RDC Data Alternatives: Conducting Research During COVID-19

Kristi Thompson  
*Western University*, kthom67@uwo.ca

Elizabeth Hill  
*Western University*, ethill@uwo.ca

Follow this and additional works at: https://ir.lib.uwo.ca/wlpres

Part of the Design of Experiments and Sample Surveys Commons, Library and Information Science Commons, and the Longitudinal Data Analysis and Time Series Commons

Citation of this paper:  
https://ir.lib.uwo.ca/wlpres/73
RDC Data Alternatives: Conducting Research During COVID-19

Elizabeth Hill & Kristi Thompson
April 29, 2020
Who we are

Elizabeth Hill
Data Librarian

Kristi Thompson
Research Data Management Librarian

Members of the Research and Scholarly Communication (RSC) team
rsclib@uwo.ca
Access to Data

1. Overview of access levels to Statistics Canada data
2. What type of access is available at Western?
3. Statistics Canada Continuum of Data Access
   1. Open
   2. Data Liberation Initiative
   3. RTRA
   4. RDC
4. Alternatives to Statistics Canada: Survey by Survey Recommendations
Access to Statistics Canada Data

- Web site http://statcan.gc.ca
- Data extractors (Data Liberation Initiative – DLI)
  - Odesi.ca
  - Statistics Canada Nesstar –
- Real Time Remote Access
- Research Data Centre (Currently closed due to Covid-19)
Access to Statistics Canada Data

The Continuum of Data Access Model helps users identify the various levels of data access at Statistics Canada, of which the DLI is one avenue. The model uses cost, the conditions attached to the terms of use, and the content of the statistical information as it applies to the various dissemination channels.
Continuum of Microdata Access

Access to microdata

Statistics Canada recognizes that researchers require access to microdata at the individual business, household or person level for research purposes. To preserve the privacy and confidentiality of respondents, and to encourage the use of microdata, Statistics Canada offers a wide range of options through a series of online channels, facilities and programs.
Categories within the Continuum

• Unrestricted access to microdata (on campus)
  • Public Use Microdata Files Collection
  • The Data Liberation Initiative

• Remote access to microdata
  • Real time Remote Access (RTRA)

• Restricted access to microdata
  • Research Data Centre (RDC) Programs
Access at Western

• Open data (statcan.gc.ca)

• Data Liberation Initiative (DLI)
  • Academic Data License (Post-secondary, voluntary subscription, most Universities, some College members)

• Real Time Remote Access (RTRA)
  • allows for remote submission of SAS code in order to produce tables based on microdata. Output is vetted by the RTRA system and provided to the researcher.

• Research Data Centre (RDC)
  • provide researchers with access, in a secure university setting, to microdata from population and household surveys, administrative data holdings and linked data
Access

• Open – http://statcan.gc.ca
Access

• Open – http://statcan.gc.ca
Access

• Browse by topic displays a list of available products.

• Survey of Statistical program can provide an insight as to what survey the product was derived from.
Access

• odesi – access to Aggregate data
A data portal for researchers, teachers and students; inspiring, developing and supporting research excellence.

Un portail de données à l'intention des chercheurs, professeurs et étudiants, qui vise à inspirer, à développer et à appuyer l'excellence en matière de recherche.

Note: In many cases, it is advisable to weight analysis results before reporting them. Correct weighting requires careful consideration, please always consult the weighting procedures of the study before applying the weights. To apply weights, select the Weight icon and choose the weight variable to be used. All results need careful interpretation. The data collectors and the data producers bear no responsibility for the analysis and interpretation of the data.

Note: Dans la plupart des cas, il est recommandé de pondérer les résultats d'analyse avant d'en rendre compte. Une pondération correcte nécessite une attention particulière, veuillez toujours consulter les méthodes de pondération d'une étude avant d'appliquer des pondérations. Tous les résultats nécessitent une interprétation attentive. Les collecteurs de données et les producteurs de données ne portent aucune responsabilité pour l'analyse et l'interprétation des données.

Hierarchical File

Abstract

This Hierarchical File, 2016 Census Public Use Microdata File (PUMF) product provides access to non-aggregated data covering a sample of 1% of the Canadian households. It is a comprehensive social, demographic and economic database about Canada and its people, and contains a wealth of characteristics on the population. The file enables the study of individuals in relation to their census families, economic families and households. Geographic identifiers have been restricted to the provinces, the three territories grouped into a region called Northern Canada and selected metropolitan areas (Toronto, Montreal, Vancouver, Edmonton and Calgary) to ensure respondents’ anonymity. This comprehensive file is excellent tool for policy analysts, pollsters, social researchers and anyone interested in modeling and performing statistical regression analysis using 2016 Census microdata.

Hierarchical File

Variable MarStH: Marital status (de facto)

DESCRIPTIVE TEXT
Marital status (de facto).

Values Categories
1 Never legally married (and not living common law) 137770 13,828,100.3 40.1%
2 Legally married (and not separated) 132051 13,255,537.6 38.5%
3 Living common law 34794 3,490,683.9 10.1%
4 Separated, divorced or widowed (and not living common law) 38166 3,830,595.0 11.1%
8 Not available 549 55,147.2 0.2%

SUMMARY STATISTICS
Valid cases 343330 3.4460064E7
Missing cases 0 0.0
This variable is numeric
Welcome to the NESSTAR data portal, a web-based exploration, extraction and analysis tool for social science data. The NESSTAR data portal consists of several catalogues of Statistics Canada data:

Statistics Canada Public Use Microdata Files (PUMF)
- English catalogue of public use microdata files, which contain anonymized, non-aggregated data. Users can download the data file and associated metadata.
- French catalogue of public use microdata files, which contain anonymized, non-aggregated data. Users can download the data file and associated metadata.

Statistics Canada metadata for Master Files (RDC)
- English catalogue of detailed microdata files, which contains most of the original information collected during the survey interview. Users can only browse or search the survey metadata.
- French catalogue of detailed microdata files, which contains most of the original information collected during the survey interview. Users can only browse or search the survey metadata.

The metadata in this tool does not represent the complete collection of PUMFs and RDC master files. For more information regarding the difference between a PUMF and master file, see About Microdata. We are continuously updating and improving our collections. Please contact us if you encounter any issues.
Nesstar at Statistics Canada
Nesstar at Statistics Canada

Dataset: Absence from Work Survey, 1999 [Canada]

**Variable Q12: LEAVE JOB/ABSENT FOR 2+ CONSECUTIVE WKS?**

**LITERAL QUESTION**
In 1998, did respondent leave a job or was absent from work for 2 or more consecutive weeks due to illness, accident or pregnancy?

<table>
<thead>
<tr>
<th>Values</th>
<th>Categories</th>
<th>N</th>
<th>NW</th>
<th>NW %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Yes</td>
<td>1324</td>
<td>1,020,501.6</td>
<td>7.2%</td>
</tr>
<tr>
<td>2</td>
<td>No</td>
<td>16339</td>
<td>13,135,447.2</td>
<td>92.8%</td>
</tr>
</tbody>
</table>

**SUMMARY STATISTICS**
Valid cases: 17663, 1.415594877287
Missing cases: 0, 0.0

This variable is alphanumeric.
Real Time Remote Access
Real Time Remote Access

The RTRA system is an online remote access facility allowing users to run SAS programs, in real-time, against microdata located in a central and secure location. A full range of descriptive statistics are available through the Real Time Remote Access tool.

Western Libraries has subscribed to RTRA, and has a 2-user license.

Due to Covid-19 restrictions and the closure of Western’s campus, and in-turn the RDC, Statistics Canada is providing access to RTRA for approved RDC researchers.

If you are a RDC Researcher, access to RTRA will be available through the RDC.
Real Time Remote Access

› Access from any computer with Internet access using a secure username and password
› Submissions made in the form of SAS programs using an EFT site
› Confidentiality is automated
Real Time Remote Access

Fast access to a pool of data sets

1 week for account creation

Deemed employee status not required
Access to data sets aren’t restricted to approved projects

Vetting of outputs is fully automated

Accessible at any time through the internet.

Full range of descriptive statistics available through the RTRA.
Real Time Remote Access

1. Rapid access:
   › 24/7 access to documentation, dummy data, RTRA accounts
   › RTRA results within an hour

2. Expands research data options
   › e.g., continuous variables

3. All points in the research lifecycle
   › e.g., strengthens research proposals and grant applications, and data - light end users

4. Building end-user statistical analysis skills
   › Supported by researcher-centred RTRA support and training
   › Basic level statistical analysis experience required
Real Time Remote Access

Notes from the User Guide (https://www.statcan.gc.ca/eng/rtra)

• Your SAS program has to be written in a standard format in order to automate the confidentiality process within RTRA

• Frequencies are rounded, this is explained within the User Guide

• Some variables are renamed, others are deleted (if disclosure risk is an issue)

• 3 types of Statistics are calculated within RTRA
  • Basic statistics, including frequency, mean, percentiles, percent distribution, proportions, ratio and share
  • Higher order statistics these statistics calculate differences between the basic statistics available in the RTRA system
  • Sequential over time statistics, compare the value of every cell with the value of the cell directly below it in the table in a sequential manner over time
Real Time Remote Access

Using RTRA

• Certain SAS commands are not allowed through remote access
• There are no limit to the number of submissions that you can perform during your RTRA contract, but you are limited to 10 successful submissions per day
• A maximum of four or five variables can be specified in the class variable list, depending on the RTRA procedure macro.
• Class variables cannot contain any missing values.
Real Time Remote Access – Available Data

Over 200 Survey Cycles

For a full list see: http://www.statcan.gc.ca/eng/rtra/surv

Surveys of Particular interest include:

• Labour Force Survey (monthly additions)
• Canadian Survey on Disability
• Aboriginal Peoples Survey
• Participation and Activity Limitation Survey
• Canadian Community Health Survey
• Post-secondary Student Information System
Real Time Remote Access

- Review of RTRA surveys not available as public microdata
- Canadian Cancer Registry
- Homicide Survey
- Postsecondary Student Information System
- Registered Apprentice System
- Uniform Crime Reports
- Canadian Mental Health Survey
- Health Services Access Survey
- Longitudinal Survey of Immigrants to Canada
- Maternity Experiences Survey
- National Apprenticeship Survey
- Post-Secondary Education Participation Survey
- Survey of Young Canadians
- Survey on Living with Chronic Diseases in Canada (SLCDC): Arthritis
Real Time Remote Access – Administrative Data

- Uniform Crime Reporting
- Canadian Cancer Registry
- Registered Apprenticeship Information System
- Post-secondary Student Information System
- Homicide Data
Real Time Remote Access

1. Rapid access:
   › Quick account application via RTRA contact to Statistics Canada, or through RDC
   › 24/7 access to documentation, dummy data, RTRA accounts
   › RTRA results within an hour

2. Expands research data options
   › e.g., continuous variables

3. All points in the research lifecycle
   › e.g., strengthens research proposals and grant applications, and data - light end users

4. Building end-user statistical analysis skills
   › Supported by researcher-centred RTRA support and training
   › Basic level statistical analysis experience required
Real Time Remote Access

General disclosure rules & controls applied to all outputs:

• Smallest level of geography is provincial
• Only weighted data are released
• All logs are scanned and modified for protection
• Maximum number of successful submissions per day: 10
• Maximum number of tables per request: 10
• No more than 500 categories can be produced for a given variable
• The RTRA System applies Additive and Controlled Rounding to all frequencies. The rounding base is survey-driven

  (e.g. CCHS 2016 rounding base is 200; Canadian Financial Capability Survey 2014 rounding base is 1000; Canadian National Immunization Coverage Survey rounding base is 50)
Real Time Remote Access

Using the system

• User Guide available on the RTRA site
  • [https://www.statcan.gc.ca/eng/rtra/programming](https://www.statcan.gc.ca/eng/rtra/programming)
• Create a SAS program
  • This involves examining the data documentation to setup the syntax for a data extraction
• 10 Remote submissions per user per day
RTRA Benefits

- Anytime access
- Few conditions on access and use of data
- Full Master files available
- Deemed employee status not required
- Up to 100 successful table outputs per day (10 per submission)
- Available program examples in RTRA documentation
RTRA Limitations

- Job submission have to be made in SAS only
- Only certain statistics available
- Not all data sets are available
- Time investment required to understand the dataset and documentation in advance of running an RTRA submission
How to Access RTRA

• If you are a RDC researcher, you can ask for access through the RDC Program

• If you are a non-RDC researcher, contact Liz or Kristi at rsclib@uwo.ca and we’ll assist you with obtaining access
Research Data Centre
Help with Data at Western

• Research Guides (Western Libraries)
  • Data and Statistics - https://guides.lib.uwo.ca/dataandstatistics
  • RTRA - https://guides.lib.uwo.ca/rtra
  • GIS - https://guides.lib.uwo.ca/gis

• Research Data Centre – https://rdc.uwo.ca/

• Statistical Consulting Service -
  https://www.lib.uwo.ca/madgic/statisticalconsulting.html

• Access to Statistical and GIS Software
  • https://wts.uwo.ca/sitelicense/
Alternatives To Statistics
Canada
Survey by survey recommendations
General Notes

• U.S. data can be a better replacement for RDC files than Canadian PUMFs because U.S. public use variables are generally released at a level of detail close to that which in Canada is only accessible in the RDC.

• U.K. data is also often quite detailed.

• U.S. has no equivalent of Statistics Canada – data is collected by the Census Bureau, Federal Reserve, Bureau of Labor Statistics, Center for Health Statistics, Center for Education Statistics...

• This leads to some duplication; other areas fall through the cracks

• Some major surveys are done by universities or other organizations (e.g. the U.S. GSS).
General Notes continued...

• Using U.S. data in place of Canadian is obviously problematic for anything that looks at health system factors and health outcomes due to the differences in health insurance.

• U.K. like Canada has national health insurance, but systems are not identical

• U.S. and Canada are closer demographically and economically than Canada and U.K.

• Relationships among social and health factors are likely to be comparable, prevalence estimates less so

• Sometimes the country is less important than other aspects of the research question
Economic and Labour surveys

Labour Force Survey, Canadian Income Survey, Household Spending and others
Labour Force and Income Surveys

• The Current Population Survey (Census Bureau)—like LFS, used to gauge unemployment rate, but asks about a number of other topics
  • NBER Supplements Coverage Chart
• Survey of Consumer Finances (Federal Reserve) includes an oversample of wealthy households
• Survey of Income and Program Participation (SIPP) (Census Bureau)
• Panel Study of Income Dynamics (University of Michigan) oversamples low income households
• National Survey of Unbanked and Underbanked Households is from the Federal Deposit Insurance Corporation and conducted as a supplement to the CPS
• U.K. Labour Force Survey releases longitudinal as well as cross-sectional data
Consumer and Expenditure Surveys


• Food expenditure supplement to the CPS is similar to Canada’s discontinued Family Food Expenditure survey. Also...

• Department of Agriculture Food Acquisition and Purchase Survey, combination of expenditure survey and a nutrition/consumption survey

• Survey of Consumer Expectations from the New York Fed – questions focusing on personal prospects and expectations
  • Available annually 2013 - now

• University of Michigan Surveys of Consumers (aka Survey of Consumer Attitudes and Behavior) – questions focusing more on expectations for the economy in general
  • Available monthly 1991 - now

• U.K. Living Costs and Food Survey – goes back to 1957
Health surveys

Canadian Community Health Survey, Addiction Surveys, and others
Canadian Community Health Survey

- **National Health Interview Survey** (NHIS) is probably most comparable to the CCHS. Includes modules on alcohol use and physical activity and has good coverage of mental health.

- The **National Health and Nutrition Examination Survey** (NHANES) is comparable to the CCHS nutrition modules.

- UCLA’s **California Health Interview Survey** is also a useful source with some questions that aren’t in the others.

- The U.K. has several health surveys; I’ll just mention the **Health Survey for England**.
Addiction Surveys

• The Behavioral Risk Factor Surveillance System (BRFSS) and the Youth Risk Behavior Surveillance System (CDC)
• National Adult and Youth Tobacco Survey (CDC)
• Monitoring the Future – high school cross-sectional that focuses on drug use
• National Survey on Drug Use and Health (NSDUH) and other data from the Substance Abuse and Mental Health Archive
• Of possible interest – a public use file for a Health Canada survey of young adults’ drug use is queued for release into ODESI
Others

• National Health Interview Survey on Disability is similar to Canada’s Survey on Disability / Participation and Activity Limitation Survey / etc. Last conducted in 1994. Currently the CDC suggests using the limited set of disability questions in the American Community Survey (Census bureau) to estimate disability prevalence.

• Listening to Mothers Survey (modules conducted between 2002 and 2018) has questions in common with Canada’s Maternity Experiences survey
GSS, Specific GSS modules
• Obviously, the U.S. General Social Survey
• World Values Surveys – some cycles include Canada
• Pew Center for People and the Press, particularly the Social Trends surveys
• The American National Election Study covers some social topics quite well – more than the Canadian Election Study
• Best analog to the time use module is the American Time Use Survey from the BLS
• The National Crime Victimization Survey is comparable to the GSS victimization modules
• CPS supplement on civic engagement has some content in common with the Canadian GSS Giving, Volunteering and Participating module. Also consider the American National Election Study
• The British Social Attitudes Survey is probably the best UK bet but has more of a political focus and is closer to the ANES / CES than the GSS.
Longitudinal Surveys

NLSCY, LSIC, NGS
NLSCY Alternatives

• Longitudinal data from the Bureau of Labour Statistics:
  • National Longitudinal Surveys of Youth (NLSY) 1979 and 1997
  • Children of the 79 cohort is a fascinating add-on
  • NLSY is strongly focused on labour market outcomes where the NLSCY covered social and health topics as well

• National Centre for Education Statistics data may be better for education-focused research questions or for looking at younger respondents
  • Early Childhood Longitudinal Studies
  • Secondary Longitudinal Studies Program

• Also consider the National Longitudinal Study of Adolescent to Adult Health (from UNC) particularly for research into physical or mental health, or social networks

• The Wisconsin Longitudinal Study (1957-2011) is another survey following adolescents into adulthood and has unique features including psychological and personality testing, sibling data, and genetic data

• The UK Data Service houses a number of longitudinal surveys, including the Millennium Cohort Study and the Child Development Study.
Immigration and Graduates

- Princeton’s Population Research department conducts some immigration-focused longitudinal studies:
  - New Immigrant Survey is the closest equivalent I’ve found to Canada’s longitudinal survey of immigrants.
  - Highly Skilled and Educated Immigrants
  - Children of Immigrants Longitudinal Study
- The CPS has an International Migration supplement from 2008, but it is a cross-section
- Also note the U.K. Longitudinal Study of Immigrant Children
- Also, though recent modules of Canada’s National Graduates Survey aren’t longitudinal, I wanted to toss a couple of graduate surveys in here
- The Baccalaureate and Beyond Longitudinal Study (B&B) is the closest equivalent to Canada’s National Graduates survey
- Princeton’s National Longitudinal Survey of Freshmen may also be of interest – looks specifically at minority achievement
Link Compendium
Government Data Sets

• U.S. Government Data Collection Agencies
  • Bureau of Labour Statistics
    • National Longitudinal Surveys site
  • Census.gov – not just the census!
  • National Center for Education Statistics
  • National Center for Health Statistics under the CDC
  • Bureau of Justice Statistics

• UK Data Service
Nongovernmental Starting Points

- **ICPSR – Interuniversity Consortium for Political and Social Research**
  - houses some (not all) government surveys, Monitoring the Future, thousands more

- **Pew Center for People and the Press**
  - I haven’t mentioned their Internet and social media surveys...

- **IPUMS, University of Minnesota**
  - U.S. and international census data, CPS

- **Institute for Social Research** at the University of Michigan

- Princeton University **Office of Population Research**
For more information

• For more information or to book an appointment to discuss data, contact Western Libraries Research and Scholarly Communications Team, wlrsc@uwo.ca
  • Data Librarian Liz Hill, ethill@uwo.ca
  • Research Data Librarian Kristi Thompson, kthom67@uwo.ca