Working Paper

Health Region Peer Groups-
Working Paper, 2018

Release date: December 14, 2018

Statistics   Statistique
Canada   Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at STATCAN.infostats-infostats.STATCAN@canada.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- Statistical Information Service   1-800-263-1136
- National telecommunications device for the hearing impaired   1-800-363-7629
- Fax line   1-514-283-9350

Depository Services Program

- Inquiries line   1-800-635-7943
- Fax line   1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Table of contents

Health Region Peer Groups – Working paper, 2018

Purpose

The purpose of this document is to define the concept of peer groups, to give an overview of how they are created and to demonstrate their usefulness. This paper presents the 2018 classification of the peer groups. More detailed technical information on the formation of the peer groups can be found in the working papers *Health Region Peer Groups* and *Health Region Peer Groups 2003* written by the Health Statistics Division of Statistics Canada.

1. Introduction

The launch of the Canadian Community Health Survey (CCHS) in 2000, along with the expansion of existing data products at the health region level, lead to the desire for a method of comparing regions with similar socio-economic determinants of health. The reasoning behind the development of such a method is that once the effects of the various social and economic characteristics known to influence health have been removed, it is possible to compare regions by measures of health status. It is also possible to compare the relative effectiveness of health promotion and prevention activities across regions. Thus, the health regions have been placed into groups with similar socio-economic characteristics using a clustering technique, and these groups are referred to as 'peer groups'.

Development of the criteria used to define peer groups required careful consideration of their intended use. The requirement that peer groups be used as a method for comparing health related issues ultimately eliminated all variables directly describing health as potential candidates in the creation of the groups. Further, it was desired that all variables used must be reliable and available for all health regions. As well, the need for objectivity required that peer groups be developed using empirical techniques. Finally, consideration of the need for simplified and relevant comparisons also required that peer groups have approximately five to ten health regions per group and that there be representation across the country within each group. In the application of the above parameters, several limiting factors resulted which required some modifications. All criteria were followed to the extent possible and any deviations are explained in detail throughout this document.

The original 2000 Peer Group Classification was released in 2002, and was based on the 1996 Census information as well as the health region boundaries as defined by the provinces and territories in 2000. In order to remain current with respect to data availability and the health region boundary changes, it is necessary to update the peer group classification over time. These updates have taken place in the form of the 2003 Peer Group Classification, the 2007 Peer Group Classification and the 2011 Peer Group Classification. The latest update to the peer groups is based on 2016 Census data and the health region boundaries as of December 2018. The final result of this classification was the creation of peer groups, representing all health regions across Canada.

This document will give an overview of how the peer groups are created. The 2018 Peer Group Classification is presented and the results are compared with the peer groups created in the past. Finally, the use of the peer groups in the analysis of health related issues will be demonstrated through an example.

2. Data

Typically, 24 variables describing the socio-economic and socio-demographic determinants of health within the health regions across Canada are used in the clustering algorithm to produce the peer groups. The variables chosen for this task cover a wide range of areas including demographic structure, social and economic status, ethnicity, Aboriginal status, housing, urbanization, income inequality and labour market conditions. Note that health-related variables were deliberately not used in the creation of the peer groups.

There have been some modifications made over time; however the majority of the variables have remained consistent since the peer groups were first created in 2002. The latest Peer Group classification used 23 out of 24 variables used in the creation of the 2011 peer groups. These variables are based on the 2016 Census. The variables used for the analysis, along with their respective sources, are outlined in Appendix A.

3. Methodology

A non-hierarchical cluster analysis was the method chosen to create the peer groups. Generally speaking, cluster analysis attempts to assign observations to groups (clusters) based on a measure of their distance from each other so that observations within each group are similar to one another with respect to variables or attributes of interest. In other words, the goal is to group the observations into homogeneous and distinct clusters. Non-hierarchical algorithms attempt to partition a set of observations into a pre-defined set of disjoint groups using a specified optimization criterion. This approach appeared best suited to meet the original objectives of the peer group project, mainly to use an empirical technique to create a pre-defined number of peer groups with approximately 5 to 10 health regions within each group.

The peer groups were created in SAS using the FASTCLUS procedure. This procedure uses a *k*-means algorithm to assign observations to a pre-defined set of *k* clusters. A description of *k*-means clustering and several variants of the method can be found in Johnson and Wicheren (2002). The basic steps for placing observations into *k* clusters are as follows:

1. Select *k* observations as cluster seeds (the initial centers of the clusters).
2. Assign observations to the nearest cluster seed. After all observations are assigned, cluster seeds are replaced by their respective cluster means. This step is repeated until the change in cluster seeds becomes or approaches zero.
3. Form final clusters by assigning each observation to its nearest cluster seed.

Complete details of the FASTCLUS procedure can be found in the SAS OnlineDoc®, Version 9.

3.1 Number of Clusters

One of the major challenges with cluster analysis is selecting the appropriate number of initial clusters. Several criteria have been suggested (Everitt et al., 2001) which generally involve the optimization of one or more test statistics. From a practical perspective it is generally left up to the analyst to determine the number that best suits a given need. For the purpose of the 2018 Peer Group classification a maximum of 16 clusters was chosen. This would give an average number of seven health regions to each peer group,[1] which is in line with the study objectives. The maximum number of clusters used in 2018 was the same as in 2014.

4. Results

4.1 Standardization of Variables

Variables measured on different scales, or on a common scale with differing variances, are often standardized in order to mitigate the effect of these differences among the variables. For this exercise, all 24 socio-economic variables were standardised (mean 0, variance 1) prior to performing the cluster analysis.

Some variables contained missing or zero values to indicate that this information was not available in certain health regions. The proportion of low income persons in private households (LowPop) and the proportion of low income children (LowKids) contained missing values, because the Census does not derive low income data for the three territories and Indian reserves. The average dwelling value (AVGDWL), the housing affordability (HOUAFF) and the immigration (IMMPER) variables contained missing or zero values for some regions, because the concept of a variable didn't apply or the question wasn't asked in regions consisting mostly of Indian reserves. The regions containing missing or zero values for these variables were 2418, 4604, 4714 and the Territories. All missing values were set to zero for the analysis, before standardizing the variables.

---

1. Note that the terms peer group and cluster are used interchangeably to refer to the classification of health regions into groups with similar socio-economic characteristics.

4.2 Creation of Peer Groups

To establish a starting point, the clustering algorithm was instructed to group the health regions into 16 clusters. Six of the resulting clusters contained only one health region. This indicated that 16 clusters were too many given that the objective of assigning peer groups is to be able to compare like health regions. The cluster analysis was rerun with a reduced number of cluster seeds.

The results of the final cluster analysis using PROC FASTCLUS can be seen in Table 4.2.1. The table shows the number of health regions contained in each peer group, as well as several statistics related to the clusters. The root mean square standard deviation is a measure of the variability in the data points around the cluster centre. The radius displays the largest Euclidean distance from the cluster centre to any observation within the cluster. The nearest cluster refers to the closest peer group in terms of Euclidean distance. Finally, the last column of the table displays the distance between the current cluster centre and that of its closest neighbour. For each of these statistics, the cluster centre is the point having coordinates that are the means of all the observations in the cluster. Euclidean distance is a statistical measure of distance between two points.

Table 4.2.1
Results of Final Cluster Analysis using PROC FASTCLUS

Cluster	Frequency	Root Mean Square Standard Deviation	Radius	Nearest Cluster	Distance Between Cluster Centers
A	6	0.60	3.71	D	3.63
B	21	0.52	3.72	C	3.13
C	37	0.45	3.65	D	2.20
D	30	0.53	3.49	C	2.20
E	7	0.48	2.72	I	4.47
F	5	1.21	6.79	I	8.78
G	4	0.78	4.29	H	7.15
H	9	0.66	4.32	J	4.10
I	2	0.73	2.46	E	4.47
J	3	0.50	2.53	H	4.10

There were two clusters that contained the majority of health regions (C and D). Both of these clusters were comprised of regions that were very similar (as both clusters were large in terms of the number of health regions and had low standard deviations). As well, these clusters were nearest neighbours and the distance between their cluster centers was small, demonstrating that the health regions in both clusters were also similar. Therefore, although these clusters did not meet the objective of having approximately five to ten regions per peer group, there did not appear to be a valid reason to split them into smaller groups.

Note that the total number of health regions in Table 4.2.1 no longer equals 110.[2] This is due to a new development that was added to the 2011 Peer Group Classification. There are two levels of geography in Ontario: there are 14 Local Health Integration Networks (LHINs) and 36 Public Health Units (PHU). Due to the relationship between the two levels of geography it was possible to incorporate both into the peer group classification. The information at the PHU level was used to create the peer groups. At the final stage in the cluster analysis, the LHIN level geography was added to the existing clusters. The LHINs did not have an impact on the placement of the other health regions into the final peer groups. In an analysis involving the peer groups, only one level of geography in Ontario should be used.

4.3 Collapsing Small Clusters

The results from section 4.2 (specifically Table 4.2.1), represent clusters that are approximately evenly spaced and have minimal within cluster variance given the parameters used by the clustering algorithm. The results in the table show that 10 clusters were formed that range in size from two to 37 health regions (excluding the LHINs). However, having a cluster with less than five regions is not practical as it does not provide many options for comparison. In order to provide more peers for comparison, clusters with less than five members were combined with their nearest neighbour. The exception was cluster G (Montréal, Toronto and Vancouver). Cluster G was not combined with another cluster since these health regions tend to be very different than other regions across the country.

---

2. For more details on the health regions used see section 5.5.

There were two clusters that were joined with their closest neighbour. Cluster I (health regions 1014 and 2409) was combined with its nearest neighbour cluster E (health regions 1012, 1013, 1203, 1305, 1306, 1307, 2411). The collapsing of clusters I and E produced a cluster with 9 health regions, so no additional collapsing was required. This combined cluster was labelled Cluster K. As well, Cluster J (health regions 3536, 4832 and 4834) was combined with cluster B (health regions 1204, 2403, 2407, 2413, 2416, 3530, 3537, 3544, 3551, 3565, 3568, 4601, 4704, 4706, 5921 and 5941). This combined cluster was labelled Cluster B. The result of collapsing the smaller clusters was that the 10 peer groups produced from the final cluster analysis using the FASTCLUS procedure and presented in Table 4.2.1 were reduced to eight groups. A list of Health regions (excluding LHINs) by the final peer groups can be found in Appendix D.

5. Discussion

5.1 Strongest Predictors

In order to determine which variables played a key role in defining the health region peer groups, the final clusters were run against all 23 variables in a stepwise discriminant analysis. Partial R^2 statistics for entry and removal were set at 0.15. Any variable which had an R^2 value of 0.5 or higher when regressed against a variable already in the model was removed from the analysis. Overall, five variables were appeared to be the most important predicators. Table 5.1.1 displays a summary of the results.

Table 5.1.1
Stepwise discriminant analysis of final health region groupings on all 23 variables

Step	Variable		Partial R-SQ	R-SQ Variables in Model
1	**PopDen**		**0.9105**	
(Population Density)	Removed			
		None		
2	**VisMin**		**0.8746**	
(Visible Minority)	Removed			
		None		
3	Pop20		**0.8414**	
(Population under 20)	Removed			
		AVGDWL		0.6461
		GOVTRAN		0.5813
		GROWTH		0.5699
		HOUAFF		0.5982
		IMMPER		0.8332
		LOW15		0.5148
		LOWKIDS		0.5065
		MIZ		0.5174
		MedInc		0.5458
		POP16		0.5072
4	**LTUNEMP**		**0.7771**	
(Long Term Unemployment Rate)	Removed			
		ABOPER		0.7092
		LNEPRNT		0.5009
		MEDSHR		0.5007
		OWNDWL		0.5777
		POP65		0.7963
		POSTSEC		0.5501
5	**MIGMOB**		**0.2479**	
(Internal Migrant Mobility)	Removed			
		EMP		0.7943
		UNEMP		0.9572

The strongest predictors of the final peer groups were population density and the proportion of the population belonging to a visible minority group. No additional variables were removed from the analysis when regressed against population density, whereas the average dwelling value was removed from the analysis when regressed against visible minority.

5.2 Principal Component Analysis

Principal component analysis is a multivariate technique which aims to reduce the number of variables in the data to a few factors called principal components. Principal components are linear combinations of the original variables and are uncorrelated. They are derived in decreasing order of importance, so that as much of the total variance in the data can be explained in as few factors as possible. Therefore, the first principal component is the most important factor since it explains the largest proportion of the total variance in the data.

A principal component analysis was performed on the 23 socio-economic variables used in the cluster analysis. The first two principal components accounted for just over 57% of the total variability. The first principal component appears to be measures of "urbanicity" (housing affordability, proportion of visible minorities, proportion of immigrants, average dwelling value, population living in census metropolitan areas, etc.). The second principal component seems to be measures of family profile (proportion of the population aged 65 and over, proportion of lone-parent families, proportion of the population aged 0 to 19, proportion of dwellings in which the owner also lives and proportion of aboriginal population). The third principal component could be interpreted as income inequality (median household income, proportion of all income that came from government transfers, proportion of low-income children, proportion of low-income persons in private households, etc.). The first six principal components accounted for over 86% of the total variability in the data, showing that 23 variables can be reduced to six factors without losing much information. These results are similar to the previous peer group classification, which indicates that the variables which drive the analysis are remaining fairly consistent over time.

5.3 Peer Group Description

The five key variables determined by the stepwise discriminant analysis were used to represent each of the clusters. The mean values of these five variables for each peer group can be found in Appendix B. For each of the five variables, several percentiles were calculated and used to classify the peer groups. Values were classified based on the following ranges.

Very High: X > 85th percentile

High: 65th percentile < **X** ≤ 85th percentile

Medium: 35th percentile < **X** ≤ 65th percentile

Low: 15th percentile < **X** ≤ 35th percentile

Very Low: X ≤ 15th percentile

The results from this classification can be found in Table 5.3.1. While the methodology is crude as a descriptive tool, it does help to distinguish one peer group's characteristics from another. As shown in the table below, there are no two peer groups which fall into the same category for all five variables. For example, peer group G (which consists of Toronto, Vancouver and Montréal) is the only group with a very high population density, a high percent of visible minorities and a very low 5-year internal migration.

The results from this classification were used to derive a written summary of the eight peer groups based on the five key variables from the discriminant analysis. This summary is presented in Appendix C.

Table 5.3.1
Final peer grouping descriptions based on five factors resulting from the stepwise discriminant analysis

Cluster	Population Density	Visible Minority	Population under 20 years old	Long term Unemployment Rate	Five-year Internal Migration
A	Very Low	High	Very High	High	High
B	High	High	Medium	Low	Low
C	Medium	Medium	Low	Medium	Medium
D	Medium	Medium	Medium	Low	High
E	Medium	Very Low	Very Low	Very High	Low
F	Very Low	Very Low	Very High	Very High	Low
G	Very High	Very High	Low	Medium	Very Low
H	Very High	Very High	Medium	Low	Medium

5.4 Geographic Limitations

Each province and territory defines the geographic boundaries for a health region based on administrative preference and these boundary definitions change over time. Health regions can be strictly urban or rural or some combination of the two. There may be considerable variability within health regions in regards to health measures due to the lack of geographic homogeneity and this should be taken into consideration when inferences are being made about a certain region. For instance, even though the health indicators in Vancouver compare favourably with the national averages, this should not be interpreted as meaning that the residents of the downtown core in Vancouver have better than average health. This lack of homogeneity in defining health region boundaries makes the exercise of assigning health regions to peer groups much more difficult, as it can have a large impact on how well a certain variable represents the entire region and in some cases important defining factors may be missed.

It should also be noted that there may be considerable variability amongst the health regions within a peer group in regards to the socio-economic factors used in the cluster analysis. This should be considered when comparing regions within a certain peer group. This variability can be seen for the 2018 peer groups in Appendix B for the five key variables determined by the stepwise discriminant analysis.

5.5 Collapsing Health Regions

There is one instance where the CCHS combines smaller health regions to increase the sample size for reportability. This occurs in northern Saskatchewan where health region 4713 (Athabasca Health Authority) is combined with 4711 (Mamawetan Churchill River Regional Health Authority) and 4712 (Keewatin Yatthé Regional Health Authority) to form 4714 (Mamawetan/Keewatin/Athabasca).

The decision was made to use the combined health region (4714) in the creation of the peer groups since the CCHS is one of the principal data sources used in an analysis of health related data by peer groups.

5.6 Geographic Representation of Final Peer Groups

The map below is a good visual representation of the geographic clustering of the health regions into the final nine peer groups. Montréal, Toronto and Vancouver form the smallest cluster because in terms of the size and the diversity of their populations, they are too different from the other health regions to be combined with any other peer group.

There are some definite clusters of health regions that formed based on common characteristics due to their location within Canada. The northern regions have clustered based on the Aboriginal make-up of their communities. All peer groups have representation across provincial and/or territorial borders.

Map 14
Health Regions and Peer Groups in Canada, 2018
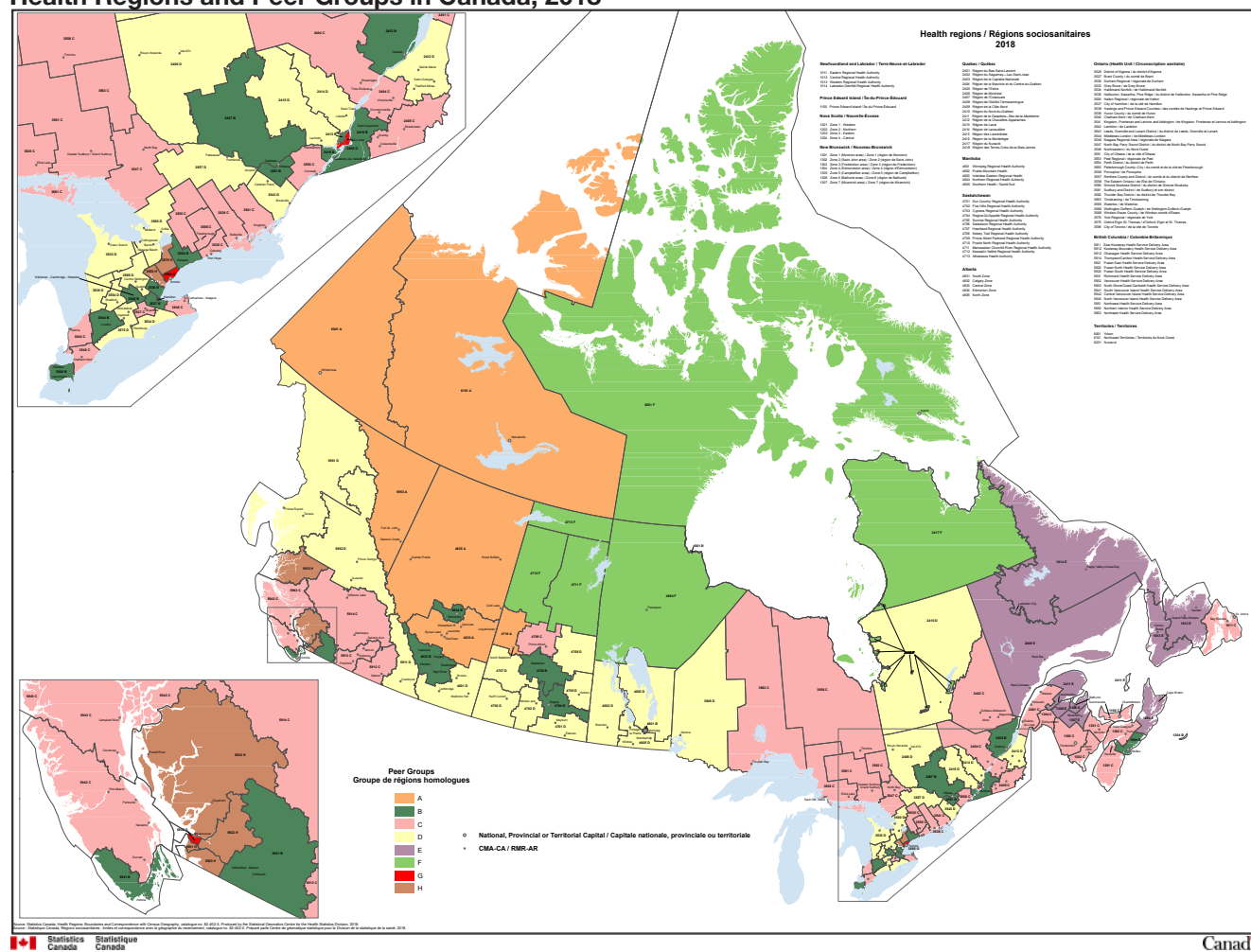
A larger version of the map is available.

6. Peer Groups in Action

The purpose of this section is to demonstrate the usefulness of the peer groups. There are two valuable, yet different, analyses possible with the peer groups; health-related indicators can be compared between and within peer groups. Since peer groups are formed based on regions that have similar socio-economic characteristics, it is expected that differences between peer groups will arise. Peer groups with better socio-economic status indicators are likely to have better health status measures. Estimates of a single peer group can also be compared with national averages in order to ascertain how well the group of regions fairs as a whole. The second analysis possible, one that may be of more relevant importance, is the comparison of health regions within a peer group. Once the effects of the various social and economic characteristics known to influence health status have been removed, a more useful comparison of regions by measures of health status is possible.

The example illustrated in Section 6.1 is a simple demonstration of how and when peer groups can be used. The example uses the 2018 Peer Group Classification and 2015-2016 CCHS data. A similar analysis will be possible with 2017-2018 CCHS data and the 2018 Peer Group Classification once the data is released. A more detailed analysis involving the peer groups can be found in the paper "The Health of Canada's Communities', written by Margot Shields and Stéphane Tremblay of Statistics Canada (2002).

6.1 Example: Heart Disease

This example focuses on the prevalence of heart disease in the population 18 years of age and over in the different regions across the country. Every CCHS respondent is asked if he/she has heart disease. The national rate of heart disease for the adult population is 4.9%. The missing rate for this health indicator is less than 0.5% and in this example the missing values have been excluded.

The rate of heart disease in each peer group is shown in Table 6.1.1, along with the description of the peer group. The prevalence of heart disease in Peer Group H is 1% lower than the national average. It is also 4.5% lower than the rate of heart disease in Peer Group E. Both of these differences are significant (p-value<0.01). Peer Group H is composed of mainly urban centers in Ontario and British Columbia with high population density. This group has a low smoking rate (13%), a low heavy drinking rate (15.1%) and a slightly below average exercise rate (67.5%). On the other hand, Peer Group E is composed of mainly rural eastern health regions that have a very high rate of government transfer income. This group has higher smoking rate (23.2%), higher heavy drinking rate (23.1%) and a lower physical activity rate (57.7%). The differences in the rates of these three risk factors between Peer Groups H and E are all significant (p-value<0.01).

Table 6.1.1
Prevalence of Heart Disease Rates by Peer Group

Peer Group	Number of Health Regions	Principal Characteristics	Heart Disease Rate
A	6	• Mainly rural and remote regions in the Western provinces and the Territories • Low population density • Higher proportion of population below 20 years of age • Average population growth from 2011 to 2016 • Average employment rate • High Aboriginal population	3.8%
B	19	• Mainly urban centers with moderately high population density • Rapid population growth from 2011 to 2016 • High proportion of people recently moved to or within these regions since last census • Very high employment rate • High immigrant and visible minority population	4.4%
C	33	• Sparsely populated urban-rural mix from coast to coast • Low percentage of visible minority population • Very low population growth from 2011 to 2016 • Moderately high proportion of population receiving government transfer income • High population of seniors aged 65 and above	6.5%
D	29	• Mainly rural regions in Quebec, Ontario and the Prairies • Moderate Aboriginal population • Low population growth from 2011 to 2016 • High employment rate	5.6%
E	9	• Mainly rural Eastern regions with low population density • Very high percentage of population receiving government transfer income • Negative population growth from 2011 to 2016 • Very high population of seniors aged 65 and above • High long-term unemployment rate	8.4%
F	5	• Northern and remote regions with very low population density • Very high Aboriginal population • High population growth from 2011 to 2016 • Low proportion postsecondary graduates • Very high proportion of lone-parent families • Very low immigrant and visible minority population • Very high proportion of children under the age of 20	4.1%[c]
G	3	• Largest metro centers with an average population density of 4389 people per square kilometre • Very low Aboriginal population • Moderate percentage of government transfer income • High immigrant and visible minority population • Very high proportion of postsecondary graduates • High proportion of households spending more than 30% of the income on shelter	4.1%

Table 6.1.1
Prevalence of Heart Disease Rates by Peer Group

Peer Group	Number of Health Regions	Principal Characteristics	Heart Disease Rate
H	6	• Mainly urban centers in Ontario and British Columbia with high population density • High population growth from 2011 to 2016 • Low Aboriginal population • Very high average dwelling value • Very high immigrant and visible minority population	3.9%

Note that when the CV associated with the estimate is between 15% and 35%, the estimate in the table has a 'C' or 'D' beside it, which indicates general unrestricted release and is a warning cautioning users of the high sampling variability. When the CV associated with the estimate is above 35%, the estimate in the table is replaced by an 'E', which indicates that the estimate cannot be released.

There are six health regions that make up Peer Group H. Table 6.1.2 shows the prevalence of heart disease in each of these regions. There are five out of six regions that have a prevalence rate equal to or below the national rate of 4.9%. The highest incidence of heart disease is 5.1% and occurs in health region 5933. On the other hand, the lowest heart disease rate is 3.2% and occurs in health region 3570.

Table 6.1.2
Prevalence of Heart Disease in the Health Regions Belonging to Peer Group B

Health Region	Name	Heart Disease Rate
3553	Peel Regional Health	3.5% C
3570	York Regional Health Unit	3.2% C
5922	Fraser North Health Service Delivery Area	4.3%
5923	Fraser South Health Service Delivery Area	4.9%
5931	Richmond Health Service Delivery Area	3.5% C
5933	North Shore/Coast Garibaldi Health Service Delivery Area	5.1% C

When it is desired to perform an analysis of a rare event at the health region level, it is often the case that the associated CV does not allow for general release of the information. Table 6.1.2 shows that 4 of the 6 regions in Peer Group H have an estimate of heart disease that is associated with a high sampling variability (CV between 15% and 25%). The majority of the health regions belonging to this group have a moderately high population density. For other peer groups that contain more remote health regions, it may not be possible to conduct the same analysis due to corresponding CVs above 35%. Typically in these cases, the results are published at the province level in order to obtain more sample size and more reliable estimates. The peer groups offer an alternative to the provinces in these types of situations.

7. Summary

As a result of health region boundary changes as of December 2018, and the availability of 2016 Census data, it was necessary to update the 2014 Peer group classification. In keeping with the original working paper, the goal was to produce a classification which would cluster health regions with similar social and economic health determinants into peer groups. Twenty-three variables covering a wide range of social, economic and demographic areas were used to cluster the health regions.

Starting with an initial set of 16 clusters, and ensuring that each cluster contained at least two health regions, the results indicated that the regions naturally grouped themselves into 10 distinct peer groups. Peer groups with fewer than five health regions were combined with their nearest neighbour. This was done to provide enough health regions within a peer group for comparison purposes. Cluster G containing Montréal, Toronto and Vancouver was not forced to join another cluster as these health regions tend to have more in common with themselves than with other health regions. The final result was eight peer groups ranging in size from three to 33 (not including the LHINs in Ontario).

Stepwise discriminant analysis was used to determine which variables had the most influence on the final peer groupings. The five most important variables were population density, proportion of visible minority, proportion of population aged 0 to 19 years old, long term unemployment rate and 5-year internal migration. Each peer group has at least one distinguishing factor in terms of these five variables.

Peer groups can be useful in an analysis of health-related indicators since once the effects of the various social and economic characteristics known to influence health status have been removed, a more useful comparison of regions is possible. Health indicators can be compared between and within peer groups. As well, the peer groups offer an alternative to the provinces when the results of an analysis cannot be presented at the health region level due to insufficient sample size or high sampling variability.

8. References

Andberg, M. R. (1973). *Cluster Analysis for Applications.* New York: Academic Press.

Everitt, B. S., Landau S., Leese M. and Stahl, D. (2001). *Cluster Analysis, 5th Edition*, John Wiley and Sons, Ltd: Chichester, UK.

Johnson, R. and Wicheren, D. (2002). *Applied Multivariate Statistical Analysis*, Prentice Hall.

Sarafin, C. (2009). 2007 Health Region Peer Groups Methodology Guide, Internal document, Health Indicators, Statistics Canada.

SAS Institute Inc., *SAS/STAT® User's Guide, Version 8*, Cary, NC: SAS Institute Inc., 1999.

Shields, M and Tremblay, S. 2002, "The Health of Canada's Communities." *Health Reports* (Statistics Canada), Catalogue Number 82-003-XIE.

Wannell B. (2009). 2007 Health Region Peer Groups, *Internal document, Health Indicators*, Statistics Canada.

Appendix A
Variable Definitions
Variables used in the creation of the 2018 Peer Groups are from Census 2016.

Variable	Description
POP16	2016 population (based on population and dwelling counts not randomly rounded but adjusted)
ABOPER	Aboriginal percentage
AVGDWL	Average value of dwelling -owner-occupied, non-farm, non-reserve)-(Canadian dollars)
EMP	Employment rate (persons aged 25 to 54)
GOVTRAN	Government transfer income in 2015, as a proportion of total income (percent)
GROWTH	Growth rate (% change in regions population between 2011 and 2016)
HOUAFF	Households spending 30% or more of household income on shelter, proportion of total shelter-cost households
IMMPER	Immigrants who arrived between 2006 and 2016, proportion of immigrant population (percent)
MEDSHR	Income share held by households whose incomes fall below the median household income in 2015 (percent)
LNEPRNT	Lone-parent families, proportion of census families (percent)
LTUNEMP	Long-term unemployment rate, labour force aged 15 and over
MedInc	Median income (household/family)
OWNDWL	Owner-occupied private non-farm, non-band, non-reserve dwellings
POP65	Population 65 years and older
POP20	Population aged 0 to 19 years, proportion of total population
POPDEN	Population density (population per square kilometer) (number)
MIZ	Population living within a Census Metropolitan Area, a Census Agglomeration or a strong Census Metropolitan Area and Census Agglomeration Influenced Zone (number)
POSTSEC	Post-secondary graduates, proportion of population aged 25 to 54 (percent)
LOWPop	Prevalence of low income before tax in 2015 for persons in private households (percent)
LOWKIDS	Prevalence of persons aged 17 years and under living in low income economic families before tax in 2015 (percent)
UNEMP	Unemployment rate 15 years and over
VISMIn	Visible minority population, proportion of total population (percent)
MIGMOB	5-year internal migrants, proportion of population aged 5 years and over (percent)

Appendix B
Descriptive Statistics for Final Peer Groups (Excluding LHINs)

Cluster		Population Density	Visible minority	Population under 20 years old	Long term Unemployment Rate	5-year internal migration
A	N	6	6	6	6	6
	MIN	0	6.6	22.9	4.4	16.7
	MAX	5	9.6	30.6	6.6	19.9
	Mean	1.52	8	27.0167	5.4333	18.2
	St. Dev.	1.96	1.3372	2.5733	0.871	1.1489
B	N	19	19	19	19	19
	MIN	10.7	5	18	2.8	5.4
	MAX	1710.9	31	26.2	5.2	18.1
	Mean	307.49	19.2421	23.0895	3.8316	12.7526
	St. Dev.	434.02	7.4286	1.9661	0.5628	3.4339
C	N	33	33	33	33	33
	MIN	0.3	1	17.9	3.4	8.4
	MAX	241.5	8.9	28.1	7.6	20.4
	Mean	25.25	3.797	20.6697	4.9879	14.6455
	St. Dev.	44.73	1.9045	1.9324	0.9509	3.4499
D	N	29	29	29	29	29
	MIN	0	1.4	20	2	11.8
	MAX	68.6	11.6	31.2	7	21.7
	Mean	17.27	4.2241	23.6621	3.6552	16.6172
	St. Dev.	21.02	2.4784	2.2187	1.1099	2.5553
E	N	9	9	9	9	9
	MIN	0.1	0.7	16.2	7.6	7
	MAX	16.3	2.6	23.7	14.8	13.8
	Mean	4.99	1.3222	18.8889	11.0889	10.3444
	St. Dev.	5.16	0.7207	2.6012	2.439	2.1847
F	N	5	5	5	5	5
	MIN	0	0.7	38.6	8.7	7.9
	MAX	3.1	3.2	43.4	16.7	12.9
	Mean	0.68	1.64	40.72	12.58	9.9
	St. Dev.	1.36	1.1718	1.7641	3.1925	1.9352
G	N	3	3	3	3	3
	MIN	3889.8	32.9	15.9	3	7.1
	MAX	4942.6	51.9	20.7	5.5	11.3
	Mean	4388.93	45.4333	18.8333	4.4333	8.5333
	St. Dev.	528.51	10.856	2.5716	1.2897	2.3965
H	N	6	6	6	6	6
	MIN	5.5	22.2	19.6	3	8.2
	MAX	1534.1	76.3	25.4	4.7	17.3
	Mean	740.3	50.4333	22.5	3.6	12.75
	St. Dev.	569.32	18.1312	2.3597	0.6197	3.2005

Appendix C
Descriptive summary of Final Peer Groups

Peer Group	Number of Health Regions	Percent of Canadian Population	Principal Characteristics
A	6	3.20%	• Mainly rural and remote regions in the Western provinces and the Territories • Low population density • Higher proportion of population below 20 years of age • Average population growth from 2011 to 2016 • Average employment rate • High Aboriginal population
B	19	34.90%	• Mainly urban centers with moderately high population density • Rapid population growth from 2011 to 2016 • High proportion of people recently moved to or within these regions since last census • Very high employment rate • High immigrant and visible minority population
C	33	18.10%	• Sparsely populated urban-rural mix from coast to coast • Low percentage of visible minority population • Very low population growth from 2011 to 2016 • Moderately high proportion of population receiving government transfer income • High population of seniors aged 65 and above
D	29	13.70%	• Mainly rural regions in Quebec. Ontario and the Prairies • Moderate Aboriginal population • Low population growth from 2011 to 2016 • High employment rate
E	9	2.00%	• Mainly rural Eastern regions with low population density • Very high proportion of population receiving government transfer income • Negative population growth from 2011 to 2016 • Very high population of seniors aged 65 and above • High long-term unemployment rate
F	5	0.50%	• Northern and remote regions with very low population density • Very high Aboriginal population • High population growth from 2011 to 2016 • Low proportion postsecondary graduates • Very high proportion of lone-parent families • Very low immigrant and visible minority population • Very high proportion of children under the age of 20
G	3	15.10%	• Largest metro centers with an average population density of 4389 people per square kilometre • Very low Aboriginal population • Moderate percentage of government transfer income • High immigrant and visible minority population • Very high proportion of postsecondary graduates • High proportion of households spending more than 30% of the income on shelter
H	6	12.50%	• Mainly urban centers in Ontario and British Columbia with high Population density • High population growth from 2011 to 2016 • Low Aboriginal population • Very high average dwelling value • Very high immigrant and visible minority population

Appendix D
Health Region Peer Groups

Peer group A

4710 Prairie North Regional Health Authority

4833 Central Zone

4835 North Zone

5953 Northeast Health Service Delivery Area

6001 Yukon

6101 Northwest Territories

Peer group B

1204 Zone 4 - Central

2403 Région de la Capitale-Nationale

2407 Région de l'Outaouais

2413 Région de Laval

2416 Région de la Montérégie

3530 Durham Regional Health Unit

3536 Halton Regional Health Unit

3537 City of Hamilton Health Unit

3544 Middlesex-London Health Unit

3551 City of Ottawa Health Unit

3565 Waterloo Health Unit

3568 Windsor-Essex County Health Unit

4601 Winnipeg Regional Health Authority

4704 Regina Qu'Appelle Regional Health Authority

4706 Saskatoon Regional Health Authority

4832 Calgary Zone

4834 Edmonton Zone

5921 Fraser East Health Service Delivery Area

5941 South Vancouver Island Health Service Delivery Area

Peer group C

1011 Eastern Regional Health Authority

1100 Prince Edward Island / Île-du-Prince-Édouard

1201 Zone 1 - Western

1202 Zone 2 - Northern

1301 Zone 1 (Moncton area)

1302 Zone 2 (Saint John area)

1303 Zone 3 (Fredericton area)

1304 Zone 4 (Edmundston area)

2401 Région du Bas-Saint-Laurent

2402 Région du Saguenay--Lac-Saint-Jean

2404 Région de la Mauricie et du Centre-du-Québec

2405 Région de l'Estrie

Appendix D
Health Region Peer Groups

Peer group C

3526 The District of Algoma Health Unit

3527 Brant County Health Unit

3535 Haliburton, Kawartha, Pine Ridge District Health Unit

3538 Hastings and Prince Edward Counties Health Unit /

3540 Chatham-Kent Health Unit

3541 Kingston, Frontenac and Lennox and Addington Health Unit

3542 Lambton Health Unit

3546 Niagara Regional Area Health Unit

3547 North Bay Parry Sound District Health Unit

3555 Peterborough County-City Health Unit

3556 Porcupine Health Unit

3558 The Eastern Ontario Health Unit

3561 Sudbury and District Health Unit

3562 Thunder Bay District Health Unit

3563 Timiskaming Health Unit

4709 Prince Albert Parkland Regional Health Authority

5912 Kootenay-Boundary Health Service Delivery Area

5913 Okanagan Health Service Delivery Area

5914 Thompson/Cariboo Health Service Delivery Area

5942 Central Vancouver Island Health Service Delivery Area

5943 North Vancouver Island Health Service Delivery Area

Peer group D

2408 Région de l'Abitibi-Témiscamingue

2410 Région du Nord-du-Québec

2412 Région de la Chaudière-Appalaches

2414 Région de Lanaudière

2415 Région des Laurentides

3533 Grey Bruce Health Unit

3534 Haldimand-Norfolk Health Unit

3539 Huron County Health Unit / Circonscription sanitaire du comté de Huron

3543 Leeds, Grenville and Lanark District Health Unit

3549 Northwestern Health Unit

3554 Perth District Health Unit

3557 Renfrew County and District Health Unit

3560 Simcoe Muskoka District Health Unit

3566 Wellington-Dufferin-Guelph Health Unit

3575 Oxford Elgin St. Thomas Health Unit

4602 Prairie Mountain Health

4603 Interlake-Eastern Regional Health Authority

Appendix D
Health Region Peer Groups

Peer group D

4605 Southern Health

4701 Sun Country Regional Health Authority

4702 Five Hills Regional Health Authority

4703 Cypress Regional Health Authority

4705 Sunrise Regional Health Authority

4707 Heartland Regional Health Authority

4708 Kelsey Trail Regional Health Authority

4831 South Zone

5911 East Kootenay Health Service Delivery Area

5951 Northwest Health Service Delivery Area

5952 Northern Interior Health Service Delivery Area

Peer group E

1012 Central Regional Health Authority

1013 Western Regional Health Authority

1014 Labrador-Grenfell Regional Health Authority

1203 Zone 3 - Eastern

1305 Zone 5 (Campbellton area)

1306 Zone 6 (Bathurst area)

1307 Zone 7 (Miramichi area)

2409 Région de la Côte-Nord

2411 Région de la Gaspésie--Îles-de-la-Madeleine

Peer group F

2417 Région du Nunavik

2418 Région des Terres-Cries-de-la-Baie-James

4604 Northern Regional Health Authority

4714 Mamawetan/Keewatin/Athabasca

6201 Nunavut

Peer group G

2406 Région de Montréal

3595 City of Toronto Health Unit

5932 Vancouver Health Service Delivery Area

Peer group H

3553 Peel Regional Health Unit

3570 York Regional Health Unit

5922 Fraser North Health Service Delivery Area

5923 Fraser South Health Service Delivery Area

5931 Richmond Health Service Delivery Area

5933 North Shore/Coast Garibaldi Health Service Delivery Area