Statistics Canada Statistique Canada

Canada

Statistics Canada
www.statcan.gc.ca

2006 Census Linked to the Discharge Abstract Database: New linked data for hospital research

Gisèle Carrière,
Health Analysis Division
Statistics Canada

November 2016

Canadian Research Data Centre Network
2016 National Conference

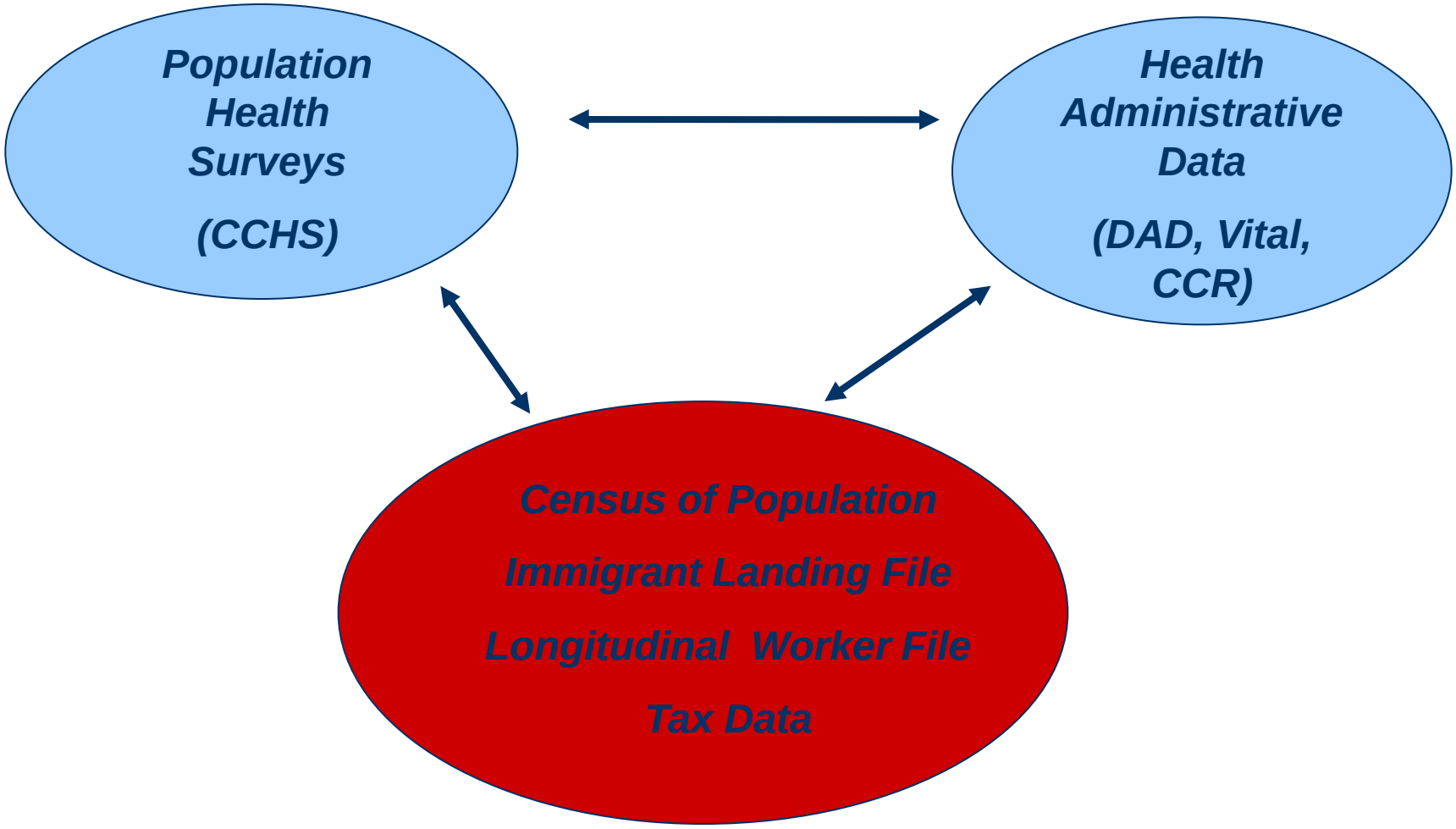
Outline

- Context
  - Record linkage -what is it?
  - Linked health data - why?
  - Record Linkage Process Model
- 2006 Census linked to the Discharge Abstract Database
  - Methodology
  - Linkage results
  - Some research highlights
- Accessing the data
  - 2006 Census-DAD Pilot Project in RDC

Context: What is record linkage?

- The process of matching records across or within datasets that refer to the same entity, such as a person
- Linkage conducted using **unique identifying** information common to datasets - name, date of birth, postal code, personal health number
- Linkage methods are primarily “exact” approaches – Linkage of data for the same units rather than a similar unit
  - Made using **unique identifying** information or similar characteristics that discriminate between units

Health Data linkages: Input Files



Context: Linked health data – Why?

- Enhance the capacity of health data to address complex questions with "value added" information - fill data gaps
  - Survey data – lots of socio-economic, risk factor information but no outcomes;
  - Administrative data – outcome information (hospitalization, mortality) but limited individual information, no health determinants (e.g. smoking behaviours)
- Linked data allow for "population health" lens to the study of health care services and outcomes
  - Used to study a wider range of determinants of health care use and outcomes of care (i.e. Socio-economic status, lifestyle factors etc.)
- Population based studies on a representative sample of Canadians
  - Large sample sizes - study specific populations and "rare" events
  - Fills gaps for specific sub-populations (e.g. Immigrants, Aboriginal peoples)
- Opportunity for cross-jurisdictional comparisons

Challenges

- Comparable data variables for linkage – what if there are limited common variables?
  - Commonly used to link data – name, date of birth, sex, postal code, personal health number
  - Not all variables available on all data sources

- Validation of linkage results – what is a “good” linkage rate?
  - 100% expected when the same population represented in all data sources
  - Health outcomes – not everyone experiences a hospitalization, cancer or death (well eventually!)

- Validation of linked data – are the data “good” for research?
  - Coverage analyses to assess the extent to which the linked data represent the “expected” health outcomes of the population of interest

Record Linkage Process Model

Phase 1: Project Planning

- Specify Needs
- Design record linkage strategy
- Approval of record linkage project

Phase 2: Record Linkage

- Data Preparation – standardize linkage variables
- Record Linkage
- Assess quality – internal and external validation

Phase 3: Integrate, Analysis, Access

- Create and document linked data file; conduct analysis
- Access and Archive
- Evaluate the record linkage project

2006 Canadian Census and Discharge Abstract Database Linkage

Funding:
Longitudinal Health and Administrative Data Initiative (LHAD)

Context

- Longitudinal Health and Administrative Data Initiative (LHAD)
  - Collaboration with provincial Ministries of Health to facilitate record linkage and guide research;

- To better understand the health outcomes and healthcare use of specific sub-populations –
  - Immigrants, Aboriginal groups
  - Identify and quantify differences
  - Understand differences in the context of other social determinants of health

Research areas

- **Immigrant research**
  - Comparative analysis of hospitalizations by immigrant status, source country and time since immigration;
  - Use of hospital services among immigrant seniors;
  - Multi-generational analysis of cardiovascular related hospitalizations – is the health advantage lost among second generation?

- **Aboriginal research**
  - Comparative analysis of hospitalization rates among Aboriginal groups, on and off reserve
  - Impact of housing condition on respiratory related hospitalizations among First Nations on reserve

Source Files

- 2006 Census
  - Short form – used for record linkage
  - Long-form – used for validation and analysis
    - 20% representative sample of the Canadian household population
    - Demography, labour market, income, education, language, disabilities, housing, immigration, ethno-cultural, Aboriginal identity, Registered Indian

- Discharge Abstract Database (DAD) (CIHI):
  - DAD 2005/06 through 2008/09 used for pre-processing
  - DAD 2006/07 through 2008/09 used for record linkage
  - Census of discharges from acute care hospitals (~3 million records per yr) (excludes Quebec)
  - Clinical diagnostic and intervention information, limited demographic

- T1Personal Masterfile (T1PMF)
  - T1 tax returns - historical
  - Annual place of residence (postal codes) to tract mobility over time
  - *No income data used!*

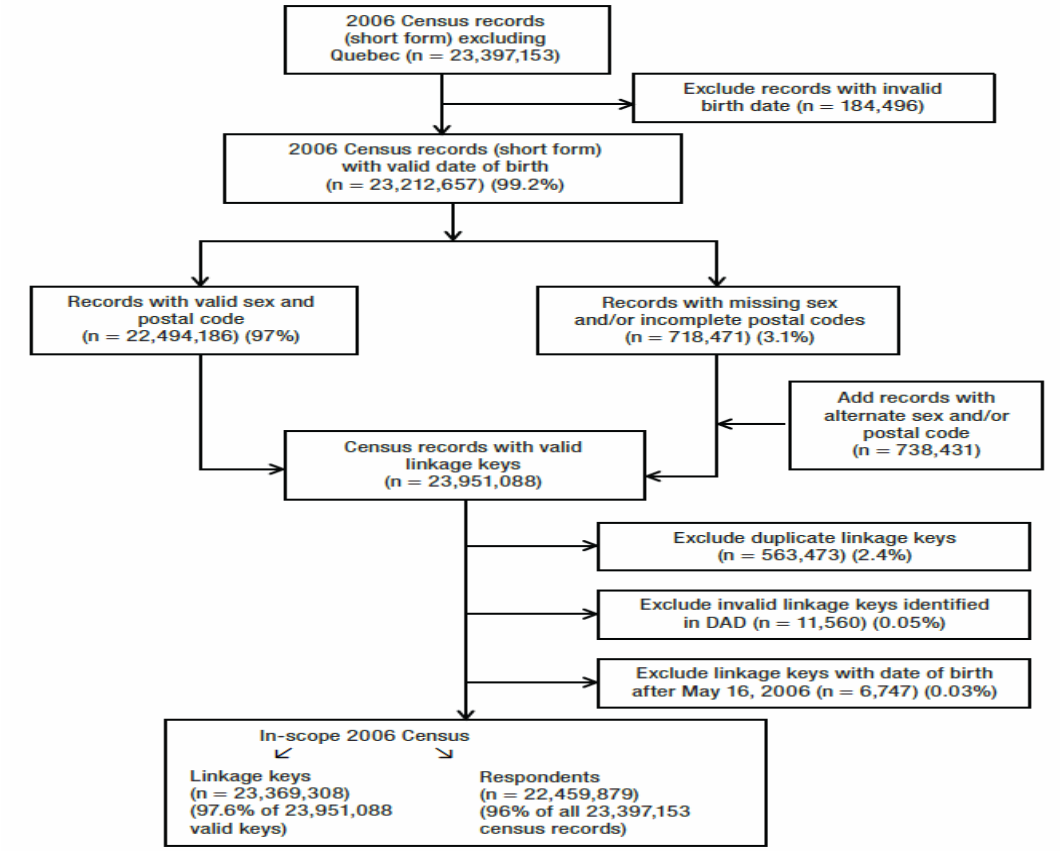
Step 1: Data Preparation

- **Unique statistical linkage key** – *date of birth, sex, postal code*
  - Only common variables on both the Census and DAD for linkage
  - Previously demonstrated that this was a feasible approach to data linkage (Rotermann, M et al., **Two approaches to linking census and hospital data.** *Health Reports* October 2014**)**

- *Eligibility of records for linkage*:
  - DAD records of stillbirths, cadavers and non-residents removed;
  - Complete (non-missing) date of birth in both Census and DAD;
  - Statistical linkage key must be unique in Census - no duplicates (e.g. twins removed)
  - Statistical linkage key associated with only one Health Insurance Number (HIN) in DAD

Step 1: Data Preparation

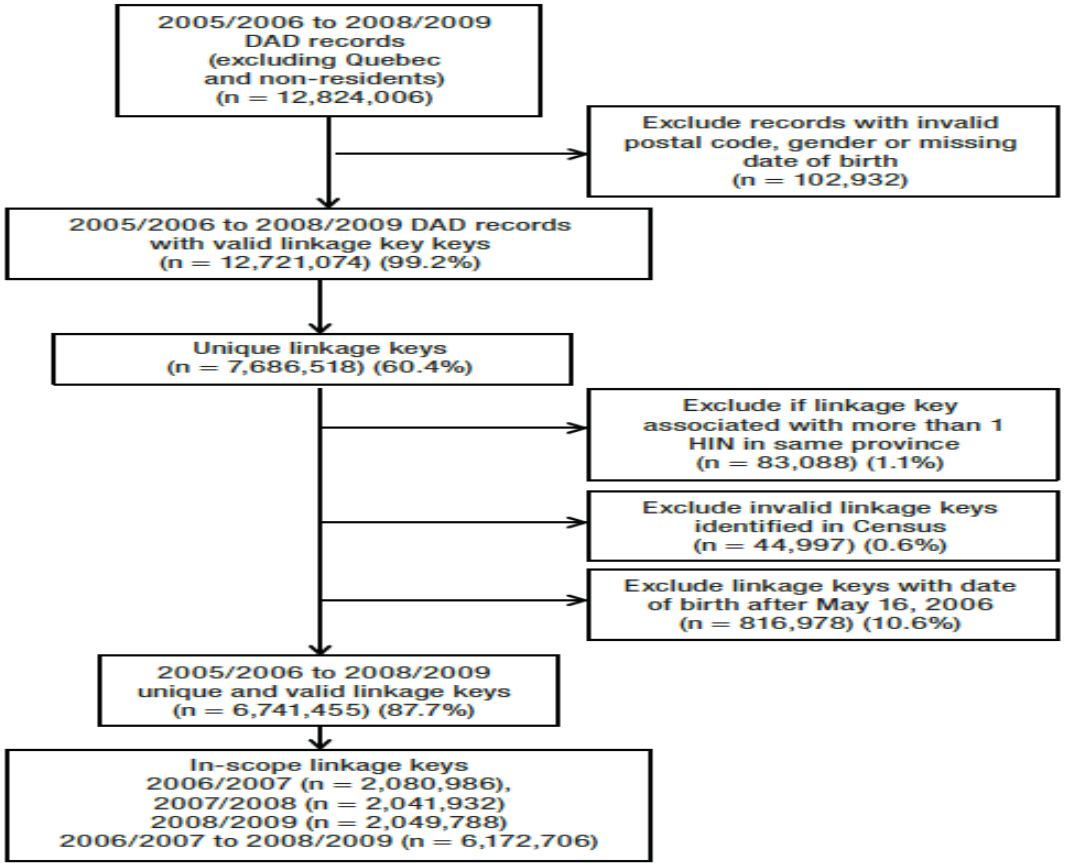
Processing of the 2006 Census for record linkage



Source: Rotermann et al. Linking 2006 Census and hospital data in Canada. Health Reports, October 2015

Step 1: Data Preparation

Processing of the 2005/06 to 2008/09 Discharge Abstract Database for record linkage



Source: Rotermann et al. Linking 2006 Census and hospital data in Canada. Health Reports, October 2015

Step 2: Data Linkage

- *Hierarchical Deterministic Linkage*
  - **Unique statistical linkage key** – *date of birth, sex, postal code*
    - Used postal code information from historical tax files as alternative to capture change in address overtime
  - Series of exact matches -conservative approach but appropriate given lack of **unique** identifying information

Pass1: required an exact match between census birth date, sex, postal code and DAD linkage keys

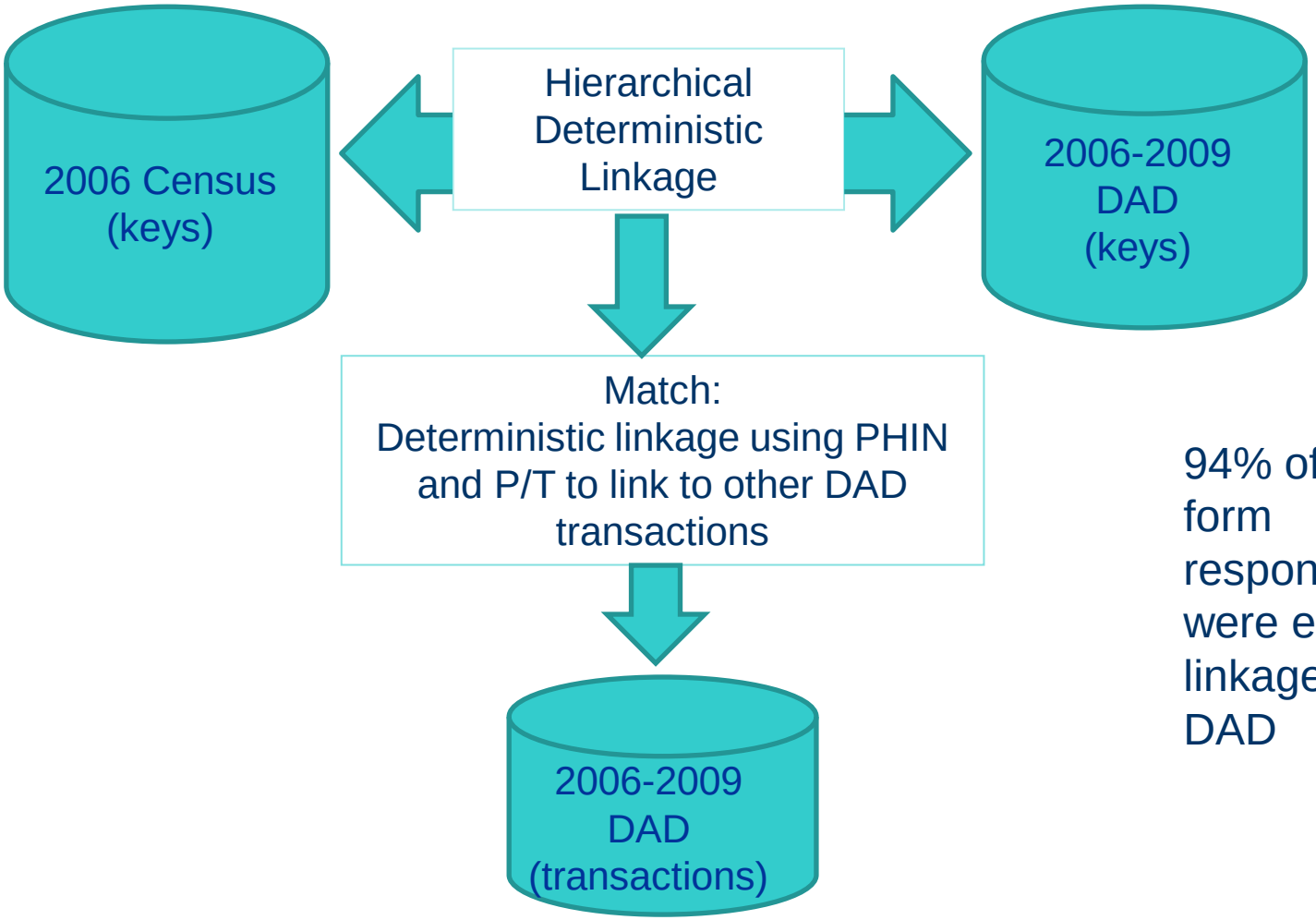
Pass 2 - 4: exact match on birth date, sex and postal code from the T1PMF and the DAD linkage keys

Pass 5-10:relaxed the rules for postal code, allowing one of the 6 characters to census-reported postal code to be dropped.

Passes 11-28: repeated the processes but also used the T1PMF postal codes.

Step 2: Record Linkage

2006 Census-DAD Record Linkage Approach



94% of long-form respondents were eligible for linkage to the DAD

Step 3: Quality assessment/ Validation

- Purpose: to assess the quality of the record linkage process and "fitness for use" of the linked data
  - Internal validation– quality of the linkage

    Do the linked pairs represent good links?

    Are there any missed links among the non-linked pairs?

  - External validation – quality of the linked data (representativeness of analytical file)

    Do the outcomes in the linked data file represent the experiences of the population of interest?

Step 3 : Quality Assessment/Validation

(Rotermann M, Sanmartin C et. (2015) *Health Reports* (October)

	Long-form census respondents - Eligible for Linkage		% of Long-form respondents linked to DAD		
			"2006/07"	"2007/08"	"2008/09
	'000	%	%	%	%
Canada excluding Quebec	4,652.70	94	5.6	5.4	5.2
Province/Territory					
Newfoundland and Labrador	97.8	94	6.1	6	5.8
Prince Edward Island	24.5	94	7.2	6.9	6.6
Nova Scotia	171.6	94	5.6	5.5	5.4
New Brunswick	141.2	96	7.1	6.8	6.4
Ontario	2,254.50	94	5.1	4.8	4.7
Manitoba	263.7	94	6.6	6.3	6.3
Saskatchewan	225.2	94	7.8	7.3	7.1
Alberta	640.7	94	5.8	5.6	5.4
British Columbia	769.1	92	5.4	5.3	5.1
Yukon	13	94	5.3	5.5	5.7
Northwest Territories	24.6	93	6.8	6.5	6.2
Nunavut	26.8	91	6.5	6.4	6.4
Sex					
Male	2,284.40	94	4.5	4.3	4.2
Female	2,368.30	94	6.7	6.4	6.2
Age group					
Younger than 1	51.8	91	16.2	3.4	2.1
1 to 4	213	93	2.7	2	1.6
5 to 9	284.4	93	1.5	1.3	1.2
10 to 14	321.7	93	1.5	1.7	1.9
15 to 19	330.7	93	2.9	3.2	3.3
20 to 24	307.7	93	4.5	4.6	4.5
25 to 34	587.4	94	7.1	6.8	6.3
35 to 44	715.5	95	4.1	3.7	3.5
45 to 54	725	94	4	4	4
55 to 64	525.3	94	6.3	6.5	6.5
65 to 74	325.9	95	10.6	10.8	10.8
75 or older	264.2	95	17.4	16.8	15.9

Source: Census of Population 2006, Census-linked Discharge Abstract Database 2006/2007, 2007/2008, 2008/2009 pooled.

Looking for

-consistently high levels of eligibility across respondent characteristics;

-"reasonable" linkage rates to DAD

Step 3: Quality Assessment

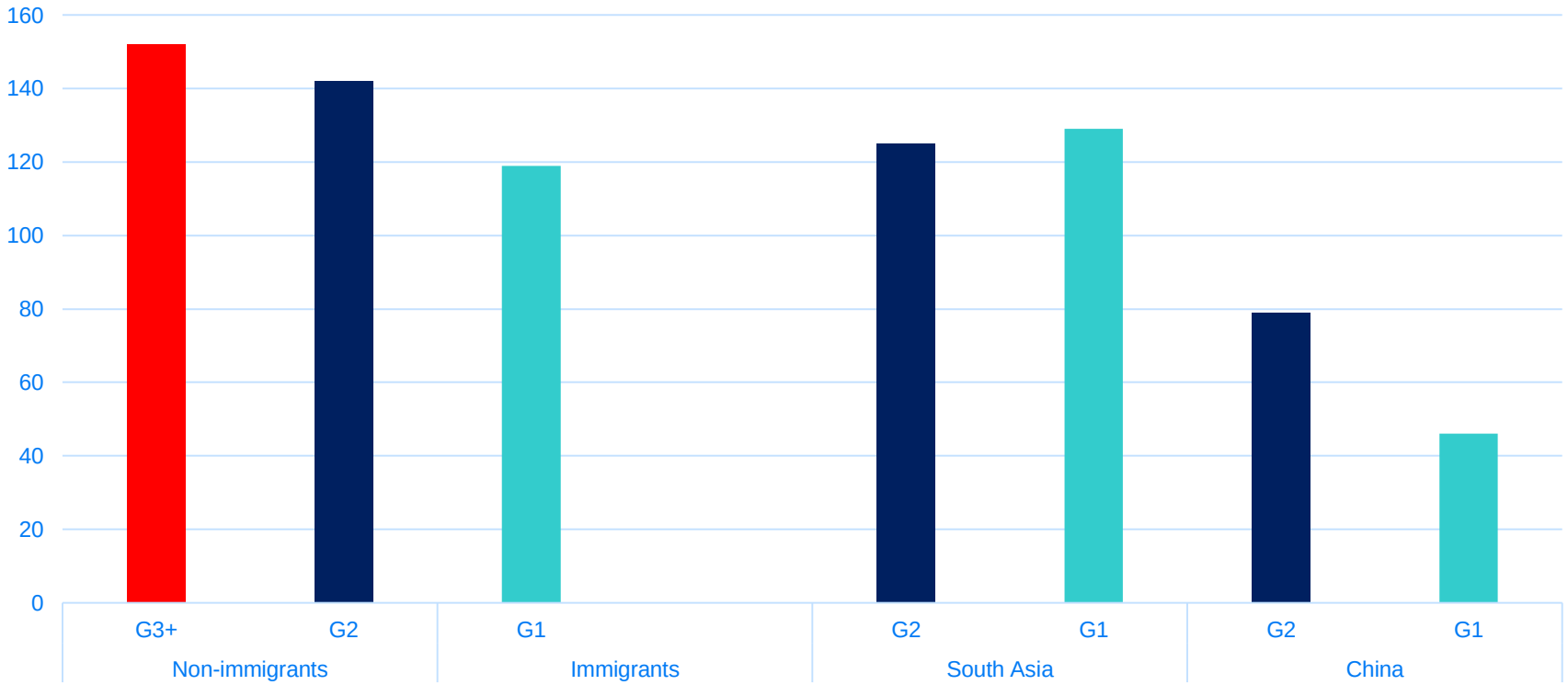
Unweighted and weighted coverage rates for acute-care hospital discharges, by province, sex and age, Canada excluding Quebec, 2006/2007, 2007/2008 and 2008/2009

	2006/2007					2007/2008					2008/2009				
	Hospital discharges (denominator)	Long-form census respondents hospitalized (numerator)				Hospital discharges (denominator)	Long-form census respondents hospitalized (numerator)				Hospital discharges (denominator)	Long-form census respondents hospitalized (numerator)			
		Unweighted		Weighted			Unweighted		Weighted			Unweighted		Weighted	
	'000	'000	%	'000	%	'000	'000	%	'000	%	'000	'000	%	'000	%
Canada excluding Quebec	**2,045.00**	**351.8**	**17.2**	**1,629.40**	**79.7**	**1,967.50**	**338.2**	**17.2**	**1,566.00**	**79.6**	**1,942.10**	**327**	**16.8**	**1,513.40**	**77.9**
Province/Territory															
Newfoundland and Labrador	52.9	8.4	16	40.5	76.5	51	8.4	16.4	39.8	78	49.4	8	16.3	38.3	77.6
Prince Edward Island	15.8	2.5	16	12.9	81.6	15.1	2.5	16.5	12.6	83.7	14.8	2.3	15.5	11.7	79.6
Nova Scotia	80.4	13.3	16.5	65.8	81.8	76.7	12.7	16.6	63.1	82.3	77	12.5	16.2	62.4	81
New Brunswick	87.1	14.2	16.3	69.7	80	82.2	13.4	16.3	66.1	80.4	79.3	12.6	15.9	61.8	78
Ontario	917.7	149.5	16.3	741.8	80.8	876.3	143	16.3	711.3	81.2	863.7	138.6	16	688	79.7
Manitoba	113.8	25.3	22.2	91.6	80.4	109.6	24	21.9	86.6	79	109	23.6	21.7	84.9	77.9
Saskatchewan	121.6	25.9	21.3	98.8	81.3	115.7	24.5	21.2	92.7	80.1	113.3	23.6	20.8	89.5	79
Alberta	291	49.7	17.1	227.4	78.2	282.9	48	17	219.4	77.6	279.3	46	16.5	210.9	75.5
British Columbia	352.6	57.4	16.3	273.1	77.4	346.5	56.1	16.2	266.6	76.9	345	54.4	15.8	257.8	74.7
Yukon	3.3	1	29	2	62.4	3.3	1	31.4	2.2	67.7	3.4	1	30.6	2.4	71.5
Northwest Territories	5.1	2.3	45.8	3.3	65.7	5.1	2.3	45.7	3.3	65.8	4.7	2.1	45	3.3	69.5
Nunavut	3.7	2.3	63	2.4	65.7	3.2	2.2	67.4	2.2	69.4	3.2	2.2	69	2.3	71.6

Data use example 1: Research Results

(Ng E, Sanmartin C et al. ***Health Reports***, October 2015)

Age standardized rate of circulatory disease related hospitalizations (per 10,000 pop) by immigrant and generation status, Canada (excluding Quebec), 2006-2009

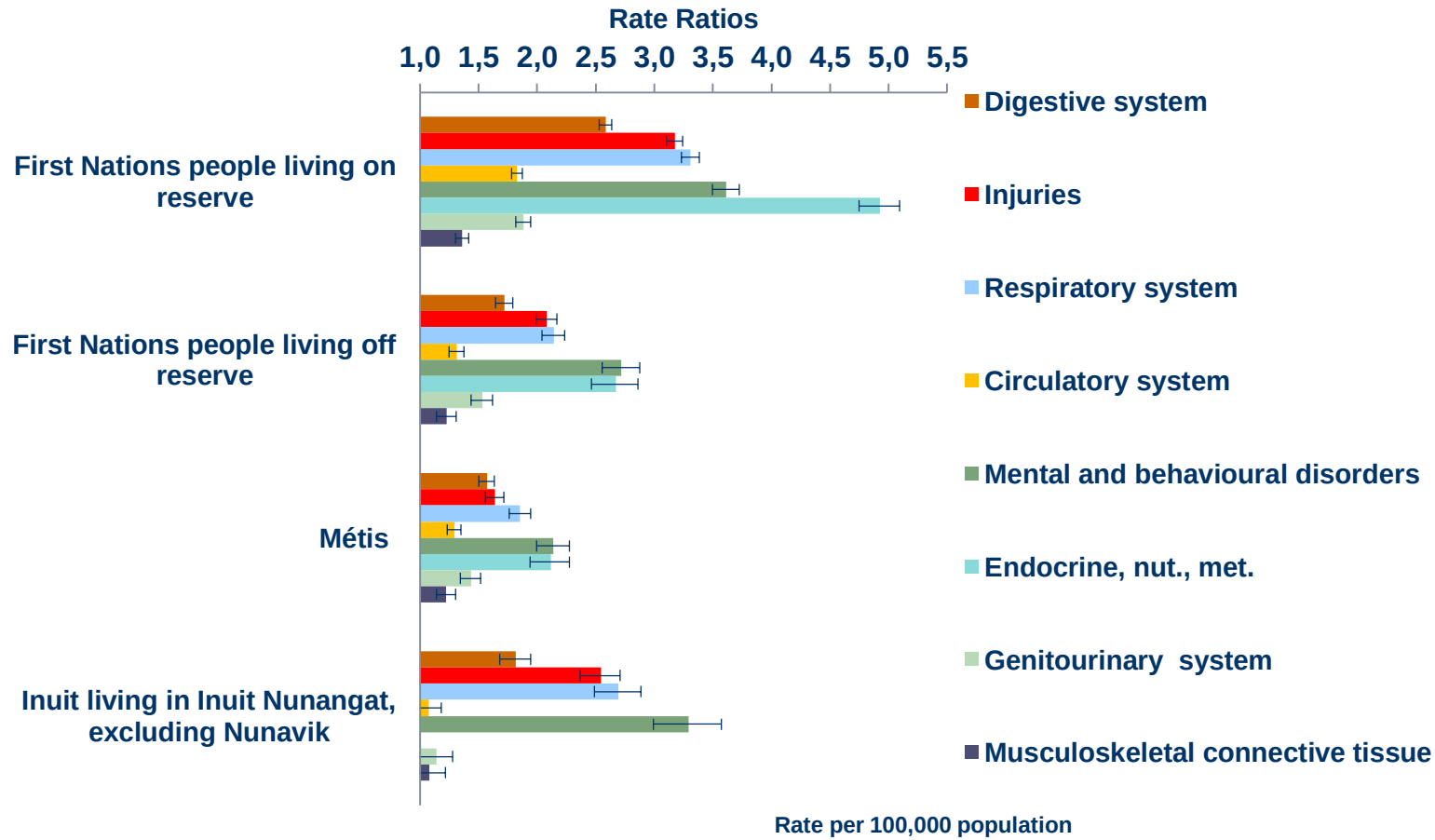


Source: Census of Population 2006, Census-linked Discharge Abstract Database 2006/2007, 2007/2008, 2008/2009 pooled.

Data use example 2: Research Results

(Carriere G, Bougie E et al. Health Reports *Health Reports*, August 2016)

Age-standardized hospitalization rate ratios by Aboriginal identity group, and by diagnostic chapter, Canada (excluding Quebec), combined fiscal 2006/2007 through 2008/2009



Source: Census of Population 2006, Census-linked Discharge Abstract Database 2006/2007, 2007/2008, 2008/2009 pooled.

Accessing the data – Partners and roles

- Statistics Canada
  - Sets security standards
  - Provides data
  - Staffs and manages the centres
- Universities
  - Provide a secure laboratory
  - Provide financing for the operations
- Funding Agencies
  - Provide funding to Universities
  - Coordinate non-government requests for access to data

Accessing the data- Statistics Canada:

Sets security standards

- Developing processes to move linked data files to the RDC (Some examples include vetting rules, user documentation)

Provides data

- Examples of important things for you to consider:
  - *What is your population of interest?*
  - *Has your project objective been clearly defined?*
  - *Can the requested data address the research questions?*

Accessing the data

Challenges –determining what information and what level of support is required for use of new linked data.

Statistics Canada, together with The Canadian Institute for Health Information (CIHI), is conducting a **pilot project** to provide access to the 2006 Census linked to the Discharge Abstract Database (2006/07-20008/09).

2006 Census-DAD Pilot Project

Purpose:

- give researchers access to these linked data in a Statistics Canada Research Data Centre (RDC)
- aim to evaluate and improve confidentiality vetting rules and help to improve documentation for researchers.

How/where to apply:

CRDCN website: http://www.rdc-cdr.org

Submit proposals to 2006 Census DAD pilot to:

Kelly Cranswick, Regional Manager, RDC Program

kelly.cranswick@canada.ca

Summary

- Lots of record linkage activity – health
- Developing processes to move linked data files to the RDCs
- 2006 Census-DAD Pilot project coming soon.
- Goal: to provide sufficient information and support to users

Contact

Gisèle Carrière
Health Analysis Division
Statistics Canada

gisele.carriere@canada.ca

Email inquiries: HAD-DAS@statcan.gc.ca

Record Linkage: http://www.statcan.gc.ca/health-sante/link-coup-eng.htm

Thank- you!