Chapter 3

Generating Rural Health Indicators in the Canadian Context

3.1 Introduction

Addressing the conceptual issues of rural health indicators is only part of the challenge. There are also practical and methodological problems that need to be dealt with. The ability to construct health indicators inevitably depends on the availability of data and the quality of such data. It is, therefore, necessary to assess the feasibility of developing rural health indicators. This will be accomplished in two stages:

(1)   Take an inventory of existing health indicators and the datasets used to derive these indicators.

(2)   Examine the extent to which it is feasible to assign the records in those datasets to "communities" or "regions" in order to employ the OECD and/or Statistics Canada definitions of "rural" to construct health indicators for rural communities in Canada.

To this end, a database of health indicators has been created (hereinafter referred to as the Rural Health Indicators Inventory Database). Through our analysis of relevant literature, we identified an extensive list of health indicators. We found a great deal of overlap in the health indicators being used. That is, many of the same health indicators appear over and over again as they are commonly used in Canada and throughout the world. The most obvious of these are measures of mortality and morbidity or the sociodemographic and socioeconomic indicators derived from the Census.

In fact, when we collapsed the extensive list of health indicators to reduce the redundancies, we found that the majority of the health indicators that are commonly employed in Canada are listed in four published sources. These four sources provide the focus for our construction of the Rural Health Indicators Inventory Database. In addition to simply listing the indicators, our Database gathers together information about the raw data from which these indicators were derived. By linking these two pieces of information, it is possible to see which dataset(s) has (have) been used to derive a specific health indicator.

The four publications that list the majority of health indicators used in Canada are as follows:

(1) *Report on the Health of Canadians* published by Health Canada

(2) *Statistics Canada Health Indicators Database* published by Statistics Canada

(3) *Community Health Indicators: Definitions and Interpretations* put out by the Canadian Institute for Health Information (CIHI)

(4) *The Institute of Clinical Evaluative Sciences Practice Atlas / (Second Edition)*.

In addition to the fact that these four publications list the major health indicators, each of the publishing agencies is a leader in health indicator development in Canada. The first three organizations deal with health indicators from a national perspective. The fourth was chosen as an example of an agency that focuses on provincial health indicators, but has a national profile. Each of the publications contains a wide range of health indicators that are based on current and available datasets. The diversity of health indicators and datasets provides a good opportunity to assess the feasibility of developing rural health indicators. It is worth noting that virtually all the indicators and datasets used to describe the health of Canadians in the June 7, 1999 issue of *Maclean's* magazine are described in the Rural Health Indicators Inventory Database.

3.2   The Four Sources of Health Indicators

Below is an outline of each of the four published sources from which the health indicators contained in the Rural Health Indicators Inventory Database were taken.

3.2.1   *Report on the Health of Canadians (Health Canada)*

This report was prepared by the Federal/Provincial/Territorial Advisory Committee on Population Health in 1996. The report contains 87 health indicators designed to provide a comprehensive overview of the state of health of Canadians and to also serve as an important national source of information for health planning and policy. The health indicators found in this report are based on a number of criteria:

(i)   They fit into the framework of the Evans-Stoddard model and begin to address some of the challenges laid out in the *Strategies for Population Health: Investing in the Health of Canadians* (ACPH, 1994);
(ii)   have national coverage;
(iii)   current data are available to derive them;
(iv)   allow for time series and international comparisons; and
(v)   can be broken down into various subgroups (e.g., age, sex, province).

All indicators listed in Appendix D of the *Report on the Health of Canadians*, which are classified into two main categories - determinants of health and health status - are included in the Rural Health Indicators Inventory Database. The *Technical Appendix to the Report on the Health of Canadians*, which was published under separate cover, identifies the data sources for each indicator and, wherever possible, the exact method of calculation. It should be noted that the technical appendix contains some health indicators that are not included in the *Report on the Health of Canadians*. Since these indicators are by and large variations of indicators found in the report, they were not included in the Rural Health Indicators Inventory Database.

3.2.2 Community Health Indicators: Definitions and Interpretations (Canadian Institute for Health Information)

This publication presents a set of 60 community health indicators. Many of the indicators in the CIHI publication are taken from an earlier publication titled *User's Guide to Community Health Indicators* published by Health Canada and the National Health Information Council in 1992. In turn, these are based on a set of health indicators developed in Quebec for the Community Health Department in 1988. The stated purpose of this publication is to develop a broad range of health indicators that can be used at both the national and sub-provincial level. The 60 indicators fall into the several major categories: Determinants of health, health status, and consequences of health problems. Each of the 60 indicators is described in some detail, including the method of calculation and the datasets used.

3.2.3 Health Indicator Database (Statistics Canada)

This electronic database contains 56 health indicators grouped into four categories: Health determinants, health status, health resources, and health care utilization. These indicators are designed as much as possible to reflect health rather than morbidity. The database uses the 20/20 Browser which, in most cases, allows the user to manipulate a specific health indicator in a number of ways (e.g., examine provincial or age/sex breakdowns). This database has been under development since 1986 and originally contained 125 indicators, but has since been pared down to the 56 now included. Some of the indicators found in the *Report on the Health of Canadians* relies on data contained in this Statistics Canada database.

3.2.4 Institute Clinical Evaluative Sciences (ICES) Practice and Electronic Atlas (Second Edition)

This publication was chosen to see whether health indicators developed at the provincial level could also be useful at a national level. The publication has 38 health indicators which fall into the categories of health determinants, health behaviours, and health status. A large part of the atlas is devoted to health resources and utilization, but these data are not presented in a way that makes them easily accessible. Thus, they are not included in the Rural Health Indicators Inventory Database.

The datasets used to derive the indicators were also inventoried and linked to the indicators. Each health indicator and its corresponding dataset(s) were then examined to see whether the dataset records could be assigned to a community and/or region. If it can be done, then the health indicator can be used to describe rural health.

3.3   Structure of the Rural Health Indicators Inventory Database

The Rural Health Indicators Inventory Database has two main components:

(1)   health indicator information; and
(2)   dataset information sheets.

Because of the large amount of information included, the Database has been stored on a CD-ROM. Information can be retrieved using *Microsoft Access*©. Instructions on how to use the Database itself is found in Appendix III.

3.3.1   Health Indicator Information

The following information is found in the Database for each health indicator.

- Indicator Name: The name given to each indicator as it appears in the source from which it was taken.

- Indicator Publishing Source: This identifies the source from which the indicator is taken (Statistics Canada, CIHI, *Report on the Health Canada*, or ICES)

- Indicator Classification: This field classifies the indicator into one of the five categories (health determinant, health behaviour, health status, health resources, health services utilization) of health indicators that have been described in Chapter 2. This is done to provide some consistency as each published source classifies its indicators in a slightly different way.

- Health Indicator Type: Indicators in the Database are classified into three types:

  (i)   Database indicator: an indicator can be taken directly from a dataset and does not require mathematical manipulation beyond calculating a simple frequency or percentage (e.g., number of smokers).
  (ii)   Derived variable: an indicator that must be calculated, requiring a numerator and a denominator source for its computations (e.g., infant mortality rate – number of infant deaths per live births).
  (iii)   Index variable: a complex indicator which combines several variables (some of which may be health indicators themselves) into a single index (e.g., Health Utilization Index).

- Numerator Sources: This field lists the dataset(s) from which the numerator of the health indicator is taken. There are three fields in the Database because it is possible that the numerator is from more than one source of data. Multiple numerator sources occur mainly with survey data. In some instances, a health indicator may have initially been established using a survey that is now discontinued, and has been replaced by another survey (e.g., the Health Promotion Survey being replaced by the National Population Health Survey). The older survey is still important as it provides the possibility for comparisons over time. The most recent data source is listed as "Numerator 1," the second most recent as "Numerator 2," and the third as "Numerator 3."   All indicators in the database will have at least one numerator source.

- Denominator Sources: This field is used for derived or index variables only. Database indicators do not require a denominator source, hence this field is blank. There are also three denominator fields in the Database as the denominator may be taken from more than one dataset. As with Numerator Sources, "Denominator 1" refers to the most recent source of data, "Denominator 2", the second most recent source.

- Description/Method of Calculation: This field provides a brief description of the health indicator and/or its method of calculation.

- Identifiable at Region Level?: This field is coded "Yes" or "No," depending on whether the indicator can be applied at the geographical level of a region (or CD). For those derived variables that are based on multiple datasets and if one of the sources cannot be used at a region level, then a "No" is shown. This assessment is based on the most recent sources of information (i.e., Numerator 1 and/or Denominator 1).

- Identifiable at Community Level?: This field is coded "Yes" or "No," depending on whether the indicator can be used to describe the health conditions at the geographical level of a community (CCS or groupings of CSDs). For those derived variables that are based on multiple datasets and if one of the sources cannot be used at a community level, then a "No" is shown. This assessment is based on the most recent sources of information (i.e., Numerator 1 and/or Denominator 1).

As outlined in Chapter 2 and Appendix II, region-level and community-level identification is key to assessing whether the health indicator can be applied or used to describe the health of Canadians living in rural areas.

3.3.2 Dataset Information Sheets

The second component in the Rural Health Indicators Inventory Database is a series of brief information sheets on the main datasets (e.g., the Census, the NPHS) from which the health indicators are derived. These dataset information sheets can be accessed by clicking on the Database tab (see Appendix III for more detailed instructions on how to use the Rural Health Indicators Inventory Database). When clicked, a brief one- or two-page description of the dataset will appear on the screen. There are over 25 datasets listed in the Rural Health Indicators Inventory Database. It should be noted at this point that datasets that contain no new data since 1990 have not been included. Information on each dataset was obtained from interviews with individuals knowledgeable about the dataset and from the relevant organization's website. Data and documentation available to the research team through the Data Liberation Initiative, as well as government publications, were also used. In no way is the information on each of the datasets meant to be exhaustive. Rather, the emphasis is placed on geographical and data release issues, the foci of the present study.

The following information is provided for each of the main dataset sources:

- Name of Dataset: Provides the official name of the data source and the organization that maintains it.

- Description: A brief description of the dataset, noting its objectives and the type of information collected.

- Start Date: The year data collection began. This information is important for those who are interested in historical comparisons.

- Frequency of Data Collection: Indicates whether data are collected annually, biannually, etc.

- Release Date: The approximate amount of time required for the data to be released after they are collected.

- Data Collection: Provides some brief information on how the data are collected (e.g., telephone survey, mailed survey, administrative data, etc.).

- Sample Size: Lists the approximate size of the sample contained in the database. This is provided for one cycle of data collection.

- Geographic Coverage: This section lists the provinces/territories from which data are collected. It also describes groups that are excluded from data collection (e.g., aboriginals, people in institutions).

- Lowest Geographical Level of Data Collection: Identified here is the lowest level of geography that is contained in the dataset. This does not mean that the data are releasable at that level. Rather, it is the starting point from which higher levels of geography can potentially be derived.

- Lowest Level of Geography for Data Release: This section identifies the geographic level at which data could potentially be released, subject of course to any data suppression or confidentiality criteria which are noted as well. Release criteria are not hard and fast and can depend on the variable in question. Given small sample sizes, or particularly sensitive issues, data may only be released at a higher geographic level than listed here. What is important for sub-provincial rural analysis is that the data be released minimally at the CD level, so that the information can be mapped or otherwise analyzed at least at the region level.

- Existing Rural Variable: Some datasets already contain a derived rural variable or a rural/urban flag. This section describes how the rural variable is defined.

- Data Elements: Gives a brief list of the kinds of data found in the dataset.

- Notes: This last section provides additional information that may be pertinent in the context of this study.

3.4   Summary Chart of Major Datasets

Table 3.1 lists, in summary fashion, the main datasets used to derive the various health indicators contained in the Rural Health Indicators Inventory Database. The chart breaks down the datasets by the organization that maintains them (e.g., Statistics Canada, CIHI, and Other Sources).

- Column 1: Gives the name of the dataset.

- Column 2: Lists the lowest geographic level at which the data are collected. This does not imply that the data are releasable at this level. Rather this is the starting point from which the data can be aggregated.

- Column 3: Describes the geographic level at which the data could potentially be released. As discussed above, the geographic level of data release can vary depending on a number of factors.

- Column 4: Indicates whether records in the dataset can be identified at the level of regions (i.e., CDs). If so, rural and non-rural assessments may be made at that general level of aggregation. However, although the records have the potential of being identified at the region level, there may be restrictions (e.g., confidentiality, statistical significance criteria) that would not permit analysis at this level of aggregation.

- Column 5: Indicates whether records in the database can be identified at the level of communities (i.e., CCSs and/or groupings of CSDs). If so, rural and non-rural assessments may be made at a more detailed level of aggregation. However, although the records have the potential of being identified at the community level, there may be restrictions (e.g., confidentiality, statistical significance criteria) that would not permit analysis at this level of aggregation.

- Column 6: Indicates whether the dataset already provides a rural/urban flag. This rural/urban designation may or may not be compatible with the identification of regions and communities (Columns 4 and 5).

- Column 7: Gives additional information regarding the applicability of the dataset to the present study.

Table 3.1: **Summary Chart of Major Datasets**

Statistics Canada						
Dataset	**Geographic Level of Data Collection**	**Geographic Level of Data Release**	**Identifiable at Region Level?**	**Identifiable at Community Level?**	**Existing Rural / Urban Flag?**	**Notes**
Annual Demographic Statistics (Post/Inter-Censal Estimates)	CSD	CSD	Yes	Yes	No	
Census of Canada	Enumeration Area	Enumeration Area	Yes	Yes	Yes	Suppression of cells with observations less than 40. For economic data, suppression when less than 250 observations.
General Social Survey (GSS)	Postal Code	Province	Yes	Yes	No	Unknown how stable estimates are at region or community levels.
Health Promotion Survey	Postal Code	Province	Yes	Yes	No	Survey is discontinued, but its data are still used. Unlikely to be releasable at the region or community level because of small sample sizes for those geographical units.
Labour Force Survey	Households within cluster sampling units	Economic Region (aggregated CD's)	No	No	No	Data also available for Census Metropolitan Area and Employment Insurance Regions. Because Economic Regions are aggregated CDs there is a potential for rural/urban designation using schemes not discussed in this study.
National Cancer Incidence Database	Postal code for patient, census tract for cancer type	CSD	Yes	Yes	No	Cells with less than 3 observations are suppressed.

Table 3.1: (continued) Summary Chart of Major Datasets

Statistics Canada						
National Population Health Survey (NPHS)	Postal Code	Province	Yes	Yes	Yes	Under normal circumstances data are not released at region or community level, especially the latter. Further study needed to investigate how stable estimates are at those geographical levels.
Survey of Residential Care Facilities	Facility	CSD	Yes	Yes	No	Suppression of cells with less than 3 facilities and where 25% of facilities making up 75% or more of total counts in a cell.
Vital Statistics (Births and Deaths)	City/Town, etc. of occurrence	CSD	Yes	Yes	No	Theoretically ideal for rural and non-rural community level analyses. However, we have been advised of possible coding biases or errors which makes use of data at region level (CD) advisable. CMAs/CAs are available. Cells with less than 3 observations are suppressed.

Table 3.1: (continued) **Summary Chart of Major Datasets**

Canadian Institute of Health Information Datasets						
Dataset	**Geographic Level of Data Collection**	**Geographic Level of Data Release**	**Identifiable at Region Level?**	**Identifiable at Community Level?**	**Existing Rural / Urban Flag?**	**Notes**
Annual Hospital Survey	Facility	Subprovincial [1]	Yes	Yes	No	Currently undergoing extensive recoding.
Discharge Abstract Database (DAD)	Postal Code	Subprovincial [1]	Yes	Yes	No	Potential region and community level analysis BUT data only available for 85% of hospitals in Canada.
Hospital Mental Health Database	Postal Code	Subprovincial [1]	Yes	Yes	No	HMD (see below) should be used in conjunction to get 100% coverage.
Hospital Morbidity Database (HMD)	Postal Code	Subprovincial [1]	Yes	Yes	No	Again, region and community level analyses may only be theoretically possible. Collection and completeness of postal code varies (e.g., Quebec first 3 digits only, thus cannot be used with either rural definition). Does not include ambulatory care and same-day surgeries.
National Health Expenditures Database	Province	Province	No	No	No	
National Physicians Database	Postal Code	Subprovincial [1]	Yes	Yes	No	All data are subject to a provincially imposed access policy and the NPDB Transfer Plan
Southam Medical Database	Postal Code	Subprovincial [1]	Yes	Yes	No	
Therapeutic Abortion Database	Province	Province	No	No	No	

---

[1] All data release is subject to CIHI's Privacy and Confidentiality Guidelines. A non-disclosure agreement must also be signed.

Table 3.1: (continued) **Summary Chart of Major Statistics Canada Datasets**

Other Datasets						
Dataset	**Geographic Level of Data Collection**	**Geographic Level of Data Release**	**Identifiable at Region Level?**	**Identifiable at Community Level?**	**Existing Rural / Urban Flag?**	**Notes**
Centre for Justice Statistics Database	3 digit police detachment code	Province	No	No	No	
National Work Injuries Program (NWISP)	Province	Province	No	No	No	Provincial data must be obtained to conduct subprovincial analyses.
Transport Canada Accident Database (TRAID)	Accident location	Province	No	No	Yes	Data must be obtained from provincial ministries in order to conduct subprovincial analyses.
Laboratory Centre for Disease Control (LCDC) Notifiable Disease Database	Province	Province	No	No	No	Provincial data must be obtained to conduct subprovincial analyses.

3.5   Issues Pertaining to the Rural Health Indicators Inventory Database

There are several observations and problems in relation to the Rural Health Indicators Inventory Database:

1)   There are 265 health indicators found in the Database. This does not include the potentially hundreds of mortality/morbidity indicators that could be associated with specific diseases/injuries (e.g. associated with International Classification of Diseases codes) or the potentially thousands of indicators that might be associated with the sociodemographic and socioeconomic measures available from the Census of Population. Of the 265, not all are unique. There are overlaps, because the same indicators are used by different organizations. However, caution should be taken when comparing two seemingly identical indicators as the definition or calculation may be slightly different and, in some cases, it may be difficult to tell if they are the same or not.

In terms of types of indicators contained in the Database, 44% of the measures are classified as health status indicators. This is not surprising in view of the fact that data relating to health status are most readily available and, as noted earlier, that many people take health indicators to mean health status indicators. The second largest category is health determinant indicators (31%). This reflects the emphasis that each of the four sources places on population health issues. Also, the determinants of health are wide ranging and thus become almost a catch-all category. Yet, there are no health determinant indicators in relation to the physical environment (e.g., air, water, and soil pollution). Health resources indicators (9%), health behaviour indicators (9%), and health services utilization indicators (7%) make up the rest.

2)   A number of problems were encountered when reviewing the four major sources of health indicators. First, it is difficult in some cases to determine how the health indicators are calculated. This is especially true with more complex indicators such as Standardized Mortality Ratios (SMR), Potential Years of Life Lost (PYLL), and Life Expectancy. For example, in the calculation of SMRs, the "standard population" is often not identified. In future, definitions and calculations should be stated explicitly and in a standardized manner. This will ensure replicability and comparability. In some cases, the datasets need to be more clearly described. For example, in the CIHI publication, *Community Health Indicators: Definitions and Interpretations,* only sources of data from Quebec are provided for some indicators. Where possible, we have chosen the equivalent national datasets and matched them with the indicators in the CIHI publication. As well, many of the national datasets used to derive the health indicators in the CIHI database are out of date (e.g., Canada Health Survey conducted in 1978). In future editions of this publication, emphasis needs to be placed on providing more up-to-date national datasets.

3)   Most of the health indicators in the Rural Health Indicators Inventory Database are derived from four sources of data: Vital Statistics, the Census, the National Population Health Survey, and the Laboratory Centre for Disease Control. The most often used dataset from CIHI is the Hospital Morbidity Database. With the exception of the information from the Laboratory Centre for Disease Control, all of the datasets have the potential for assigning

records to regions and/or communities and, therefore, are amenable to rural/non-rural analyses.

4) The survey-based datasets (e.g., the National Population Health Survey) are generally collected at a low enough level of geography that would allow for, through custom tabulations, record identification at region and/or community levels. However, given the relatively small sample sizes of most of the surveys, subprovincial estimates would be difficult. On the other hand, rural and non-rural summaries at a provincial level would be possible.

5) Other than the Statistics Canada urban/rural EA codes for the Census, only two datasets have existing rural/urban flags: The National Population Health Survey and the Transport Canada Accident Injury Database (TRAID). A health indicator constructed from each of these two datasets using the existing rural/urban flag has been used as examples and will be discussed in greater detail in Chapter 4. Suffice it to say here that the two definitions of rural contained in these two datasets are not consistent with each other. The TRAID definition appears to have very limited utility and cannot be easily compared to the rural/urban designations used in the National Population Health Survey. The existing rural/urban flag on the National Population Health Survey is not available for all provinces on the public release file and, thus, national data must be obtained through custom tabulations.