Catalogue no. 89-648-X
ISBN 978-0-660-05733-0

Longitudinal and International Study of Adults Research Paper Series

Historical Data Linkage Quality: The Longitudinal and International Study of Adults, and Tax Records on Labour and Income

by James Hemeon

Release date: August 18, 2016

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

email at STATCAN.infostats-infostats.STATCAN@canada.ca

telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following toll-free numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-877-287-4369

Depository Services Program

- Inquiries line 1-800-635-7943
- Fax line 1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Standard table symbols

The following symbols are used in Statistics Canada publications:

. not available for any reference period
.. not available for a specific reference period
... not applicable
0 true zero or a value rounded to zero
0^s value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded
^{p} preliminary
^{r} revised
x suppressed to meet the confidentiality requirements of the *Statistics Act*
^{E} use with caution
F too unreliable to be published
* significantly different from reference category ($p < 0.05$)

Published by authority of the Minister responsible for Statistics Canada

An HTML version is also available.

Cette publication est aussi disponible en français.

Table of Contents

1. Introduction

Statistics Canada has collected administrative data for statistical purposes since its inception as the Dominion Bureau of Statistics in 1918. The linkage of survey data to administrative sources is becoming increasingly common as a means to reduce respondent burden, to replace survey questions with data that could otherwise be subject to respondent recall bias, and to collect data that a respondent may not feel comfortable disclosing during a survey interview. By nature, it can reduce the costs associated with survey collection. Statistical agencies around the world have been using administrative data to replace questionnaires for decades (Economic and Social Council, 2009).

Survey-collected data from the Longitudinal and International Study of Adults (LISA) is linked to tax and other administrative data sources for each year of survey collection. In addition, the LISA performs a historical linkage to tax files preceding the first year of LISA collection, providing a significant amount of longitudinal data with no added response burden, and at no extra collection cost. Though cross-sectional socioeconomic data, which represents a moment in time, can be extremely useful, the availability of a high-quality longitudinal dataset such as the LISA allows for the analysis of trends over the course of people's lives, which provides additional insight for public policy decision-making.

The purpose of this paper is to explore the data quality of the LISA historical linkage data. Heisz et al (2013), using data from a pilot study, analyzed tax data linkage rates and data accuracy, and presented the benefits of historical linkage data. The current paper will apply some of the same methods, and expand upon these findings, using the LISA dataset. More specifically, this paper will analyze the linkage rate, the degree to which it decreases as the administrative data years go back in time, and the potential of historical linkage data in analyzing phenomena that require a longitudinal data series.

2. Sample

The LISA is a sample survey, with a stratified multi-stage, multi-phase design. The sample was drawn in 2011 by selecting dwellings from 2011 Canadian Census of Population data, and is therefore a representation of the population at that time. The first LISA interviews took place in late 2011 and early 2012. For simplicity, this first collection wave is referred to as LISA 2012, and thus the LISA 2012 database is used in this study. The sample included dwellings in Canada's ten provinces, and excluded regular members of the Canadian Forces, individuals living in institutions, and individuals living on reserves and other Aboriginal settlements in the provinces. The data includes information on respondent demographics, family and household composition, literacy numeracy and problem solving skills, education and training, health, income and wealth, and labour market participation (Statistics Canada, 2014).

The database contains data from 23,926 respondents, aged 15 years and older. The file also contains 2,943 non-respondents and 5,264 non-responding children (under the age of 15). Upon completing the LISA interview, respondents were informed of the data replacement plans to link their survey data to administrative sources - a practice referred to as "informed replacement".

Following collection, Social Insurance Numbers (SIN) were retrieved for respondents from the tax databases of 2010 and 2011, using the respondent's first name, last name, date of birth, sex, marital status, address and postal code (Social Insurance Numbers are not collected directly from respondents). If no direct match was found, a SIN was linked using probabilistic linkage with the aforementioned auxiliary variables.

Once a SIN was identified, the LISA data was linked with different tax files of individuals: (i) the T1 Family File (T1FF), (ii) the statement and summary of compensation paid by employers (T4 file), and (iii) the Pension Plans in Canada file[1]. Two different types of linkage were carried out: (i) a yearly linkage (renewable for each new wave of the survey) and (ii) a historical linkage of tax data for all years going back to 1982 (T1 Family File) or 2000 (T4 file, Pension Plans in Canada file).

1. The Pension Plans in Canada is a complete annual survey of registered pension plans in Canada. Additional linkages are planned for future cycles.

3. Results

3.1 Linkage rate between 1982 and 2011

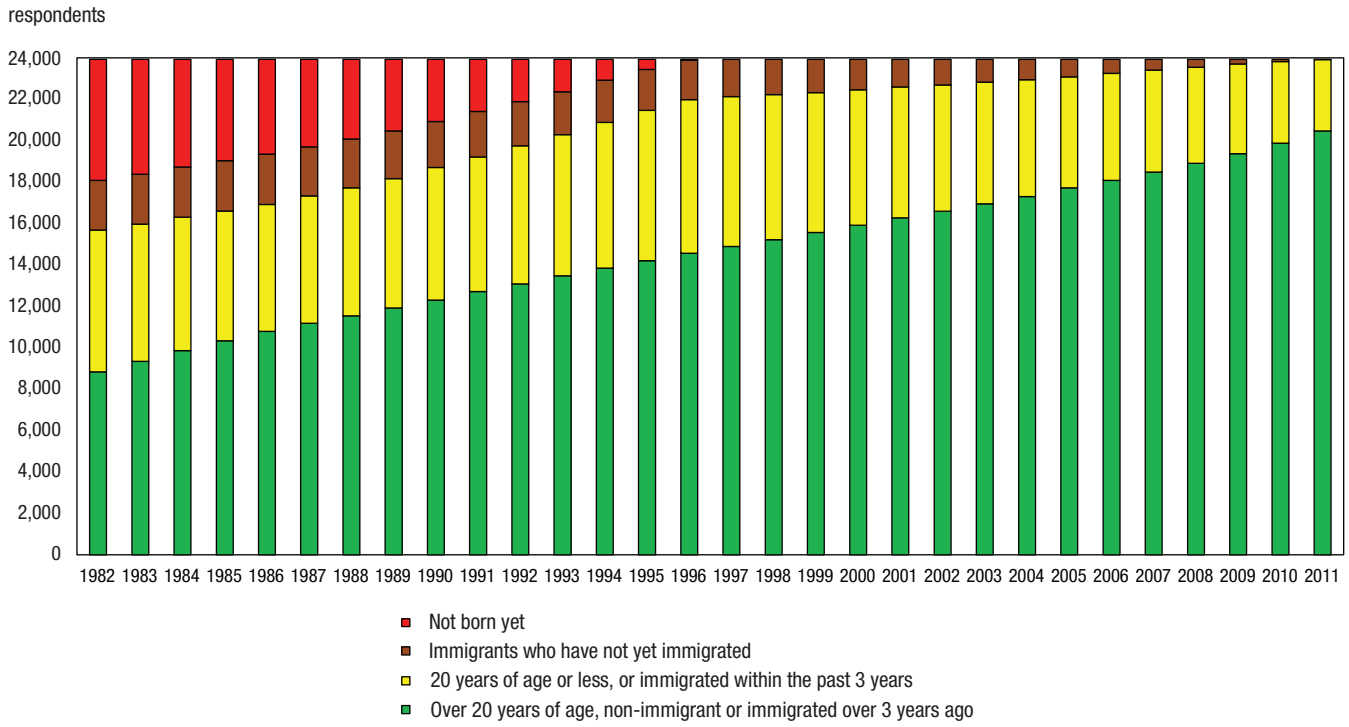
Data linkage fails when no linkage key can be established, or the linkage key does not find a match in the administrative data file.

As indicated above, SIN codes were retrieved for individuals using the tax databases of 2010 and 2011. If no SIN was found during this linkage process, this may be attributed to a respondent having not filed a personal tax return during those two years, or it may be due to no suitable link being found in the probabilistic linkage. For 7.5% of LISA respondents, a linkage was attempted but no SIN was found. Of these, 55.1% were 17 years of age or less, and 64.9% were 20 years of age or less. Therefore, the majority of these cases can likely be attributed to respondents being young, and not having established a need to file a tax return.

In historical data linkage, there is the additional problem that a person's SIN may change over time. If the respondent's SIN was not consistent over time, the linkage will fail when the SIN can no longer be found. Social Insurance Number (SIN) is a relatively stable linkage key; however, in some cases, it changes over time. For example, immigrants to Canada are assigned a temporary SIN on their arrival in Canada, and are then assigned a permanent SIN. If a SIN cannot be found for this reason, the longitudinal dataset could be missing information, since there may have been an income tax return filed during the earlier years, but it could not be associated with the respondent.

Moreover, in historical linkage there is the additional problem that a person, as we go back to older administrative files, might fall out of the administrative file because they become too young to file, or if the respondents were immigrants, they might have been living in another country, or had not yet established stable filing patterns using a permanent SIN code in Canada. The LISA sample includes respondents as young as 15 years of age (as of 2011), while the historical linkage to T1FF data includes 30 years of tax data, in the 2012 LISA release. Therefore, the availability of tax data precedes the year of birth for some LISA respondents, and precedes the year of immigration to Canada for others. For example, in 1982, 24.4% of LISA respondents were not yet born, 10.0% had not yet immigrated to Canada, and 28.6% were 20 years of age or less or had immigrated to Canada within the past 3 years. By 1997, all LISA respondents had been born, 7.5% had not yet immigrated to Canada, and 30.2% of respondents were 20 years of age or less or had immigrated within the past 3 years (see Figure 3.1-1). Therefore, a linkage is impossible, or unlikely, for a subset of the LISA sample during some years, indicating that we would expect historical linkage rates to decline as we go back in time.

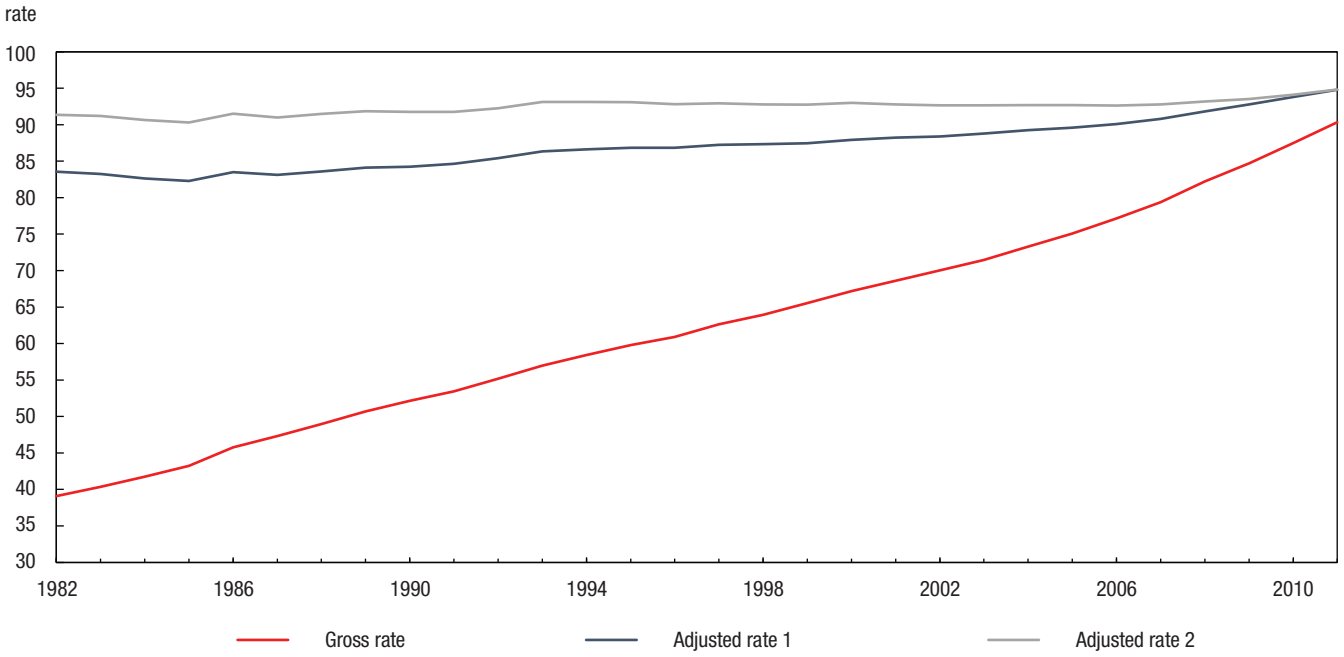
Figure 3.1-1
Characteristics of the LISA sample by year, 1982 to 2011



Source: Longitudinal and International Study of Adults (2012).

A linkage rate was calculated between survey respondents and the T1FF data files for all years from 1982 to 2011, to determine the extent of its decline going back over time. Three different linkage rates were calculated, including those respondents for whom no SIN was found: (1) a gross rate (using all respondents in the sample), (2) an adjusted rate using a sample excluding respondents under 20 years of age in a given tax year, and (3) a second adjusted rate using a sample excluding respondents under 20 years of age in a given tax year, as well as immigrants landed in the three years preceding a given tax year (see Figure 3.1-2). The adjustment based on age reflects the fact that this group is less likely to produce an income tax return during a given year. The adjustment based on immigrant status reflects the fact that this group is unlikely to have filed Canadian taxes in pre-immigration years, and three years is chosen to give the immigrant respondents time to establish stable filing patterns with a permanent SIN.

Figure 3.1-2
Linkage rate of the LISA to the T1 Family File (T1FF), 1982 to 2011



Source: LISA (2012) and linked data from the T1FF (1982 to 2011).

Results in Figure 3.1-2 indicate that the linkage rate decreases going back in time, regardless of the sample used for the calculation. However, the most significant reduction in the linkage rate occurs when the calculation is based on a sample without exclusions, in which case it decreases significantly (1.76% per year, on average), from 90.3% in 2011 to only 39.1% in 1982. Excluding those respondents under the age of 20 in a given tax year, the rate decreases much less (0.39%, on average), from 94.8% in 2011 to 83.6% in 1982, remaining over 82% for all years. When also excluding respondents who immigrated to Canada in the 3 years prior to a given tax year, the decrease in linkage rate is small (0.12% per year, on average) from 94.8% in 2011 to 91.3% in 1982, remaining over 90% for all years. In other words, when the sample is limited to the population that is likely to file a tax return and have a constant SIN over time, the linkage rate remains high across all years.

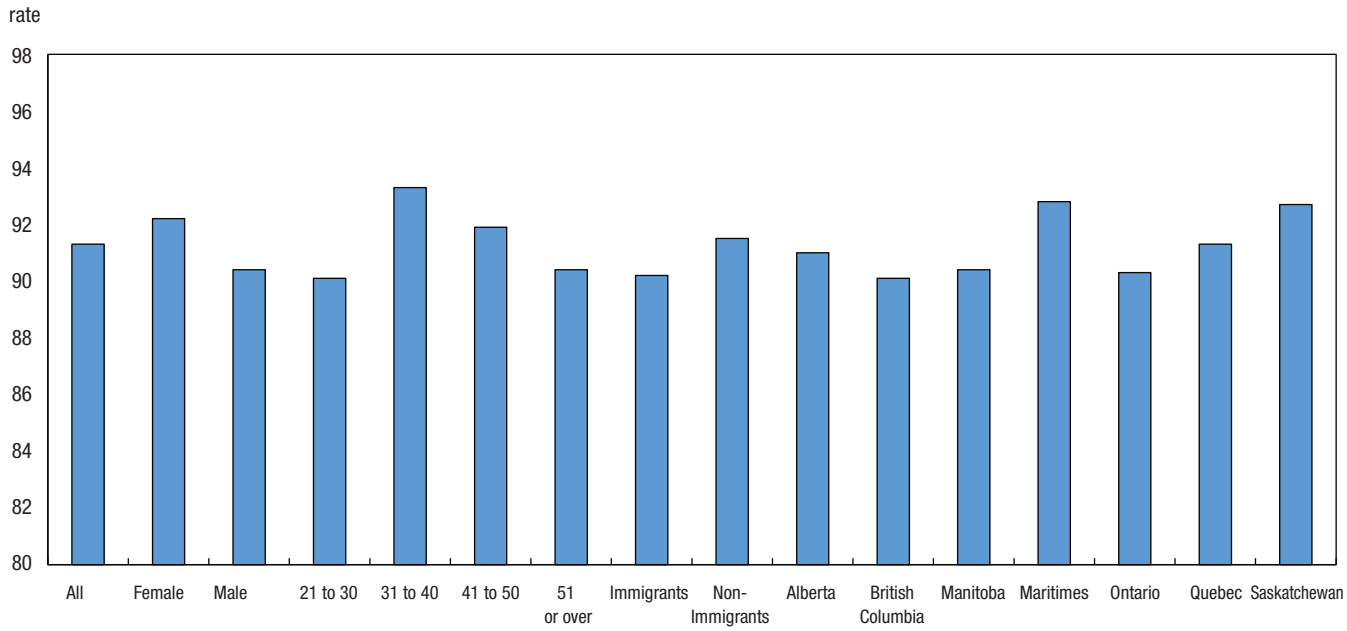
In the LISA sample, 8.5% of respondents were not linked to tax data for any year between 1982 and 2011. This group is comprised of respondents for whom no SIN was found, and those who opted out of linkage[2].

3.2 Linkage rate examined

To analyze whether the linkage data is representative of the sample, several linkage rates were calculated for respondents who were over the age of 20 in a given tax year and were non-immigrants or who had immigrated over 3 years prior to the given tax year (see 'Adjusted rate 2' in Figure 3.1-2). In addition to the overall 'Adjusted 2' linkage rate, linkage rates were calculated for sub-samples by age (in a given tax year), sex, immigrant status, and province of residence (as of 2011) for the tax years 1982, 1996, and 2011. Because of the slight decrease in linkage rate in 1985, as shown in Figure 3.1-2, 1985 was also included (see Figure 3.2-1 – Figure 3.2-4). Note that these rates are based on unweighted frequency counts, as the objective of the present paper is to analyze linkage quality, rather than representativeness to the population. For the total number of observations linked in 1982, 1985, 1996, and 2011, see Table 3.2-5.

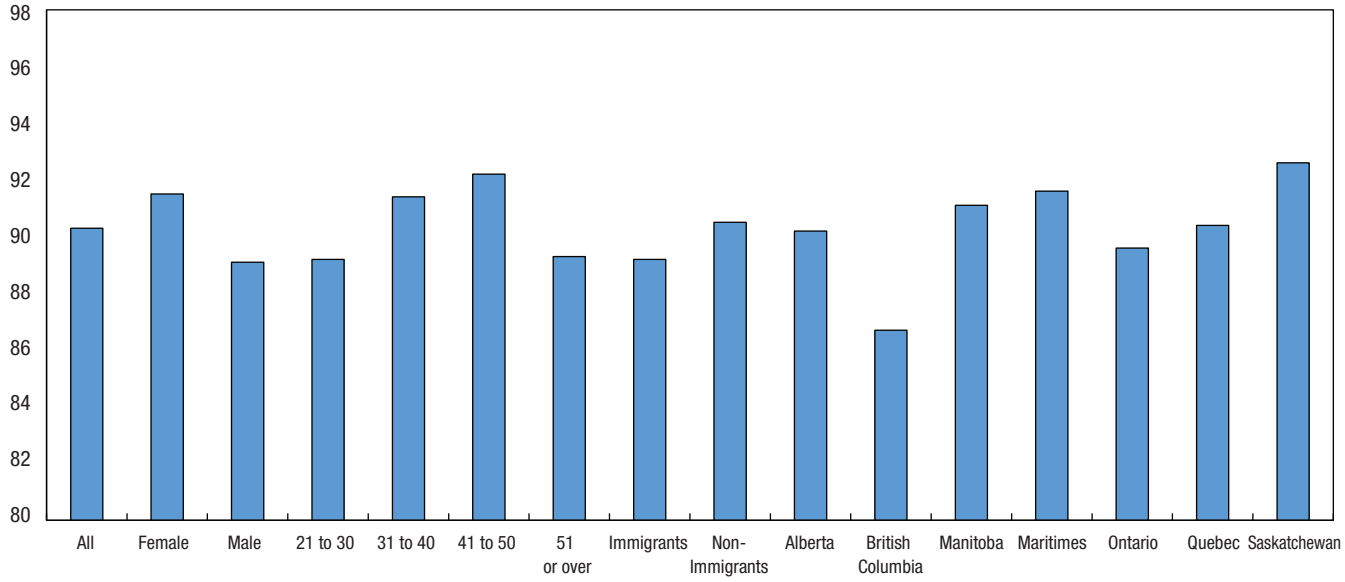
2. According to Statistics Canada's policy on record linkage, respondents may opt out of record linkage. This affected a small number of LISA respondents.

Figure 3.2-1
Adjusted linkage rate 2 of LISA demographic sub-groups, 1982



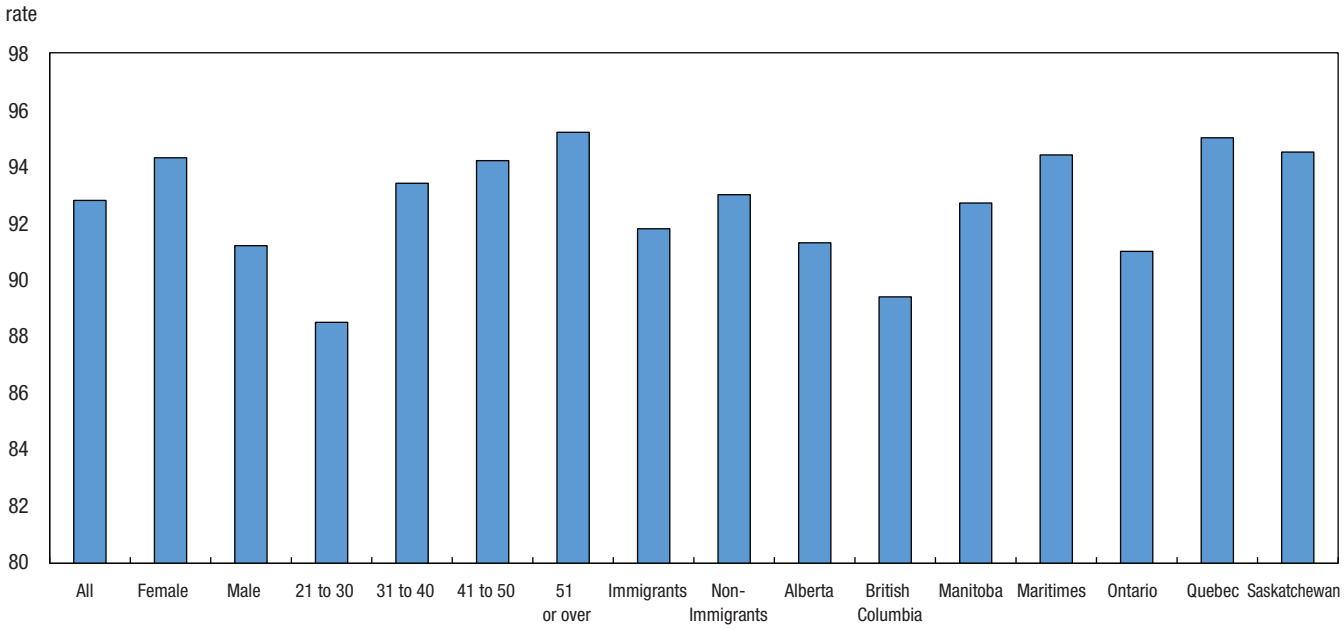
Source: LISA (2012) and linked data from the T1FF (1982).

Figure 3.2-2
Adjusted linkage rate 2 of LISA demographic sub-groups, 1985



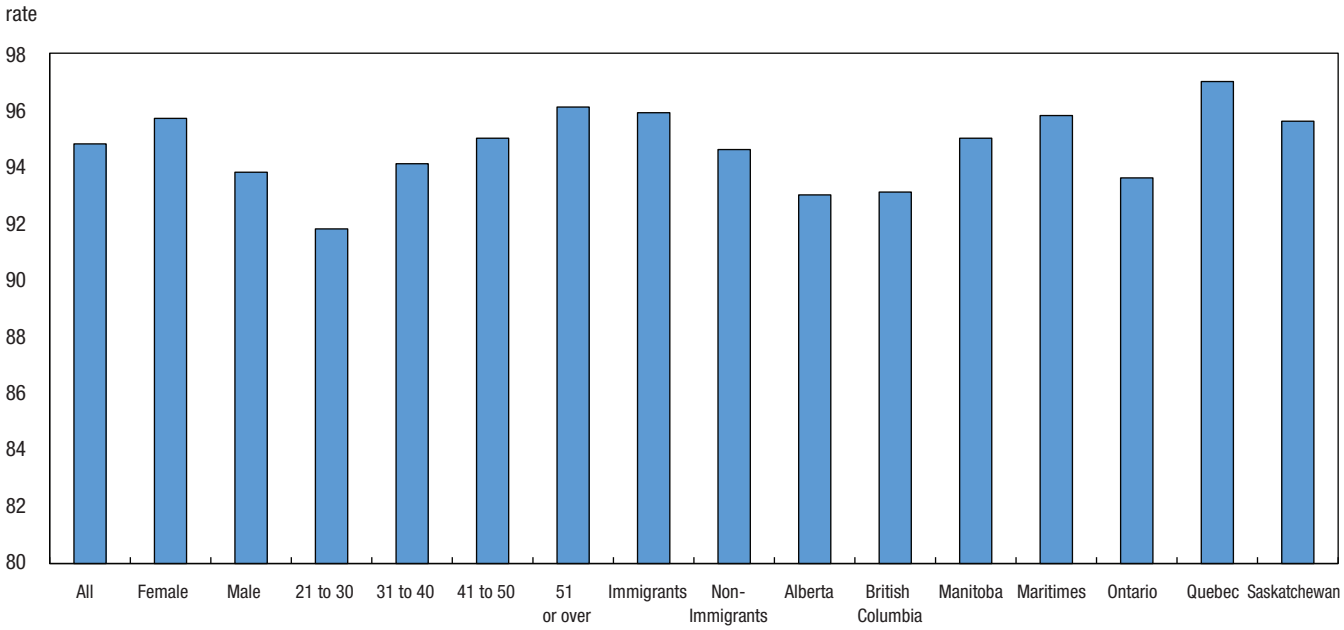
Source: LISA (2012) and linked data from the T1FF (1985).

Figure 3.2-3
Adjusted linkage rate 2 of LISA demographic sub-groups, 1996



Source: LISA (2012) and linked data from the T1FF (1996).

Figure 3.2-4
Adjusted linkage rate 2 of LISA demographic sub-groups, 2011



Source: LISA (2012) and linked data from the T1FF (2011).

The overall linkage rates for 1982, 1985, 1996, and 2011 are 91.3%, 90.3%, 92.8%, and 94.8%, respectively.

Males have a slightly lower linkage rate than females in all years, with 95.7% of females linked in 2011, compared to 93.8% of males.

The results suggest that the linkage rate generally increases with age, similar to findings by Li et al (2006); however, the rate for the youngest respondents, while lower, is still reasonably high at 91.8%. Respondents aged 51 or above in a given tax year have the highest linkage rate of all age groups in 2011, but this rate is lower in earlier years, as the number of respondents in that age group in a given tax year decreases sharply from 8,879 in 2011 to 544 in 1982. This is to be expected, as the '51 or over' age group in 1982 was aged 81 or over at the time of their LISA interview, and may have been less likely to file a tax return in 2011 or 2010 (and therefore, also less likely to find a linkage SIN).

Immigrants have a slightly higher linkage rate than non-immigrants in 2011. Respondents who resided in the provinces of Ontario and British Columbia in 2011 generally have slightly lower linkage rates, when compared to other provinces. This is particularly apparent in 1985, where the linkage rate for British Columbia was 86.7%, a 3.5% drop from 1982. The reason for this decrease is unclear. Upon analysis of the 312 total respondents who had a linkage in 1982 but a missed linkage in 1985, no trend was found among age, sex, or immigrant status.

Table 3.2-5
LISA sub-group linkage observations (1982, 1985, 1996, 2011)

Category	1982	1985	1996	2011
All	8,068	9,334	13,514	19,403
Female	4,153	4,878	7,220	10,247
Male	3,915	4,456	6,294	9,156
21-30	3,991	4,168	2,911	3,040
31-40	2,460	3,102	4,630	3,043
41-50	1,073	1,246	3,581	4,441
51 or over	544	818	2,392	8,879
Immigrants	889	1,038	1,829	3,558
Non-Immigrants	7,179	8,296	11,685	15,845
Alberta resident 2011	735	860	1,313	2,056
BC resident 2011	860	948	1,396	2,129
Manitoba resident 2011	576	679	963	1,373
Maritimes resident 2011	2,043	2,379	3,278	4,138
Ontario resident 2011	1,716	1,963	2,930	4,493
Quebec resident 2011	1,593	1,869	2,734	3,920
Saskatchewan resident 2011	545	636	900	1,294

Source: LISA (2012) and linked data from the T1FF (1982, 1985, 1996, 2011)

3.3 Balanced Panels

A longitudinal dataset requires data on its sample over a period of time. A break in the continuity of data could limit its usability for researchers for some purposes. A longitudinal dataset is said to be a "balanced panel" when all observations (respondents) are present in the dataset in all periods (in the case of LISA, each year). For historical linkage, a balanced panel requires the linkage of a tax record for each year.

If a researcher required balanced panels from those who were likely to file a tax return and likely to have a constant SIN over time (see 'Adjusted rate 2' in Figure 3.1-2), a 30-year panel could be constructed with a 74.3% linkage rate, and would contain 6,564 respondents (using years 1982-2011). If a researcher required a 25-year panel, it could be constructed with a 78.1% linkage rate (1987-2011), containing 8,735 respondents. A 20-year panel could be constructed with an 82.1% linkage rate (1992-2011), containing 10,733 respondents. A 15-year panel could be constructed with an 84.5% linkage rate (1997-2011), containing 12,579 respondents. A 10-year panel could be constructed with an 86.7% linkage rate (2002-2011), containing 14,371 respondents. If a researcher required only a 5-year panel, it could be constructed with an 89.7% linkage rate (2007-2011), and would contain 16,568 respondents (see Appendix A). Thus, LISA can be used to create long, balanced panels of a size sufficient for many analyses.

3.4 Comparison of earnings from T1FF and T4 files

One way to verify the reliability of administrative files is to compare their data to the values in administrative files from another source.

Earnings amounts in T1FF data files and in T4 data files were compared for the period from 2000[3] to 2011. The results show that the majority of cases - 97% each year, on average - present a similar earnings situation in both the T1 Family File (T1FF) and the employer file (T4) (see Table 3.4 1). In other words, only 3% (approximately) of cases show earnings in one file without showing earnings in the other. Approximately 71% of cases show earnings in both the T1FF and T4. Another 26% of respondents had $0 earnings in the T1FF and no T4 information. The number of cases with no T1FF information and a T4 earnings value of $0 is insignificant.

Table 3.4-1
Source of earnings, T1FF and the T4 file, 2000 to 2011

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
	%											
Both sources - earnings (T1FF and T4 >= $0)	72.7	72.2	71.9	71.6	71.0	71.3	71.0	70.9	70.5	69.7	69.7	70.1
Single source - earnings (either T1FF or T4 > $0)	3.1	3.2	3.1	3.3	3.5	3.2	3.6	3.6	3.5	2.9	2.3	1.3
Single source - no earnings (T1FF = $0, no T4)	24.2	24.6	25	25.2	25.4	25.5	25.4	25.5	26	27.4	28.1	28.6
Single source - no earnings (no T1FF, T4 = $0)	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0

Source: LISA (2012) and linked data from the T1FF and T4 files (2000-2011).

The vast majority of cases show earnings in both the T1FF and T4 files, or $0 earnings in the T1FF filing and no T4 information (indicating agreement between the two files). When earnings values are reported in both the T1FF and T4 files, approximately 95% of cases show a difference of no more than one dollar between data sources (see Table 3.4-2). It should be noted that, while T4 earnings values contain cents values, T1FF earnings do not contain cents. Approximately 98% of cases show a difference of no more than one thousand dollars between data sources.

3. Information from T4 files was not available for years prior to 2000.

Table 3.4-2
Difference in earnings reported in the T1FF and the T4 file, 2000 to 2011

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
	%											
$0.01 to $1.00	92.7	95.0	94.2	94.6	94.5	94.4	94.2	94.7	95.1	95.7	96.1	96.5
$0.01 to $100.00	2.2	1.9	1.7	1.6	1.6	1.8	1.8	1.4	1.3	1.0	0.8	1.0
$100.01 to $1000	1.8	1.5	1.8	1.9	1.8	1.7	1.8	1.9	1.6	1.5	1.3	1.1
T1FF < T4 by over $1000	2.6	0.9	1.5	1.2	1.3	1.3	1.5	1.3	1.3	1.2	1.2	0.9
T1FF > T4 by over $1000	0.7	0.7	0.8	0.6	0.9	0.9	0.7	0.7	0.8	0.6	0.6	0.5

Source: LISA (2012) and linked data from the T1FF and T4 files (2000-2011).

From 2000 to 2011, the difference in median employment earnings calculated from the two data sources is, on average, $116 (see Table 3.4-3). The median earnings, when present in the T1FF and T4 sources, are very similar, which suggests that the T1FF linkage data is accurate, and that the T4 is also present where expected.

The median earnings, when found in one data source only, are significantly lower than median earnings when found in both data sources. Examining this in more detail shows that the majority of values from a single source are attributed to T1FF values of $0, in which case a T4 may not be available. The majority of single source earnings values greater than $0 are attributed to T4 earnings values with no T1FF information.

Table 3.4-3
Median employment earnings[1] of the T1FF and T4 file

	Both sources			Single source, >$0			
	T1FF		T4	T1FF		T4	
Year	N	Median	Median	N	Median	N	Median
2000	12,000	31,979	32,584	84	7,248	430	6,294
2001	12,192	33,109	33,152	78	11,396	462	6,822
2002	12,377	32,517	32,634	67	11,545	461	8,273
2003	12,601	32,727	32,815	57	6,559	515	5,374
2004	12,808	32,814	32,851	141	26,926	494	5,940
2005	13,162	33,317	33,330	89	9,903	495	5,901
2006	13,523	33,541	33,653	98	11,574	588	5,311
2007	13,917	33,816	33,942	78	12,624	628	4,787
2008	14,287	33,983	34,052	123	12,612	589	5,564
2009	14,473	33,844	34,009	113	14,004	489	5,728
2010	14,853	33,322	33,410	94	10,724	390	8,912
2011	15,290	33,073	33,133	96	11,457	181	19,543

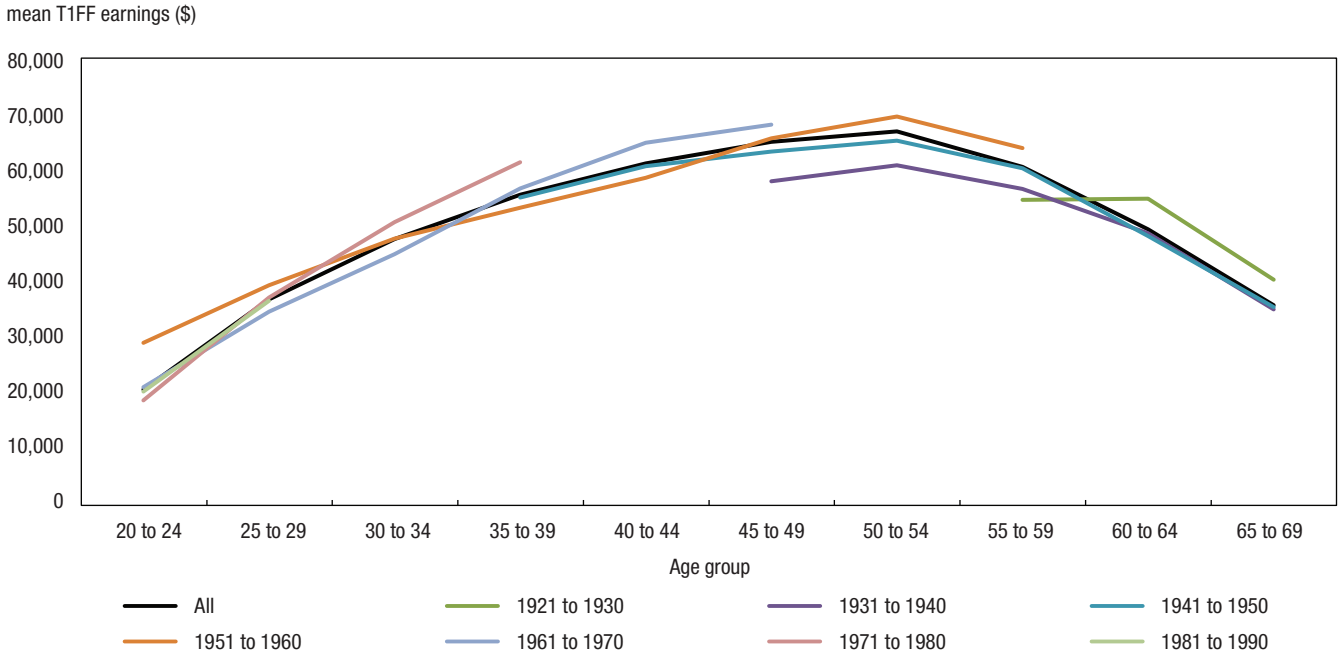
1. Earnings expressed in 2011 constant dollars
Source: LISA (2012) and linked data from the T1FF and T4 files (2000-2011).

3.5 Profiles of earnings by age and sex

To demonstrate the potential of LISA linked data in creating a long data series, an age-earnings profile for each sex was created for different birth cohorts.

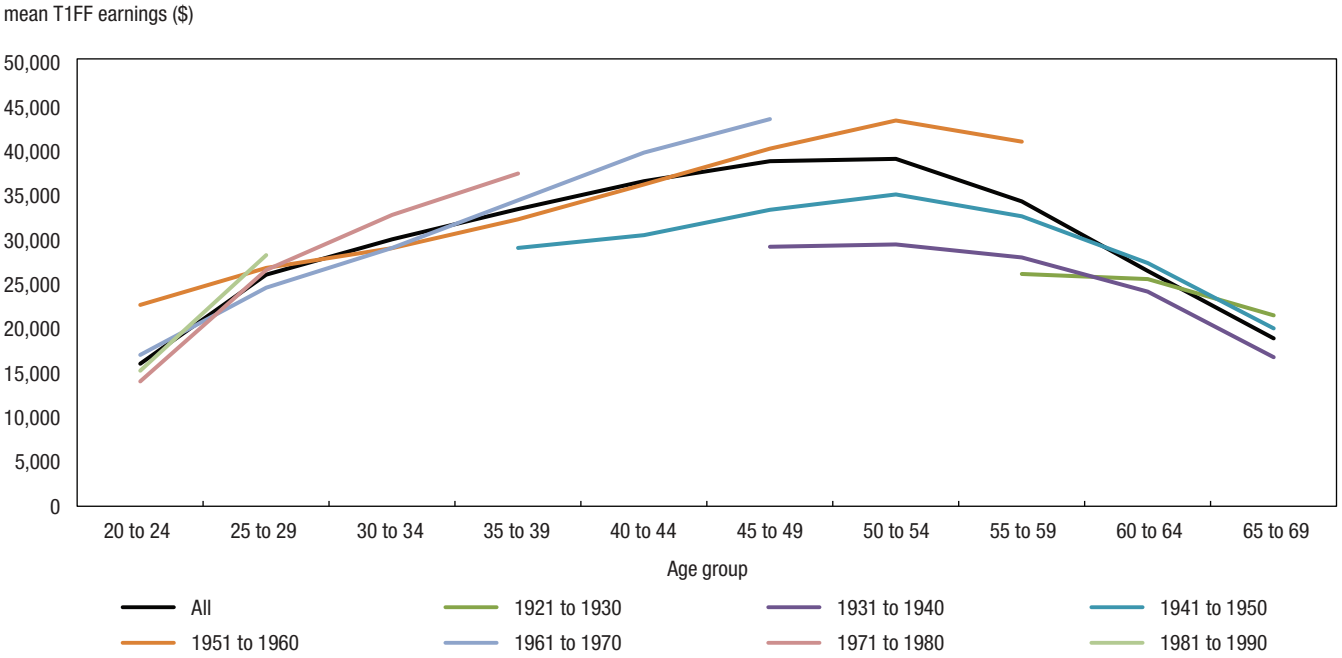
Due to the similar earnings found when comparing the T1 Family File (T1FF) and employers' (T4) files, the earnings used were those from the T1FF, thus providing a longer data series. The sample was divided into seven birth year groups, at five-year intervals, for which the change in employment earnings was tracked by age.

Figure 3.5-1
Earnings profile for males, by age group and birth year cohort



Source: LISA (2012) and linked data from the T1FF from 1982 to 2011.

Figure 3.5-2
Earnings profile for females, by age group and birth year cohort



Source: LISA (2012) and linked data from the T1FF from 1982 to 2011.

The age-earnings profiles by birth year cohort show a trend of lower earnings at the beginning of career, and a faster growth in earnings, for workers in more recent cohorts (e.g., 1971-1980 cohort versus 1961-1970 cohort). There is also a trend of higher peak earnings for workers in more recent cohorts (e.g., 1951-60 cohort versus 1941-50 cohort). These trends are similar to those found in related literature (Vijay et al, 2014; Beach and Finnie, 2004).

Most notably, females show higher career earnings progression for each successive cohort. Females aged 50-54 in the 1951-1960 cohort earned 47% more than females in the 1931-1940 cohort did when they were aged 50-54, compared to a 14% increase in earnings between the respective cohorts for males aged 50-54. These support previous findings of a greater increase in female earnings, relative to males (Williams, 2010; Suh, 2010; Blau and Kahn, 2006).

4. Conclusions

This study provides a partial assessment of the quality of administrative data from 1982 to 2011 which was linked to 2012 LISA data. In particular, linkage rates were analyzed, the data was compared across administrative sources, and the ability to use the linkage data to analyze selected phenomena requiring a longitudinal data series was assessed.

Linkage rates to administrative data were examined in multiple ways, and the results indicate that the linkage rates are high, with more than 90% of LISA respondents aged 15 and over being linked in 2011. Linkage rates to prior years were also high, especially when rates were calculated for respondents who were aged 20 and over and who had immigrated at least three years prior to the year linked. Among key demographic sub-groups, the linkage rate remains high. However, data users must consider that certain sub-groups may not have as many observations historically. For example, immigrants will only have linkage data on or after the year in which they entered Canada.

The results also suggest that the data obtained by the historical linkage produces data that is coherent across administrative data sources, and can be used for observing phenomena that require a longitudinal data series, as well as long panel datasets.

As the data is based a sample drawn in 2011, it is most appropriately used for studies describing the life-course histories of that particular cohort, as opposed to cross-sectional referencing to individual years. The linkage allows for analysis of retrospective income data that would otherwise have not been possible without 30 years of survey collection, or without introducing a significant recall bias. Furthermore, upcoming data releases will be coupled with additional years of LISA survey data, which will increase the analytical potential of the dataset.

Bibliography

Beach, C. and Finnie, R. (2004), "A Longitudinal Analysis of Earnings Change in Canada", Analytical Studies Branch. Research Paper, Statistics Canada.
http://www.statcan.gc.ca/pub/11f0019m/11f0019m2004227-eng.pdf

Blau, Francine D.; Kahn, Lawrence M. (2006) "The US gender pay gap in the 1990s: slowing convergence", IZA Discussion Papers, No. 2176
www.econstor.eu/dspace/bitstream/10419/34046/1/51436131X.pdf

Economic and Social Council. (2009) "Main Results Of The UNECE-UNSD Survey On The 2010 Round of Population and Housing Censuses", Economic Commission for Europe, Conference of European Statisticians. Twelfth Meeting, 28-30 October 2009.
http://unstats.un.org/unsd/censuskb20/Attachment459.aspx

Gill, Vijay, Knowles, James, Stewart-Patterson, David. (2014). "The Buck Stops Here: Trends in Income Equality Between Generations", Ottawa: The Conference Board of Canada.

Heisz, Andrew, Langevin, Manon, Randle, Jeffrey. (2013). "Historical data linkage of tax records on labour and income: The case of the Living in Canada Survey pilot". Statistics Canada Catalogue no. 89-648-X (2).
http://www.statcan.gc.ca/pub/89-648-x/89-648-x2013002-eng.htm

Li, Bing, Quan, Huge, Fond, Andrew, Lu, Mingshan. (2006) "Assessing record linkage between health care and Vital Statistics databases using deterministic methods", BioMed Central Health Services Research 2006, 6:48.
http://www.biomedcentral.com/1472-6963/6/48/

Sakshug, Joseph W., Couper, Mick P., Ofstedal, Mary B., Weir, David R. (2012) "Linking Survey and Administrative Records: Mechanisms of Consent", Sociological Methods & Research, 41(4) 535-569.
http://smr.sagepub.com/content/41/4/535.full.pdf

Statistics Canada. (2014). LISA Detailed information for 2014 (Wave 2).
http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5144

Suh, Jingyo. (2010) "Decomposition of the Change in the Gender Wage Gap", Research in Business and Economics Journal, 2-18.
http://www.aabri.com/manuscripts/08076.pdf

Williams, Cara. (2010) "Women in Canada: A Gender-based Statistical Report. Sixth Edition". Economic Well-being. Statistics Canada Catalogue no. 89-503-X. p. 32-33.
http://www.statcan.gc.ca/pub/89-503-x/2010001/article/11388-eng.pdf

Appendix A.

LISA Balanced Panels

		5yr	10yr	15yr	20yr	25yr	30yr
1982	%	85.3%	80.6%	78.9%	77.3%	75.9%	74.3%
	N	7,531	7,123	6,966	6,831	6,700	6,564
1987	%	85.7%	83.2%	81.5%	79.7%	78.1%	
	N	9,582	9,307	9,112	8,918	8,735	
1992	%	88.8%	86.2%	84.0%	82.1%		
	N	11,606	11,265	10,978	10,733		
1997	%	89.3%	86.6%	84.5%			
	N	13,305	12,889	12,579			
2002	%	89.2%	86.7%				
	N	14,798	14,371				
2007	%	89.7%					
	N	16,568					

Source: LISA (2012) and linked data from the T1FF from 1982 to 2011.