)	0.97(0.96-0.98)	0.99(0.98-0.99)
<b>Black</b>	0.90(0.88-0.92)	0.97(0.97-0.98)	0.81(0.78-0.84)	0.98(0.98-0.99)
<b>Filipino</b>	0.90(0.86-0.93)	0.99(0.99-0.99)	0.90(0.87-0.93)	0.99(0.99-0.99)
<b>Latin American</b>	0.93(0.90-0.95)	0.96(0.96-0.97)	0.64(0.60-0.68)	0.99(0.99-0.99)
<b>West Asian</b>	0.69(0.62-0.75)	0.98(0.98-0.98)	0.59(0.53-0.65)	0.98(0.98-0.99)
<b>South-East Asian</b>	0.29(0.23-0.36)	0.98(0.98-0.99)	0.47(0.38-0.56)	0.97(0.97-0.98)
<b>Korean</b>	0.96(0.89-0.99)	0.99(0.99-1.00)	0.97(0.91-0.99)	0.99(0.99-0.99)
<b>Arab</b>	0.87(0.82-0.91)	0.98(0.98-0.99)	0.70(0.64-0.76)	0.99(0.99-0.99)
<b>Japanese</b>	0.90(0.69-0.98)	0.99(0.99-1.00)	0.95(0.75-0.99)	0.99(0.99-1.00)

CCHS: Canadian Community Health Survey

**Figure 1.** World distribution of major ethnic groups associated with countries of birth of the immigrants in Ontario CIC LIDS (1985-2010)



CIC LIDS: Citizenship and Immigration Canada Landed Immigrant Database System. The countries' boundaries shown on the map is not an expression of the authors' views on the legal status of territories or definition of the boundaries.

