# Urban Physical Environments and Health Inequalities

## Data and Analysis Methodology

Factors Influencing Health

Canadian Institute for Health Information
Institut canadien d'information sur la santé

## Who We Are

Established in 1994, CIHI is an independent, not-for-profit corporation that provides essential information on Canada's health system and the health of Canadians. Funded by federal, provincial and territorial governments, we are guided by a Board of Directors made up of health leaders across the country.

## Our Vision

To help improve Canada's health system and the well-being of Canadians by being a leading source of unbiased, credible and comparable information that will enable health leaders to make better-informed decisions.

# Table of Contents

# 1.0 Synopsis

The analyses of the Canadian Population Health Initiative (CPHI) explore patterns of health within and between population groups to foster a better understanding of factors that affect the health of individuals and communities. We also seek out and summarize evidence about what works at a policy and program level to contribute to the development of policies that reduce inequities and improve the health and well-being of Canadians.

We know from previous research that inequalities in health exist between and within Canada's cities. Many of these inequalities have been associated with differences in socio-economic status, the availability and quality of social networks, as well as how the built environment is designed. While less explored in a population health setting, the physical environment also plays an important role in the overall health and well-being of urban Canadians.

The report *Urban Physical Environments and Health Inequalities* builds on previous research to explore two aspects of the urban physical environment that are changing as a result of increased urbanization: air pollution and urban heat islands. These topics were chosen based on their documented impact on respiratory and cardiovascular health. They were also chosen because major sources of air pollution (such as industrial pollutant emitters and high-traffic roadways) and intensified hot weather events due to the effects of the urban heat island can be influenced through policies, programs and urban design. The report presents new analyses based on a number of sources, including data from CIHI's hospitalization records (Discharge Abstract Database), the Canadian census, Environment Canada's National Pollutant Release Inventory and daily climate data, and Natural Resources Canada's thermal satellite imagery.

This data and analysis methodology paper is organized in three major sections. The first section presents a summary of the data sources. The second section provides a detailed description of the geographic information system methodology and data preparation. The final section summarizes the statistical analyses and summary statistics calculated for the report. Limitations are also highlighted.

The overall intention of this methods paper is to give an overview of how the analysis was completed. Attempts were made to include enough detail so comparable results could be generated. However, if you require more information or would like to discuss the approaches taken for these analyses with one of our team members, please contact us at cphi@cihi.ca.

# 2.0 Data and Information Sources

The analyses presented in this report are based on a number of sources, such as health databases from within CIHI and information maintained by other organizations. These include the following:

- Canadian census
- Deprivation Index (derived from the Canadian census)
- Postal Code Conversion File
- Discharge Abstract Database
- MED-ÉCHO
- National Population Health Survey
- National Pollutant Release Inventory
- Toxic equivalency potentials
- GeoBase
- Satellite imagery
- Daily climate data

## 2.1 Census of Canada

Statistics Canada conducts a census once every five years. The census of Canada is used to provide reliable estimates of the Canadian population based on the demographic, social and economic characteristics of the population at a specific point in time. The census captures data on population and dwelling counts, income, age, sex, marital status, employment and education. This information provided a basis for developing some of the analytical tools used in this report, such as the Deprivation Index, and for the results presented.

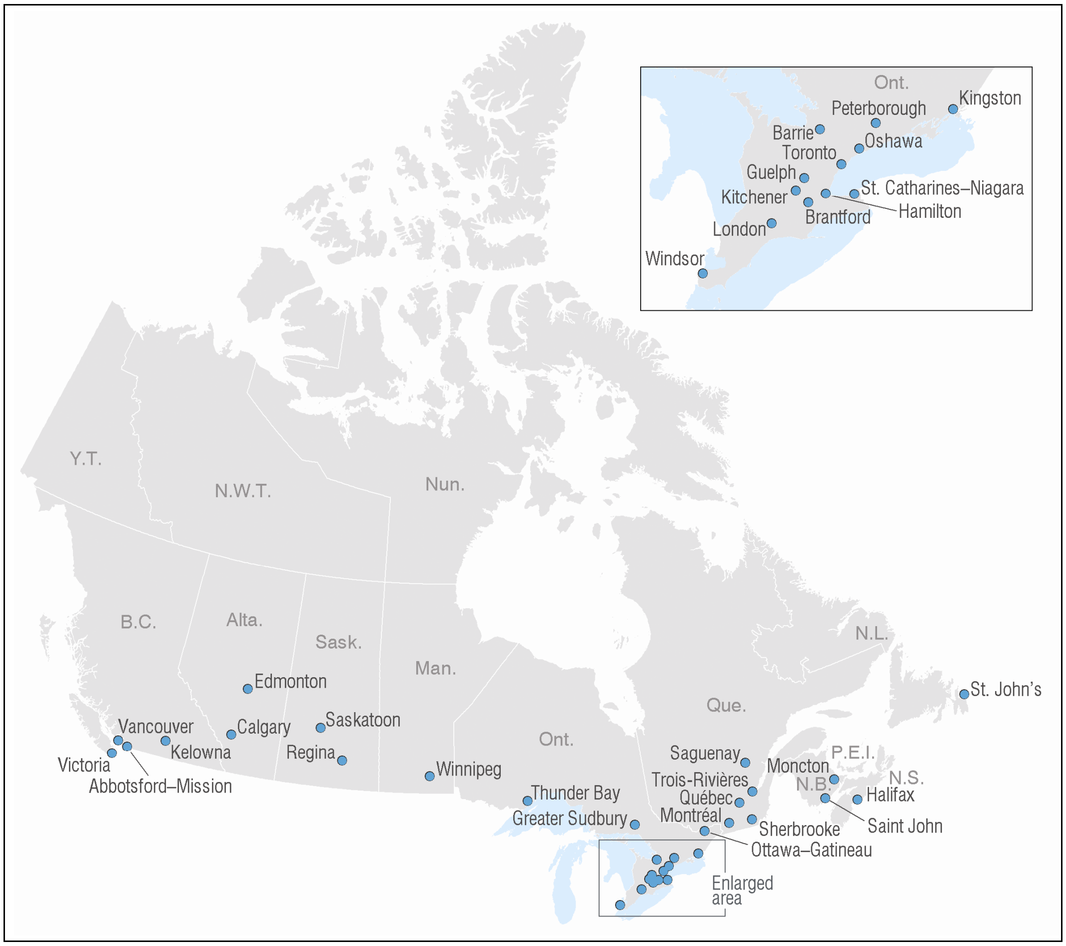
The smallest geo-statistical units of the census used in this report are dissemination areas (DAs). These were linked to hospitalization data and aggregated to higher units of geography, particularly census metropolitan areas (CMAs). According to Statistics Canada, a **dissemination area** is a small geographical area, typically with a population of 400 to 700 people, consisting of one or more blocks of houses in close proximity. A **census metropolitan area** is defined as a geographical area that contains one or more neighbouring municipalities positioned around a major urban core. More specifically, the population of the urban core must be a minimum of 100,000.[1]

The report specifically looks at socio-economic status and health in Canada's urban environment; therefore, rural locations were not taken into consideration. All of Canada's 33 CMAs (see Figure 1), as designated by Statistics Canada, were chosen for further examination to give a sufficient geographical representation of Canada's urban areas.

The Postal Code Conversion File (PCCF) March 2008 version was used to link six-character postal codes to the 2006 census DAs.

CMA and DA digital boundary files were obtained from Statistics Canada for the 2006 census year and used for geospatial analysis.

Figure 1: Geographical Location of Canada's 33 Census Metropolitan Areas



## 2.2 Institut national de santé publique du Québec Deprivation Index

To operationalize area-level socio-economic status, the Institut national de santé publique du Québec (INSPQ) Deprivation Index was used to assign geographical areas in each of Canada's 33 CMAs into one of five groups. Each group represents approximately 20% of the population, ranked into quintiles by the socio-economic status of the area. The 2006 INSPQ Deprivation Index includes both material and social components shown to be related to health and allows data to be presented at the Statistics Canada DA level. For further information on how these components were identified and calculated, please see Pampalon et al., 2009.[2]

Components of the INSPQ Deprivation Index include the following:

**Material component:**

- Percentage of the population without high school graduation
- Employment ratio
- Average income

**Social component:**

- Percentage of families headed by a single parent
- Percentage of the population that lives alone
- Percentage of the population that is separated, divorced or widowed

Each DA in the 33 CMAs was assigned an individual score on both the material and social components of the INSPQ Deprivation Index. In total, 32,765 DAs were assigned scores, ranging from 1 (the highest socio-economic status group) to 5 (the lowest group). The process for combining the material and social components of the INSPQ Deprivation Index to calculate an overall or combined score is depicted in Table 1.

Table 1: Combination of Material and Social Components of the INSPQ Deprivation Index

| | | Social Components | | | | |
|---|---|---|---|---|---|---|
| | | Quintile 1 | Quintile 2 | Quintile 3 | Quintile 4 | Quintile 5 |
| Material Components | Quintile 1 | Highest | Highest | Highest | Upper-Middle | Middle |
| | Quintile 2 | Highest | Upper-Middle | Upper-Middle | Middle | Lower-Middle |
| | Quintile 3 | Highest | Upper-Middle | Middle | Lower-Middle | Lowest |
| | Quintile 4 | Upper-Middle | Middle | Lower-Middle | Lower-Middle | Lowest |
| | Quintile 5 | Middle | Lower-Middle | Lowest | Lowest | Lowest |

Table 2 presents the distribution of DAs by socio-economic status (SES) in each of Canada's 33 CMAs. Note that even though the number of DAs is not the same in each SES group, the assignment of quintiles was weighted by population such that each CMA is set up to have approximately 20% of its population in each group.

Table 2: Population Distribution by Socio-Economic Status in Canada's 33 Census Metropolitan Areas

| Census Metropolitan Area | Socio-Economic Status Group | | | | | | Total Population |
|---|---|---|---|---|---|---|---|
| | Lowest | Lower-Middle | Middle | Upper-Middle | Highest | Missing | |
| **Abbotsford–Mission** | 20% | 19% | 17% | 21% | 19% | 4% | **159,020** |
| **Barrie** | 19% | 21% | 21% | 19% | 18% | 2% | **177,061** |
| **Brantford** | 26% | 13% | 12% | 21% | 25% | 2% | **124,607** |
| **Calgary** | 19% | 20% | 22% | 20% | 17% | 2% | **1,079,310** |
| **Edmonton** | 21% | 18% | 19% | 19% | 21% | 3% | **1,034,945** |
| **Greater Sudbury** | 20% | 20% | 14% | 24% | 18% | 5% | **158,258** |
| **Guelph** | 24% | 17% | 15% | 23% | 20% | 2% | **127,009** |
| **Halifax** | 18% | 20% | 19% | 20% | 20% | 3% | **372,858** |
| **Hamilton** | 24% | 14% | 16% | 21% | 22% | 3% | **692,911** |
| **Kelowna** | 23% | 16% | 19% | 17% | 21% | 4% | **162,276** |
| **Kingston** | 22% | 15% | 14% | 22% | 21% | 6% | **152,358** |
| **Kitchener** | 23% | 15% | 18% | 20% | 21% | 3% | **451,235** |

Table 2: Population Distribution by Socio-Economic Status in Canada's 33 Census Metropolitan Areas (cont'd)

| Census Metropolitan Area | Socio-Economic Status Group | | | | | | Total Population |
|---|---|---|---|---|---|---|---|
| | Lowest | Lower-Middle | Middle | Upper-Middle | Highest | Missing | |
| **London** | 23% | 17% | 15% | 18% | 24% | 3% | **457,720** |
| **Moncton** | 24% | 17% | 15% | 20% | 22% | 1% | **126,424** |
| **Montréal** | 21% | 19% | 17% | 19% | 21% | 3% | **3,635,571** |
| **Oshawa** | 22% | 15% | 18% | 22% | 21% | 2% | **330,594** |
| **Ottawa–Gatineau** | 22% | 16% | 16% | 22% | 21% | 3% | **1,130,761** |
| **Peterborough** | 24% | 14% | 14% | 19% | 24% | 4% | **116,570** |
| **Québec** | 19% | 17% | 21% | 21% | 19% | 3% | **715,515** |
| **Regina** | 22% | 18% | 15% | 21% | 22% | 2% | **194,971** |
| **Saguenay** | 22% | 16% | 16% | 24% | 20% | 2% | **151,643** |
| **Saint John** | 22% | 21% | 17% | 11% | 26% | 3% | **122,389** |
| **Saskatoon** | 20% | 17% | 19% | 22% | 19% | 2% | **233,923** |
| **Sherbrooke** | 22% | 16% | 14% | 21% | 21% | 5% | **186,952** |
| **St. Catharines–Niagara** | 24% | 16% | 12% | 22% | 22% | 4% | **390,317** |
| **St. John's** | 19% | 19% | 18% | 23% | 19% | 2% | **181,113** |
| **Thunder Bay** | 22% | 19% | 12% | 19% | 24% | 4% | **122,907** |
| **Toronto** | 18% | 18% | 23% | 21% | 17% | 2% | **5,113,149** |
| **Trois-Rivières** | 19% | 18% | 20% | 12% | 25% | 4% | **141,529** |
| **Vancouver** | 14% | 22% | 25% | 19% | 15% | 3% | **2,116,581** |
| **Victoria** | 20% | 17% | 18% | 21% | 19% | 5% | **330,088** |
| **Windsor** | 25% | 14% | 13% | 21% | 25% | 2% | **323,342** |
| **Winnipeg** | 24% | 16% | 15% | 21% | 22% | 3% | **694,668** |
| **Total CMA Population** | **20%** | **18%** | **19%** | **20%** | **19%** | **3%** | **21,508,575** |

## 2.3 Postal Code Conversion File

The PCCF provides a linkage between Statistics Canada's standard geographic areas and Canada Post's six-character postal code. The PCCF uses the single-linkage indicator methodology, where a single DA is selected for a 1:1 link with a postal code.[3] A region assignment macro, developed by CIHI, was used to assign geographic information to patient records using postal codes.

## 2.4 Discharge Abstract Database

The Discharge Abstract Database (DAD) is one of CIHI's data holdings. It contains demographic, administrative and clinical data related to hospital discharges for various acute and chronic conditions.[4] The data is provided to CIHI directly from all participating hospitals in every province and territory, excluding Quebec.[5] Hospitalization data presented in this report is acute care cases extracted from the DAD based on the most responsible diagnosis or the diagnosis responsible for the greatest portion of the patient's stay in hospital. Data was extracted for the years 2005 to 2008 for diseases of the circulatory and respiratory systems.

The following ICD codes for diseases of respiratory and circulatory systems were used to select the cases and calculate the hospitalization rates.

- Diseases of the respiratory system
  Any of the following diagnosis codes with a diagnosis type M (the most responsible diagnosis):
  - ICD-9
    460 to 466, 470 to 478, 480 to 487, 490 to 496, 500 to 508, 510 to 519
  - ICD-10-CA
    J01 to J06, J09 to J18, J20 to J22, J30 to J47, J60 to J70, J80 to J86, J90 to J99
- Diseases of the circulatory system
  Any of the following diagnosis codes with a diagnosis type M and diagnosis coding class not equal to 0:
  - ICD-9
    393 to 398, 401 to 405, 410 to 417, 420 to 438, 440 to 448, 451 to 459
  - ICD-10-CA
    I00 to I02, I05 to I13, I15, I20 to I28, I30 to I52, I60 to I74, I77 to I89, I95, I97 to I99

## 2.5 MED-ÉCHO

Data for hospitalizations in the province of Québec was provided by the ministère de la Santé et des Services sociaux. Hospitalization counts for respiratory and circulatory conditions as specified above were summarized and provided for this analysis by the INSPQ.

## 2.6 National Population Health Survey

The National Population Health Survey began in 1994; it collects information about the health of Canadians every two years. This Statistics Canada survey includes household and institutional residents in all provinces and territories, excluding persons living on Aboriginal reserves, on Canadian Forces bases and in remote areas.[6] The analysis uses cycle 7 (2006–2007) of the longitudinal square file, which contains records for all responding members of the original panel whether or not information about them was obtained in subsequent cycles. The analysis was limited to adults age 18 and older due to the measurement of the variables of interest. The square weight was used in the analysis, which applies to the responding members of the original panel and was post-stratified to the 1994–1995 population estimates based on the 1996 census counts by age group and sex within each province to account for non-response. Tests of significance, coefficients of variation and confidence intervals were performed using the bootstrap technique, which accounts for survey design effects.

## 2.7 National Pollutant Release Inventory

The National Pollutant Release Inventory (NPRI) is Canada's legislated, publicly accessible inventory of all pollutant releases to the air, water and land and acts as a resource for regulating and reducing toxic pollutant emissions across the country.[7] Established by Environment Canada, it contains information reported by facilities and extensive emission reports and trends for specific air pollutants that arise from sources such as motor vehicles, residential heating, forest fires and agriculture.[8] Only NPRI data from 2007 was used in this report.

## 2.8 Toxic Equivalency Potentials

Using a scoring system, toxic equivalency potentials (TEPs) provide a way to assess differences in the relative human health risk associated with the release of various chemicals to the air or water. Scores are provided for more than 350 chemicals and take into account information about each chemical's toxicity and exposure potential.[9] This approach was developed by scientists at the University of California at Berkley as a screening tool for relative risk ranking and was subsequently reviewed by the Science Advisory Board of the U.S. Environmental Protection Agency.[10]

## 2.9 GeoBase

GeoBase provides online access to quality geospatial data for all of Canada at no cost and without restrictions on use. It represents an initiative by federal, provincial and territorial agencies and is a key component of the Canadian Geospatial Data Infrastructure, which is part of the GeoConnections program led by Natural Resources Canada.[11] National Road Network data (edition 2.0)—used for the proximity-to-highways analysis—and land cover data, circa 2000—used in the heat extremes and health section—were obtained from GeoBase. Both data sets were in vector format.

## 2.10 Satellite Imagery

Satellites orbiting Earth allow information to be collected from space. For the heat extremes and health section of the report, Landsat 5 Thematic Mapper thermal satellite imagery (band 6) was used to illustrate and analyze land surface temperature. Imagery for Toronto was obtained from Natural Resources Canada, while the Montréal imagery was obtained from the Université du Québec à Montréal. Both images are from 2008, with the Toronto image captured on September 3 and the Montréal one on July 5. The spatial resolution of the images for Toronto and Montréal are 60 metres and 30 metres, respectively. CPHI used only one thermal image for each CMA that was captured during sunny weather conditions and deemed to be representative of how land surface temperatures vary across a CMA during a typical sunny summer day.[12]

High-resolution multispectral satellite imagery available in Google Earth was also utilized as a means to visually examine aspects of the urban physical environment.

## 2.11 Daily Climate Data

Environment Canada's National Climate Data and Information Archive contains historical climate data from all of Canada's meteorological stations, many of which are located in urban areas. CPHI obtained hourly temperature data from the stations at Toronto's Pearson International Airport and Montréal's Pierre Elliott Trudeau International Airport for the months of May through September for the years 2005 to 2008. This data was used in the heat extremes and health section of the report.

# 3.0 GIS Methodology

The CPHI report *Urban Physical Environments and Health Inequalities* is split into three main sections:

1. Air pollution and health;
2. Heat extremes and health; and
3. Directions for policy and potential for action.

Below are detailed explanations pertaining to the geographic information system (GIS) methods utilized to produce results for sections 1 and 2 of the report. The GIS software used was MapInfo Professional V.10.

## 3.1 Air Pollution

This section explains the steps taken to prepare the data and the methods used to analyze air pollution presented in the report, beginning with proximity to NPRI facilities and followed by proximity to highways.

### 3.1.1 Proximity to Polluting Facilities

#### Data Preparation

Beginning with all facilities that reported to the NPRI in 2007, the first step was to remove those that did not report an air release, since air pollution was the focus. Next, facilities that released substances known to negatively affect respiratory or circulatory health were isolated. Table 3 lists the six substances selected for analysis based on an extensive literature review summarized in the report.

Six files (one for each substance) were created that included a unique facility identification number, the total amount released to air for each facility (in tonnes) and a TEP rank. The TEP rank was used to assess the relative health risk posed by each substance, since not all substances pose an equal risk.

Table 3 shows the TEP rank for each of the six substances analysed in the report.[9] TEP scores were calculated by multiplying the release amount of each substance by its TEP rank and then cumulating scores for each facility. This yielded a single score for each facility.

Table 3: Substance, Chemical Abstracts Service Code and Toxic Equivalency Potential Rank for Substances Linked to Respiratory and Circulatory Health

| Substance | CAS Code | TEP Rank (Non-Carcinogenic) |
|---|---|---|
| **Nitrogen Dioxide** | 11104-93-1 | 2.2 |
| **Sulphur Dioxide** | 7446-09-5 | 3.1 |
| **PM2.5** | NA–M10 | 17 |
| **VOCs** | NA–M16 | 1[13] |
| **Toluene** | 108-88-3 | 1 |
| **Carbon Monoxide** | 630-08-0 | 0.14 |

**Notes**
CAS: Chemical Abstracts Service.
TEP: toxic equivalency potential.

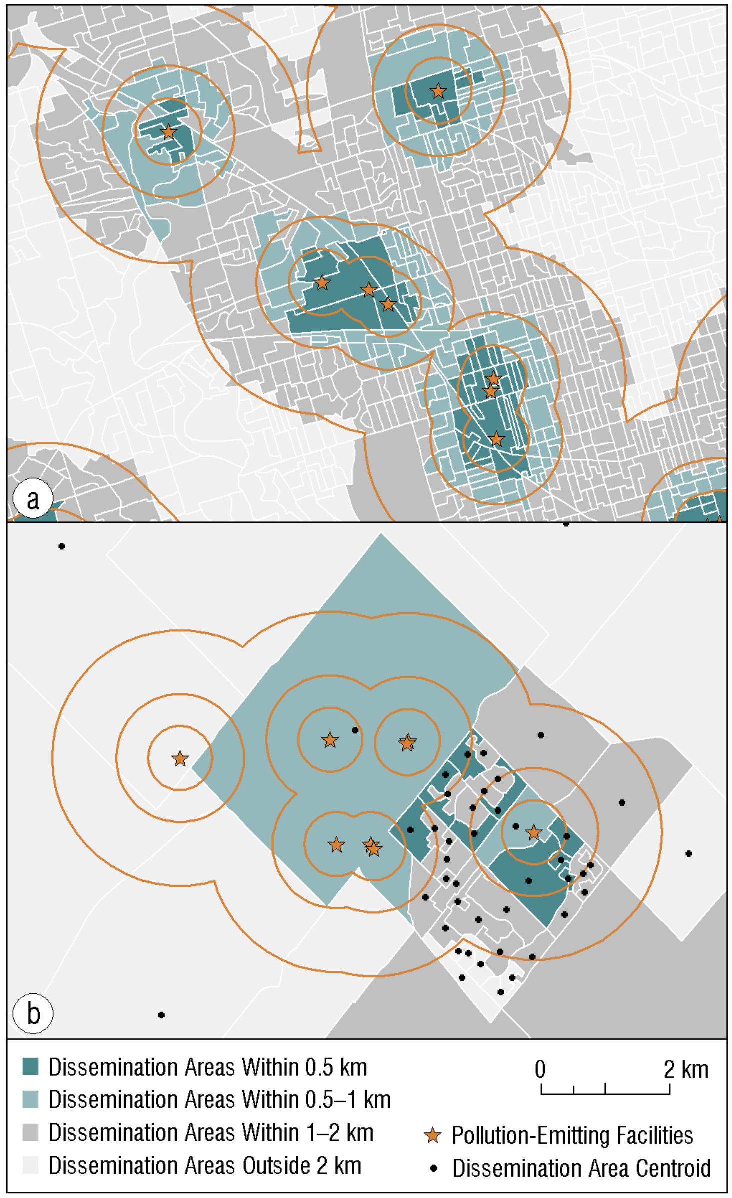
Using the GIS, the latitude and longitude coordinates, obtained from a locations file available on the NPRI website, were mapped. Next, the TEP scores from the six substance files were linked to the mapped facility locations using the identification numbers. A link success rate of 99.9% was achieved. Facilities that did not release any of the six substances previously identified were removed, as were facilities that were not located within the CMA boundaries. The final number of facilities included in the analysis was 1,450.

To assess whether or not the magnitude of a facility's TEP score affected the health or socio-economic status analysis, a test was performed using only the facilities from the highest and lowest quintiles. Specifically, the top 20% and bottom 20% of facilities (in terms of TEP scores) were analyzed separately to examine whether there were significant differences in health outcomes. This analysis revealed no significant differences between the highest and lowest quintile groups for respiratory or circulatory hospitalizations and led to the conclusion that facilities should not be differentiated based on TEP scores. Thus, all facilities were treated equally.

## Creating Buffer Zones

To examine the relationship between proximity to the selected facilities, socio-economic status and respiratory and circulatory health, we derived a distance variable that could be applied at the DA level. This was achieved by creating buffers of a specified distance around each of the 1,450 facilities. Three buffer zones (0.5 kilometres, 0.5 to 1 kilometre and 1 to 2 kilometres) were created to assign DAs to each distance interval using spatial queries. The choice of buffer sizes was guided by similar research conducted by Toronto Public Health.[13] DAs were assigned to the buffer zone in which their centroid, the geometric centre of a polygon, was located (see Figure 2b). All DAs not within one of the three buffer zones comprised the comparison group for analysis purposes (see Section 4.1).

Figure 2: Assigning Dissemination Areas to Buffer Zones




There are limitations to this type of analysis. In some cases, especially in higher population density areas where DAs tend to be smaller (see Figure 2a), the whole DA, or most of it, falls entirely within a single buffer zone. In other cases, however, where population density is lower and DAs are larger, a single DA might cover two or even all three buffer zones (see Figure 2b). Given that each DA can be assigned to only one buffer zone, and that assignment is based on the location of the DA centroid, some spatial misrepresentation occurs. That is, some DAs will be assigned to a particular buffer zone even though a large proportion of the DA may be in another buffer zone. Evidence of this can be seen in Figure 2b, where DA centroids are shown along with the buffer zone to which they were assigned. This problem is especially evident in the urban fringes, as population densities decline and DAs can be very large compared with more central regions of CMAs.

Another limitation or potential source of error is that the accuracy of the locations used (latitude and longitude coordinates) to map the facilities could not be verified. The latitude and longitude values were obtained directly from the NPRI website, and a visual examination of selected facilities overlaid on satellite imagery in Google Earth revealed that some locations were not completely accurate. The extent to which spatial accuracy issues with the facility locations affected the analysis has not been addressed.

Prevailing winds also have an impact on regional and local air pollution levels. However, wind speed and trajectory are complex variables to incorporate at a pan-Canadian scale and therefore were not included in this analysis.

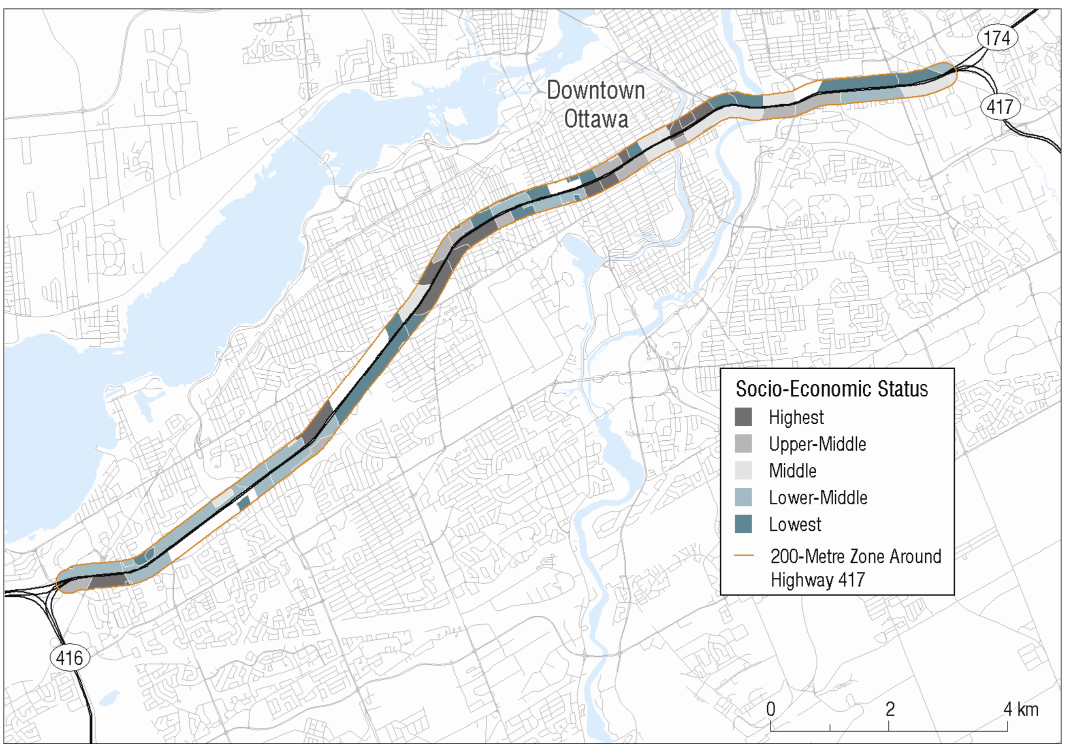
### 3.1.2 Proximity to Highways

CPHI analysis examined the socio-economic status distribution of the DAs that were within a 200-metre highway buffer zone. The buffer size selection of 200-metres was guided by other research in this area.[14–16]

One continuous high-traffic volume section of highway for each CMA was chosen for analysis based on average annual daily traffic (AADT) data obtained from provincial transportation ministries. Since AADT values varied significantly within and between each CMA, no single cut-off, or traffic volume threshold, could be used to select highway sections with comparable AADT values. In addition, AADT data reporting varied between provinces. For instance, data for Ontario was available on a highway section basis (for example, Highway 401 between Dixon Road and Islington Avenue), whereas AADT data for British Columbia was restricted to a limited number of permanent monitoring sites.

As a result of differences in AADT values and data availability, as well as road configuration, the highway selection criteria were CMA-specific. Consequently, the length of highway used for analysis varied between CMAs, with Toronto being the longest and Montréal the shortest, with distances of 36 kilometres and 13 kilometres, respectively. Figure 3 illustrates the section of highway chosen for the Ottawa–Gatineau CMA, as well as the socio-economic status of the DA portions that are within the 200-metre buffer.

Figure 3: Ottawa–Gatineau Proximity to Highways Map



Examination of the distribution of socio-economic status within the 200-metre buffer required the DAs to be split. Using a GIS, the DAs were split along the 200-metre buffer line, and the land area of each portion within the buffer was calculated. Finally, the proportion of land area from each socio-economic status group was calculated and graphed in the report.

Although the proportion of land area by socio-economic status within the 200-metre buffer was calculated, we cannot determine how many people live there. As such, it is impossible to know how the population distribution compares with the land area distribution within each socio-economic status group, and it is also impossible to comment on the relationship between proximity to roadways and circulatory and respiratory hospitalization rates.

## 3.2 Heat

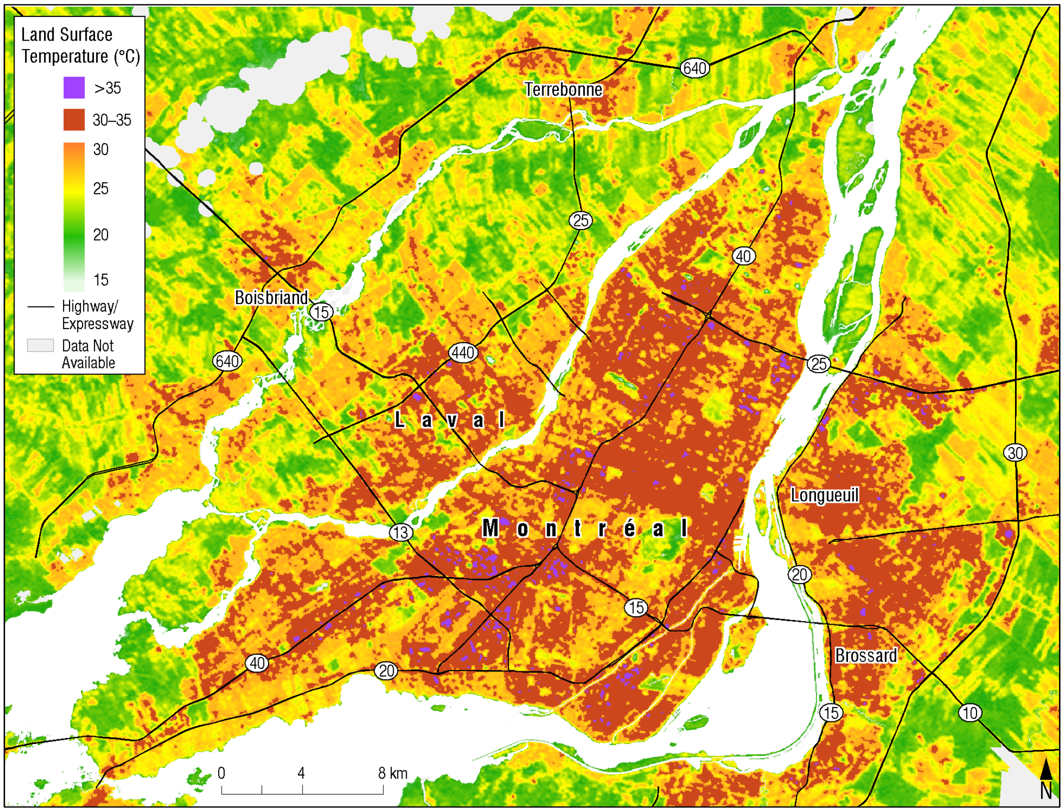
To examine the characteristics of the urban physical environment that contribute to heat islands and assess the extent to which lower socio-economic status DAs are characterized by hotter temperatures, new analyses were conducted using land surface temperature data derived from thermal satellite imagery (see Figure 4). Although land surface temperature is not directly correlated with air temperature, it is an indicator of urban heat island intensity, particularly during nighttime periods.[17] High-resolution satellite imagery from Google Earth was also utilized in these analyses to visually inspect specific areas of the urban physical environment.

The two CMAs of Toronto and Montréal were chosen for analysis on a data availability basis. Using Vertical Mapper, an add-on to MapInfo that supports raster-based data formats, the land surface temperature data was mapped and then analyzed at the DA level.

In both Montréal and Toronto, the land surface temperature data did not cover the entire CMA. When the land surface temperature data did not provide complete DA coverage, the DA was removed from the subsequent analysis. In Montréal, 5,079 out of 6,082 DAs were used for analysis; in Toronto, 6,210 out of 7,012 were used.

Mean, minimum, maximum and range of land surface temperature were calculated for each DA. DAs were separated into two categories—those with mean temperatures below and above 30°C—and each category was then graphed based on socio-economic status. The mean land surface temperature statistic was also utilized in the odds ratio analysis, which is further discussed in the upcoming analysis methodology section.

Figure 4: Land Surface Temperature Map of Montréal




### 3.2.1 Neighbourhood Analysis

In the heat section of the report, neighbourhoods were also used as a unit of analysis. The neighbourhood selection process was based on two information sources. First, the Canadian Mortgage and Housing Corporation (CMHC) uses proximity to the downtown core and type of home to create three distinct zones within urban areas: central area, inner suburbs and outer suburbs.[18] Using GIS, neighbourhoods were chosen from each of these zones, with two selected from the central area due to the range of neighbourhood types in that zone. Second, city staff from both Toronto and Montréal were consulted regarding neighbourhood selection; in both cases, selections were altered based on the feedback received.

Neighbourhood boundaries were obtained from both the City of Toronto and the City of Montréal in a spatial data format. Census data at the DA level was aggregated to the neighbourhood level to derive additional contextual information to compare neighbourhoods. This process was aided by the fact that neighbourhood boundaries align perfectly with DA boundaries.

### 3.2.2 Green Space Analysis

Land cover data, circa 2000, was used to create a green space representation for GIS analysis. It was chosen because it provided consistent and comparable data across Canada's CMAs. The land cover data consisted of several land use classes, the following of which were considered green space:

- Shrubland
- Wetland
- Grassland
- Annual crops
- Perennial crops and pasture
- Coniferous
- Broadleaf
- Mixed wood

Although there are various definitions of green space, the land use classes chosen to represent green space in this report are consistent with other literature.[19]

To create the green space graph seen in the report, a GIS was used to calculate the percentage of green space in each DA; the percentages were then averaged based on socio-economic status groups.

# 4.0 Analysis Methodology

This section will outline the analysis completed in the report using multiple data sources and derived variables outlined in the preceding sections.

## 4.1 Statistical Analyses

To determine if low socio-economic status areas were more exposed to adverse conditions specifically related to urban air pollution and heat islands, this report examined the population distribution across socio-economic status quintiles within close proximity to polluting facilities, as well as based on exposure to elevated land surface temperature and green space. Population counts were all based on the 2006 census. To obtain the distribution of socio-economic status groups within 200 metres of highways, total area, as opposed to population, was used because it was impossible to estimate the proportion of the population in a DA that was within the boundary or just outside of it.

As part of air pollution analyses, hospitalizations, socio-economic status group and buffer zone were all merged at the DA level. This was based on the assignment of each DA to a buffer zone through GIS analysis. Age-standardized hospitalization rates were calculated for each buffer zone, based on respiratory and circulatory hospitalizations from the DAD for 2006–2007 and the 2006 census population aggregated by buffer zone. Confidence intervals for age-standardized rates were also calculated to demonstrate the variability in the results, in particular for the smaller population within 500 metres of facilities. Formulas for confidence intervals are provided in Section 4.2. Next, this same analysis was repeated when further stratified by socio-economic status group.

To test the hypothesis that proximity to polluting facilities was associated with poorer health, each buffer zone was compared with the area more than two kilometres away from any facility. This comparison was further stratified by socio-economic status group, comparing the hospitalization rates by proximity for each socio-economic status group to determine if the relationship was consistent for all groups. Since both socio-economic status groups and buffer zones are non-overlapping, or independent, a standard comparison of two rates was done. Further detail on the tests of the hypothesis is provided in Section 4.3.

Heat-related analysis included a test of the number of hospitalizations on hot days compared with non-hot days to determine if a relationship existed. There were only small differences in hospitalization counts for Montréal and Toronto for hot versus non-hot days overall and within socio-economic areas. Heat analysis also employed logistic regression models to examine the odds of elevated land surface temperature (higher than 30°C) of each socio-economic status group compared with the highest group separately for Montréal and Toronto.

## 4.2 Calculating the Confidence Interval for Age-Standardized Rates

For many analyses conducted within this report, rates are expressed with an accompanying 95% confidence interval. This represents the degree of certainty or reliability for all rates and odds calculated.

Confidence intervals are based on age-standardized rates. Six age groups were used to standardize results to the 1991 Canadian population (as was done in the previous *Reducing Gaps* report, used as a benchmark for comparison). The six age groups were defined by populations younger than age 10, age 10 to 20, age 21 to 35, age 36 to 55, age 56 to 69 and age 70 and older.

The confidence interval calculation involves the use of Bernoulli distribution. To conduct the calculation, the indicator in question, symbolized with (*r*), is taken as a weighted sum of Bernoulli distribution. The mean and the variance are as follows:

$$E(\hat{p}) = p, var(\hat{p}) = \frac{p(1-p)}{n}, where$$

$$\hat{p} = \frac{s}{n}, the\ estimated\ proportion$$

$$s = the\ number\ of\ successes$$

$$n = the\ number\ of\ trials$$

$$p = the\ true\ proportion\ of\ successes$$

For all analyses conducted for the report, each $\frac{count_i}{population_i}$ is an estimated proportion $(\hat{p}_i)$, with *count* representing the *successes* and $population_i$ representing the *number of trials*.

The variance of a weighted sum of random independent variables is the weighted sum of the variance of the variables, with the original weights squared.

Example:

$$Var(Age\ Standardized\ Rate\ (r)) = \sum_{i=1}^{j} Wi^2 Var * Ri(1-Ri) / n$$

$$Var(r) = \sum_{i=1}^{k}\left[\left(\frac{cdn\_standard\_1991_i}{cdn\_standard\_1991}\right)^2 * \left(\frac{count_i}{population_i}\right) * \left(1 - \frac{count_i}{population_i}\right) / population_i\right]$$

The substitution of $\hat{p}_i$, for $p_i$, must be made.

In addition, note that the central limit theorem

$$Z = \frac{(r-R)}{\sqrt{Var(r)}}, where$$

$$R = E(r)$$

has approximately a standard normal distribution and therefore is (more or less) a pivot variable for R. Therefore, an approximate 95% confidence interval for R (the true value of the indicator) is as follows:

$$\left[r - 1.96\sqrt{Var(r)}, r + 1.96\sqrt{Var(r)}\right]$$

The final formula used to make all confidence interval calculations for this report is shown below. It represents the 95% confidence interval for all of the age-adjusted indicators used and is represented by $r \pm 1.96 \times SD$.

$$SD = \sqrt{\sum_{i=1}^{n} \left(\frac{cdn\_standard\_1991_i}{cdn\_standard\_1991}\right)^2 * \left(\frac{count_i}{population_i}\right) * \left(1 - \frac{count_i}{population_i}\right) / \, population_i}$$

## 4.3 Calculating Significance Testing

In a significance test, the difference between the two rates is compared against each rate. If the 95% confidence interval does not include 0, we can conclude with 95% certainty that the two rates in question are statistically different.

One example of conducting significance tests is comparing the rates of one deprivation group against another deprivation group. In this case, each deprivation group and age group is independent of one another; therefore, the covariance is zero.

Let $r_1$ and $r_2$ represent the estimates of two deprivation group rates. The 95% confidence interval for the difference between the true values of these rates is

$$r_1 - r_2 \pm 1.96 \sqrt{\sum_i w_{si}^2 \, m_{1i}(1 - m_{1i}) \, / \, n_{1i} + \sum_i w_{si}^2 \, m_{2i}(1 - m_{2i}) \, / \, n_{2i}}$$

where

$w$ are the standardizing weights $\frac{cdn\_standard\_1991_i}{cdn\_standard\_1991}$, so that $\sum_i w_{si} = 1$;

$m1$ is the specific rate for the *ith* age group of the first deprivation group;

$m2$ is the specific rate for the *ith* age group of the second deprivation group;

$n1$ is the population size for the *ith* age group of the first deprivation group; and

$n2$ is the population size for the *ith* age group of the second deprivation group.

For all CPHI analyses conducted within the report, the asterisk (*) was applied to variables that have been shown to be significant with 95% confidence.

# 5.0 General Limitations

As with any analysis, the methodology employed in this report is subject to limitations because of the choices made during the analytical process. In particular, for this report

- The analyses were performed at the CMA level rather than the city or municipality level. The CMA boundaries may not overlap with political or other administrative boundaries.
- The measure of socio-economic status used in this report was an area-level measure as opposed to an individual measure. As such, there may be some level of error when an individual's socio-economic status does not coincide with the socio-economic status of the area in which he or she lives (for example, a lower socio-economic status DA may also include individuals with higher socio-economic status). Similarly, previous studies have shown that area-level socio-economic status measures tend to underestimate the health gap between deprived and affluent areas, when compared with individual measures of socio-economic status.[20]
- Many factors influence socio-economic status and health. The INSPQ's Deprivation Index is calculated from six variables that comprise the material and social factors, each of which has been shown to be relevant to health. Other variables also influence one's socio-economic status and could be used either in isolation or combination to measure socio-economic status.
- The hospitalization data presented in this report does not necessarily reflect overall health and health status on its own. Multiple factors can influence hospitalization rates, such as prevalence of underlying conditions, access to primary health care, preventive community services and health behaviours like smoking and physical activity. Likewise, hospitalization rates may or may not coincide with mortality statistics (available through the Vital Statistics databases).
- The analysis conducted for this report is cross-sectional, so the time spent within a certain area cannot be taken into account (that is, we do not know whether a person lived in the same area for a long or short period of time, nor whether this person spent most of his or her time in a particular area).

# References


1. Statistics Canada, *Geographic Units: Census Metropolitan Area (CMA) and Census Agglomeration (CA)*, last modified December 17, 2002, accessed from <http://www12.statcan.ca/english/census01/products/reference/dict/geo009.htm>.

2. R. Pampalon et al., "A Deprivation Index for Health Planning in Canada," *Chronic Diseases in Canada* 29, 4 (2009): pp. 178–191, accessed August 12, 2010, from <http://www.phac-aspc.gc.ca/publicat/cdic-mcc/29-4/ar_05-eng.php>.

3. Statistics Canada, *Postal Code Conversion File (PCCF), Reference Guide* (Ottawa, Ont.: Statistics Canada, 2008), accessed from <http://www.statcan.gc.ca/bsolc/olc-cel/olc-cel?catno=92-153-G&chropg=1&lang=eng&asc=1>.

4. Canadian Institute for Health Information, *Quick Reference Guide* (Ottawa, Ont.: CIHI, 2009), accessed from <http://secure.cihi.ca/cihiweb/en/downloads/DAD_Quick_Reference_e.pdf>.

5. Canadian Institute for Health Information, *Discharge Abstract Database*, accessed from <http://www.cihi.ca/cihi-ext-portal/internet/en/document/types+of+care/hospital+care/rehabilitation/services_dad>.

6. Statistics Canada, *National Population Health Survey (NPHS)* (general information and products and services), last updated 2008, accessed from <http://www.statcan.gc.ca/concepts/nphs-ensp/index-eng.htm>.

7. Environment Canada, *National Pollutant Release Inventory*, last updated 2010, accessed from <http://www.ec.gc.ca/inrp-npri/default.asp?lang=en>.

8. Environment Canada, *About the National Pollutant Release Inventory (NPRI)*, accessed from <http://www.ec.gc.ca/inrp-npri/default.asp?lang=En&n=CA8D2224-1>.

9. Environmental Defense Scorecard, *Pollution Locator: Risk Scoring System*, last updated 2005, accessed from <http://scorecard.org/env-releases/def/tep_gen.html>.

10. Commission for Environmental Cooperation, *Taking Stock: 2005 North American Pollutant Releases and Transfers*, last updated 2009, accessed from <http://www.cec.org/takingstock/Default.aspx>.

11. Natural Resources Canada, *About GeoBase Initiative*, last updated August 6, 2009, accessed from <http://www.geobase.ca/geobase/en/about/index.html>.

12. Phone call with Matthew Maloley, Environmental Analyst, Natural Resources Canada, May 6, 2010.

13. S. Briggs, M. Campbell and S. Gower, *Geospatial Distribution of Air Pollution and Socioeconomic Status in Toronto* (Toronto, Ont.: City of Toronto, 2009).

14. E. Garshick et al., "Residence Near a Major Road and Respiratory Symptoms in U.S. Veterans," *Epidemiology* 14, 6 (2003): pp. 728–736.

15. L. Bayer-Oglesby et al., "Living Near Main Streets and Respiratory Symptoms in Adults: The Swiss Cohort Study on Air Pollution and Lung Diseases in Adults," *American Journal of Epidemiology* 164, 12 (2006): pp. 1190–1198.

16. M. Généreux et al., "Neighbourhood Socioeconomic Status, Maternal Education and Adverse Birth Outcomes Among Mothers Living Near Highways," *Journal of Epidemiology & Community Health* 62, 8 (2008): pp. 695–700.

17. M. J. Maloley, *Thermal Remote Sensing of Urban Heat Islands: Greater Toronto Area* (Ottawa, Ont.: Natural Resources Canada, 2009): p. 32.

18. Canada Mortgage and Housing Corporation, *Comparing Neighbourhoods for Sustainable Features*, last updated 2010, accessed on June 23, 2010, from <http://www.cmhc-schl.gc.ca/en/co/buho/sune>.

19. J. Maas et al., "Green Space, Urbanity, and Health: How Strong Is the Relation?," *Journal of Epidemiology & Community Health* 60, 7 (2006): pp. 587–592.

20. R. Pampalon et al., *A Comparison of Individual and Area-Based Socio-Economic Data for Monitoring Social Inequalities in Health* (Ottawa, Ont.: Statistics Canada, 2009), accessed from <http://www.statcan.gc.ca/pub/82-003-x/2009004/article/11035-eng.htm>.

Production of this report is made possible by financial contributions from Health Canada and provincial and territorial governments. The views expressed herein do not necessarily represent the views of Health Canada or any provincial or territorial government.



Canadian Institute for Health Information
495 Richmond Road, Suite 600
Ottawa, Ontario  K2A 4H6

Phone: 613-241-7860
Fax: 613-241-8120
www.cihi.ca
copyright@cihi.ca

ISBN 978-1-55465-885-5



How to cite this document:
Canadian Institute for Health Information, *Urban Physical Environments and Health Inequalities—Data and Analysis Methodology* (Ottawa, Ont.: CIHI, 2011).

Cette publication est aussi disponible en français sous le titre *Environnements physiques en milieu urbain et inégalités en santé — méthodologie de collecte et d'analyse des données*.
ISBN 978-1-55465-887-9

# Talk to Us


**CIHI Ottawa**
495 Richmond Road, Suite 600
Ottawa, Ontario  K2A 4H6
Phone: 613-241-7860

**CIHI Toronto**
4110 Yonge Street, Suite 300
Toronto, Ontario  M2P 2B7
Phone: 416-481-2002

**CIHI Victoria**
880 Douglas Street, Suite 600
Victoria, British Columbia  V8W 2B7
Phone: 250-220-4100

**CIHI Montréal**
1010 Sherbrooke Street West, Suite 300
Montréal, Quebec  H3A 2R7
Phone: 514-842-2226

**CIHI St. John's**
140 Water Street, Suite 701
St. John's, Newfoundland and Labrador  A1C 6H6
Phone: 709-576-7006



March 2011

www.cihi.ca

Canadian Institute for Health Information
Institut canadien d'information sur la santé