



Statistics
Canada

Statistique
Canada

Canada



Statistics Canada
www.statcan.gc.ca

LAD THE LONGITUDINAL ADMINISTRATIVE DATABANK

Western Data Day

Nell Hamalainen

Income Statistics Division

March 24, 2017



Outline of Presentation

1. LAD construction and content
2. Research applications
3. Accessing the LAD and confidentiality
4. Summary



Section 1: LAD Construction and Content

LAD - Sampling Frame

- The LAD is based on administrative tax data drawn from the T1 Family File (T1FF)
- T1FF - universe and coverage
 - Persons who completed a T1 tax return for the year of reference or who received CCTB (Canada Child Tax Benefits)
 - Records for non-filing spouses and non-filing children are constructed from administrative sources
 - Allows for family as the unit of analysis
 - 96% coverage rate when including dependents

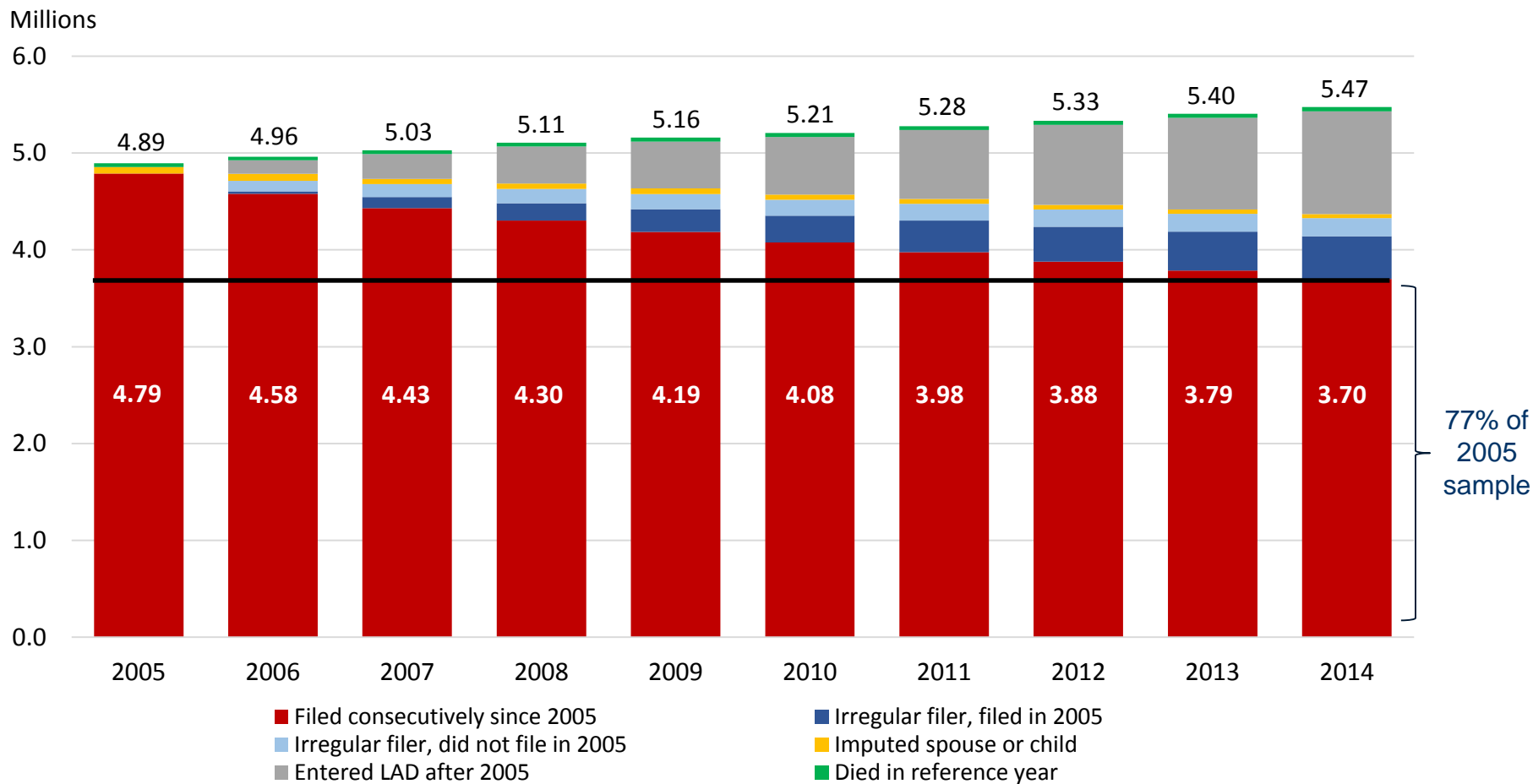
LAD - Sampling Scheme

- 20% random sample of *individual* records on T1FF
- Only T1FF records with a SIN are sampled
- Constant unique SIN via SIN cross-referencing
- No imputation for non filers, late filers
- 33 Years of data: 1982 - 2014
- Large final sample: 3.0 million in 1982, 5.5 million in 2014



Filing Persistency on LAD

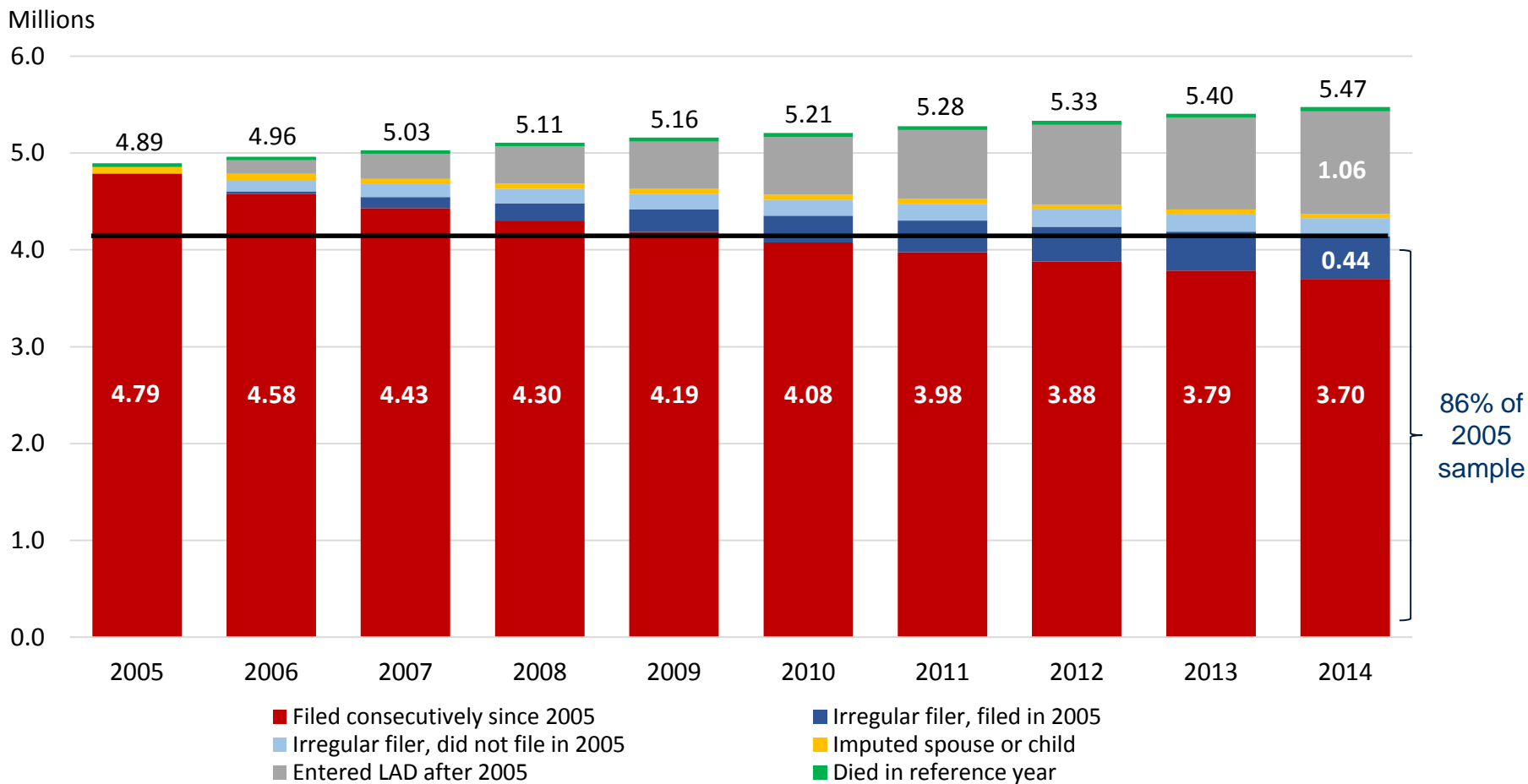
Number of LAD records by entry and filing persistency, 2005 to 2014





Filing Persistency on LAD

Number of LAD records by entry and filing persistency, 2005 to 2014



Demographic Variables (e.g.)

- Individual Demographics
 - age, sex, year of death
- Family Demographics attached to each selected individual
 - family type (couple, lone parent, person not in census family)
 - number & age of children
- Geography
 - Derived from postal codes, allow for analysis of small areas

All income and tax variables are available at four levels of aggregation

- Although LAD is an *individual* level file, income and tax information aggregated over an individual's family is brought over from the T1FF
 1. Individual level information
 2. Family level aggregate
 3. Couples
 4. Kids
- Allows for analysis of the family income situation of individuals

Market income variables (e.g.)

- Employment Income
 - wages, salaries and commissions
 - self-employment (available by type)
- Private Retirement Income
 - private pension income
 - RRSP income
- Investment Income
 - dividends
 - rental income
 - limited partnership income

Government Transfers (e.g.)

- Government Transfers
 - OAS/GIS
 - Canada/Quebec Pension Plan
 - Employment Insurance
 - Social Assistance
 - Workers' Compensation
 - Child Benefits
 - Provincial refundable tax credits
 - Those not found on the T1 are calculated

Tax variables (e.g.)

- Federal and provincial income tax
 - Federal and Provincial tax payable
 - Deductions (e.g. child care expenses, split-pension amount)
 - Federal non-refundable tax credits (e.g. charitable giving, children's fitness and arts credit)
 - RRSP Contributions
 - Quebec taxes are calculated

- Federal payroll taxes
 - EI
 - CPP/QPP

Other variables (e.g.)

- TFSA Information
 - Account holders / Contributions / Market value
- Industry of employment
 - Three-digit NAICS (T4 linked to business register)
 - NAICS for two separate jobs (two highest T4 incomes)
- Immigration variables drawn from the Longitudinal Immigration Data Base (IMDB)
 - Immigrants at time of landing, 1980 to 2014
 - Country of origin, mother tongue, Education, intended occupation, marital status, etc.



LAD data dictionary

Longitudinal Administrative Data Dictionary, 2014



Section 2: LAD Research

Lad Research

- Longitudinal analysis
 - Lifetime dynamics
 - Event impacts
- Individual tax filer is the preferred unit of analysis
- Record linkage
 - Require linkage agreements
 - Linkage performed at head office

Examples of research questions that can be addressed using the LAD

- What are the income dynamics of individuals and their families?
- Does interprovincial migration affect earnings?
- What characteristics distinguish continuous from occasional RRSP contributors?
- What is the impact of marital separation on family income?
- Do the same people experience low income year after year?
- How do children affect a woman's income?

Recently published LAD Research

- Top-End Progressivity and Federal Tax Preferences in Canada (Murphy, Veall and Wolfson, 2015)
 - Relied on tax deduction and credit variables in LAD
 - The majority of federal tax preferences increase progressivity of the personal income tax system at the very top of the income spectrum
- Piercing the Veil: Private Corporations and the Income of the Affluent (Wolfson, Veall, Brooks and Murphy, 2016)
 - Data linkage between LAD and T2 income tax returns filed by Canadian-controlled private corporations (CCPC)
 - Top income shares are significantly higher when CCPC incomes are included

LAD CANSIM tables

- High income tables (introduced in 2013)
 - High income trends of tax filers in Canada
 - A Profile of High Income Canadians (Murphy, Roberts, and Wolfson 2007)
- Low income dynamics tables (introduced 2015)
 - Longitudinal measures, e.g. two-year low-income transitions, low income persistence over 8-year periods
 - What can we learn about low-income dynamics in Canada from the Longitudinal Administrative Databank? (Zhang, 2014)
- Income mobility (in development)
 - The evolution of income mobility in Canada: Evidence from the Longitudinal Administrative Databank, 1982 to 2012 (Zhang, 2016)



Section 3: LAD Access and Confidentiality



LAD in RDCs

- Following a successful pilot project that allowed deemed government researchers to access the LAD from within Statistics Canada's FRDC, LAD is now being gradually rolled-out to the larger StatCan RDC network
- McMaster, Laval, UQAM, Carleton, UNB, UofT
- Speed of roll-out managed by the Microdata Access Division at Statistics Canada

Confidentiality and Security

- Researchers using the RDC must protect confidential information:
 - Are deemed employees of Statistics Canada
 - Swear the oath of office
 - Must have security clearance
 - Only vetted output can be removed from the RDC
- Disclosure analysis is performed by RDC analysts on all output leaving the RDC

Disclosure Control Techniques

- Rules to prevent disclosure
 - Addition of noise
 - Rounding
 - Dominance tests
 - Collapsing of cells with low counts
 - No residual disclosure
- LAD researchers must integrate these rules into analysis codes
- *Vetting guide for RDC Analysts - “Confidentiality Guidelines for the LAD”*

Summary

- The LAD is a good tool for studying many longitudinal socio-economic dynamics
- There are currently 33 years of reliable, comprehensive income data
- The LAD, as a 20% sample of the T1FF, can describe very small regions ($N > 5,000,000$)
- 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



For questions regarding LAD contact

Brian Murphy
Special Advisor
Income Statistics Division
brianb.murphy@canada.ca
613-715-9481

Paul Roberts
Senior Research Analyst
Income Statistics Division
paul.roberts@canada.ca
613-716-8481