



Statistics
Canada

Statistique
Canada

Canada



Statistics Canada
www.statcan.gc.ca



The T1 Family File (T1FF) and the Longitudinal Administrative Databank (LAD): Administrative Datasets for Small Area Analysis

**Prepared for the CRRF – RPLC Rural Conference,
Building Vibrant Rural Futures,
Guelph, Ontario**

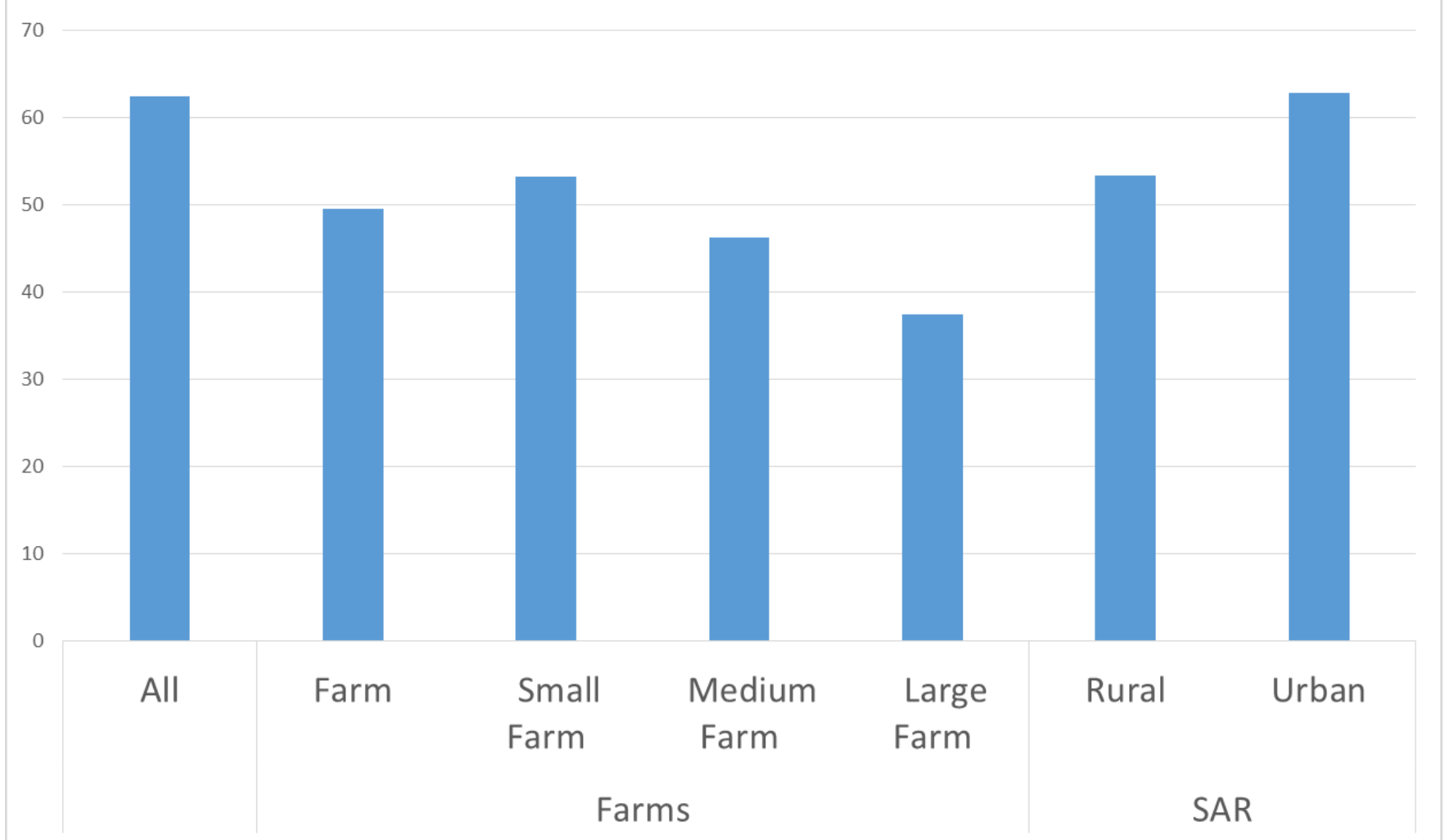
Brian Murphy

Income Statistics Division

October 13, 2016



Average percentage of family income from wages and salaries



T1FF and LAD are not new

- Administrative data development program, 1979
 - Produce yearly migration data in order to support the Population Estimate Program of Demography division
 - Migration estimates
 - Supply personal and family income data for low levels of Geography
 - T1FF
- First LAD completed 1989 (1982-1986)
 - Economic Council PSID inspired analysis - T1FF/SA files



T1FF: Universe and Coverage

Target Population

- Persons who completed a T1 tax return for the year of reference and their dependants
 - Their non-filing spouses (including wage and salary information from the T4 file)
 - Their non-filing children identified from three sources (the CCTB file, the births files, and an historical file)
 - Filing children who reported the same address as their parent

File Creation: Input Files

- T1FF
 - T1 Income tax returns
 - T4 Statement of earnings
 - Canada Child Tax Benefit (CTB) file
 - Business Register
 - Addresses and postal walks
 - Postal Code Conversion File (PCCF)
 - Birth file
- LAD
 - T1FF
 - Longitudinal Immigration Database (IMDB)
 - SIN_TTY Cross reference file
 - Tax Free Savings Account (TFSA)

T1FF – T1 Information

- Personal identifiers (Used for file creation only)
 - Social Insurance Number - SIN
 - Mailing address
 - Birth date, gender and marital status
- Sources of income
- Deductions, exemptions and tax credits
- Information about family members included on individual tax forms



T1FF - Processing

- Coverage enhancement
- Creation of census families
- Geography components
- Imputation of Income variable

T1FF Strengths

- 100% of Canadian tax filers
- Not longitudinal, but good for cross-sectional, single year analysis
- More than 70% filing rate compared to population estimates of Canadians
- 96% coverage rate when including the dependents
- Exists since 1982



Section 3: LAD Content and Construction

Longitudinal Administrative Databank

Coverage and content

- 20% longitudinal sample of the T1FF
- 1982 to most recent year of T1 data (2013)
- Selection of variables from the T1FF
- Information available at the individual, spouse/parent and family level
- The primary source for variable information is the LAD Data Dictionary

Demographic Variables

- Individual Demographics
 - age, sex, marital status, language, etc.
- Family Demographics attached to each selected individual
 - type of family (Couple , Lone Parent, Person not in census family)
 - number & age of children
- Spouse or parent information
- Geography

Income and other variables

- Employment Income
 - Wages, Salaries, Commissions, Tips
 - Self-employment
- Investment Income and other Income
 - Net rental income
 - Alimony
 - Other Pensions
 - RRSP
 - Limited Partnership
- Tax credit such as tuition fees
- Transfer Payments incl.
 - Old Age Security
 - Net Federal Supplement
 - Canada/Quebec Pension
 - Employment Insurance
 - Social Assistance
 - Workers' Compensation
 - Child Tax Benefits
- Other variables such as disability amount
- Two-digits NAICS (since 1999 – from the Business Register)

Immigration variables

- Since 2002, the LAD contains information on recent immigrants at time of landing – 1980 to 2012. Variables include:
 - Official languages ability indicator
 - Country of citizenship at landing
 - Country of last permanent residence
 - Country of birth
 - Level of education at landing
 - Landing year
 - Marital status at landing
 - Native language (or mother tongue)
 - Intended place of destination
 - Intended occupation

Tax Free Savings Account Variables

The LAD currently has three TFSA variables:

- TFSA contributions
- TFSA calendar year end
- TFSA withdrawals

Annual information for these variables exists from 2009 to 2013

LAD Strengths

- The LAD is a good tool for studying many longitudinal socio-economic dynamics
- There are currently 31 years of reliable, comprehensive income data
- The LAD, as a 20% sample of the T1FF, can describe very small regions
- Having customizable geography and many economic variables makes the LAD an extremely versatile research tool.
- The result is a very useful research databank gradually being rolled-out to the RDCs

Geographies on T1FF and LAD

- Canada
- Provinces and Territories
- Postal Geography (T1FF)
 - City Totals
 - Urban Forward Sortation Areas (excludes Rural Routes and Suburban)
 - Services, and Other Urban Areas within City)
 - Rural **Postal Code** Areas (Within City)
 - Other Urban Areas (Non-residential within city)
 - Rural Communities (not in City)
 - Other Provincial Totals

Census Geographies

- PCCF mapping of postal codes to Census Geography
 - 50% rule. Where a Postal Code area overlaps a census boundary, the entire postal code population is placed in the Census geographic area in which 50% or more of the population of the postal code resides
- Rural postal codes have a zero in the second digit
- Rural postal codes can be released at postal code level – subject to suppression
 - Need 100 records in postal code and 15 records in a category
- Non-rural postal codes must be aggregated for dissemination

Geographies on T1FF and LAD

- Census Geography
 - Economic Regions
 - Census Divisions
 - Census Metropolitan Areas
 - Census Agglomerations
 - Census Tracts
 - Federal Electoral Districts (2003 Representation Order)
 - Area Size of Residence (ASR)
- Custom Geographies based on Postal Code

Protecting Confidentiality

- Rules to prevent disclosure
 - Addition of noise
 - Suppression
 - Dominance
 - Residual disclosure avoidance
 - Rounding

Accessing T1FF and LAD

- T1FF (No direct access to microdata)
 - Annual Tables
 - CANSIM 111-* series
 - Neighbourhood Information Databank
 - Custom Tabulation
- LAD
 - CANSIM 204-* series
 - Custom Tabulation, linkage
 - Microdata through Research Data Centres



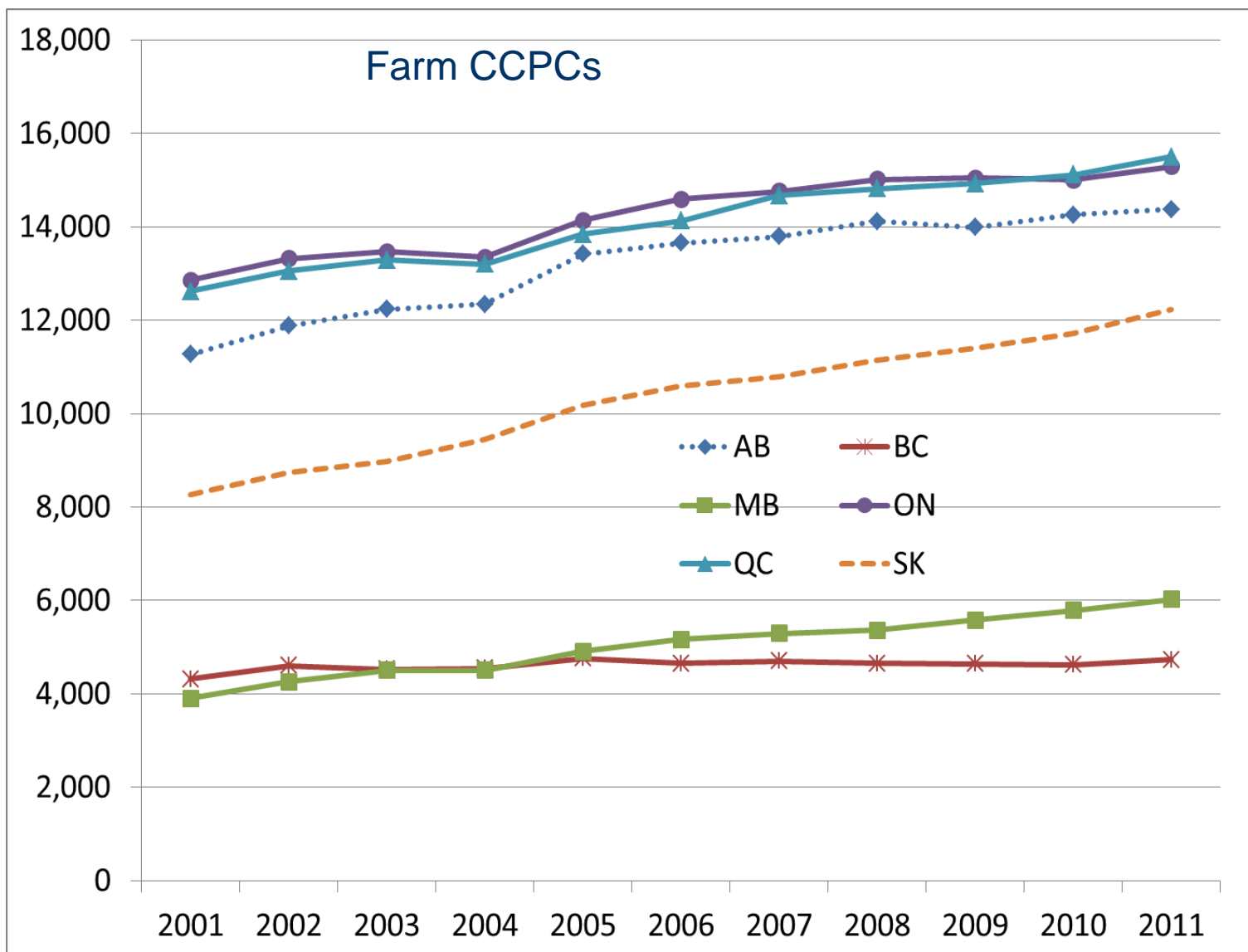
LAD Data Access

- Very controlled access
 - Only StatCan or deemed employees
 - No release of micro-data (no PUMFs)
 - Confidential aggregate data must remain on-site
 - Secure physical environment (Statcan or RDC)
- Research assistant service on a cost recovery basis
- Record linkage requests



Selected Research Areas

- Incomes of Farm Families
- Income and earnings dynamics of individuals and their families
- Dynamics of low income Canadians
- Characteristics of those in social-assisted housing
- Small business starts via patterns and levels of self-employment
- Estimating the number of marriages and divorces in Canada
- Life after Service: Post military retirement income patterns
- Post immigration emigration
- High Income taxation progressivity
- Returns to education
- Labour migration patterns



Questions?

- For more information, please contact:
- Pour plus d'information, veuillez contacter :

Brian Murphy (613-617-9481)

BrianB.Murphy@Canada.ca

Income Statistics Division Client Services Section
STATCAN.incomerevenu.STATCAN@Canada.ca