**International Population Data Linkage Conference**

**The LIDIC Hackathon: LInked Data Innovation Challenge**

**Workshop Date and Time:** September 11, 2018 afternoon

**Data description:**

The synthesized data base links socioeconomic and mortality data representing the Canadian population. The data based was derived from existing linked data available at Statistics Canada. This linked data base contains rich demographic, socioeconomic (income, education, employment), identity (immigration, ethnicity) and geographic (urban rural) information linked to mortality data (including cause of death).

**Exploratory data analysis:**

Here is the tabulation of some important variables in the dataset:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Age group** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **19-24** | 370193 | 8.52 | 370193 | 8.52 |
| **25-29** | 357121 | 8.22 | 727314 | 16.73 |
| **30-34** | 370583 | 8.53 | 1097897 | 25.26 |
| **35-39** | 406491 | 9.35 | 1504388 | 34.61 |
| **40-44** | 479630 | 11.03 | 1984018 | 45.64 |
| **45-49** | 485782 | 11.18 | 2469800 | 56.82 |
| **50-54** | 437395 | 10.06 | 2907195 | 66.88 |
| **55-59** | 388211 | 8.93 | 3295406 | 75.81 |
| **60-64** | 295752 | 6.80 | 3591158 | 82.62 |
| **65-69** | 227771 | 5.24 | 3818929 | 87.86 |
| **70-74** | 192595 | 4.43 | 4011524 | 92.29 |
| **75-79** | 156532 | 3.60 | 4168056 | 95.89 |
| **80-84** | 106865 | 2.46 | 4274921 | 98.35 |
| **85-89** | 51146 | 1.18 | 4326067 | 99.53 |
| **90 plus** | 20582 | 0.47 | 4346649 | 100.00 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **sex** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **Female** | 2242681 | 51.60 | 2242681 | 51.60 |
| **Male** | 2103968 | 48.40 | 4346649 | 100.00 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Rural urban** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **Rural** | 984994 | 22.66 | 984994 | 22.66 |
| **Urban** | 3361655 | 77.34 | 4346649 | 100.00 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **province** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **Alberta** | 444249 | 10.22 | 444249 | 10.22 |
| **British Columbia** | 557466 | 12.83 | 1001715 | 23.05 |
| **Manitoba** | 176994 | 4.07 | 1178709 | 27.12 |
| **New Brunswick** | 105097 | 2.42 | 1283806 | 29.54 |
| **Newfoundland and Labrador** | 74281 | 1.71 | 1358087 | 31.24 |
| **Northwest Territories** | 15454 | 0.36 | 1373541 | 31.60 |
| **Nova Scotia** | 128755 | 2.96 | 1502296 | 34.56 |
| **Nunavut** | 14409 | 0.33 | 1516705 | 34.89 |
| **Ontario** | 1604077 | 36.90 | 3120782 | 71.80 |
| **Prince Edward Island** | 18372 | 0.42 | 3139154 | 72.22 |
| **Quebec** | 1047824 | 24.11 | 4186978 | 96.33 |
| **Saskatchewan** | 150432 | 3.46 | 4337410 | 99.79 |
| **Yukon** | 9239 | 0.21 | 4346649 | 100.00 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Income Decile** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **1 - lowest** | 181990 | 4.19 | 181990 | 4.19 |
| **2** | 396994 | 9.13 | 1106116 | 25.45 |
| **3** | 428154 | 9.85 | 1534270 | 35.30 |
| **4** | 439574 | 10.11 | 1973844 | 45.41 |
| **5** | 449077 | 10.33 | 2422921 | 55.74 |
| **6** | 457898 | 10.53 | 2880819 | 66.28 |
| **7** | 470822 | 10.83 | 3351641 | 77.11 |
| **8** | 486826 | 11.20 | 3838467 | 88.31 |
| **9** | 508182 | 11.69 | 4346649 | 100.00 |
| **10 - highest** | 527132 | 12.13 | 709122 | 16.31 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Aboriginal status** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **Aboriginal ID** | 245694 | 5.65 | 245694 | 5.65 |
| **Non-Aboriginal ID** | 4100955 | 94.35 | 4346649 | 100.00 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Citizen status** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **Canadian** | 4119958 | 94.78 | 4119958 | 94.78 |
| **non-Canadian** | 226691 | 5.22 | 4346649 | 100.00 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Mortality status** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **Dead** | 339237 | 7.80 | 339237 | 7.80 |
| **Not Dead** | 4007412 | 92.20 | 4346649 | 100.00 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ADL disability difficulty** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **No** | 3359763 | 77.30 | 3359763 | 77.30 |
| **Not stated** | 45969 | 1.06 | 3405732 | 78.35 |
| **Often** | 423068 | 9.73 | 3828800 | 88.09 |
| **Sometimes** | 517849 | 11.91 | 4346649 | 100.00 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Visible minority** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **Aboriginal self-reporting** | 241183 | 5.55 | 241183 | 5.55 |
| **Arab** | 27776 | 0.64 | 268959 | 6.19 |
| **Black** | 79628 | 1.83 | 348587 | 8.02 |
| **Chinese** | 141501 | 3.26 | 490088 | 11.28 |
| **Filipino** | 46079 | 1.06 | 536167 | 12.34 |
| **Japanese** | 10427 | 0.24 | 546594 | 12.58 |
| **Korean** | 13813 | 0.32 | 560407 | 12.89 |
| **Latin American** | 36556 | 0.84 | 596963 | 13.73 |
| **Multiple visible minority** | 12218 | 0.28 | 609181 | 14.01 |
| **Not a visible minority** | 3582222 | 82.41 | 4191403 | 96.43 |
| **Other visible minority** | 8341 | 0.19 | 4199744 | 96.62 |
| **South Asian** | 106406 | 2.45 | 4306150 | 99.07 |
| **Southeast Asian** | 24144 | 0.56 | 4330294 | 99.62 |
| **West Asian** | 16355 | 0.38 | 4346649 | 100.00 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **First language** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **English** | 3209323 | 73.83 | 3209323 | 73.83 |
| **English & French** | 40041 | 0.92 | 3249364 | 74.76 |
| **French** | 1032652 | 23.76 | 4282016 | 98.51 |
| **Other** | 64633 | 1.49 | 4346649 | 100.00 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Employment status** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **Didn't work** | 1271988 | 29.26 | 1271988 | 29.26 |
| **Full-time** | 2512562 | 57.80 | 3784550 | 87.07 |
| **Part-time** | 562099 | 12.93 | 4346649 | 100.00 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **generation** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **1st Generation** | 967248 | 22.25 | 967248 | 22.25 |
| **2nd Generation** | 667811 | 15.36 | 1635059 | 37.62 |
| **3rd Generation** | 2711590 | 62.38 | 4346649 | 100.00 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Immigration status** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **Immigrant** | 926718 | 21.32 | 926718 | 21.32 |
| **Non-immigrant** | 3389112 | 77.97 | 4315830 | 99.29 |
| **Non-permanent resident** | 30819 | 0.71 | 4346649 | 100.00 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Income after tax** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **Concept not applicable** | 257094 | 5.91 | 257094 | 5.91 |
| **low income** | 401292 | 9.23 | 658386 | 15.15 |
| **non-low income** | 3688263 | 84.85 | 4346649 | 100.00 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Marriage status** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **Divorced** | 376637 | 8.66 | 376637 | 8.66 |
| **Married** | 2304941 | 53.03 | 2681578 | 61.69 |
| **Never married** | 1264326 | 29.09 | 3945904 | 90.78 |
| **Separated** | 140009 | 3.22 | 4085913 | 94.00 |
| **Widowed** | 260736 | 6.00 | 4346649 | 100.00 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Official language** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **English** | 2941427 | 67.67 | 2941427 | 67.67 |
| **English & French** | 800365 | 18.41 | 3741792 | 86.08 |
| **French** | 538988 | 12.40 | 4280780 | 98.48 |
| **Other** | 65869 | 1.52 | 4346649 | 100.00 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Place of birth** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **Home province** | 2751371 | 63.30 | 2751371 | 63.30 |
| **Other province** | 628030 | 14.45 | 3379401 | 77.75 |
| **Outside Canada** | 967248 | 22.25 | 4346649 | 100.00 |

Age distribution shows that nearly 30% of the individuals are within 40 to 59 years of age. 77% of the population lives in urban region while 22% of the population lives in rural region. Only 5.65% of the population has the aboriginal status. 95% of the population is Canadian citizen. 22% of the population has some kind of ADL disability. 18% of the population is visible minority. English is the first language for 73% of the population while French is the first language for 24% of the population. Nearly 62% of the population is the third generation Canadian. 21% of the population has immigrant status. 63% of the individuals were born in the home province, while 22% were born outside Canada.

**Question 1: What socioeconomic factors contribute to the greatest inequalities in mortality?**

We can notice from the tables above that 7.8% of the individuals are classified as dead. First, we assess the greatest inequalities in mortality in relation to the following individual and socioeconomic characteristics: (1) age group, (2) sex, (3) income decile, (4) residing in urban or rural region, (5) occupation, (6) province, (7) income before tax, (8) income after tax, (9) marriage status. We examine the univariate associations between these socio-economic predictors and the mortality outcome:

|  |
| --- |
| Table 1: Mortality with respect to individual and geographical characteristics |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Mortality status** | | | | **Total** |
| **Dead** | | **Not Dead** | |
| **N** | **Percent (%)** | **N** | **Percent (%)** | **N** |
| **Age group** | 2719 | 0.7% | 367474 | 99.3% | **370193** |
| 19-24 |
| 25-29 | 2689 | 0.8% | 354432 | 99.2% | **357121** |
| 30-34 | 3022 | 0.8% | 367561 | 99.2% | **370583** |
| 35-39 | 4753 | 1.2% | 401738 | 98.8% | **406491** |
| 40-44 | 8542 | 1.8% | 471088 | 98.2% | **479630** |
| 45-49 | 15233 | 3.1% | 470549 | 96.9% | **485782** |
| 50-54 | 15620 | 3.6% | 421775 | 96.4% | **437395** |
| 55-59 | 33817 | 8.7% | 354394 | 91.3% | **388211** |
| 60-64 | 52503 | 17.8% | 243249 | 82.2% | **295752** |
| 65-69 | 45316 | 19.9% | 182455 | 80.1% | **227771** |
| 70-74 | 44310 | 23.0% | 148285 | 77.0% | **192595** |
| 75-79 | 42996 | 27.5% | 113536 | 72.5% | **156532** |
| 80-84 | 35841 | 33.5% | 71024 | 66.5% | **106865** |
| 85-89 | 21700 | 42.4% | 29446 | 57.6% | **51146** |
| 90 plus | 10176 | 49.4% | 10406 | 50.6% | **20582** |
| **Sex** | 160849 | 7.2% | 2081832 | 92.8% | **2242681** |
| Female |
| Male | 178388 | 8.5% | 1925580 | 91.5% | **2103968** |
| **Income decile** | 14072 | 7.7% | 167918 | 92.3% | **181990** |
| 1 - lowest |
| 2 | 46193 | 11.6% | 350801 | 88.4% | **396994** |
| 3 | 58183 | 13.6% | 369971 | 86.4% | **428154** |
| 4 | 47282 | 10.8% | 392292 | 89.2% | **439574** |
| 5 | 39374 | 8.8% | 409703 | 91.2% | **449077** |
| 6 | 33399 | 7.3% | 424499 | 92.7% | **457898** |
| 7 | 29318 | 6.2% | 441504 | 93.8% | **470822** |
| 8 | 26298 | 5.4% | 460528 | 94.6% | **486826** |
| 9 | 23554 | 4.6% | 484628 | 95.4% | **508182** |
| 10 - highest | 21564 | 4.1% | 505568 | 95.9% | **527132** |
| **Residing region** | 78658 | 8.0% | 906336 | 92.0% | **984994** |
| Rural |
| Urban | 260579 | 7.8% | 3101076 | 92.2% | **3361655** |
| **Occupation** | 2044 | 2.2% | 91821 | 97.8% | **93865** |
| Art & Culture |
| Business | 12757 | 2.2% | 570348 | 97.8% | **583105** |
| Govt & Religion | 5688 | 1.9% | 286603 | 98.1% | **292291** |
| Health | 3322 | 1.8% | 180147 | 98.2% | **183469** |
| Management | 8390 | 2.6% | 314865 | 97.4% | **323255** |
| Manufacturing | 4524 | 2.6% | 172732 | 97.4% | **177256** |
| Not Applicable | 257393 | 21.5% | 938604 | 78.5% | **1195997** |
| Primary industry | 6764 | 5.2% | 122706 | 94.8% | **129470** |
| Sales & Service | 17816 | 2.6% | 657107 | 97.4% | **674923** |
| Science | 3646 | 1.8% | 204677 | 98.2% | **208323** |
| Trades & Transport | 16893 | 3.5% | 467802 | 96.5% | **484695** |
| **Province** | 29046 | 6.5% | 415203 | 93.5% | **444249** |
| Alberta |
| British Columbia | 45702 | 8.2% | 511764 | 91.8% | **557466** |
| Manitoba | 16290 | 9.2% | 160704 | 90.8% | **176994** |
| New Brunswick | 9083 | 8.6% | 96014 | 91.4% | **105097** |
| Newfoundland and Labrador | 6258 | 8.4% | 68023 | 91.6% | **74281** |
| Northwest Territories | 1017 | 6.6% | 14437 | 93.4% | **15454** |
| Nova Scotia | 12099 | 9.4% | 116656 | 90.6% | **128755** |
| Nunavut | 727 | 5.0% | 13682 | 95.0% | **14409** |
| Ontario | 122618 | 7.6% | 1481459 | 92.4% | **1604077** |
| Prince Edward Island | 1776 | 9.7% | 16596 | 90.3% | **18372** |
| Quebec | 79657 | 7.6% | 968167 | 92.4% | **1047824** |
| Saskatchewan | 14308 | 9.5% | 136124 | 90.5% | **150432** |
| Yukon | 656 | 7.1% | 8583 | 92.9% | **9239** |
| **Income before tax** | 30937 | 12.0% | 226157 | 88.0% | **257094** |
| Concept not applicable |
| low income | 55159 | 10.0% | 495166 | 90.0% | **550325** |
| non-low income | 253141 | 7.2% | 3286089 | 92.8% | **3539230** |
| **Income after tax** | 30937 | 12.0% | 226157 | 88.0% | **257094** |
| Concept not applicable |
| low income | 31496 | 7.8% | 369796 | 92.2% | **401292** |
| non-low income | 276804 | 7.5% | 3411459 | 92.5% | **3688263** |
| **Marriage status** | 31628 | 8.4% | 345009 | 91.6% | **376637** |
| Divorced |
| Married | 169294 | 7.3% | 2135647 | 92.7% | **2304941** |
| Never married | 41947 | 3.3% | 1222379 | 96.7% | **1264326** |
| Separated | 9873 | 7.1% | 130136 | 92.9% | **140009** |
| Widowed | 86495 | 33.2% | 174241 | 66.8% | **260736** |
| **Total** | **339237** | **7.8%** | **4007412** | **92.2%** | **4346649** |

The univariate analysis does not control for other explanatory variables. However, we can still notice some interesting relationships. For example, mortality increases with respect to age group (e.g. 0.7% mortality in age group 19-24 and 49.4% in age group 90+). Mortality is higher among males than females (8.5% versus 7.2%). Mortality tends to decrease with respect to increase in income decile. Mortality is higher in rural regions than urban regions (8% versus 7.8%). Occupation involving trades and transport (3.5%), management (2.6%), manufacturing (2.6%) tend to have higher mortality. Manitoba, Nova Scotia, PEI have higher mortality than Alberta, NWT, Ontario. Individuals with low income (before or after tax) tend to have higher mortality than individuals with non-low income. Divorced individuals have higher mortality than married individuals. 33% mortality among widowed individuals is likely to be confounded with respect to age (e.g. spouse is more likely to die if they are elderly).

Next, we fit logistic regression for mortality with the respect to the SES variables listed in Table 1. Since income before tax and income after tax tend to be highly correlated with each other, we only adjust for income after tax in the logistic regression model. This will safeguard the final model against redundant information which can lead to inflation of standard errors of the odds ratios (i.e. multicollinearity). We do not adjust for occupation because it is not recorded for 1.19 million individuals (27%). Furthermore, there is evidence that occupation information is not missing at random. This is because individuals who are dead are more likely to have occupation information missing. We report the adjusted odds ratios in Table 2 and expected probabilities in Table 3. The expected probabilities are estimated using LSMEANS statement in PROC LOGISTIC. Figure 1 shows the chloropeth map for 13 Canadian provinces by overlaying the adjusted probabilities of mortality. In addition, the adjusted probability shown in Table 3 are plotted in Figure 2.

Table 2: Adjusted odds ratio for mortality with respect to individual characteristics and socioeconomic characteristics

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Effect** | **Index group** | **Reference group** | **Odds ratio** | **95% confidence interval** | | **P-value** |
| Age group (years) | 25-29 years | 19-24 years | 1.064 | 1.008 | 1.122 | 0.0242 |
| 30-34 years | 1.193 | 1.132 | 1.257 | <.0001 |
| 35-39 years | 1.818 | 1.733 | 1.907 | <.0001 |
| 40-44 years | 2.931 | 2.805 | 3.062 | <.0001 |
| 45-49 years | 5.496 | 5.271 | 5.730 | <.0001 |
| 50-54 years | 6.366 | 6.106 | 6.638 | <.0001 |
| 55-59 years | 15.561 | 14.945 | 16.202 | <.0001 |
| 60-64 years | 31.394 | 30.161 | 32.678 | <.0001 |
| 65-69 years | 35.170 | 33.780 | 36.618 | <.0001 |
| 70-74 years | 41.463 | 39.820 | 43.174 | <.0001 |
| 75-79 years | 51.351 | 49.308 | 53.479 | <.0001 |
| 80-84 years | 66.544 | 63.861 | 69.340 | <.0001 |
| 85-89 years | 93.321 | 89.385 | 97.429 | <.0001 |
| 90+ years | 120.437 | 114.753 | 126.402 | <.0001 |
| Sex | Male | Female | 1.767 | 1.753 | 1.782 | <.0001 |
| Income deciles | 1 - lowest | 10 - highest | 2.263 | 2.200 | 2.328 | <.0001 |
| 2 | 2.604 | 2.549 | 2.661 | <.0001 |
| 3 | 2.413 | 2.371 | 2.456 | <.0001 |
| 4 | 2.124 | 2.086 | 2.162 | <.0001 |
| 5 | 1.849 | 1.816 | 1.883 | <.0001 |
| 6 | 1.627 | 1.597 | 1.657 | <.0001 |
| 7 | 1.481 | 1.453 | 1.510 | <.0001 |
| 8 | 1.334 | 1.308 | 1.360 | <.0001 |
| 9 | 1.182 | 1.159 | 1.206 | <.0001 |
| Residing region | Urban | Rural | 1.043 | 1.033 | 1.053 | <.0001 |
| Province | Alberta | Ontario | 0.988 | 0.974 | 1.003 | 0.1071 |
| British Columbia | 0.974 | 0.962 | 0.986 | <.0001 |
| Manitoba | 1.167 | 1.144 | 1.190 | <.0001 |
| New Brunswick | 1.061 | 1.034 | 1.088 | <.0001 |
| Newfoundland and Labrador | 1.098 | 1.066 | 1.131 | <.0001 |
| Northwest Territories | 0.889 | 0.826 | 0.957 | 0.0017 |
| Nova Scotia | 1.153 | 1.128 | 1.179 | <.0001 |
| Nunavut | 1.166 | 1.073 | 1.268 | 0.0003 |
| Prince Edward Island | 1.209 | 1.144 | 1.278 | <.0001 |
| Quebec | 0.943 | 0.933 | 0.953 | <.0001 |
| Saskatchewan | 1.178 | 1.153 | 1.202 | <.0001 |
| Yukon | 0.791 | 0.723 | 0.865 | <.0001 |
| Income after tax | Concept not applicable | non-low income | 1.605 | 1.578 | 1.632 | <.0001 |
| low income | 0.857 | 0.840 | 0.873 | <.0001 |
| Marriage status | Divorced | Married | 1.139 | 1.124 | 1.154 | <.0001 |
| Never married | 1.295 | 1.279 | 1.312 | <.0001 |
| Separated | 1.195 | 1.168 | 1.222 | <.0001 |
| Widowed | 2.433 | 2.406 | 2.460 | <.0001 |

Table 3: Expected probability of mortality with respect to individual and socioeconomic characteristics

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Effect** | **Level** | **Probability** | **95% confidence interval** | |
| Age group | 19-24 | 0.008 | 0.008 | 0.008 |
| 25-29 | 0.009 | 0.008 | 0.009 |
| 30-34 | 0.010 | 0.009 | 0.010 |
| 35-39 | 0.015 | 0.014 | 0.015 |
| 40-44 | 0.024 | 0.023 | 0.024 |
| 45-49 | 0.043 | 0.042 | 0.044 |
| 50-54 | 0.050 | 0.049 | 0.051 |
| 55-59 | 0.113 | 0.112 | 0.115 |
| 60-64 | 0.205 | 0.202 | 0.208 |
| 65-69 | 0.224 | 0.221 | 0.227 |
| 70-74 | 0.254 | 0.251 | 0.258 |
| 75-79 | 0.297 | 0.293 | 0.301 |
| 80-84 | 0.354 | 0.349 | 0.358 |
| 85-89 | 0.434 | 0.429 | 0.440 |
| 90 plus | 0.498 | 0.490 | 0.505 |
| sex | Male | 0.110 | 0.109 | 0.112 |
| Female | 0.065 | 0.065 | 0.066 |
| Income decile | 1 - lowest | 0.110 | 0.107 | 0.112 |
| 2 | 0.124 | 0.122 | 0.126 |
| 3 | 0.116 | 0.114 | 0.118 |
| 4 | 0.104 | 0.102 | 0.105 |
| 5 | 0.092 | 0.090 | 0.093 |
| 6 | 0.081 | 0.080 | 0.083 |
| 7 | 0.075 | 0.073 | 0.076 |
| 8 | 0.068 | 0.066 | 0.069 |
| 9 | 0.061 | 0.059 | 0.062 |
| 10 - highest | 0.052 | 0.051 | 0.053 |
| Residing region | Urban | 0.087 | 0.086 | 0.088 |
| Rural | 0.084 | 0.082 | 0.085 |
| province | Alberta | 0.081 | 0.080 | 0.083 |
| British Columbia | 0.080 | 0.079 | 0.081 |
| Manitoba | 0.095 | 0.093 | 0.096 |
| New Brunswick | 0.087 | 0.085 | 0.089 |
| Newfoundland and Labrador | 0.090 | 0.087 | 0.092 |
| Northwest Territories | 0.074 | 0.069 | 0.079 |
| Nova Scotia | 0.094 | 0.092 | 0.096 |
| Nunavut | 0.095 | 0.088 | 0.102 |
| Prince Edward Island | 0.098 | 0.093 | 0.103 |
| Quebec | 0.078 | 0.077 | 0.079 |
| Saskatchewan | 0.095 | 0.094 | 0.097 |
| Yukon | 0.066 | 0.061 | 0.072 |
| Ontario | 0.082 | 0.081 | 0.083 |
| Income after tax | Concept not applicable | 0.118 | 0.117 | 0.120 |
| low income | 0.067 | 0.066 | 0.068 |
| non-low income | 0.077 | 0.076 | 0.078 |
| Marriage status | Divorced | 0.073 | 0.072 | 0.075 |
| Never married | 0.083 | 0.081 | 0.084 |
| Separated | 0.077 | 0.075 | 0.079 |
| Widowed | 0.145 | 0.143 | 0.147 |
| Married | 0.065 | 0.064 | 0.066 |

Figure 1: Adjusted mortality rate with respect to each province

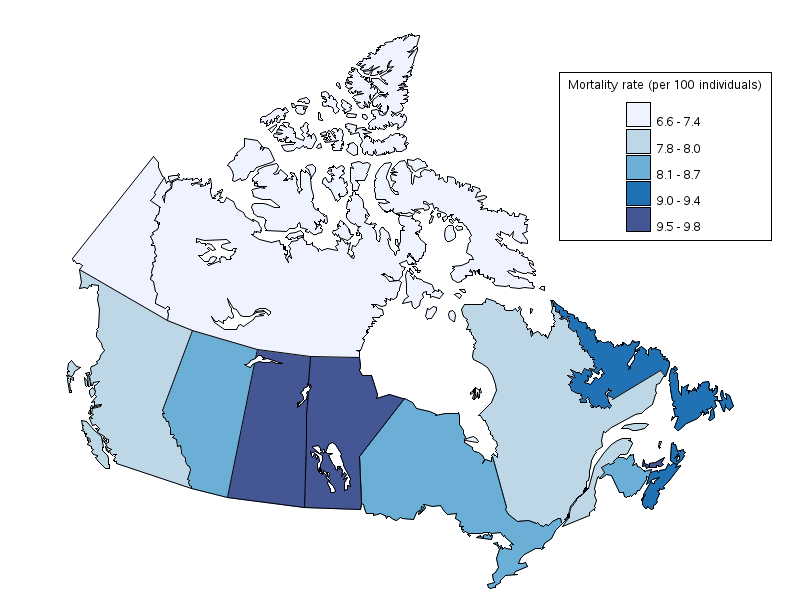


Figure 2: Adjusted probabilities with respect to individual characteristics

|  |  |
| --- | --- |
| The SGPlot Procedure | The SGPlot Procedure |
| The SGPlot Procedure | The SGPlot Procedure |
| The SGPlot Procedure | The SGPlot Procedure |

**Question 2: Does the healthy immigrant advantage extend to the second generation?**

We define healthy immigrant as an individual without any disabilities. Hence, we define a new variable using the information of immigration status and disability status:

|  |
| --- |
| if immigration\_stat='Immigrant' and  adl\_disab\_difficulty='No' then health\_immigration\_status='Healthy immigrant';  if immigration\_stat='Immigrant' and  adl\_disab\_difficulty in('Often','Sometimes') then health\_immigration\_status='Unhealthy immigrant';  if immigration\_stat in('Non-immigrant','Non-permanent resident') and  adl\_disab\_difficulty='No' then health\_immigration\_status='Healthy non-immigrant';  if immigration\_stat in('Non-immigrant','Non-permanent resident') and  adl\_disab\_difficulty in('Often','Sometimes') then health\_immigration\_status='Unhealthy non-immigrant';  if adl\_disab\_difficulty='Not stated' then health\_immigration\_status=''; \* health status unknown; |

Now, we examine the proportion of immigrants and non-immigrants who are healthy with respect to the three generations as:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Health and immigration status** | **Generation** | | | | | | **Total** |
| **1st Generation** | | **2nd Generation** | | **3rd Generation** | |
| **N** | **Percent (%)** | **N** | **Percent (%)** | **N** | **Percent (%)** | **N** |
| **Healthy immigrant** | 692834 | 100.0% | . | . | . | . | 692834 |
| **Healthy non-immigrant** | 34215 | 1.3% | 492118 | 18.5% | 2140596 | 80.3% | 2666929 |
| **Unhealthy immigrant** | 224336 | 100.0% | . | . | . | . | 224336 |
| **Unhealthy non-immigrant** | 5700 | 0.8% | 169300 | 23.6% | 541581 | 75.6% | 716581 |
| **Missing information** | 10163 | 22.1% | 6393 | 13.9% | 29413 | 64.0% | 45969 |
| **Total** | 967248 | 22.3% | 667811 | 15.4% | 2711590 | 62.4% | 4346649 |

Note:

* All healthy/unhealthy immigrants belong to first generation. There are no individuals to compare in the second or third generation.
* However, healthy/unhealthy non-immigrants can belong to first, second and third generation…..

…not sure how to answer this question properly….maybe I need to think about this question from different prospective….

….Let’s say an individual immigrates to Canada….either they are healthy or unhealthy…..now we need to determine if the kids will have the healthy immigrant advantage from their parents….but the problem is that we do not have detailed data on kids to answer this question….

Possible alternative:

One solution is to use the information of 1st , 2nd and 3rd generation as the outcome variable and then compare the proportion of individuals with ADL disability. Hence, we can create the following table:

Table 4: Generation with respect to individual characteristics

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **generation** | | | | | | **Total** |
| **1st Generation** | | **2nd Generation** | | **3rd Generation** | |
| **N** | **Row Percent (%)** | **N** | **Row Percent (%)** | **N** | **Row Percent (%)** | **N** |
| **Age group (years)** | 60574 | 6.3% | 61934 | 9.3% | 247685 | 9.1% | **370193** |
| 19-24 |
| 25-29 | 61029 | 6.3% | 60101 | 9.0% | 235991 | 8.7% | **357121** |
| 30-34 | 68445 | 7.1% | 62918 | 9.4% | 239220 | 8.8% | **370583** |
| 35-39 | 92203 | 9.5% | 60238 | 9.0% | 254050 | 9.4% | **406491** |
| 40-44 | 106480 | 11.0% | 65101 | 9.7% | 308049 | 11.4% | **479630** |
| 45-49 | 104649 | 10.8% | 58380 | 8.7% | 322753 | 11.9% | **485782** |
| 50-54 | 96171 | 9.9% | 52366 | 7.8% | 288858 | 10.7% | **437395** |
| 55-59 | 90855 | 9.4% | 49205 | 7.4% | 248151 | 9.2% | **388211** |
| 60-64 | 79123 | 8.2% | 43009 | 6.4% | 173620 | 6.4% | **295752** |
| 65-69 | 62280 | 6.4% | 38877 | 5.8% | 126614 | 4.7% | **227771** |
| 70-74 | 52826 | 5.5% | 38777 | 5.8% | 100992 | 3.7% | **192595** |
| 75-79 | 43096 | 4.5% | 34644 | 5.2% | 78792 | 2.9% | **156532** |
| 80-84 | 30364 | 3.1% | 24341 | 3.6% | 52160 | 1.9% | **106865** |
| 85-89 | 13673 | 1.4% | 12795 | 1.9% | 24678 | 0.9% | **51146** |
| 90 plus | 5480 | 0.6% | 5125 | 0.8% | 9977 | 0.4% | **20582** |
| **sex** | 504348 | 52.1% | 345363 | 51.7% | 1392970 | 51.4% | **2242681** |
| Female |
| Male | 462900 | 47.9% | 322448 | 48.3% | 1318620 | 48.6% | **2103968** |
| **Income decile** | 57213 | 5.9% | 22000 | 3.3% | 102777 | 3.8% | **181990** |
| 1 - lowest |
| 2 | 107891 | 11.2% | 48153 | 7.2% | 240950 | 8.9% | **396994** |
| 3 | 117703 | 12.2% | 59070 | 8.8% | 251381 | 9.3% | **428154** |
| 4 | 113352 | 11.7% | 63036 | 9.4% | 263186 | 9.7% | **439574** |
| 5 | 106015 | 11.0% | 67969 | 10.2% | 275093 | 10.1% | **449077** |
| 6 | 99705 | 10.3% | 71706 | 10.7% | 286487 | 10.6% | **457898** |
| 7 | 94340 | 9.8% | 76310 | 11.4% | 300172 | 11.1% | **470822** |
| 8 | 90192 | 9.3% | 81077 | 12.1% | 315557 | 11.6% | **486826** |
| 9 | 88859 | 9.2% | 86395 | 12.9% | 332928 | 12.3% | **508182** |
| 10 - highest | 91978 | 9.5% | 92095 | 13.8% | 343059 | 12.7% | **527132** |
| **ADL difficulty** | 796409 | 82.3% | 529243 | 79.3% | 2285051 | 84.3% | **3610703** |
| No |
| Not stated | 12047 | 1.2% | 7723 | 1.2% | 34289 | 1.3% | **54059** |
| Often | 63142 | 6.5% | 53130 | 8.0% | 165730 | 6.1% | **282002** |
| Sometimes | 95650 | 9.9% | 77715 | 11.6% | 226520 | 8.4% | **399885** |
| **ADL disability difficulty** | 727049 | 75.2% | 492118 | 73.7% | 2140596 | 78.9% | **3359763** |
| No |
| Not stated | 10163 | 1.1% | 6393 | 1.0% | 29413 | 1.1% | **45969** |
| Often | 105084 | 10.9% | 73953 | 11.1% | 244031 | 9.0% | **423068** |
| Sometimes | 124952 | 12.9% | 95347 | 14.3% | 297550 | 11.0% | **517849** |
| **Different types of ADL difficulty** | 27796 | 2.9% | 22544 | 3.4% | 70391 | 2.6% | **120731** |
| ADL |
| ADL & Home | 17383 | 1.8% | 14143 | 2.1% | 34996 | 1.3% | **66522** |
| ADL & Home & Work/School | 4142 | 0.4% | 2783 | 0.4% | 10509 | 0.4% | **17434** |
| ADL & Home & Work/School & other | 28011 | 2.9% | 24809 | 3.7% | 95697 | 3.5% | **148517** |
| ADL & Home & other | 72216 | 7.5% | 59516 | 8.9% | 156994 | 5.8% | **288726** |
| ADL & Work & other | 1645 | 0.2% | 1325 | 0.2% | 5235 | 0.2% | **8205** |
| ADL & Work/school | 1896 | 0.2% | 1215 | 0.2% | 5077 | 0.2% | **8188** |
| ADL & other | 5703 | 0.6% | 4510 | 0.7% | 13351 | 0.5% | **23564** |
| Home | 8930 | 0.9% | 6037 | 0.9% | 19725 | 0.7% | **34692** |
| Home & Work/school | 3631 | 0.4% | 2214 | 0.3% | 8767 | 0.3% | **14612** |
| Home & Work/school & other | 9578 | 1.0% | 8628 | 1.3% | 36222 | 1.3% | **54428** |
| Home & other | 10694 | 1.1% | 7958 | 1.2% | 29330 | 1.1% | **47982** |
| None | 727049 | 75.2% | 492118 | 73.7% | 2140596 | 78.9% | **3359763** |
| Not stated | 10163 | 1.1% | 6393 | 1.0% | 29413 | 1.1% | **45969** |
| Other | 10276 | 1.1% | 5260 | 0.8% | 21042 | 0.8% | **36578** |
| Work/school | 14415 | 1.5% | 4939 | 0.7% | 20634 | 0.8% | **39988** |
| Work/school & other | 13720 | 1.4% | 3419 | 0.5% | 13611 | 0.5% | **30750** |
| **Employment status** | 334563 | 34.6% | 205024 | 30.7% | 732401 | 27.0% | **1271988** |
| Didn't work |
| Full-time | 514479 | 53.2% | 368922 | 55.2% | 1629161 | 60.1% | **2512562** |
| Part-time | 118206 | 12.2% | 93865 | 14.1% | 350028 | 12.9% | **562099** |
| **Occupation** | 18246 | 1.9% | 17531 | 2.6% | 58088 | 2.1% | **93865** |
| Art & Culture |
| Business | 114016 | 11.8% | 96859 | 14.5% | 372230 | 13.7% | **583105** |
| Govt & Religion | 52677 | 5.4% | 49687 | 7.4% | 189927 | 7.0% | **292291** |
| Health | 39340 | 4.1% | 26030 | 3.9% | 118099 | 4.4% | **183469** |
| Management | 69399 | 7.2% | 55989 | 8.4% | 197867 | 7.3% | **323255** |
| Manufacturing | 52423 | 5.4% | 18095 | 2.7% | 106738 | 3.9% | **177256** |
| Not Applicable | 312854 | 32.3% | 194905 | 29.2% | 688238 | 25.4% | **1195997** |
| Primary industry | 21955 | 2.3% | 17878 | 2.7% | 89637 | 3.3% | **129470** |
| Sales & Service | 141339 | 14.6% | 95088 | 14.2% | 438496 | 16.2% | **674923** |
| Science | 58654 | 6.1% | 32711 | 4.9% | 116958 | 4.3% | **208323** |
| Trades & Transport | 86345 | 8.9% | 63038 | 9.4% | 335312 | 12.4% | **484695** |
| **Number of kids** | 415228 | 42.9% | 254113 | 38.1% | 1082912 | 39.9% | **1752253** |
| 1 or 2 |
| 3 or more | 102056 | 10.6% | 57036 | 8.5% | 239411 | 8.8% | **398503** |
| None | 449964 | 46.5% | 356662 | 53.4% | 1389267 | 51.2% | **2195893** |
| **Total** | **967248** | **100.0%** | **667811** | **100.0%** | **2711590** | **100.0%** | **4346649** |

We can notice that the first generation individuals tend to be younger than second or third generation individual. More first generation individuals tend to live in low-income areas as compared to second or third generation individuals. A higher proportion of first generation individuals (82.4%) do not have any ADL difficulty than second generation (79.3%). A higher proportion of first generation individuals (34.6%) did not work as compared to second (30.7%) or third (27.0%) generation individuals. A higher proportion of first generation individuals (5.4%) work in manufacturing sector as compared to second (2.7%) or third (3.9%) generation individuals. A lower proportion of first generation (5.4%) individual work in the government and religion as compared to second (7.4%) or third (7.0%) generation individuals.

We can probably fit an ordinal logistic regression model (e.g. proportional odds model) to assess if healthy immigration advantage extends to second generation.

We will discuss this further on Tuesday afternoon.

**Question 3: Does risk of cardiovascular mortality vary among immigrants from different source countries?**

The information on cardiovascular mortality is available using *cause\_dead\_2* variable. We define a new variable named mortality as:

|  |
| --- |
| if cause\_death\_2='CVD' then mortality='Death due to CVD';  if cause\_death\_2='Did not Die ' then mortality='Did not die';  if cause\_death\_2 ne 'CVD' and cause\_death\_2 ne 'Did not Die ' then mortality='Death due to other competing risk'; |

Now we examine the relationship between type of mortality and source country:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **birth\_country** | **mortality** | | | | | | **Total** |
| **Death due to CVD** | | **Death due to other competing risk** | | **Did not die** | |
| **N** | **Percent (%)** | **N** | **Percent (%)** | **N** | **Percent (%)** | **N** |
| **Australia, NZ, Oceania & Greenland** | 628 | 0.2% | 9034 | 3.3% | 262339 | 96.4% | 272001 |
| **E Asia** | 2 | 0.1% | 80 | 2.1% | 3810 | 97.9% | 3892 |
| **Eastern Europe** | 1144 | 0.8% | 17779 | 12.6% | 121633 | 86.5% | 140556 |
| **Latin America & Caribbean** | 541 | 0.3% | 8903 | 5.4% | 156326 | 94.3% | 165770 |
| **N Africa, SW Asia & Middle East** | 844 | 0.7% | 12807 | 10.8% | 104554 | 88.5% | 118205 |
| **S Asia** | 20 | 0.2% | 279 | 3.2% | 8353 | 96.5% | 8652 |
| **SE Asia** | 1 | 0.0% | 64 | 1.9% | 3255 | 98.0% | 3320 |
| **Sub-Saharan Africa** | 161 | 0.2% | 2804 | 3.5% | 77212 | 96.3% | 80177 |
| **Western Europe** | 1129 | 0.7% | 17522 | 10.6% | 146313 | 88.7% | 164964 |
| **non-Immigrant** | 13637 | 0.4% | 251858 | 7.4% | 3123617 | 92.2% | 3389112 |
| **Total** | 18107 | 0.4% | 321130 | 7.4% | 4007412 | 92.2% | 4346649 |

We can notice that a higher proportion of individuals from N Africa, SW Asia & Middle East (0.7%) and Western Europe (0.7%) die due to CVD related problems.

We may need to collapse categories for birth country. For example, there is only 1 individual from southeast Asia who died from CVD, 2 individuals from east Asia who died from CVD. Maybe it is best to collapse southeast Asia, east Asia, south Asia into Asia.

We will discuss this further on Tuesday afternoon.