



Statistics
Canada

Statistique
Canada

THE LONGITUDINAL ADMINISTRATIVE DATABANK

Workshop: New developments in CRDCN dataset portfolio

www.statcan.gc.ca



CANADA 150

Telling Canada's
story in numbers

Paul Roberts
Income Statistics Division.

November 16, 2017

Canada 



Outline of Presentation

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

Section 1: LAD construction and content



LAD – Data Sources

- The LAD is based on administrative data files, mainly tax data
 - T1 family file (T1FF)
 - T1 Annual individual tax return
 - T4 Statement of Remuneration Paid (T4)
 - Canada Child Tax Benefit
 - Tax-free savings account file (TFSA)
 - Postal code conversion file (PCCF)
 - Longitudinal immigration file (IMDB)
 - Crossref file



LAD – Universe and Coverage

- T1FF - universe and coverage
 - Persons who completed a T1 tax return for the year of reference or filers or family members who received CCTB (Canada Child Tax Benefits) or T4
 - Dependents identified in T1 or CCTB files
- LAD
 - Persons who completed a T1 tax return for the year of reference or family members of filers who received CCTB (Canada Child Tax Benefits) or T4



LAD - Sampling Scheme

- Only T1FF records with a SIN are sampled
- 20% random sample of these *individual* records on T1FF
- Constant unique SIN via SIN cross-referencing
- New individuals are sampled and added to the LAD each year from the T1FF. For example, young adults who enter into the labor market, and the arrival of new immigrants
- Some individuals stop having information added to their LAD record because they die or because they leave the country (emigration)
- Some individuals may be missing or have a sporadic presence on the LAD because they are late filers, non filers or non-residents

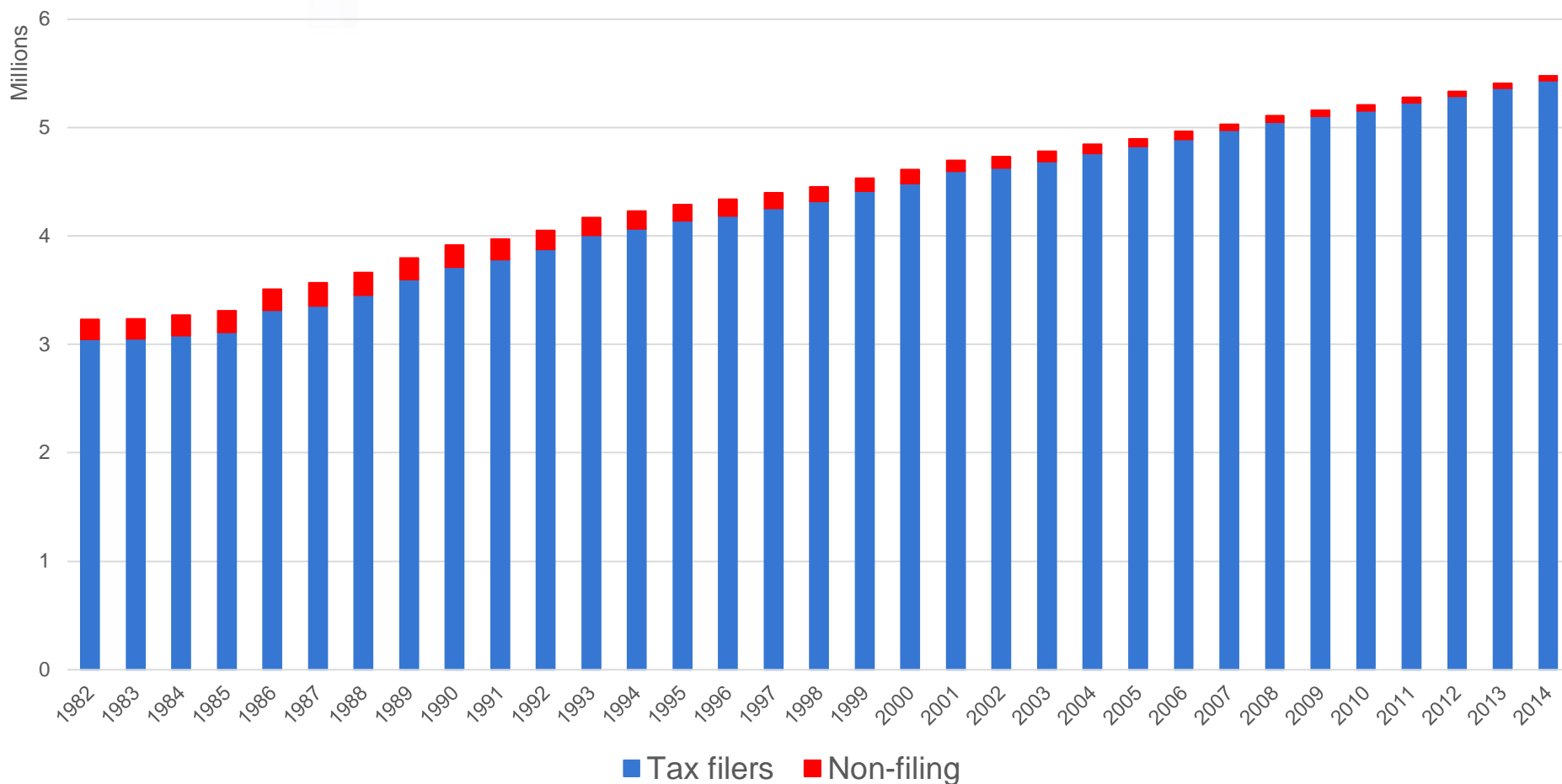


LAD - Sampling Scheme (cont')

- Tax filing rates increased through 1980s, stable after 1992
 - Incentives for individuals to file taxes
 - Federal Sales Tax Credit (1986)
 - Goods and Services Tax Credit (1992)
 - Child Tax Credit and Benefits
- These changes had a marked impact on the proportion of Canadians filing tax returns
- The quality of the T1FF (LAD) has been increasing steadily since 1982, concurrent with these changes

LAD - Sampling Scheme (cont')

Number of LAD records by year, 1982-2014





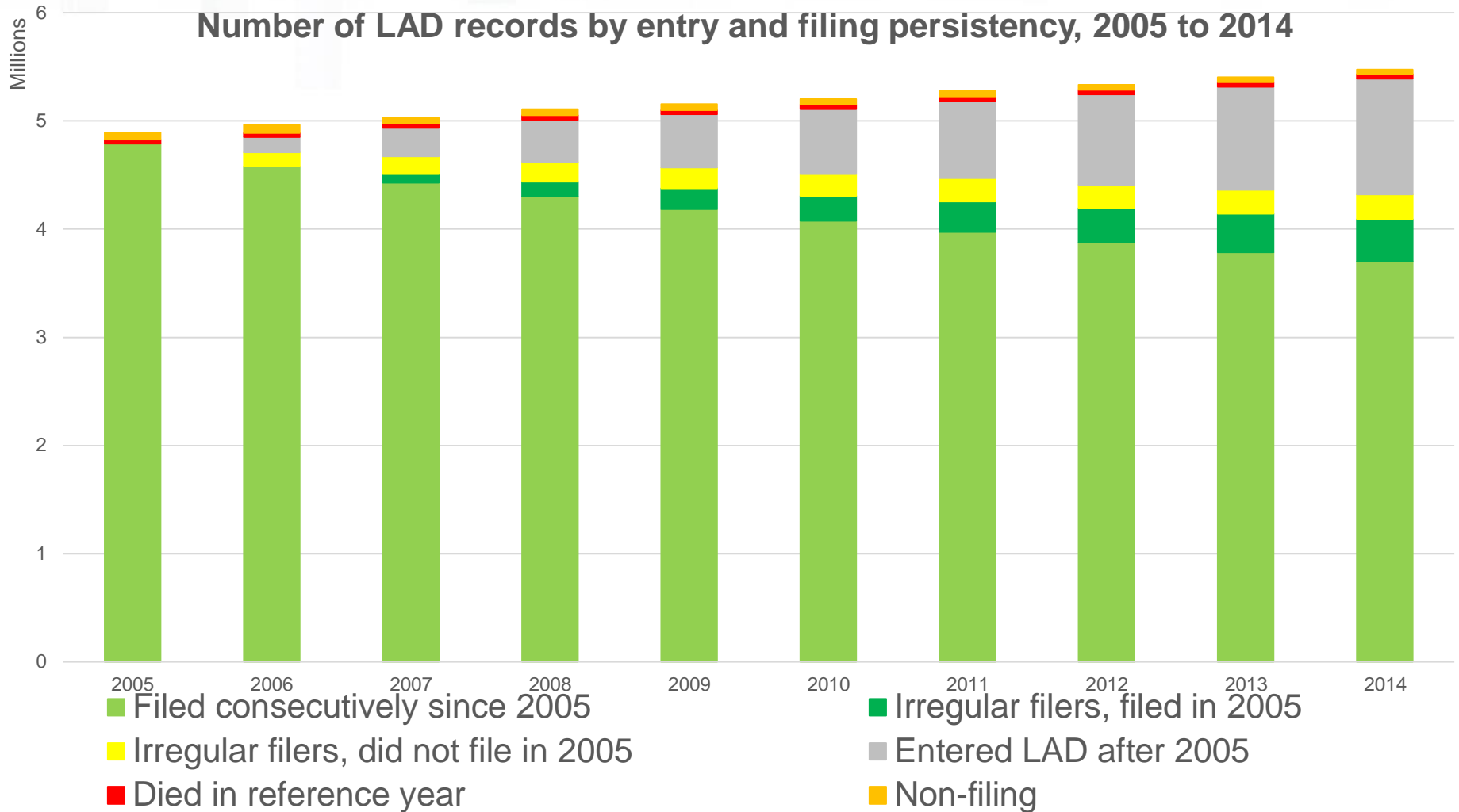
LAD - Sampling Scheme (cont')

- From 1982 to 2014:
 - 7.6 million different individuals appear in the LAD
 - 2.1 million individuals who appeared in the LAD prior to 2014, do not appear in 2014.
 - 1.14 million individuals are reported as dead, with 90 thousand reporting the year of death to be in 2014 or 2015.
- Also, about 6% of the LAD sample in 1985 were non-filers with data generated from spousal or other administrative sources. By 2014 this had dropped to less than 1%.



Filing Persistency on LAD

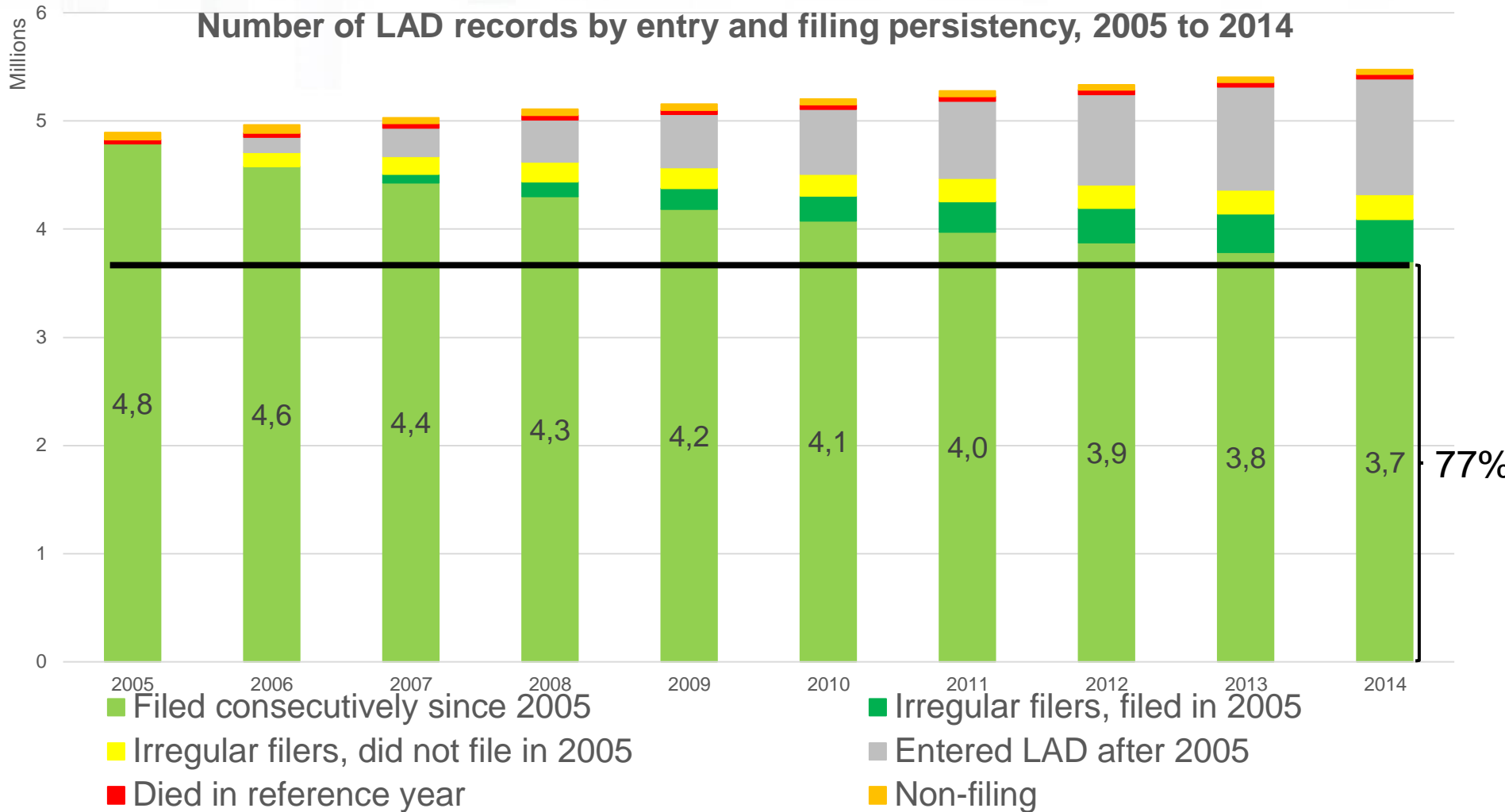
Number of LAD records by entry and filing persistency, 2005 to 2014





Filing Persistency on LAD(cont')

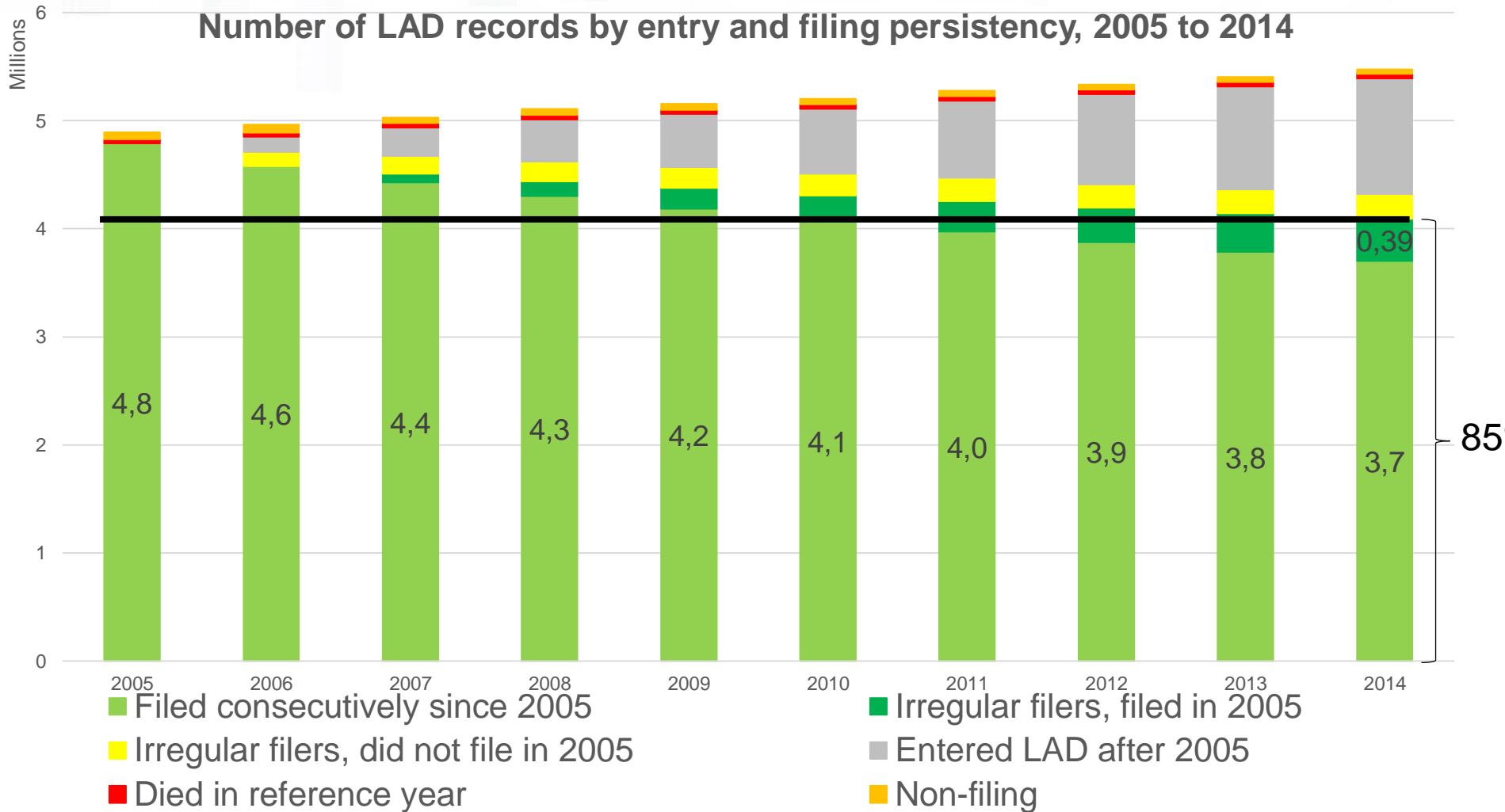
Number of LAD records by entry and filing persistency, 2005 to 2014





Filing Persistency on LAD(cont')

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

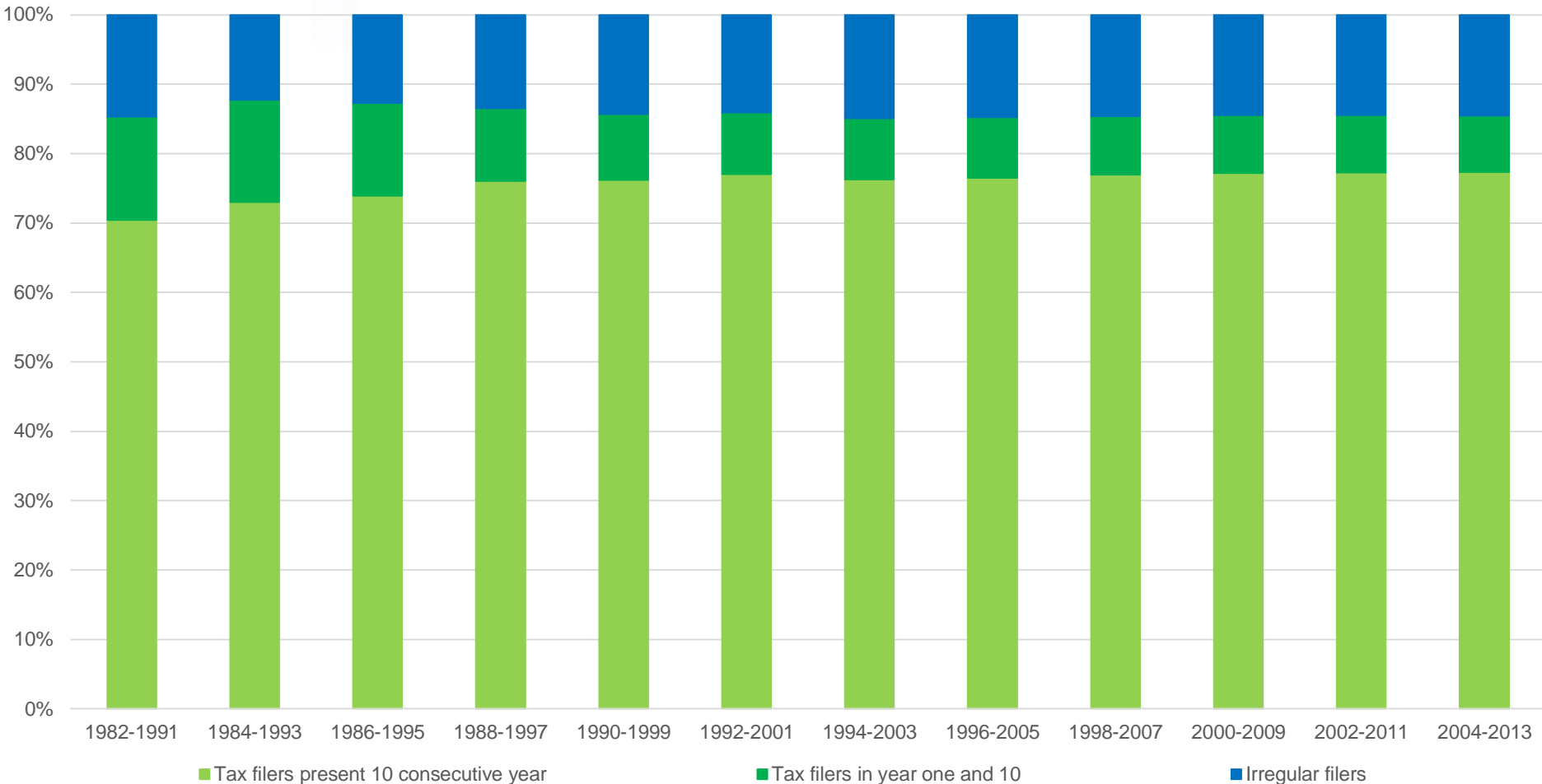


85



Filing Persistency on LAD, 1982-2014

Number of LAD records by filing persistency pattern, 2005 to 2014





Demographic Variables

- Individual demographics
 - Age, sex, year of death, marital status
- 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



Selected Market Income Variables

- 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

- 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

- Federal and provincial income tax
 - Federal and provincial income 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

- **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, 2000->)
- **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](#)

<http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=4107>



LAD data in CANSIM

- High income tables (introduced in 2013):
 - **Table 204-0001 : High income trends of tax filers in Canada, provinces, territories and census metropolitan areas (CMA), national thresholds; and**
 - **Table 204-0002 : High income trends of tax filers in Canada, provinces, territories and census metropolitan areas (CMA), specific geographic area thresholds.**

A Profile of High Income Canadians (Murphy, Roberts, and Wolfson 2007)

<http://www.statcan.gc.ca/pub/75f0002m/75f0002m2007006-eng.pdf>

- Low income dynamics tables (introduced 2015):
 - **Table 204-0101 : Low income entry and exit rates of taxfilers in Canada, provinces and census metropolitan areas (CMA);**
 - **Table 204-0102 : Low income persistence of taxfilers in Canada, provinces, census metropolitan areas (CMA); and**
 - **Table 204-0103: Low income duration of taxfilers in Canada, provinces and census metropolitan areas (CMA).**

What can we learn about low-income dynamics in Canada from the Longitudinal Administrative Databank? (Zhang, 2014)

<http://www.statcan.gc.ca/pub/75f0002m/75f0002m2014002-eng.pdf>

LAD data in CANSIM (cont')

- Family income mobility (in development)
 - **Table 204-0201: One-year income mobility of taxfilers, Canada, provinces, territories, and census metropolitan areas (CMA); and**
 - **Table 204-0202: Five-year income mobility, Canada, provinces, territories, and census metropolitan areas (CMA).**

The evolution of income mobility in Canada: Evidence from the Longitudinal Administrative Databank, 1982 to 2012 (Zhang, 2016)

<http://www.statcan.gc.ca/pub/75f0002m/75f0002m2016001-eng.pdf>

Section 2: LAD research applications



LAD Research

- Longitudinal analysis
 - Lifetime dynamics
 - Event impacts
- Cross sectional/time series
- Individual tax filer is the recommended unit of analysis
- Record linkage
 - Require linkage agreements
 - Linkage performed at head office

Examples of research questions that can be addressed using the LAD

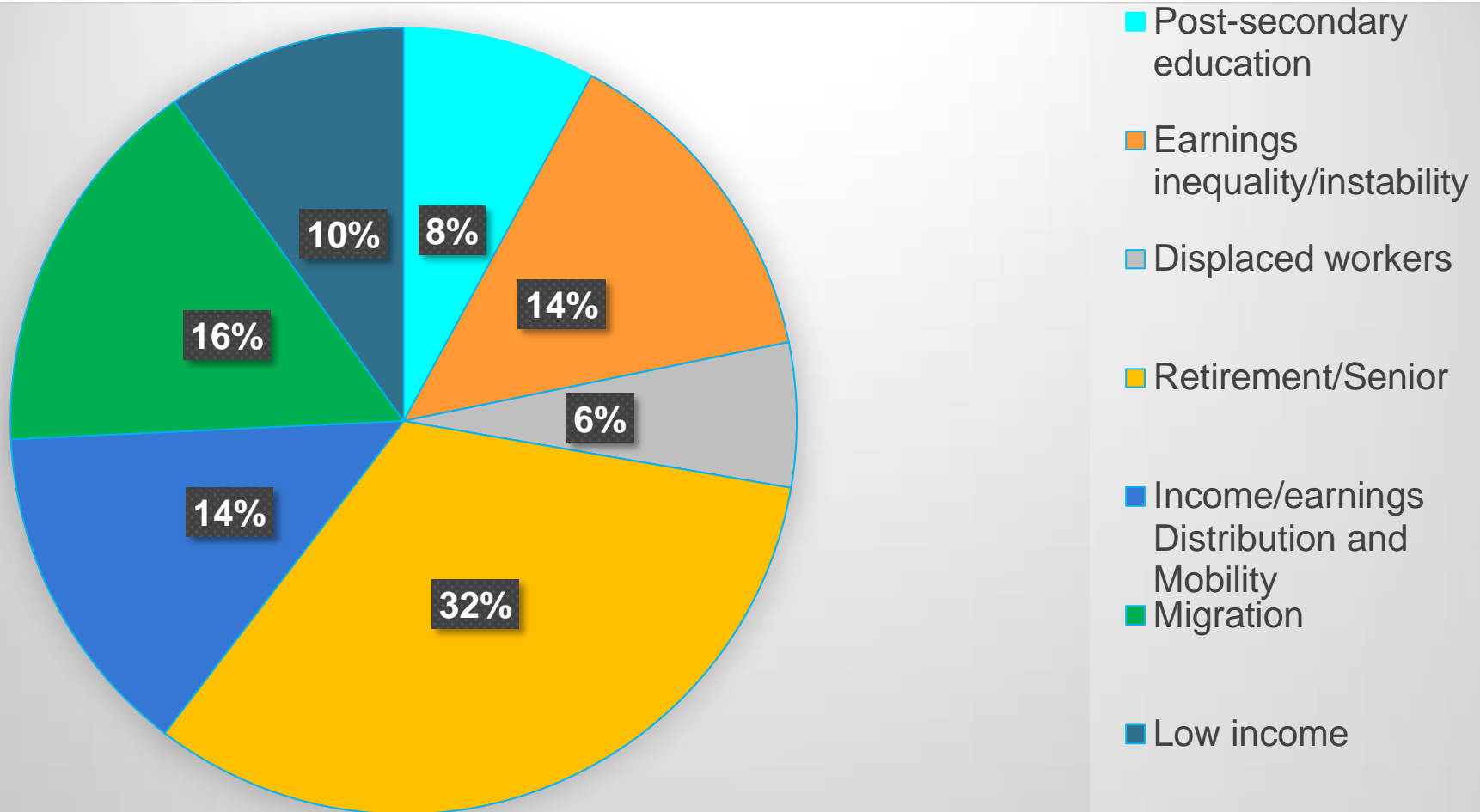
- What are the income dynamics of individuals and their families?
- Does interprovincial migration affect earnings?
- What characteristics distinguish continuous from occasional RRSP contributors?
- What is the impact of marital separation on family income?
- Do the same people experience low income 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

Labour & income related publications with the LAD



LAD in RDCs – 28 proposals

Immigration (9)

- Migration dynamics
- Ethnic enclaves
- Live-in caregivers/ economic integration

- Skills
- Domestic work experience
- Dependent applicant status

Labour Market (5)

- Maternity leave/ labour force participation
- Older workers/education
- Self-employment

Labour Market outcomes

Tax/transfer (8)

- Charitable tax credit
- Tax planning
- tax deductions (medical/disability)/ mortality
- Tax credits/fertility treatments

- WITB
- Disability benefits

Other (6)

- Retirement savings
- Inter-provincial migration
- Marriage/divorce



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 Federal Research Data Centre (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”*

Section 4: 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

brianb.murphy@canada.ca

613-617-9481

Paul Roberts

paul.roberts@canada.ca

613-716-8481

Benoît St-Jean

Benoit.st-jean@canada.ca

613-852-6475

Appendix



How the four levels of aggregation Work

Identification number :	Family identification number :	Description of individual: INDFLI	Agration level			Kids file: "K"
			Individual : "I"	Parents/spouse : "P"	Family : "F"	
1	1	8	32000	32000	32000	0
2	2	1	31000	67000	67000	0
3	2	3	36000	67000	67000	0
4	4	1	33000	62000	67000	0
5	4	3	29000	62000	67000	0
6	4	5	5000	62000	67000	5000
7	7	7	28000	28000	37000	0
8	7	5	5500	28000	37000	5500
9	7	5	3500	28000	37000	3500