Overview of the Landscape of Rural Health Research with Secondary Data Sources
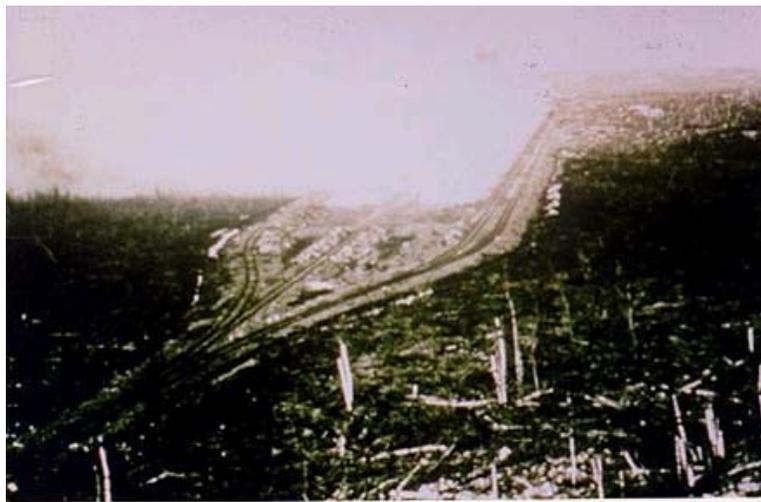
Denis Heng

Centre for Rural and Northern Health Research, Laurentian University

Marie Desmeules

Centre for Chronic Disease Prevention and Control, Health Canada

Rural Health: Celebrating Diversity & Strength
Doing Rural Health Research Using Secondary Data Sources
October 19-23, 2003

Roast yards – 1920's

Roast yards - 1989

Background & Introduction

- Research dependent on availability of secondary data

- Important for background material and setting research in a wider context
  - Effectively used to enhance primary data collection

Rural should not be ignored

- Two tier health:

Rural vs Urban?

- Much research has described the health gap
  - Kirby Panel report (2002)
  - Romanow Commission report (2003)
  - Ministerial Advisory Council on Rural Health report (2002)

Uses of Secondary Data Sets

- Monitoring the health of the population
- Disease surveillance, estimating incidence and prevalence
- Determining the effectiveness of treatment
- Quantifying prognosis and survival
- Monitoring the adequacy of care

- Risk factor / health determinants surveillance to improve primary prevention strategies
- Needs assessments to support administrative functions
- Assessing the usefulness of preventive strategies, diagnostic tests and screening programs
- Informing health policy through studies on cost-effectiveness

"Essential Apparatus" (Wolfson 1996)

- Large-scale, population level databases
  - Canadian Institute for Advanced Research Program in Population Health, Working Paper Series
- Numerous sources of secondary data
  - Some is relevant to our research, most is not
  - Finding suitable data often requires a considerable amount of "detective work"

Data Sources: Defined

- Primary data
  - Data that you or your colleagues collect specifically for the purpose of answering your research question.

- Secondary data
  - Data that has been collected for another purpose, which is employed to answer a different research question.
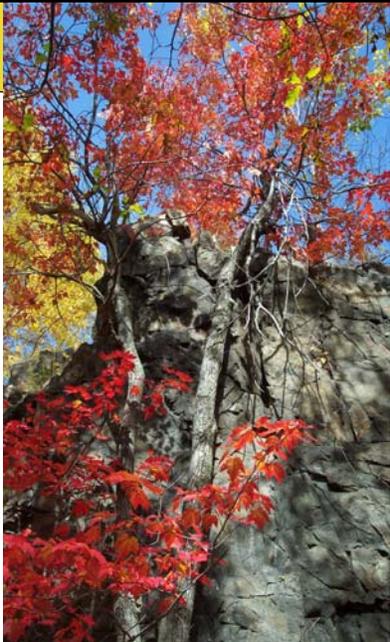
- **Epidemiologic data**
  - Census
  - Vital statistics
  - Health surveys
  - Social surveys
  - Registries

- **Administrative data**
  - Provincial health insurance plans
  - Health Human Resources
  - Drug benefit plans

Advantages of Secondary Data

1. Resource implications – usually less expensive and quicker to obtain compared to primary data collection.
2. If you are looking for a small effect size, it may be impossible to collect primary data on a sufficient number of cases.
3. Collecting primary data prospectively may be unethical if a therapy is "standard of care."
4. The accessible population for primary data may be less representative of the target population than that for secondary data.
5. You may not need to worry about informed consent, human subjects restrictions, etc.

Limitations of Secondary Data

1. Secondary data is not gathered for purposes of rural-specific research.
   - Data collected by other people

2. May only partially address the research questions.
   - Breadth of data is limited in some areas
   - It is impossible to collect all the factors that may be associated with your outcome of interest

Limitations of Secondary Data

3. Data definitions, categories, levels of aggregation may differ from those required.
   - E.g. Rural flags in different secondary data sources

   - Many of the data utilized are only indirect measures of concerns that might exist in the community.

Rural: Defined

- Challenge: defining the concept of 'rural' and operationalizing the term in a manner appropriate for research

- Can include features such as size, population density and location, economic, social, ecological attributes

Some of the more interesting rural flags

- Transport Canada Accident Injury Database (TRIAD)

    - Rural accidents are those that occur on primary or secondary highways, as well as local roads, or where the speed limit at the collision site exceeds 60km/h.

- National Population Health Survey

    - Rural flag is only available for 7 of the 10 provinces

What is Rural?

Definition of Rural	Building blocks
Census "Rural Areas"	Enumeration area (EA)
Rural and Small Town MIZ	Census subdivision (CSD)
OECD "rural" communities	Census consolidated subdivision (CCS)
OECD "predominantly rural regions"	Census division (CD)
Ehrensaft's "Beale codes"	Census division (CD)

Adapted from du Plessis, V et al. (2002). *Definitions of "Rural"*. Agriculture and Rural Working Paper Series, Working Paper No. 61. Ottawa: Statistics Canada, Agricultural Division.

Limitations of Secondary Data

4. Cost and access – may still be difficult in spite of resource advantages.

5. Some of the available data may be relatively dated.
    - Body Mass Index (BMI) – until recently

6. Validity of data may vary, depending on source.
    - Undersampling of rural areas
    - Rural = Small = Data suppression

Other issues of Secondary Data

Caution must be exercised when interpreting data

- Inferring causality is problematic

Integration of secondary data with each other and with primary data also an issue.

- Extreme caution must be exercised when merging data from different sources in order to make sure that they are measuring the same thing.

e.g. Manitoba Population Health Research Data Repository

Dalhousie University – Population Health Research Unit
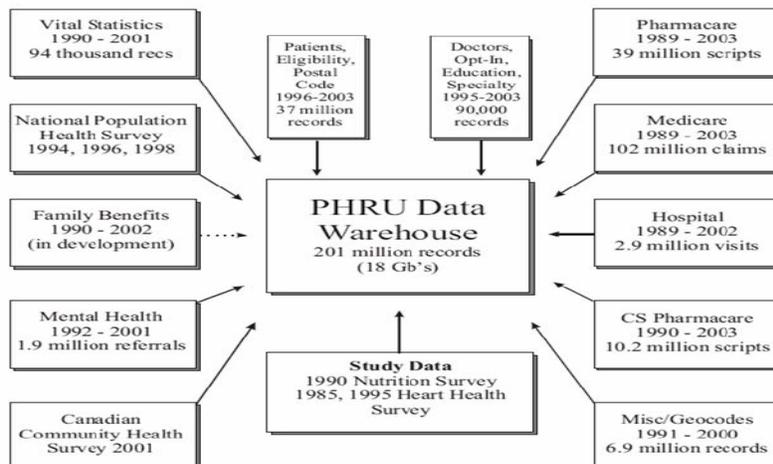
Basic Regreening Recipe

Region of Sudbury Land Reclamation

- 10 tonnes per hectare of crushed dolomitic limestone
- 400 kg/ha of 6-24-24 (NPK) fertilizer
- 40 kg/ha of seed mixture
  - 75% grasses
  - 25% legumes
- 6 million trees planted

Doing Rural Health with Secondary Data Sources

- ???

Examples of Rural Health projects using secondary data sources

- Determinants of Health of Rural Populations and Communities
  - Rural Development Institute, Brandon University
- Rural and Small Town Analysis Bulletin
  - Statistics Canada
- Canada's Rural Communities
  - Canadian Population Health Initiative

Canada's Rural Communities (Overview)

❖ Multi-year collaborative research program

❖ Involves researchers from CRaNHR, Health Canada, CIHI, Statistics Canada, ICES, Dalhousie University, and UBC.

❖ Five major components:

   ❖ Conceptual and methodological considerations

   ❖ Health status of rural and urban Canadians

   ❖ Determinants of rural health

   ❖ Health services utilization

   ❖ Feasibility study for international comparison

Inventory of Rural Health Research-Related Databases

• Goal: Build capacity within the field of rural health research

Compiled from …
• Pitblado et al. (1999) *Assessing Rural Health: Towards Developing Health Indicators for Rural Canada*

• Relevant organization's websites

• Government publications and data documentation

• Feedback from individuals knowledgeable about the data sources

Inventory Details

- Inclusion criteria:
    - Readily available national data
    - Rural and non-rural geographies
    - Data relevant to a variety of rural health issues

- Appropriate geographical levels for rural health research
    - Collected, Analyzable, Released

Inventory of Rural Health Research Related Databases	
Database name:	**Discharge Abstract Database**
Abstract: Maintained by: Availability:	This database collects data on hospital discharges. These data are also used to evaluate patient expected length of stay and resource consumption. In addition to inpatient hospitalizations, some jurisdictions / facilities provide information on same day surgery discharges. CIHI Data disclosure is determined by CIHI's Privacy and Confidentiality Policy (see www.cihi.ca). If privacy and confidentiality criteria are met, data can be released at the sub-provincial level. Costs are dependent upon the level data required, plus programming and processing fees.
Start Date:	1963 1979/80 – Present form
Release Date:	Approximately 5 months after reference period for general release
Frequency of collection:	On-going
Data Collection:	CIHI receives data directly from participating hospitals.
Sample size:	About 4.3 million records per annum
Geographic coverage:	Coverage has been increasing over time. Currently, about 85% of hospital inpatient discharges in Canada are included, consisting of all hospitals in every province and territory except Quebec and parts of Manitoba (Manitoba only submits 40% of their data) for both acute/inpatient data. It also contains Day Surgery Data for all provinces except Quebec, Alberta and parts of Manitoba. If comprehensive geographic coverage is required, the Hospital Morbidity Database may be used instead for national reporting.

Lowest geographic level collected:	Postal Code of patient. Collection and completeness of postal codes varies (e.g. Quebec reports only the Forward Sortation Area – the first 3 digits of the postal code)
Lowest geographic level of release:	Forward Sortation Area or Enumeration Area depending on request
Existing rural variable:	No

Data Elements:

Selected data elements include:
• most responsible physician/diagnosis
• principal procedure
• patient gender, date of birth, postal code
• institution/hospital number.
• admission/discharge, (e.g. admission category, length of stay).

Notes:

The advantages of the Discharge Abstract Database include its timeliness, value-added data elements (e.g. Case Mix Groups and Resource Intensity Weights), and coverage of same day surgery for selected jurisdictions/facilities.

Data are reported based on the region of the patient's residence, not region of hospitalization. Consequently, these figures reflect the hospitalization experience of residents of the region wherever they are treated as opposed to the comprehensive activity of the region's hospitals. It represents how frequently residents of a given area received hospital care rather than the volume of services provided by hospitals.

For more information, contact dad@cihi.ca

Conclusion

- It takes time for change to occur, but it is possible.

Overview of the Landscape of
Rural Health Research with
Secondary Data Sources

Marie Desmeules
Centre for Chronic Disease Prevention and Control, Health Canada

Denis Heng
Centre for Rural and Northern Health Research, Laurentian University

Rural Health: Celebrating Diversity & Strength
Doing Rural Health Research Using Secondary Data
Sources
October 19-23, 2003

Important themes/subjects

Health Status	Non-medical determinants	Health Services and resources
Mortality (all ages, Infant, adult, all cause, cause specific)	Social Support	Use and Accessibility to health care facilities and practitioners
Mental Health (stress, depression, etc)	Health-related behaviours	Satisfaction with health services/ health system
Self – Rated Health	Education, Income, Unemployment	Immunization and Preventative care
Chronic conditions	Physical activity	Distance to health care facility, transportation
Etc.	Nutrition	Etc.
	Etc.	

Examples of rural specific analysis using secondary data

Component 2: Health status of rural and urban Canadians

- Provides a systematic and comprehensive assessment of the health status of rural Canadians, generating a framework within which health determinants and potential interventions can be identified.

## Example #1: The Canadian Annual Mortality Data

- Used to assess the mortality patterns of rural populations and to compare them to their urban counterparts
- Records were aggregated at the CSD level (based on the 1996 Census boundaries) for the period of 1986-1996

Defining rural

- Each CSD was then assigned a Metropolitan Influence Zone (MIZ) category according to the degree of influence of a larger urban centre (a Census Metropolitan Area (CMA) or a Census Agglomeration (CA)).

- The four MIZ are as follows: Strong MIZ, Moderate MIZ, Weak MIZ and No MIZ.

Chronic Disease Mortality

Age-standardized mortality rates (ASMR) by cause and MIZ categories, people aged 20-44 years old, Canada 1986-1996

Legend: ■ CMA/CA ☐ Strong MIZ ■ Moderate MIZ ■ Weak MIZ ☐ No MIZ
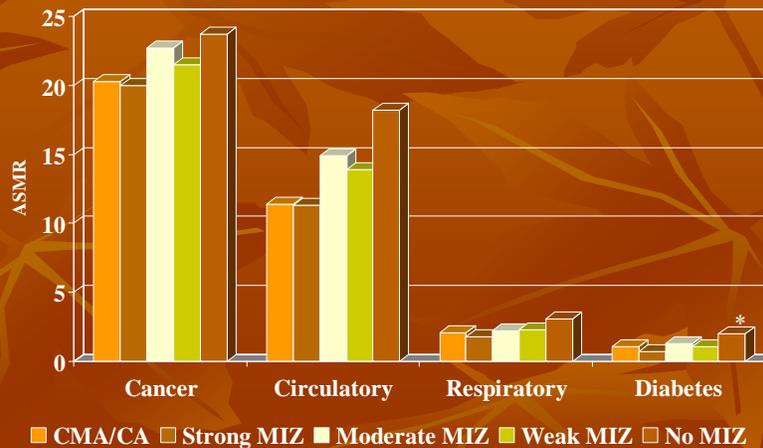
Challenges

- Merging the MIZ file with the Mortality file
- CSD boundaries change over time
- Combined at least 5 years of data together in order to get stable rates by age group, causes and MIZ categories
- Potential misclassification of deaths into incorrect areas of residence
- Only the underlying cause of death was available to us

Example #2: Regression analysis

- MIZ definition was used
- Merged mortality and 1996 Census data
  - To examine the association between all-cause mortality and selected socio-economic and demographic determinants of health

Selected determinants of health

- Income, Education
- Ethnicity, Age, Gender
- Migration
- Marital status
- Private dwellings in need of major repairs,
- % of people with an occupation in the primary industry
- unemployment rate
- % of people with a health-related occupation

Adjusted relative risk (RR) estimates for the association between place of residence and all-cause mortality, people aged 45-64 years, Canada, 1986-1996

Parameters	RR (95% CI)
CMA/CA	1.00
Strong MIZ	1.16 (1.04-1.29)*
Moderate MIZ	1.19 (1.08-1.31)*
Weak MIZ	1.12 (1.01-1.24)*
No MIZ	1.25 (1.12-1.41)*
% with completed secondary in CSD	
75-100%	1.00
50-74%	1.66 (1.37-2.00)*
25-49%	1.95 (1.54-2.48)*
0-24%	1.43 (1.19-1.72)*
Median Household Income	
≥$40,000	1.00
$20,000-<$40,000	1.12 (1.03-1.21)*
<$20,000	1.50 (1.34-1.66)*
% of Aboriginal in the CSD	
<10%	1.00
≥10%	1.28 (1.21-1/35)*
Mobility status (% movers in or out of the CSD)	
<30%	1.00
≥30%	1.54 (1.39-1.69)*

Challenges

Challenges similar to the first example were encountered for this analysis. Additional challenges were:

- Census data availability by CSD: only aggregated data by CSD
- Data suppression for income values in CSD with a population less than 250

Component #3
Determinants of rural health

❖ **Research Objective**: To identify key determinants of health can explain why some rural communities are healthy and others not.
❖ **Research Questions**:
  ❖ To what extent is residence in a rural or remote location a determinant of health above and beyond other known determinants of health?
  ❖ How do other individual and community determinants of health impact upon the rural population as compared to the urban population?

Example #3 Smoking, Obesity and Physical Activity among Rural Canadians

- Data Source: Canadian Community Health Survey 2000/2001
- Bivariate analyses: age-adjusted prevalence by MIZ categories
- Multivariate analyses- examples:
  - Smoking (or physical activity, or obesity) as dependent variable, and MIZ categories, SES variables, etc. as independent variables
  - Self-rated health as dependent variable, and MIZ categories, smoking, physical activity, obesity, SES, etc… as independent variables.

Component 4: Health Services Utilization

❖ **Research Objective**: to describe the patterns of health services use among urban and rural communities in 3 provinces:

  ❖ Nova Scotia
  ❖ Ontario
  ❖ British Columbia

Research Questions

❖ What are the patterns of overall "traditional" health services utilization in rural Canada as compared to urban Canada?
  ❖ defined by physician visits and hospitalization
  ❖ patterns by subgroups or types of rural communities

❖ What is the distribution of health human resources by place of residence?
  ❖ number of physicians and nurses

❖ Has the magnitude of postulated urban/rural difference in utilization rates changed over time?

Data Sources

- ❖ Provincial Physician Claims files
  - ❖ (NS, BC, ON)

- ❖ Hospital Morbidity Database (CIHI)

- ❖ Health Services Survey (subsample of CCHS)

Example #4: Physician Claims and Hospitalization Data Analysis

- ❖ 1996 census boundaries
  - ❖ PCCF – November 2000

- ❖ Rates of physician visits and hospitalizations during most recent 6-year period
  - ❖ 1995/1996 to 2000/2001
  - ❖ Mean bi-annual rates of visits per population
  - ❖ Where feasible, a person-oriented approach will be used (number of individuals who had one or more visits in a time period).

Indicators (Physician Claims)

- Ambulatory visit rates
  - Overall, and by diagnosis
  - By procedure

- Ambulatory visit rates
  - All physicians
  - GP vs specialists

- Stratified by
- MIZ category
- Gender
- Age

Indicators (Hospitalization data)

- Separation rates
- Length of stay
- Hospitalization rates by diagnosis
- Hospitalization rates by selected procedures

Stratified by
- MIZ category
- Gender
- Age

Discussion

- Enormous amount of information, datasets, data fields available.
- Identify focused research questions that are relevant for policy/program development, eg where potential for intervention at the following levels:
    - Primary prevention (risk factors, determinants of health)
    - Enhanced/ more timely access to health services
    - Etc.