Catalogue no. 89-648-X
ISSN 1927-0100
ISBN 978-0-660-33929-0

Longitudinal and International Study of Adults Research Paper Series

Sample selection in tax data sets of intergenerational links: Evidence from the Longitudinal and International Study of Adults

by Gaëlle Simard-Duplain and Xavier St-Denis

Release date: March 17, 2020

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 numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-514-283-9350

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.

Published by authority of the Minister responsible for Statistics Canada

© Her Majesty the Queen in Right of Canada as represented by the Minister of Industry, 2020

All rights reserved. Use of this publication is governed by the Statistics Canada Open Licence Agreement.

An HTML version is also available.

Cette publication est aussi disponible en français.

Table of contents

Sample selection in tax data sets of intergenerational links: Evidence from the Longitudinal and International Study of Adults

by **Gaëlle Simard-Duplain** and **Xavier St-Denis**

Abstract

Administrative data sets have become increasingly popular sources of information to study mobility across generations. However, the inclusion of parent-child pairs depends on the primary purpose for which the data was collected. In the case of tax records, both parents and children must have worked and filed their taxes, and the children's labour market entry must have happened before they left the parental home. This paper documents selection in samples of parent-child pairs constructed from personal income tax records from Canada, and discusses implications for intergenerational research. It takes advantage of the fact that Statistics Canada's Longitudinal and International Study of Adults (LISA) includes both survey and administrative data to inform the nature and severity of the resulting sample selection. Results show that respondents who were successfully linked to their parents are more educated, and are more likely to have grown up in better educated, nuclear families. However, correcting for sample selection suggests that there is no bias in unadjusted estimates.

1 Introduction

This paper uses a new linkage between survey data and administrative records to assess the effect of using tax data on sample selection in intergenerational studies. For example, these include studies of intergenerational mobility interested in the association between the socioeconomic status of parents (income, education, etc.) and of their children, at different points in their life course.

The long panels and availability of parent-child links necessary to investigate intergenerational dynamics have made administrative records an appealing option for researchers. As they have become increasingly accessible, the understanding of the transmission of (dis)advantage across generations has also increasingly come to rely on these sources of information. This study explores the extent to which intergenerational data sets constructed from tax records exhibit sample selection, and discusses the implications for research.

This study uses the Longitudinal and International Study of Adults (LISA), a relatively new Canadian survey, for which respondents were linked to their personal income tax records. The data were recently extended to include the personal income tax records of past and present family members of LISA respondents. As such, they bear similarities with traditional administrative sources, while allowing users to characterize sample selection due to the broad set of variables included in the survey itself. More specifically, this study contrasts LISA respondents for whom a parental link could be established to those for whom no parent was found. It focuses on individuals aged born between 1963 and 1995.

The results show that respondents included in the intergenerational sample are more educated on average, report more educated parents, and are more likely to have grown up in nuclear families. However, estimates from a simple selection correction exercise suggest that these differences do not introduce substantial bias in the estimation of intergenerational income correlation.

The paper then turns to the potential for interactions between the selection described in this study, and the various decisions researchers face when carrying intergenerational studies. Generally, researchers who engage in empirical work must choose to construct their sample and define their key variables to best answer the question of interest. In turn, these decisions can matter for the interpretation of their results. In the context of intergenerational studies based on tax data, administrative records are used both to identify parent-child links and to measure the

Authors: Gaëlle Simard-Duplain, Statistics Canada and HEC Montréal, Xavier St-Denis, Statistics Canada and University of Toronto.

outcome of said parents and children. This may have implications for the choices researchers make. In particular, three specific decisions are explored in this paper: the cohort choice, the type of parent used, and the income measure used. First, although cohorts used in intergenerational analyses depend in part on the availability of data going back in time, researchers nonetheless have some degrees of freedom in that respect. Second, children in intergenerational samples may be linked to more than two parents, and may be linked with varying certainty. The paper explores the implications of different parent types, including parents that are linked for only one vs. more than one period ("first observed" vs. "confirmed" parents) and parents that are linked when children were young, and thus more likely to correspond to birth parents. Finally, income measures used in intergenerational studies must be designed both to provide an adequate proxy for permanent income, and to limit sample exclusions.

The results show little reason for concern with respect to the choice of cohort: there is no evidence that selection is worse for respondents born closer to the bounds used to define the cohorts. On the other hand, the choice of parent used has important ramifications for the magnitude and form of sample selection. Requiring parents to be observed with their child for at least two periods appears to reinforce selection patterns exhibited by first observed parents. However, it may reduce the probability of false matches. Alternatively, birth parents feature somewhat different selection, most likely resulting from the fact that their identification relies in part on Child Tax Benefit-claiming opportunities. That part of the analysis further emphasizes the important role that family background plays in determining respondents' inclusion in the intergenerational sample. Again however, correcting for sample selection has only a negligible impact on the estimates of intergenerational income correlation. Finally, results illustrate differences in the age, income distributions and characteristics of parents included when the two leading measures of parental income are used.

The next section provides a brief overview of the literature that has documented selection in tax data sets, in the context of intergenerational mobility studies and otherwise. Section 3 provides a detailed description of the LISA family files, in particular with respect to the construction of the intergenerational sample. The paper then present baseline sample selection results in Section 4, and discusses the three issues that may interact with that selection in Section 5. Section 6 contrasts estimates of intergenerational income correlation, with and without a selection correction. Finally, Section 7 concludes the paper.

2 Literature review and contribution

Measurement issues are central to the literature on intergenerational income mobility. In their seminal work, Becker and Tomes (1986) found that children's outcomes were only mildly related to their parents'. This led them to conclude that the intergenerational transmission of income was not a major concern in the United States. Later work challenged this view, by showing that data limitations can give rise to important errors-in-variables and life-cycle biases in estimates (eg. Atkinson, Maynard, and Trinder, 1983; Solon, 1992; Zimmerman, 1992). Errors-in-variables bias refers to the bias that results from using annual income, or the average of a small number of annual income values, instead of true permanent income. Life-cycle bias refers to the bias that is introduced by measuring income for parents and children at different points in their respective life-cycles. In the Canadian context, Chen et al. (2017) show that measuring parental and child income at similar ages, well into their working years, and including at least ten years of data for parents, produces estimates of the intergenerational elasticity of income (IGE) around 0.32. This is nearly 50% higher than previous findings by Corak and Heisz (1995, 1999). The latter had measured parental income irrespective of parents' age, and used at most five years of income data.

This paper contributes to the literature that focuses on the challenges of measuring intergenerational correlation in incomes. More specifically, it takes advantage of the linkage between the LISA survey and administrative records to evaluate the selection that may arise in studies exploiting tax data. The results apply most closely to the Canadian context, because of their dependence on the particular tax system that produces the data. However, parallels can be drawn with administrative records in other countries, based on the nature of the differences in tax filing incentives.

In Canada, the research that has focused on the measurement of intergenerational income correlation has relied largely on the Intergenerational Income Database (IID) (eg. Blanden 2005; Chen et al., 2007; Connolly et al., 2018; Corak and Heisz, 1999; Oreopoulos, 2003). The data set was originally constructed by establishing father-son links through personal income tax records, for children born between 1963 and 1966. It was later extended to include links involving mothers and daughters, and children born as late as 1985. The corresponding extension in research

possibilities has made it even more relevant to better understand the potential for selection, and its implications with respect to the understanding of intergenerational income mobility. For individuals to be part of the sample, they must have lived with a parent in at least one period where they both filed income taxes. Consequently, existing results on the association between child and parent incomes rely on the adequacy of tax data for the identification of parent-child pairs, and for the measurement of their incomes.

While administrative records have grown increasingly popular for research in Canada and around the world, their coverage rate necessarily depends on the data's primary purpose and its effect on collection. In the case of tax data, the structure of the tax system affects filing behaviour. As a result, there is always concern that certain groups may be poorly represented. A number of papers have examined the selection biases that may result from the use of tax data in contexts other than intergenerational studies. Frenette, Green and Picot (2004) and Frenette, Green and Milligan (2007) note that starting in the early 1990s personal income tax records cover approximately 96% of the Canadian population identified by the Census. The high coverage rate followed the introduction of a number of policies which improved filing incentives for low-income individuals, including the Goods and Services Tax (GST) Credit in 1989 and the Child Tax Benefit in 1993. However, none of these papers document the characteristics of the 4% that does not file taxes following the changes.[1] The potential for bias is exacerbated in the context of intergenerational studies, as the selection criteria must be applied twice, first to parents, then to children. Indeed, Cook and Demnati (2000) report that the IID sample population in 1986 accounts for approximately 72% of the corresponding group in the 1986 Census.

In this context, the contribution of this paper is two-fold. First, it uses the rich set of individual and family characteristics documented by the LISA to study the selection that arises in parent-child samples constructed from tax records. Second, it investigates how other decisions made by researchers interact with baseline selection, potentially further biasing results. In particular, it considers the choice of cohort, the way parents are defined, and the consequences of decisions made to minimize errors-in-variables and life-cycle biases. As such, this study provides a detailed discussion of the choices that researchers face in intergenerational studies, clarifying the trade-offs involved, so that informed decisions can be made according to the questions of interest.

Because tax data is used both to establish parent-child links and to measure the income of parents and children, the particular way in which the sample is selected may bias results on the association in incomes. Studies which have used the IID have explored this issue, within the realm of possibilities afforded by their data. Corak and Heisz (1999) evaluate the impact of selection by re-estimating the intergenerational elasticity of earnings and income using Heckman's two-stage estimation method. They find no evidence of bias in the estimates that omit the selection correction. However, they are restricted in their analysis to the variables included in the IID. They model selection as being driven by geographical variables, and sons' age and marital status, but it is unclear that these variables satisfy the exclusion restriction. As a consequence, their results rely heavily on non-linearity. Oreopoulos (2003) compares the IID sample to a similarly aged sample from the Longitudinal Administrative Databank (LAD). The LAD was generated and has been maintained to include approximately 20% of Canadian tax filers every year from 1982 to present. For all tax filers in the sample, it records family composition, including the presence of children; where children do not themselves file taxes, they are imputed based on parents' Child Tax Benefit claims. Comparing the IID and LAD samples therefore allows to quantify and characterize the sample of children who do not file personal income taxes while living with their parents. In the context of his paper, Oreopoulos (2003) is mostly interested in the difference in selection across small and large density housing projects. He finds that the IID coverage rate, parent marital status, number of children in the family, and parental income do not differ between the two groups. However, Oreopoulos (2003) finds that parental income is lower for missing children, and Oreopoulos et al. (2008) estimates that the IID sample underrepresents low-income individuals among the 25-32 years old group, compared to the 1996 Census. Hence, all three studies are limited to the demographic variables available in administrate data, which are scarce. In this context, the LISA is particularly well suited for the evaluation of the selection problem.

Having discussed baseline selection, this paper investigates how it may interact with other decisions made by researchers. First, it considers how the choice of cohort may contribute to selection. Second, it examines the types of parents that are identified in the parent-child linkages. More specifically, the analysis presented

1. In Canada, 3.5 to 4.8% of tax filers are late filers, not included in conventional tax datasets such as the T1FF. Late filers are more likely to be young, low-income, have a tax balance close to zero, reside in certain provinces and in the territories, and be non-residents or emigrants (Messacar, 2017). Late filing is also a repeated behaviour. It is plausible that non-filers share some of these characteristics.

below compares first observed parents, which form the basis of the IID, to first confirmed parents (parents that are observed at least twice with a given respondent) and birth parents (parents that were in the respondent's household around the time they were born). This is the first paper to explicitly address this distinction. Finally, it investigates the trade-off between measurement error and sample selection in the choice of a measure for parental income. As discussed, there is ample evidence that measuring parental and child income at different stages in their respective careers and using too few years of data to do so results in the underestimation of their income correlation. However, because different measures vary in the stringency of their requirements on the data, measures that account for errors-in-variables and life-cycle biases may result in different levels of sample selection.

3 Data

This paper uses the Longitudinal and International Study of Adults (LISA), which is constituted of three sets of files: first, the longitudinal survey itself; second, the administrative records of survey respondents; and third, the administrative records of survey respondents' family members, thereafter the family files. Data for the longitudinal survey was first collected from November 2011 to June 2012 (Wave 1), then from January 2014 to June 2014 (Wave 2). The LISA sample was designed to be representative of the population of the Canadian provinces at the time of the first wave, in 2012; it is not limited to the working age population, and children of original sample members are added to future waves of the survey as they turn 15 years old.

The Wave 1 sample consisted of 32,133 sample members, of which 23,926 were respondents. Wave 2 contained 19,178 respondents, of which 16,895 were also respondents in Wave 1. Each wave of the longitudinal survey incorporates core subjects, as well as feature modules which change from one wave to the next, including questions on education and training, family background and life events, marital and fertility histories, and cognitive (PIAAC) and non-cognitive skills. Furthermore, for both respondents and their family members, survey data were linked to the following administrative records: T1 Family Files (T1FF), from 1982 to 2013, T4 files and Pension Plans in Canada (PPIC) files, from 2000 to 2013, and the Immigrant Longitudinal Database (IMDB), which started collecting data on immigrants in 1980. A technical overview of the linkage methodology and data structure is provided in the survey documentation (Statistics Canada, 2018).

The family files include data for past and current family members; that is, for spouses and common-law partners, parents, children, and siblings. The present paper focuses on the pairs formed by LISA respondents and their parents. Both the linkages between and the income information for respondents and family members are drawn from the T1 Family Files (T1FF). These are created by Statistics Canada from the personal income tax files received from the Canada Revenue Agency, or T1s. They are processed to identify individuals who belonged to a same family in a given year, producing the T1 Family Files. T1FF processing creates a Family Identification Number (FIN), which groups together the SINs of family members, and a small number of T1FF processing variables which identify the nature of the relationship of family members to each other. Two particular features of T1FF processing are worth noting. First, T1FF processing is cross-sectional: information from previous years is not used to create a given year's families.[2] Second, the concept of family used is close to the Census family, with the exception that family members must necessarily have a fiscal relationship. This includes partners, married or common-law, and dependents, such as children for whom parents claim the Child Tax benefit.

In this context, LISA respondents are successfully linked with a parent if they both filed a T1 while living together, in at least one year between 1982 and 2013. For the majority of LISA respondents, this happens between the ages of 15 and 21, a pattern documented more carefully in Section 5.1. It is worth emphasizing the implications for researchers interested in using the LISA family files or other tax-based data sets to study immigrant populations. Indeed, establishing parent-child links is substantially less likely for foreign-born individuals. If they immigrated without their parents, then it is simply impossible. For this reason, the analysis excludes foreign-born respondents who arrived at 16 years old or later in the analysis that follows.

In the LISA, respondents must both consent to the administrative linkage and be successfully found in tax records to be associated with their T1FF data. When given the opportunity to object to the administrative linkage, very few Wave 2 respondents did. In turn, 95% of those who consented were successfully linked to the T1FF for at

2. Unlike other relationships, parent-child links actually have a longitudinal component. The T1FF team maintains a historical children file, which keeps a record of past parent-child links to aid in the construction of families in subsequent years.

least one year. The linkage rate for LISA respondents is carefully described in Hemeon (2016). Finally, at least one parental link can be found for 51% of Wave 2 respondents (84% of those born between 1963 and 1995). Table 1 shows the total number of parent-child links for respondents born between 1963 and 1995, and more specifically for those born between 1963 and 1981 (aged 33 to 51 at Wave 2, in 2014), and 1982 and 1995 (aged 19 to 32 at Wave 2). These cohort choices are based on the range of years for which personal income tax records are available, and are further discussed in Section 5.1. Chart 1 illustrates how the linkage rate to the T1FF and the linkage rate to parents in the T1FF change across birth years.[3] Among LISA respondents linked to the T1FF, the percentage of LISA respondents linked to both their own T1FF record and the T1FF record of at least one parent is approximately 80% for respondents born between 1967 and 1980, and approximately 90% for respondents born in 1981 or later. Only respondents born between 1963 and 1966 have linkage rates below 80%. This is consistent with the fact that there is an increased likelihood of finding parents for younger LISA respondents, who potentially live with their parents for a larger portion of the period covered by the data. Results not presented here show that much of the variation with respect to birth year is driven by female respondents.

4 Baseline selection

Table 2a contrasts the characteristics of respondents born in 1963-1981 (excluding foreign-born respondents who arrived in Canada at 16 years old or later), based on whether or not a link could be established between them and at least one parent. It also includes variables that characterize the parents of the respondents for whom a parental link was established. Importantly, the data does not allow to confirm that the parents referred to by the respondents in answering these questions are the same as the parents found for them in administrative records. Nonetheless, these variables inform differences in the environment in which respondents with and without a parental link grew up. Table 2b presents corresponding results for the 1982-1995 cohort.

Among older individuals, results in Table 2a show that people for whom a parental link was established are 10.6 percentage points more likely to be men. This is unsurprising given that finding a parent is a function of personal income tax filing, and thus of labour market participation. Consistent with this, respondents with a parent-child link are almost 9.6 percentage points more likely to be employed in Wave 2 (86.5% compared to 76.9%).[4] They are also significantly more educated on average. For instance, they are 16.5 percentage points less likely to have at most a high school diploma (25.1% versus 41.7%), and 14.9 percentage points more likely to have a bachelor's degree or more. Finally, 66.4% of respondents for whom a parent-child link was found reported very good health, compared to 59.9% for other respondents.

Respondents with a parent-child link also come from very different family backgrounds. As reported in Table 2a, respondents with a parent-child link are almost 4.4 percentage points less likely to self-identify as Aboriginal (3.6 vs. 8.0%), but 5.9 percentage points more likely to be foreign born or to have foreign-born parents. There are also economically large differences in family living arrangements at birth: only 2.7% of those with a parental linkage didn't live with both parents, compared to 7.8% of those for whom such a link could not be established. The former are also 29.6 percentage points more likely to have lived with both parents at age 15. Furthermore, respondents with a parental link reported better-educated parents. For instance, mothers (fathers) were 9.2 (8.2) percentage points less likely to have obtained less than a high school degree and 5.3 (8.5) percentage points more likely to be university-educated.

Finally, respondents with a parent-child link differ in terms of their own family trajectory. They are more likely to be married by 10.4 percentage points, and less likely to be common-law or separated. Interestingly, although respondents with a parent-child link get married less than two years later than respondents with no link, they report having had their first child three and a half years later.

Most of these differences also hold for the younger cohort (Table 2b), although with some quantitative differences. Respondents with parent-child links are still more likely to be men, but the difference is smaller and not statistically significant. Conversely, the employment gap by success of parental linkage among the younger

3. Note that the number of respondents by birth cohort is U-shaped. This may be attributable to the fact that the sampling frame for the LISA was based on dwellings, so that younger individuals are captured both in their own households and in the households of their parents. The relatively high proportion of individuals born in the 1960s to early 1970s likely corresponds to the end of the baby boom.
4. As described in Hemeon (2016), SINs were found for Wave 1 (2012) respondents by comparing their name, surname and other identifying characteristics with those of all tax filers observed in the T1FF in 2010 and 2011. This process was repeated following Wave 2 (2014) collection, for respondents for whom a SIN was not found in the previous wave and for new respondents. As a result, the correlation between employment and the identification of a parent-child link may also stem from the linkage of respondents to their own SIN.

cohort is 15.0 percentage points, one and a half times as large as among the older cohort. Respondents with parent-child links are also more educated among those born between 1982 and 1995. In particular, they are 12.9 percentage points less likely to have less than a high school degree (15.5 vs. 28.4%). Health differences are not statistically significant.

Reflecting changes in family structure over time, the difference in living arrangements between respondents with and without parent-child links are even greater in the younger cohort. Those with a successful linkage are 9.7 percentage points more likely to have lived with both parents at birth (96.6 vs. 86.8%), and 43.5 percentage points more likely to have lived with both parents at age 15 (80.2 vs. 36.6%). However, differences in parental education are somewhat less pronounced in the younger cohort. Finally, note that respondents with a parent-child link are 16.1 percentage points more likely to live in a single detached home at the time of the Wave 2 survey. This may be because they are more likely to still live with their parents or because they have greater wealth. Alternatively, it could also be that children who leave the parental home after labour market entry have greater attachment, which translates to home ownership in adulthood.

As discussed in Section 3, parent-child pairs that are included in the sample are characterized by five features: parents have worked and filed their taxes; children have worked (or other reason for tax filing) and filed their taxes; and labour market entry (or other reason for tax filing) for children took place before they left the parental home. It is not possible to tell simply based on the comparisons in Tables 2a and 2b which channel(s) is (are) responsible for the observed disparities.

If intergenerational mobility is not affected by gender or employment status, i.e., if these are simply issues of representation, then this can easily be fixed by reweighting the observations used in the analysis. For instance, the IID includes weights derived based on gender, geography and parental income (Cook and Demnati, 2000; Corak, 2017). Other findings from Tables 2a and 2b are potentially more troublesome for the measurement of intergenerational correlation. For instance, Bloome (2017) finds that children raised outside of two-parent families are more likely to end up at the bottom of the income distribution as adults, irrespective of their parents' income; that is, they experience greater downward mobility. If these children are poorly represented in tax-based data sets of parent-child links, then intergenerational income correlation would be overestimated in studies that rely on administrative records. Even more so, the lack of mobility at the bottom of the income distribution would be underestimated given the correlation between marital dissolution and low income. Furthermore, as family structure becomes increasingly diverse, selection may change over time, complicating matters for researchers interested in making inter-cohort mobility comparisons. To provide some insight into the importance of these concerns, Section 6 contrasts estimates of intergenerational income correlation with and without a correction for selection.

As a final consideration in this discussion, baseline selection may differ for parent-child links that are established before 1993 or in 1993 and after. As mentioned before, the current high coverage rate of tax records was achieved in 1993, following the introduction of the GST and the conversion of the Child Tax Credit into the Child Tax Benefit. Until recently, much of the literature on the intergenerational transmission of (dis)advantage in Canada has focused on individuals born in 1970 or earlier, for whom most parent-child links would have been done before 1993. However, as more years are being added to the IID, it becomes increasingly important to understand if and how comparisons can be made across that threshold. For LISA respondents born between 1971 and 1981, nearly 40% of parent-child links are established before 1993. Hence, that cohort is used to compare respondent characteristics based on the timing of their first parent-child link around the 1993 threshold. Results are presented in Table 3, and suggest that the two groups are largely similar. Those for whom a parent-child link was established before 1993 are on average 5.3 years older, consistent with the fact that linkage typically happens around 16 to 19 years old. They are also more likely to be separated and less likely to be single, probably reflecting in part the age difference. Interestingly, they are 10.1 percentage points more likely to live in a single detached home. Individuals born between 1971 and 1981 are 33 to 43 years old in 2014, so that some of that difference may reflect the average age gap between the two. However, the magnitude of the difference in home ownership is fairly large, suggesting that there may be other forces at play. In particular, the difference would be consistent with a better representation of respondents from lower-income backgrounds starting in 1993. Finally, those for whom a parent-child link was established before 1993 are 7.2 percentage points less likely to report their mother as university-educated. Overall, comparisons across the entire period for which administrative data is available can be made with reasonable confidence. Indeed, with the exception of the reported dwelling, those characteristics that do differ across the two groups mostly appear to stem from the fact that respondents linked before 1993 are

born earlier on average than respondents linked in 1993 or later. Importantly, most of the characteristics found to be associated with selection in Table 2a are uniform across the two groups. For instance, they don't differ in their sex, employment status in 2014, or educational attainment.

5 Interaction between selection and other research choices

5.1 Choice of cohort

In previous sections, this study discussed selection for respondents born between 1963 and 1981, and between 1982 and 1995. Chart 2 plots the distribution of the first observation of first observed parents against child age, as well as the distribution of all observations of first observed parents (regardless of their order) against child age, for respondents born between 1963 and 1995. The modal age at which a first observed parent is observed for the first time is 18 years old. The greatest number of first observed parents are observed when children are 19 years old. The lower bound in birth years, 1963, was chosen so that respondents are 19 years old in 1982, the first year for which the T1FF is available. It also corresponds to the lower bound of the IID sample. Next, 1995 is chosen as the upper bound so that respondents are 18 years old in 2013, the last year for which the T1FF is available for the second wave of LISA. It ensures that there are sufficiently many years during which a parental linkage can be established for the youngest respondents included in the sample. Finally, respondents are divided in two groups, the first one of which is born before 1982 and the second one, in 1982 or later. Parental T1FF data for respondents born between 1982 and 1995 is therefore available from the year they are born, a feature which is of interest to explore different channels for the intergenerational transmission of (dis)advantage.

Together with the way relationships are established, the choice of cohorts can lead to truncation of parent-child links for the youngest and oldest respondents. In particular, the selection issues discussed previously are likely to be worse around the bounds, or to interact with some of the parents' or respondents' characteristics. Observing more years for the youngest and oldest respondents might have allowed for a link to be established for some respondents that are currently not associated with a parent; or for more links to be established for respondents for whom there is already at least one parent. Some of these limitations may be visible in the relationship between linkage rate and birth year. The data in Chart 1 shows that the linkage rate was below 80% only for the oldest respondents included in the intergenerational sample. To further assess the seriousness of the concern about truncation, the following exercise is conducted. Most linkages in the older cohort happen at 16 years old or later. Therefore, to evaluate what is lost from observing respondents born between 1963 and 1965 for the first time between 17 and 19 years old, the older cohort is restricted to include only people born between 1966 and 1981. In this case, all respondents could be observed at the latest at 16 years old. The results of this exercise show that a negligible fraction of parent-child links established for that cohort are observed before 1985 only. While these may be a very select group, their size is unlikely to affect overall estimates.

5.2 Choosing (defining) parents

The IID was constructed by retaining first observed parents. This paper investigates the impact of this decision by investigating the type of parent identified through tax records. The issue is in fact two-fold. First, because the T1FF is essentially processed every year independently from the previous year, parent-child links are susceptible to false matches. Second, defining parents is also a broader question about which adults are relevant to understand intergenerational dynamics: Should step-parents be included along with birth parents? Do parents who live longer with children matter more? Or parents who lived with children when they were younger? Answers to these questions have implications for selection.

5.2.1. Risk of false matches and confirmed parents

Chart 3a shows that in the older cohort approximately 15% of first observed parents were linked for only one period. In the younger cohort, that number is approximately half as large (Chart 3b). Single-period parent-child matches may reflect a number of things. First, they may arise as a natural extension of the selection discussed in Section 4; that is, they may correspond to respondents who left the parental household immediately following labour market entry. Second, they may correspond to parent-child pairs where the parent, child, or both worked and/or filed their taxes only sporadically around the time of the child's labour market entry. Finally, they may correspond to false matches. If a non-negligible portion of parent-child links observed for a single period are

false matches, then estimates computed using these links understate the correlation between parent and child incomes.

In this study, first confirmed parents are defined as the first parents that are observed with the respondent for at least two periods. Tables 4a and 4b show the characteristics of respondents for whom first observed parents are found, based on whether these are then confirmed or not. These comparisons are informative in two ways. First, they show along which margin selection is most pronounced, further revealing some of the underlying mechanisms at play. Second, they also point to groups that are potentially more sensitive to false matches, because they are more likely to be observed for only one period. If first observed parents do include false matches, then it can be advantageous for researchers to use first confirmed parents and correct for selection.

In both cohorts, requiring that a parent be observed for at least two periods with the respondent appears to strengthen existing selection issues (Tables 4a and 4b). For instance, respondents with first confirmed parents in both cohorts are more likely to be better educated on average, and those in the older cohort are more likely to be men. In both cohorts, respondents with first confirmed parents also differ from those with only first observed parents in that they are less likely to be born in Canada to Canadian parents and more likely to be born in Canada to foreign-born parents. That distinction was present to a certain extent between respondents with and without parental links, but it was both economically and statistically less significant. In other words, second-generation immigrants are more likely to be observed for a second period, conditional on being observed in at least one period. This suggests different patterns in the time at which respondents leave the family home, highlighting the role that family dynamics play in determining whether or not parent-child pairs are observed in tax records.

5.2.2. Birth parents and step-parents

Parents identified through the T1FF are adults that were present in the child's home when they were 15 to 21 years old, in most cases. They can be birth parents, adoptive or step-parents, or other legal guardians. For respondents born between 1982 and 1995, it can be determined when a parent-child pair started living together; that is whether the adult found is likely to be a birth parent or not. T1FF processing uses information from the Child Tax Benefit (CTB) to determine the family composition of adult tax filers. When they claim the benefit, parents must report the age of each dependent child, irrespective of that child's tax filing status. That information is then associated with their SIN in the T1FF. In other words, the date of birth of the child is retained as part of the parent's T1FF record. Because no child's SIN is added to the records based on the CTB, parent-child links cannot be established for those years where only CTB information is available; a T1 must be filled for the child's SIN to appear with the parent's SIN. However, data on the number and age of children present in an adult tax filer's family can be used to find when an established parent-child link started. Therefore, even if a respondent only appears with their SIN in their parent's family at age 15, it is possible to establish whether the parent found is likely to be a birth parent. To allow for the fact that not all individuals file their taxes in all years, birth parent are defined as a tax filer that appears with the LISA respondent anytime when the latter is aged 0 to 4 years old.[5] Any parent who is observed in the T1FF while the child is 0 to 4 years old, but not in the same FIN, can be seen as a step-parent.

Table 5 compares the characteristics of respondents born between 1982 and 1995, based on whether or not their first observed parents correspond to birth parents.[6] A number of differences are consistent with CTB-claiming opportunities, and therefore likely reflect selection. Mothers of respondents for whom a birth parent could be found in the tax data are 7.7 percentage points less likely to have a university degree. Respondents with a birth parent are 11.7 percentage point more likely to be Canadian born with Canadian-born parents (71.9 vs. 60.2%). There are almost no foreign-born respondents with a birth parent (likely due to the fact that only a small proportion of foreign-born respondents immigrated to Canada before 5 years old). Respondents for whom birth parents are found in the tax data are less likely to have lived with both parents at birth and at 15 years old. Finally, they are

5. The method used to identify (likely) birth parents starts from a parent-child link established in a year where both parent and child filed their taxes while living together. Then, it is possible to return back to the birth year of the child, checking in each year whether the parent's family composition confirms the presence of a child born in the same year as the respondent. While this is the best approach available given the data, it presents a number of potential weaknesses. First, it is constrained in the type of birth parent observed. The method to identify birth parents is retroactive; that is, it identifies parent-child links at the time the child first files their T1, then verify if the parent thus identified could correspond to a birth parent. In consequence, it is not possible to identify birth parents that left the family and are no longer present by the time the child files their first T1. Second, it cannot distinguish birth parents from step-parents that appear early in a child's life, or from adoptive parents. Third, if a respondent is observed with a step-sibling of the same age in the first year they file taxes, the method cannot determine whether it is them or the step-sibling that lived with the identified parent at birth. In other words, it cannot determine whether the identified parent is a birth parent to them or a step-parent. In brief, while attractive, this method only provides a rough categorization of birth and other parents.
6. The distinction between first observed and birth parents is different than that between first observed and first confirmed parents. Unlike confirming first observed parents, identifying birth parents does not tighten the requirement on respondents' filing behaviour at the beginning of their career. However, it does require parents to have filed their taxes and/or claimed the CTB.

also more likely to live in a single detached home at Wave 2, potentially reflecting a more affluent and/or stable family background.

5.3 Parental income measures

Parental income measures that have been used in intergenerational studies broadly fit in two categories. To illustrate the first one, income is averaged over a five-year period that corresponds approximately to the time of first parental link, which is operationalized as the period where the child is 15 to 19 years old.[7] This is the type of measure used by Corak and Heisz (1999), Corak (2017), and Connolly et al. (2018), referred to below as COR.[8] Parents are allowed to have no income or no tax records in one or more years (coded as zero income). However, parental observations with an average of less than $500 are dropped from the sample. All values are CPI adjusted in 2013 constant Canadian dollars. The second measure used in this study follows Chen et al. (2017), and is referred to below as COP. It is intended to represent true permanent income, or lifetime income. In this case, annual income is averaged over the 21 years when the parent is between ages 35 to 55. This approach is designed to avoid errors-in-variables bias arising from the averaging of parental income over a number of years too small to obtain an adequate measure of permanent parental income. Explicitly restricting the parents' age when their income is measured also limits life-cycle bias by comparing parental and child income at more similar ages. Parental observations with less than 10 years of income at or above $500 are excluded.[9]

While the different measures of parental income vary in the severity of errors-in-variables bias and life-cycle bias they imply, they also have an impact on the way the sample is selected. Most notably, by restricting the measurement of parental income to the years where the parent is aged 35 to 55, the COP measure imposes a relatively stricter tax filing criteria. In particular, it may distort selection for the older cohort, for whom the link was established closer to the lower period bound. This latter issue arises because the definition of income interacts with the criteria for sample selection around the bounds. On the other hand, the COR measure imposes different selection, because it is based on the child's age, and requires less years of income on the part of the parents. To see how these differences may impact estimates more generally, the two measures are compared to identify differences in terms of the parent's age when the respondent is born and when their income is measured; the distribution of parental income; and the characteristics of respondents and their parents.

The difference between the two measures is evident in Charts 4 and 5, which respectively show the age of parents when the respondent was born and the age of parents when their income is measured, by child's birth year. The COP measure implicitly requires parents to be at most 45 years old in 1982, which excludes parent-child links where parents had their child relatively late. For instance, for respondents born in 1963, the upper bound on parental age in 1982 implies that parents who had their child after 26 years old would mostly be excluded. The converse effect can also be seen for respondents born closer to the upper bound; the COP measure requires that parents be at least 45 years old in 2013, so that the parents included have typically had their children later in life. With respect to parents' age when their income is measured, the difference between the COP and COR measures is less stark, especially for the older child birth years. Among the younger cohort, COP tends to measure parents' income at a younger age.

Differences between the two measures are also seen in the distribution of parental income in Charts 6a and 6b. In both cohorts, COP under-represents low-income parents (less than $15,000) respective to COR, and to a certain extent high-income parents ($75,000 and up for the older cohort, $60,000 and up for the younger cohort). Correspondingly, it tends to select more heavily parents from the middle of the income distribution. Notably, the differences for the older cohort are largest at the bottom and in the middle of the income distribution; and for the younger cohort, at the top of the income distribution.

Next, Tables 6a and 6b report the results of a comparison the characteristics of respondents for whom a parent could only be included according to the COR measure, to the characteristics of respondents for whom both the

7. Children born between 1963 and 1966 are all older than 15 years old in 1982. For these respondents, the age range is widened to 15-21 in 1982 and 15-20 in 1983, so that parental income averages are calculated over at least three years.
8. Existing studies using the IID average parental income between 1978 and 1982, which corresponds to the time when children born in 1963 to 1966 (the cohorts typically used in these studies) are 12 to 19 (15-19 for those born in 1963, and 12 to 16 for those born in 1966) (Corak and Heisz 1999; Chen et al. 2017). Corak (2017) instead averages parental income over a five year period when children are 15 to 19. He uses the 1963-1970 and 1967-1970 birth cohorts.
9. Note that some parents of children in the oldest cohorts will be too old to meet the above criterion and will automatically be dropped if more than 45 years old in 1982. The same will occur for parents of children in the younger cohorts if they are below 45 years old in 2013. In addition, many parents have a sufficient number of years of income above $500, but since part of their 35-to-55 age interval falls before 1982 or after 2013, it is censored.

COR and COP measures could be computed. Results are shown separately for the older and younger cohort in Tables 6a and 6b. For respondents born between 1963 and 1981, using the COP measure not only selects parents who had their child relatively earlier, but also over-represents younger members of the cohort. The converse holds for the 1982-1995 cohort. Furthermore, differences in living arrangements between COR-only and COR-and-COP groups, and the comparison of these differences across the two cohorts, suggests an important consequence of constraining the age at which parents had their children. For the 1963-1981 cohort, Chart 4 shows that using the COP measure selects parents who had their children relatively earlier. Relatedly, the results show that respondents in this cohort for whom both the COP and the COR measures can be calculated are less likely to have been living with both their parents, at birth and at age 15. Conversely, the COP measure for the 1982-1995 cohort selects more heavily parents who've had their children relatively later, and respondents for whom both parental income measures can be calculated are more likely to have lived with both parents. However, the difference is not statistically significant for the younger cohort. Next, COR-and-COP respondents born between 1982 and 1995 are much more educated on average than COR-only respondents. In particular, they are 22.6 percentage points less likely to have less than a high school degree. Finally, for the older cohort, the sample of respondents for whom the COP measure could be calculated reported much more educated parents.

While the COR and COP measures are used in this analysis as points of comparison to the existing literature on intergenerational mobility in Canada, the results outlined in this section apply more generally whenever making decisions about income measurement. These choices can interact in important ways with sample selection, resulting in the exclusion of certain families who may systematically differ in their experience of mobility. Furthermore, they may do so in ways that render cross-cohort comparisons harder to interpret. Finally, requiring more periods of non-negative income also has obvious implications for the labour market attachment and related characteristics of the families included in the analysis. In practice, for LISA users the trade-off between measurement and selection bias should further reflect the fact that, on the one hand, the LISA may facilitate corrections for sample selection; and on the other hand, some models are less sensitive to measurement bias, such as rank-rank regressions (Chetty et al. 2014; Connolly et al., 2018; Corak 2017).

6 Accounting for selection

Sections 4 and 5 raise questions about the impact of selection on the estimation of intergenerational income correlation. Table 7 presents numbers that ignore selection, and contrasts them with the same figures, corrected for selection. Results are only presented for the 1963-1981 cohort, as much of the 1982-1995 group is too young to obtain an adequate measure of income. The COR income measure described in Section 5.3 is used.[10] The measure of intergenerational mobility used is the elasticity of child income with respect to parental income; that is, the logarithm of child income is regressed on the logarithm of parental income. Let β, the slope coefficient from that regression. Then, a one-percent increase in parental income is associated with a β-percent increase in child income.[11]

Estimates are adjusted for selection using the Heckman correction (Heckman, 1979). Recall that selection is problematic only if it is correlated with determinants of child income that are unaccounted for in the regression. To implement the Heckman correction, it is necessary to exploit variables that influence the probability of being in the sample, with no direct impact on child (adult) income. The results in Table 7 were obtained using a series of dummy variables reflecting the number of periods for which children were observed in the T1FF between ages 15 and 19.[12] The correction was not implemented using the family background variables discussed in Sections 4 and 5, since these factors are likely to affect child income directly. Conversely, conditional on being in the T1FF,[13] the age at which children started filing taxes and the consistency with which they did so is unlikely to affect their income today. Instead, it probably results from differences across children in the type of job they got as teenagers (e.g. babysitting vs. fast food industry). Results not shown here confirm that the probability of being in the sample is increasing in the number of periods for which children were observed.

10. Total family income is used. If a parent or child is observed with a spouse, then family income is equal to the sum of their incomes; otherwise, family income is equal to individual income. No adjustment is made for the number of family members or earners. As such, the estimates represent the transmission of total available resources.
11. Using a rank-rank regression, where the percentile income rank of children is regressed on that of their parents, produces qualitatively similar results as those presented here.
12. For individuals born before 1967, the first five years of data were used. For instance, for individuals born in 1963, ages 19 to 23 were used.
13. Recall that this is a minor condition, given the high linkage rate of LISA sample members with their own tax records.

Table 7 shows three sets of comparisons. First, it present baselines estimates to address the concerns raised in Table 2a. Second, it includes estimates that use first confirmed parents instead of first observed parents, to evaluate whether using this parent type results in worse or different bias.[14] Finally, it shows estimates using the sample of respondents for which the COP measure could be used. The COR measure is used for this comparison as well. None of the comparisons show evidence of selection bias. For each one, although the corrected estimate is slightly lower than the baseline estimate, each lies well within the other's confidence interval. In other words, the selection presented in Sections 4 and 5 appears unproblematic for researchers interested in studying intergenerational mobility using the LISA intergenerational sample and other data sets constructed from personal income tax records.

7 Conclusion

This paper presents an intergenerational data set constructed from the newly developed LISA family files. This data is used to demonstrate the sources of sample selection that may arise in data sets of parent-child links constructed from tax records. In Canada and elsewhere, administrative sources have become increasingly popular for intergenerational studies. However, they may be subject to non-random sample selection, based on the primary purpose for which they were collected. This paper finds that respondents who are part of the LISA intergenerational sample are more educated on average and have more educated parents. Furthermore, they are more likely to have grown up in a nuclear family. Depending on how these characteristics are correlated with mobility, estimates of intergenerational correlation could be biased as a result. However, estimates from a simple selection correction exercise suggest that these differences do not introduce substantial bias in the estimation of intergenerational income correlation.

Researchers interested in intergenerational income mobility must also make a number of other choices that may interact with the baseline selection described here. In particular, they must determine which birth years to consider, how to define parents, and how to measure incomes. The results show that the choice of cohort does not interact significantly with the way parent-child links are selected. However, the definition of parents chosen for the analysis matters. The IID has been constructed around the concept of first observed parent; that is, the first adult that is observed with the respondent is picked as their parent for the analysis, irrespective of the number of periods for which they are in the same family as the respondent, or of other parents that may appear later on. On the one hand, the nature of the T1FF implies that some of these parents may be false matches, resulting in underestimation of intergenerational income correlation. On the other hand, proper understanding of the mechanisms underlying mobility may require researchers to pay closer attention to the differential role of parents who are present for relatively more years, or parents who are present at birth. The analysis presented in this paper shows that this gives rise to different levels and forms of selection. Finally, results show that the interaction between baseline selection and the choice of parental income measure exacerbates selection. Again however, neither the parent type nor the income measure used appear to result in different or worse bias than the baseline.

In general, the findings from this study show that certain types of families are systematically excluded from intergenerational samples constructed from tax data. Nonetheless, these exclusions do not appear to result in bias in the estimation of mobility. This confirms that researchers who use the LISA or, more generally, who rely on intergenerational samples constructed from tax records, may do so without concern for sample selection.

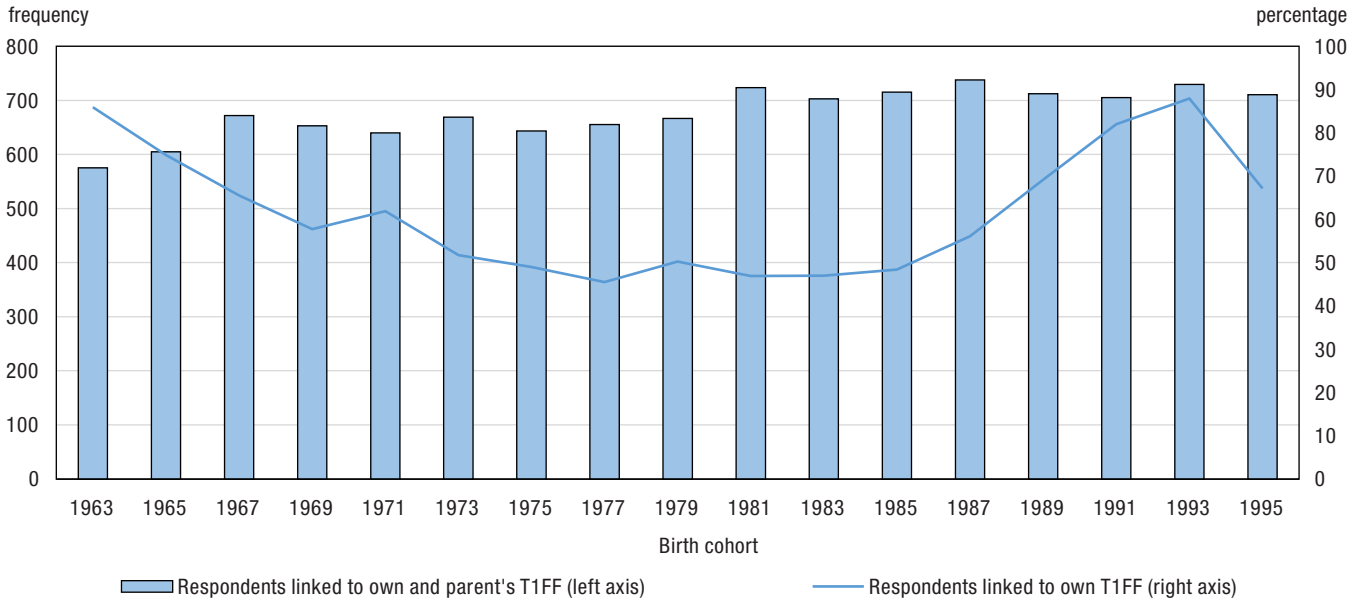
14. The comparison using birth parents is omitted, since the coverage for the 1963-1981 cohort is low.

Table 1
LISA intergenerational sample size, by respondent birth year and sex

Cohort	Sex	Frequency
1963 to 1995	Total	6,895
	Male	3,425
	Female	3,470
1963 to 1981	Total	3,617
	Male	1,757
	Female	1,860
1982 to 1995	Total	3,278
	Male	1,668
	Female	1,610

Source: Statistics Canada, Longitudinal and International Study of Adults, 2014, and T1FF, 1982 to 2013.

Chart 1
Respondents linked to the T1FF records of at least one parent, by birth cohort



Note: Respondents who immigrated to Canada at 15 years old or later are excluded from the sample. The percentage represents the share of respondents linked to their own T1FF who were also linked to the T1FF of at least one parent. All plotted statistics pool two birth years together (1963-1964, 1965-1966, etc.), except for the 1995 birth cohort (single year).
Source: Statistics Canada, Longitudinal and International Study of Adults, 2014, and T1FF, 1982 to 2013.

Table 2a
Respondent characteristics, 1963 to 1981, with and without parent-child link

Variables	No link	With link	Difference	Confidence interval
Age	42.5	40.9	-1.6*	(-2.2, -1.0)
Sex	57.7	47.1	-10.6*	(-15.8, -5.4)
Married	32.3	42.7	10.4*	(5.3, 15.5)
Common-law	30.9	25.0	-5.9*	(-10.1, -1.8)
Separated/Divorced/Widowed	14.7	11.2	-3.5	(-7.0, 0.0)
Single (never married)	21.9	21.1	-0.9	(-6.0, 4.3)
Age, first marriage	26.3	27.8	1.5*	(0.4, 2.5)
Age, first marriage/union	24.8	26.6	1.8*	(1.1, 2.5)
Age, first child	25.9	29.4	3.5*	(2.9, 4.2)
Household size	3.1	3.1	0.0	(-0.1, 0.2)
Includes 3 generations or more	3.4	3.2	-0.2	(-2.0, 1.6)
Includes 2 generations or less	77.7	80.2	2.5	(-3.2, 8.2)
Person not in a census family	18.9	16.6	-2.3	(-7.9, 3.3)
Single detached house	67.1	71.8	4.7	(-1.5, 10.9)
Atlantic	8.4	8.3	-0.1	(-1.9, 1.7)
Quebec	26.4	24.6	-1.7	(-6.0, 2.5)
Ontario	35.2	37.4	2.2	(-2.8, 7.3)
Prairies	17.4	17.7	0.3	(-3.0, 3.6)
BC	12.7	12.0	-0.7	(-4.2, 2.8)
Rural area	21.0	19.3	-1.7	(-5.4, 2.0)
Aboriginal person	8.0	3.6	-4.4*	(-7.0, -1.7)
Own education: Less than High School	15.0	6.8	-8.2*	(-11.2, -5.1)
Own education: High School	26.7	18.3	-8.3*	(-12.8, -3.8)
Own education: Trade/Vocational/Apprenticeship	16.4	12.2	-4.2*	(-7.6, -0.9)
Own education: Some postsecondary	22.8	28.6	5.8*	(1.3, 10.3)
Own education: Bachelor or above	18.9	33.8	14.9*	(10.8, 19.0)
Health: Very good/Excellent	59.9	66.4	6.4*	(1.7, 11.2)
Health: Good	28.8	26.7	-2.1	(-6.3, 2.1)
Health: Fair/poor	11.0	6.9	-4.1*	(-7.1, -1.1)
Employed	76.9	86.5	9.6*	(5.6, 13.5)
Self-employed	16.3	13.7	-2.6	(-6.4, 1.2)
Living with both parents at birth	92.2	97.3	5.1*	(2.5, 7.7)
Living with both parents at 15	53.0	82.6	29.6*	(24.8, 34.4)
Mother education: Less than High School	36.0	26.8	-9.2*	(-14.1, -4.4)
Mother education: High School	31.6	32.0	0.4	(-4.4, 5.3)
Mother education: Some postsecondary (below university)	18.3	24.0	5.7*	(2.0, 9.4)
Mother education: University (certificate, diploma or degree)	9.1	14.3	5.3*	(2.1, 8.5)
Father education: Less than High School	39.4	31.2	-8.2*	(-13.2, -3.2)
Father education: High School	22.6	21.3	-1.3	(-5.4, 2.7)
Father education: Some postsecondary (below university)	16.3	23.0	6.7*	(3.4, 10.1)
Father education: University (certificate, diploma or degree)	12.6	21.1	8.5*	(4.8, 12.2)
Visible minority	8.2	8.2	-0.1	(-3.5, 3.4)
Born in Canada, Canadian-born parents	76.7	70.7	-6.0*	(-10.4, -1.5)
Foreign-born	6.7	9.1	2.4	(-0.6, 5.4)
Born in Canada, at least one foreign-born parent	16.7	20.2	3.5	(-0.5, 7.6)
Number	**988**	**3617**	**...**	**...**

... not applicable
* difference between "no link" and "link" valus is significant ($p < 0.05$)

Note: Confidence intervals are normal-approximation confidence intervals with 95% coverage.
Source: Statistics Canada, Longitudinal and International Study of Adults, 2014, and T1FF, 1982 to 2013.

Table 2b
Respondent characteristics, 1982 to 1995, with and without parent-child link

Variables	No link	With link	Difference	Confidence interval
Age	23.7	24.4	0.8*	(0.2, 1.3)
Sex	51.8	47.9	-3.9	(-11.1, 3.3)
Ever married	13.6	11.5	-2.0	(-6.8, 2.7)
Age, first marriage	23.8	25.4	1.6*	(0.7, 2.5)
Age, first marriage/union	21.6	23.5	1.9*	(1.1, 2.7)
Age, first child	22.7	25.1	2.4*	(1.4, 3.4)
Household size	3.1	3.1	0.0	(-0.2, 0.2)
Includes 3 generations or more	6.2	3.1	-3.1*	(-5.3, -0.9)
Includes 2 generations or less	71.8	79.2	7.4	(-0.9, 15.7)
Person not in a census family	22.0	17.7	-4.3	(-12.7, 4.2)
Single detached house	49.9	66.0	16.1*	(7.9, 24.3)
Atlantic	5.7	6.5	0.8	(-1.2, 2.8)
Quebec	22.2	23.3	1.1	(-4.6, 6.8)
Ontario	38.3	38.7	0.5	(-7.3, 8.3)
Prairies	17.9	19.5	1.6	(-2.9, 6.1)
BC	16.0	12.0	-4.0	(-9.5, 1.6)
Rural area	11.8	13.8	2.0	(-4.2, 8.3)
Aboriginal person	5.9	3.1	-2.9	(-5.7, 0.0)
Own education: Less than High School	28.4	15.5	-12.9*	(-19.6, -6.1)
Own education: High School	38.8	35.1	-3.7	(-10.9, 3.6)
Own education: Trade/Vocational/Apprenticeship	9.8	8.3	-1.5	(-6.2, 3.2)
Own education: Some postsecondary	13.0	17.9	4.9	(0.0, 9.9)
Own education: Bachelor or above	10.0	23.2	13.2*	(8.7, 17.8)
Health: Very good/Excellent	71.0	72.5	1.5	(-4.6, 7.6)
Health: Good	21.7	22.8	1.1	(-4.2, 6.4)
Health: Fair/poor	7.3	4.6	-2.7	(-6.4, 1.0)
Employed	60.8	75.8	15.0*	(7.8, 22.1)
Self-employed	7.6	6.0	-1.6	(-6.7, 3.4)
Living with both parents at birth	86.8	96.6	9.7*	(4.6, 14.9)
Living with both parents at 15	36.6	80.2	43.5*	(36.6, 50.4)
Mother education: Less than High School	15.3	8.1	-7.2*	(-11.7, -2.7)
Mother education: High School	31.8	26.6	-5.2	(-12.5, 2.0)
Mother education: Some postsecondary (below university)	27.5	34.1	6.6*	(0.7, 12.5)
Mother education: University (certificate, diploma or degree)	24.3	30.2	5.9	(-1.8, 13.5)
Father education: Less than High School	15.7	12.1	-3.6	(-8.1, 0.9)
Father education: High School	26.9	24.2	-2.7	(-9.3, 4.0)
Father education: Some postsecondary (below university)	28.3	29.2	0.9	(-5.5, 7.2)
Father education: University (certificate, diploma or degree)	23.0	32.2	9.2*	(3.0, 15.4)
Visible minority	21.6	17.6	-4.0	(-10.2, 2.2)
Born in Canada, Canadian-born parents	65.6	67.3	1.7	(-5.0, 8.5)
Foreign-born	16.0	11.3	-4.7	(-10.4, 0.9)
Born in Canada, at least one foreign-born parent	18.5	21.5	3.0	(-2.1, 8.1)
Number	**517**	**3278**	**...**	**...**

... not applicable
* difference between "no link" and "link" valus is significant ($p < 0.05$)

Note: Confidence intervals are normal-approximation confidence intervals with 95% coverage.
Source: Statistics Canada, Longitudinal and International Study of Adults, 2014, and T1FF, 1982 to 2013.

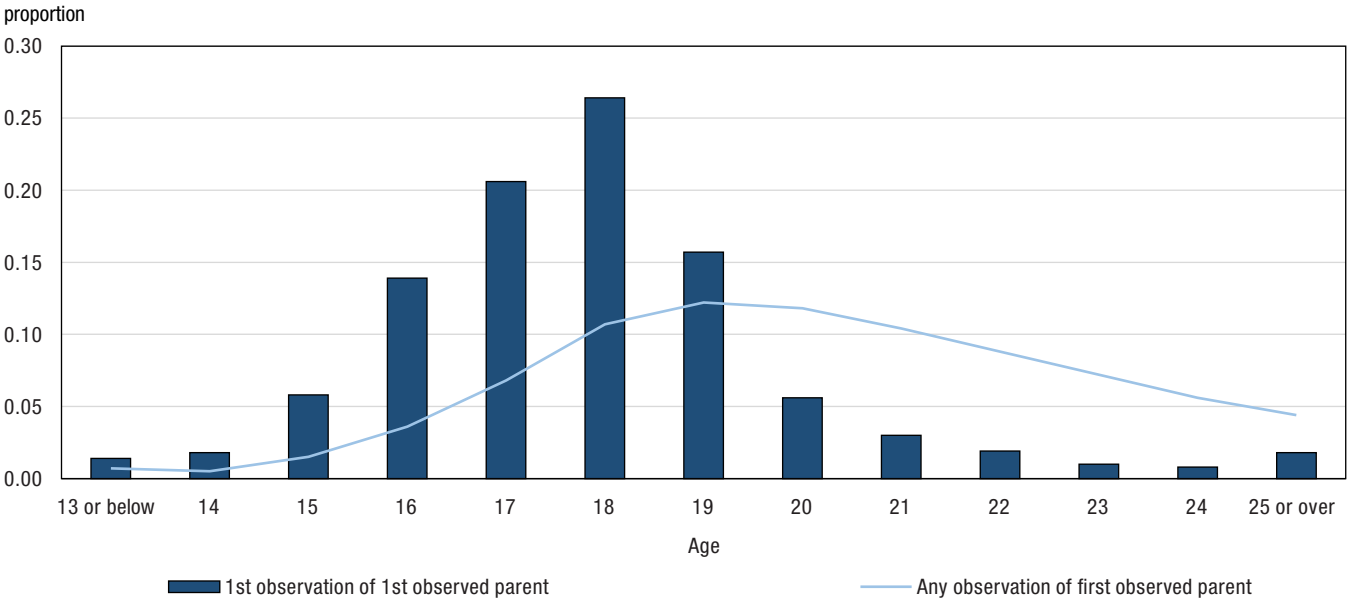
Table 3
Respondent characteristics, 1971 to 1981, linked before 1993 and linked in 1993 or after

Variables	< 1993	≥ 1993	Difference	Confidence interval
Age	40.4	35.1	-5.3*	(-5.6, -5.1)
Sex	45.0	47.4	2.4	(-3.7, 8.4)
Married	44.7	38.4	-6.4	(-13.9, 1.2)
Common-law	26.7	28.8	2.1	(-3.7, 7.8)
Separated/Divorced/Widowed	12.1	5.5	-6.6*	(-10.0, -3.2)
Single (never married)	16.4	27.3	10.9*	(3.8, 17.9)
Age, first marriage	27.6	27.6	0.0	(-0.7, 0.7)
Age, first marriage/union	26.5	25.8	-0.8*	(-1.4, -0.1)
Age, first child	29.5	28.8	-0.7	(-1.6, 0.1)
Household size	3.3	3.2	-0.1	(-0.3, 0.1)
Includes 3 generations or more	1.8	3.3	1.5	(-0.2, 3.3)
Includes 2 generations or less	85.5	79.5	-5.9	(-13.0, 1.2)
Person not in a census family	12.8	17.2	4.4	(-2.6, 11.4)
Single detached house	77.2	67.2	-10.1*	(-17.9, -2.2)
Atlantic	7.4	7.4	0.0	(-1.9, 2.0)
Quebec	23.1	26.4	3.2	(-1.5, 8.0)
Ontario	36.6	35.8	-0.7	(-8.0, 6.5)
Prairies	20.2	17.6	-2.6	(-7.2, 1.9)
BC	12.6	12.8	0.1	(-3.9, 4.1)
Rural area	18.6	17.0	-1.7	(-6.2, 2.9)
Aboriginal person	2.4	2.5	0.0	(-1.5, 1.6)
Own education: Less than High School	7.2	5.3	-1.9	(-4.7, 0.9)
Own education: High School	15.8	17.4	1.6	(-3.2, 6.4)
Own education: Trade/Vocational/Apprenticeship	11.8	10.0	-1.8	(-5.1, 1.5)
Own education: Some postsecondary	30.6	30.7	0.1	(-6.5, 6.7)
Own education: Bachelor or above	34.4	35.9	1.5	(-4.9, 8.0)
Health: Very good/Excellent	66.3	68.6	2.2	(-4.4, 8.9)
Health: Good	26.2	26.6	0.4	(-5.8, 6.5)
Health: Fair/poor	7.4	4.8	-2.6	(-5.6, 0.4)
Employed	88.8	86.2	-2.5	(-6.4, 1.3)
Self-employed	13.9	13.7	-0.2	(-4.4, 4.0)
Living with both parents at birth	98.4	95.8	-2.6*	(-4.5, -0.7)
Living with both parents at 15	84.0	79.8	-4.2	(-9.1, 0.7)
Mother education: Less than High School	24.6	19.2	-5.4	(-10.9, 0.2)
Mother education: High School	32.3	30.5	-1.8	(-8.5, 4.8)
Mother education: Some postsecondary (below university)	26.7	28.8	2.1	(-3.6, 7.7)
Mother education: University (certificate, diploma or degree)	12.5	19.8	7.2*	(1.9, 12.6)
Father education: Less than High School	28.0	22.9	-5.1	(-11.4, 1.2)
Father education: High School	21.2	21.9	0.7	(-4.3, 5.7)
Father education: Some postsecondary (below university)	23.9	28.5	4.6	(-2.1, 11.4)
Father education: University (certificate, diploma or degree)	23.9	24.4	0.5	(-5.0, 6.1)
Visible minority	7.4	11.4	4.0	(-2.3, 10.3)
Born in Canada, Canadian-born parents	66.7	70.9	4.2	(-2.7, 11.1)
Foreign-born	8.4	9.1	0.6	(-2.9, 4.1)
Born in Canada, at least one foreign-born parent	24.8	20.0	-4.8	(-11.7, 2.0)
Number	**720**	**1133**	**...**	**...**

... not applicable
* difference between "no link" and "link" valus is significant ($p < 0.05$)

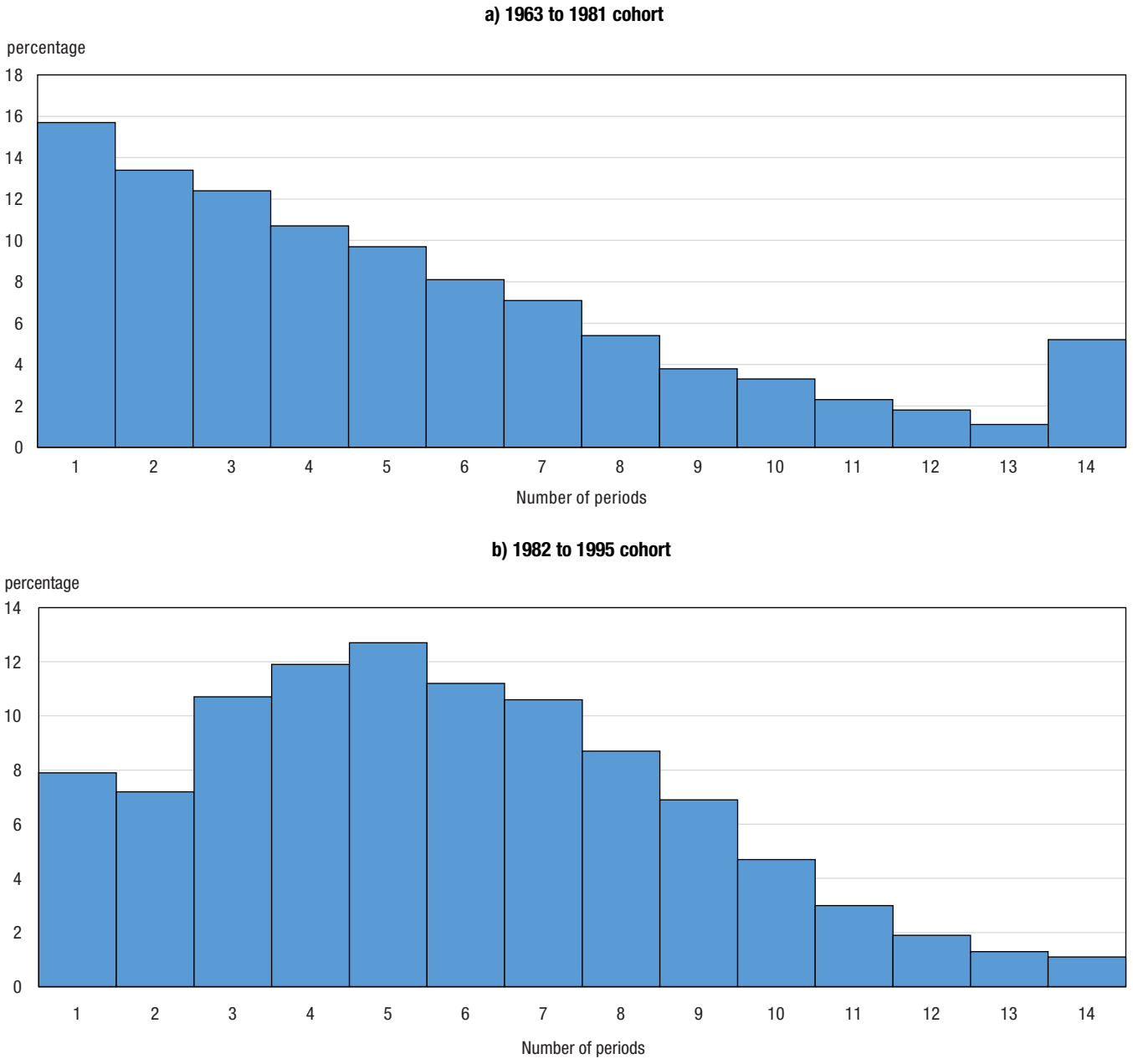
Note: Respondents linked in 1993 or after refer to those respondents for whom a parent-child link was established for the first time in 1993 or after. Confidence intervals are normal-approximation confidence intervals with 95% coverage.
Source: Statistics Canada, Longitudinal and International Study of Adults, 2014, and T1FF, 1982 to 2013.

Chart 2
Respondent age at first observation and at any observation with first observed parent



Note: Respondents who immigrated to Canada at 15 years old or later are excluded from the sample. Age is top coded at 25 (results for first observed parent at 25 years old include parents who were first observed later than 25 years old; however, results for any observation of first observed parent at 25 years old include only observations of parents at exactly 25 years old).
Source: Statistics Canada, Longitudinal and International Study of Adults, 2014, and T1FF, 1982 to 2013.

Chart 3
Number of periods respondent was observed with first observed parent



Note: Respondents who immigrated to Canada at 15 years old or later are excluded from the sample. The number of periods is top coded at 14.
Source: Statistics Canada, Longitudinal and International Study of Adults, 2014, and T1FF, 1982 to 2013.

Table 4a
Respondent characteristics, 1963 to 1981, first observed (unconfirmed) and first confirmed parents

Variables	First observed	First confirmed	Difference	Confidence interval
Age	41.5	40.9	-0.6	(-2.0, 0.7)
Sex	56.8	46.3	-10.5*	(-19.6, -1.5)
Married	40.0	43.5	3.5	(-5.5, 12.5)
Common-law	31.8	23.4	-8.4	(-18.6, 1.8)
Separated/Divorced/Widowed	12.6	11.4	-1.1	(-5.2, 3.0)
Single (never married)	15.7	21.6	6.0	(-0.5, 12.4)
Age, first marriage	26.2	28.0	1.8*	(1.0, 2.5)
Age, first marriage/union	24.7	26.8	2.2*	(1.4, 2.9)
Age, first child	27.3	29.8	2.5*	(1.5, 3.5)
Household size	3.2	3.1	-0.1	(-0.3, 0.1)
Includes 3 generations or more	3.8	3.0	-0.8	(-3.1, 1.6)
Includes 2 generations or less	85.0	79.7	-5.3	(-10.6, 0.0)
Person not in a census family	11.2	17.3	6.0*	(1.1, 11.0)
Single detached house	66.6	72.2	5.6	(-6.3, 17.6)
Atlantic	11.4	7.2	-4.2*	(-6.8, -1.5)
Quebec	25.6	24.2	-1.5	(-7.7, 4.8)
Ontario	36.1	37.3	1.3	(-8.9, 11.4)
Prairies	18.6	18.1	-0.5	(-5.1, 4.2)
BC	8.4	13.2	4.8*	(1.5, 8.2)
Rural area	28.4	23.2	-5.2	(-13.5, 3.1)
Aboriginal person	7.2	4.0	-3.2	(-9.4, 2.9)
Own education: Less than High School	10.0	6.2	-3.9*	(-7.1, -0.6)
Own education: High School	25.7	17.6	-8.1*	(-15.1, -1.1)
Own education: Trade/Vocational/Apprenticeship	12.5	12.0	-0.6	(-4.6, 3.5)
Own education: Some postsecondary	28.2	28.0	-0.1	(-6.8, 6.5)
Own education: Bachelor or above	23.6	35.9	12.3*	(2.1, 22.5)
Health: Very good/Excellent	63.7	66.8	3.1	(-6.9, 13.1)
Health: Good	32.0	25.9	-6.2	(-16.4, 4.1)
Health: Fair/poor	4.3	7.4	3.1*	(0.9, 5.3)
Employed	86.2	86.9	0.7	(-5.0, 6.4)
Self-employed	15.7	13.2	-2.5	(-7.7, 2.7)
Living with both parents at birth	96.4	97.7	1.3	(-0.6, 3.3)
Living with both parents at 15	67.2	85.3	18.1*	(10.7, 25.5)
Mother education: Less than High School	28.5	26.4	-2.1	(-8.5, 4.3)
Mother education: High School	30.2	32.1	1.9	(-5.3, 9.1)
Mother education: Some postsecondary (below university)	23.7	24.4	0.7	(-5.8, 7.2)
Mother education: University (certificate, diploma or degree)	14.9	14.3	-0.5	(-11.3, 10.3)
Father education: Less than High School	37.4	29.9	-7.5	(-16.7, 1.7)
Father education: High School	23.8	21.0	-2.8	(-9.7, 4.0)
Father education: Some postsecondary (below university)	17.8	23.7	5.9*	(0.6, 11.1)
Father education: University (certificate, diploma or degree)	19.0	22.0	3.0	(-2.7, 8.7)
Visible minority	6.7	8.3	1.6	(-2.4, 5.5)
Born in Canada, Canadian-born parents	76.8	69.4	-7.5*	(-13.5, -1.4)
Foreign-born	8.2	9.4	1.2	(-2.4, 4.8)
Born in Canada, at least one foreign-born parent	14.9	21.2	6.3*	(1.2, 11.4)
Number	**518**	**2817**	**...**	**...**

... not applicable
* difference between "no link" and "link" valus is significant ($p < 0.05$)

Note: Respondents with first observed, unconfirmed parents are respondents who were observed with the corresponding parent for only one period. Confidence intervals are normal-approximation confidence intervals with 95% coverage.
Source: Statistics Canada, Longitudinal and International Study of Adults, 2014, and T1FF, 1982 to 2013.

Table 4b
Respondent characteristics, 1982 to 1995, first observed (unconfirmed) and first confirmed parents

Variables	First observed	First confirmed	Difference	Confidence interval
Age	22.1	24.6	2.6*	(1.8, 3.3)
Sex	49.7	48.2	-1.6	(-11.2, 8.1)
Ever married	9.5	11.2	1.7	(-3.7, 7.1)
Age, first marriage	23.9	25.5	1.6*	(0.1, 3.0)
Age, first marriage/union	21.6	23.8	2.2*	(1.1, 3.3)
Age, first child	23.0	25.6	2.6*	(1.2, 4.0)
Household size	3.5	3.1	-0.4*	(-0.6, -0.1)
Includes 3 generations or more	6.0	2.8	-3.2*	(-5.8, -0.5)
Includes 2 generations or less	83.9	79.0	-4.8	(-12.1, 2.4)
Person not in a census family	10.1	18.1	8.0*	(1.0, 14.9)
Single detached house	66.6	66.8	0.2	(-12.7, 13.1)
Atlantic	5.2	6.1	1.0	(-1.3, 3.2)
Quebec	20.6	23.3	2.7	(-4.0, 9.3)
Ontario	42.5	38.5	-4.0	(-14.5, 6.5)
Prairies	18.9	19.9	1.0	(-4.7, 6.6)
BC	12.8	12.2	-0.6	(-6.7, 5.4)
Rural area	19.8	23.7	3.9	(-6.6, 14.5)
Aboriginal person	7.7	3.4	-4.3	(-16.1, 7.5)
Own education: Less than High School	40.0	12.7	-27.3*	(-35.4, -19.2)
Own education: High School	32.7	35.7	2.9	(-7.7, 13.5)
Own education: Trade/Vocational/Apprenticeship	12.7	7.1	-5.7	(-11.5, 0.1)
Own education: Some postsecondary	8.8	18.8	10.0*	(5.8, 14.3)
Own education: Bachelor or above	5.7	25.8	20.1*	(14.8, 25.4)
Health: Very good/Excellent	69.9	72.9	3.0	(-8.0, 13.9)
Health: Good	21.7	22.8	1.1	(-5.6, 7.7)
Health: Fair/poor	8.1	4.2	-3.8	(-15.6, 7.9)
Employed	56.8	77.9	21.2*	(11.3, 31.0)
Self-employed	7.9	6.0	-1.9	(-15.8, 12.0)
Living with both parents at birth	90.3	97.6	7.2*	(2.2, 12.2)
Living with both parents at 15	58.7	84.6	26.0*	(17.5, 34.5)
Mother education: Less than High School	13.5	7.1	-6.4*	(-11.0, -1.7)
Mother education: High School	23.1	26.7	3.6	(-7.5, 14.6)
Mother education: Some postsecondary (below university)	34.0	33.5	-0.5	(-9.0, 8.0)
Mother education: University (certificate, diploma or degree)	27.8	31.9	4.1	(-3.6, 11.7)
Father education: Less than High School	15.8	11.1	-4.8	(-10.1, 0.6)
Father education: High School	23.5	24.3	0.8	(-10.3, 11.9)
Father education: Some postsecondary (below university)	25.0	29.4	4.4	(-2.4, 11.2)
Father education: University (certificate, diploma or degree)	28.5	33.8	5.4	(-2.5, 13.3)
Visible minority	19.3	18.0	-1.3	(-7.6, 5.1)
Born in Canada, Canadian-born parents	71.7	65.7	-6.0	(-13.7, 1.7)
Foreign-born	12.8	11.5	-1.4	(-7.2, 4.4)
Born in Canada, at least one foreign-born parent	15.5	22.8	7.4*	(2.1, 12.6)
Number	**331**	**2730**	**...**	**...**

... not applicable
* difference between "no link" and "link" valus is significant ($p < 0.05$)

Note: Respondents with first observed, unconfirmed parents are respondents who were observed with the corresponding parent for only one period. Confidence intervals are normal-approximation confidence intervals with 95% coverage.
Source: Statistics Canada, Longitudinal and International Study of Adults, 2014, and T1FF, 1982 to 2013.

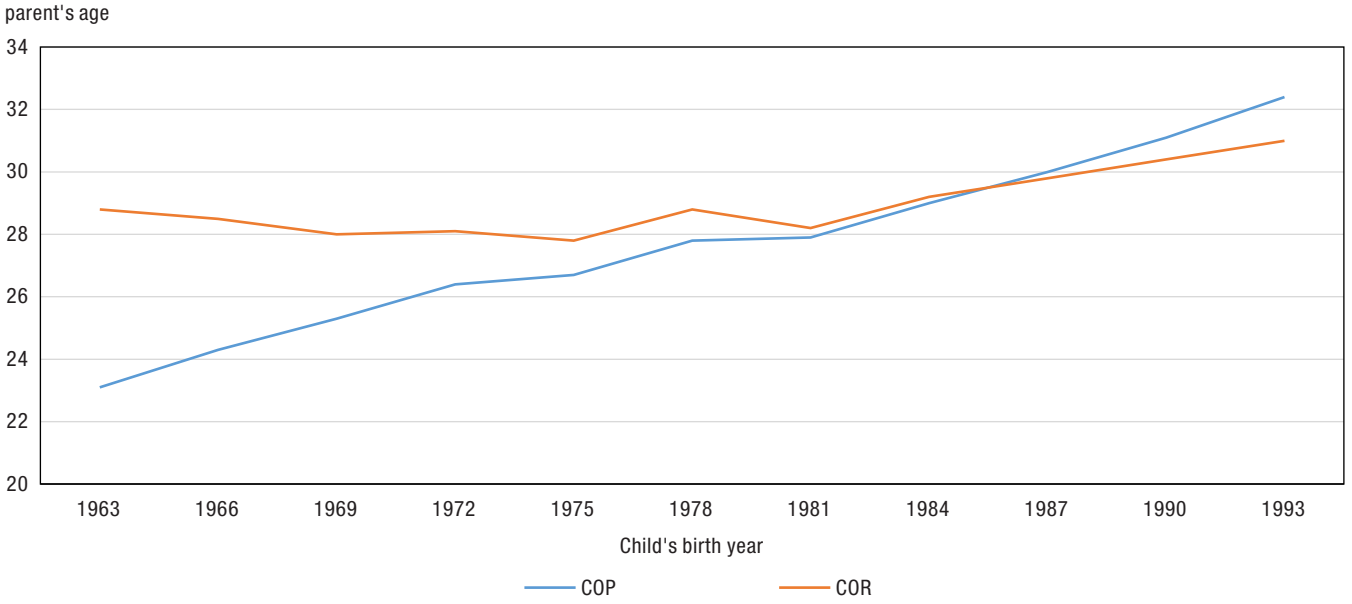
Table 5
Respondent characteristics, 1982 to 1995, first observed (non-birth) and birth parents

Variables	First observed	Birth parent	Difference	Confidence interval
Age	24.3	24.3	0.0	(-0.4, 0.4)
Sex	46.1	50.3	4.2	(-1.5, 9.9)
Ever married	10.6	11.4	0.8	(-2.7, 4.3)
Age, first marriage	25.1	25.5	0.4	(-0.3, 1.2)
Age, first marriage/union	23.4	23.7	0.3	(-0.3, 0.9)
Age, first child	25.2	25.0	-0.3	(-1.2, 0.7)
Household size	3.3	3.0	-0.2*	(-0.4, -0.1)
Includes 3 generations or more	2.9	3.5	0.6	(-0.7, 2.0)
Includes 2 generations or less	79.4	79.8	0.5	(-6.1, 7.0)
Person not in a census family	17.8	16.7	-1.1	(-7.6, 5.4)
Single detached house	68.1	65.6	-2.5	(-9.9, 5.0)
Atlantic	5.7	6.3	0.6	(-0.9, 2.2)
Quebec	21.1	24.6	3.5	(-1.3, 8.3)
Ontario	38.9	39.0	0.1	(-6.0, 6.3)
Prairies	20.8	18.9	-1.9	(-5.6, 1.8)
BC	13.5	11.1	-2.4	(-6.4, 1.6)
Rural area	24.1	22.6	-1.5	(-11.3, 8.4)
Aboriginal person	2.7	4.9	2.3	(-1.5, 6.0)
Own education: Less than High School	15.9	15.7	-0.3	(-3.3, 2.8)
Own education: High School	35.7	35.0	-0.8	(-6.1, 4.5)
Own education: Trade/Vocational/Apprenticeship	6.9	8.4	1.5	(-2.1, 5.1)
Own education: Some postsecondary	18.2	17.2	-1.1	(-5.4, 3.2)
Own education: Bachelor or above	23.2	23.8	0.6	(-5.2, 6.4)
Health: Very good/Excellent	72.4	72.6	0.3	(-5.1, 5.6)
Health: Good	22.8	22.5	-0.4	(-5.2, 4.4)
Health: Fair/poor	4.6	4.7	0.1	(-3.5, 3.6)
Employed	75.7	75.4	-0.3	(-5.1, 4.6)
Self-employed	5.5	6.7	1.2	(-3.3, 5.8)
Living with both parents at birth	97.2	96.3	-0.9	(-2.6, 0.8)
Living with both parents at 15	87.0	76.9	-10.1*	(-14.7, -5.4)
Mother education: Less than High School	7.8	7.9	0.0	(-2.3, 2.4)
Mother education: High School	26.7	25.9	-0.8	(-5.9, 4.3)
Mother education: Some postsecondary (below university)	29.4	37.4	8.0*	(2.9, 13.1)
Mother education: University (certificate, diploma or degree)	35.5	27.8	-7.7*	(-13.5, -1.8)
Father education: Less than High School	12.4	10.9	-1.5	(-4.5, 1.5)
Father education: High School	21.7	26.4	4.7	(-0.4, 9.8)
Father education: Some postsecondary (below university)	28.6	29.1	0.5	(-4.4, 5.3)
Father education: University (certificate, diploma or degree)	35.8	30.9	-4.9	(-11.0, 1.2)
Visible minority	24.5	12.5	-12.0*	(-17.0, -6.9)
Born in Canada, Canadian-born parents	60.2	71.9	11.7*	(6.1, 17.3)
Foreign-born	20.1	3.9	-16.3*	(-21.0, -11.5)
Born in Canada, at least one foreign-born parent	19.6	24.2	4.6*	(0.5, 8.7)
Number	**1454**	**1607**	**...**	**...**

... not applicable
* difference between "no link" and "link" valus is significant ($p < 0.05$)

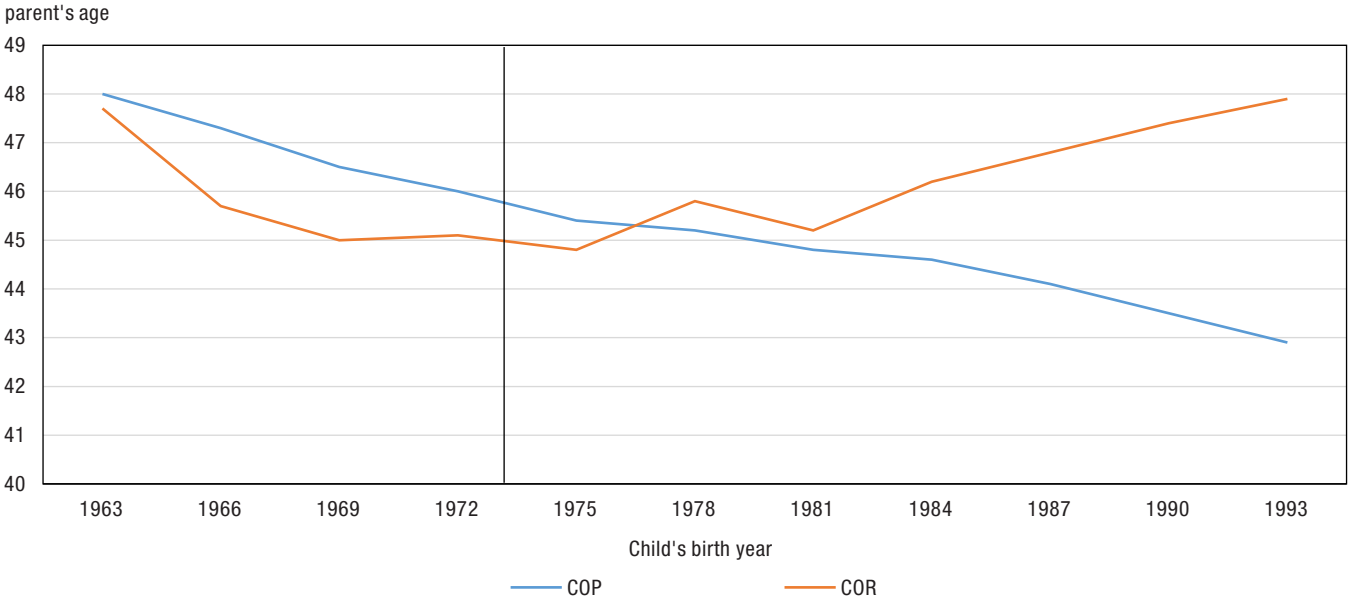
Note: Respondents with first observed, non-birth parents are respondents who were linked to a parent with whom they were not observed between ages 0 to 4 years old. Confidence intervals are normal-approximation confidence intervals with 95% coverage.
Source: Statistics Canada, Longitudinal and International Study of Adults, 2014, and T1FF, 1982 to 2013.

Chart 4
Parent's age at child's birth, by respondent's birth cohort



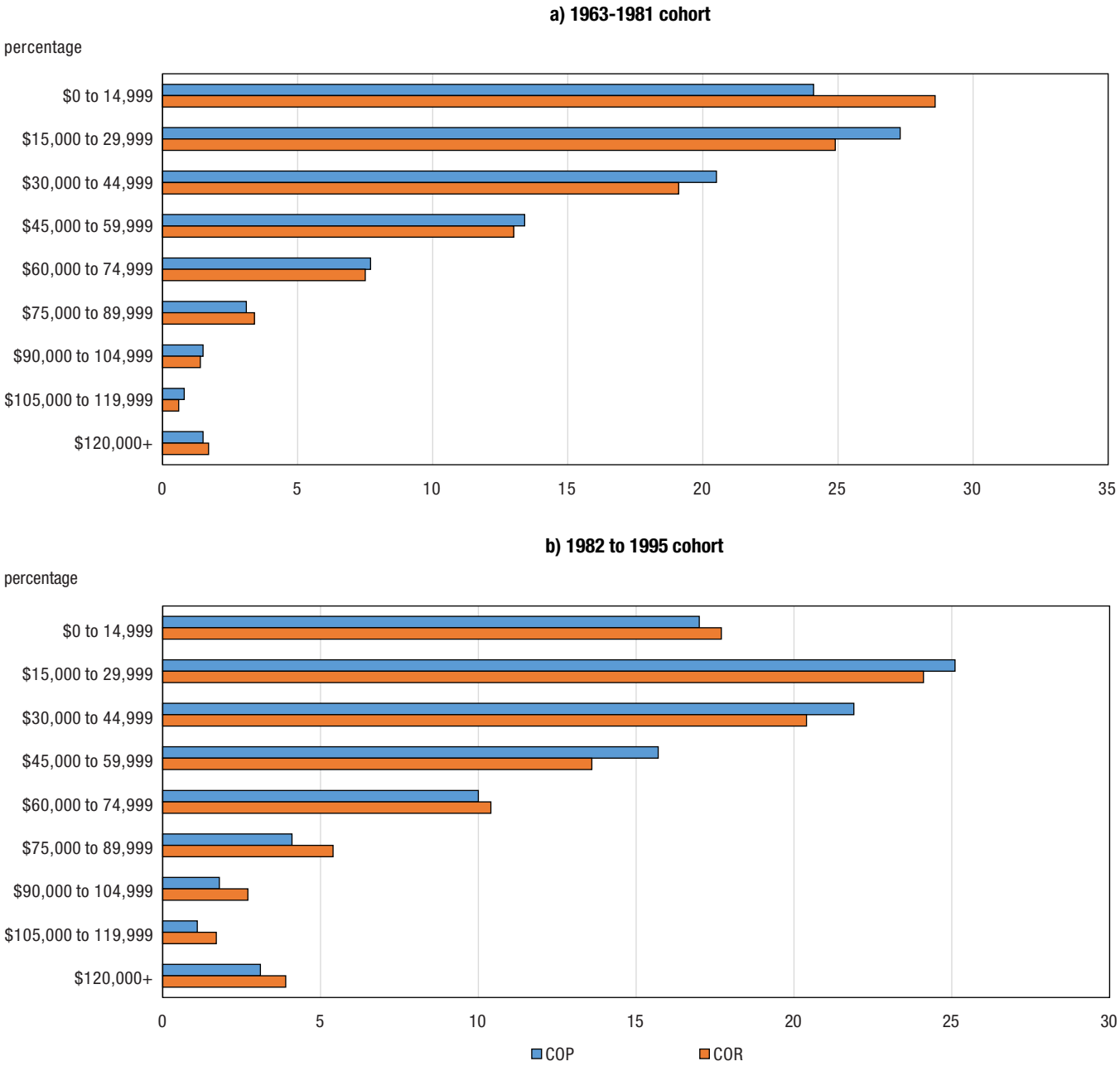
Note: Respondents who immigrated to Canada at 15 years old or later are excluded from the sample.
Source: Statistics Canada, Longitudinal and International Study of Adults, 2014, and T1FF, 1982 to 2013.

Chart 5
Parent's age when their income is measured, by respondent's birth cohort



Note: Respondents who immigrated to Canada at 15 years old or later are excluded from the sample.
Source: Statistics Canada, Longitudinal and International Study of Adults, 2014, and T1FF, 1982-2013.

Chart 6
Distribution of parental income, by parental income measure



Note: Respondents who immigrated to Canada at 15 years old or later are excluded from the sample.
Source: Statistics Canada, Longitudinal and International Study of Adults, 2014, and T1FF, 1982-2013.

Table 6a
Respondent characteristics, 1963 to 1981, with and without COP parental income measure

Variables	COR, no COP	COR and COP	Difference	Confidence interval
Age	44.9	39.6	-5.3*	(-5.8, -4.8)
Sex	48.0	46.4	-1.6	(-6.5, 3.3)
Married	45.9	42.2	-3.7	(-9.3, 1.9)
Common-law	16.9	27.6	10.8*	(7.3, 14.2)
Separated/Divorced/Widowed	14.4	9.6	-4.8*	(-8.3, -1.3)
Single (never married)	22.8	20.6	-2.3	(-7.5, 3.0)
Age, first marriage	27.9	27.7	-0.2	(-0.8, 0.4)
Age, first marriage/union	27.0	26.4	-0.6	(-1.1, 0.0)
Age, first child	29.5	29.3	-0.2	(-0.8, 0.5)
Household size	3.0	3.2	0.1	(0.0, 0.3)
Includes 3 generations or more	2.6	3.1	0.5	(-1.2, 2.2)
Includes 2 generations or less	79.7	80.9	1.2	(-4.1, 6.4)
Person not in a census family	17.7	16.0	-1.7	(-6.9, 3.4)
Single detached house	75.3	71.0	-4.3	(-10.0, 1.4)
Atlantic	8.6	8.1	-0.6	(-2.1, 1.0)
Quebec	22.8	24.5	1.8	(-2.4, 5.9)
Ontario	38.3	37.2	-1.1	(-6.7, 4.6)
Prairies	17.3	18.2	0.9	(-2.4, 4.1)
BC	12.9	12.0	-1.0	(-4.4, 2.4)
Rural area	29.7	24.5	-5.3	(-12.5, 2.0)
Aboriginal person	5.6	4.0	-1.5	(-4.8, 1.8)
Own education: Less than High School	7.7	6.3	-1.4	(-4.4, 1.6)
Own education: High School	19.5	17.8	-1.7	(-6.1, 2.7)
Own education: Trade/Vocational/Apprenticeship	12.7	12.2	-0.5	(-3.5, 2.5)
Own education: Some postsecondary	27.6	28.8	1.2	(-3.4, 5.8)
Own education: Bachelor or above	32.5	34.5	2.0	(-2.8, 6.8)
Health: Very good/Excellent	63.5	67.5	4.0	(-0.9, 8.8)
Health: Good	27.0	26.8	-0.3	(-4.8, 4.3)
Health: Fair/poor	9.4	5.7	-3.7*	(-6.9, -0.5)
Employed	86.7	87.0	0.4	(-3.7, 4.4)
Self-employed	13.6	13.9	0.2	(-3.2, 3.7)
Living with both parents at birth	99.0	96.8	-2.3*	(-3.5, -1.1)
Living with both parents at 15	87.3	80.6	-6.8*	(-10.4, -3.1)
Mother education: Less than High School	35.5	23.7	-11.8*	(-16.4, -7.2)
Mother education: High School	28.4	33.3	4.9*	(0.2, 9.5)
Mother education: Some postsecondary (below university)	20.4	24.9	4.4*	(0.1, 8.7)
Mother education: University (certificate, diploma or degree)	11.8	15.6	3.8*	(0.1, 7.5)
Father education: Less than High School	41.1	27.8	-13.2*	(-17.9, -8.5)
Father education: High School	14.9	23.2	8.3*	(4.6, 12.0)
Father education: Some postsecondary (below university)	22.3	23.1	0.9	(-3.6, 5.3)
Father education: University (certificate, diploma or degree)	18.2	22.6	4.4*	(0.2, 8.6)
Visible minority	6.7	8.8	2.1	(-1.4, 5.5)
Born in Canada, Canadian-born parents	67.7	70.8	3.1	(-1.8, 8.0)
Foreign-born	9.3	9.2	-0.2	(-3.3, 3.0)
Born in Canada, at least one foreign-born parent	22.9	20.0	-2.9	(-7.5, 1.6)
Number	**909**	**2522**	**...**	**...**

... not applicable
* difference between "no link" and "link" valus is significant ($p < 0.05$)

Note: Respondents without COP parental income measure are respondents for whom the COR parental income measure could be computed, but not the COP measure. Confidence intervals are normal-approximation confidence intervals with 95% coverage.
Source: Statistics Canada, Longitudinal and International Study of Adults, 2014, and T1FF, 1982 to 2013.

Table 6b
Respondent characteristics, 1982 to 1995, with and without COP parental income measure

Variables	COR, no COP	COR and COP	Difference	Confidence interval
Age	21.9	24.6	2.8*	(2.1, 3.4)
Sex	40.1	48.7	8.6*	(0.7, 16.5)
Ever married	12.3	11.5	-0.8	(-7.9, 6.3)
Age, first marriage	25.0	25.4	0.3	(-0.8, 1.5)
Age, first marriage/union	23.6	23.5	-0.1	(-1.4, 1.2)
Age, first child	24.1	25.1	1.0	(-0.7, 2.8)
Household size	3.8	3.1	-0.7*	(-0.9, -0.5)
Includes 3 generations or more	5.3	2.9	-2.4	(-5.3, 0.4)
Includes 2 generations or less	82.8	79.4	-3.4	(-10.1, 3.3)
Person not in a census family	11.9	17.7	5.8	(-0.2, 11.9)
Single detached house	57.6	67.0	9.4	(-1.1, 20.0)
Atlantic	7.9	6.3	-1.5	(-4.2, 1.1)
Quebec	15.2	24.3	9.2*	(3.4, 14.9)
Ontario	46.6	37.4	-9.2*	(-18.2, -0.1)
Prairies	20.2	19.6	-0.7	(-6.8, 5.5)
BC	10.2	12.4	2.2	(-2.8, 7.2)
Rural area	31.8	23.1	-8.7	(-26.4, 9.0)
Aboriginal person	1.9	4.1	2.1	(-0.6, 4.9)
Own education: Less than High School	36.2	13.6	-22.6*	(-29.9, -15.3)
Own education: High School	38.6	35.1	-3.4	(-11.5, 4.6)
Own education: Trade/Vocational/Apprenticeship	3.5	8.8	5.3*	(2.2, 8.4)
Own education: Some postsecondary	9.7	18.8	9.0*	(4.2, 13.9)
Own education: Bachelor or above	11.9	23.7	11.8*	(5.9, 17.6)
Health: Very good/Excellent	69.4	72.8	3.4	(-5.0, 11.7)
Health: Good	22.0	22.9	1.0	(-6.2, 8.1)
Health: Fair/poor	8.6	4.1	-4.5	(-10.9, 2.0)
Employed	72.3	76.0	3.7	(-3.2, 10.6)
Self-employed	3.5	6.3	2.8	(-1.0, 6.6)
Living with both parents at birth	92.1	97.1	5.0	(-0.3, 10.3)
Living with both parents at 15	75.6	80.6	5.0	(-1.9, 11.9)
Mother education: Less than High School	11.6	7.7	-4.0	(-9.0, 1.1)
Mother education: High School	29.5	26.3	-3.2	(-11.3, 4.9)
Mother education: Some postsecondary (below university)	30.2	34.8	4.6	(-3.9, 13.1)
Mother education: University (certificate, diploma or degree)	28.3	30.3	1.9	(-5.5, 9.4)
Father education: Less than High School	11.7	12.0	0.4	(-4.5, 5.2)
Father education: High School	22.8	24.7	1.8	(-5.2, 8.9)
Father education: Some postsecondary (below university)	22.0	29.6	7.5*	(1.1, 13.9)
Father education: University (certificate, diploma or degree)	40.4	31.5	-8.9	(-18.2, 0.4)
Visible minority	41.0	15.3	-25.8*	(-34.0, -17.5)
Born in Canada, Canadian-born parents	44.1	69.7	25.7*	(16.8, 34.5)
Foreign-born	42.4	8.1	-34.3*	(-42.7, -26.0)
Born in Canada, at least one foreign-born parent	13.5	22.2	8.7*	(2.5, 14.8)
Number	**322**	**2903**	**...**	**...**

... not applicable
* difference between "no link" and "link" valus is significant ($p < 0.05$)

Note: Respondents without COP parental income measure are respondents for whom the COR parental income measure could be computed, but not the COP measure. Confidence intervals are normal-approximation confidence intervals with 95% coverage.
Source: Statistics Canada, Longitudinal and International Study of Adults, 2014, and T1FF, 1982 to 2013.

Table 7
Intergenerational income correlation, without and with selection correction

		Without correction		With correction	
		Coefficient	Confidence interval	Coefficient	Confidence interval
Baseline estimates	Parental income	0.275	(0.221, 0.329)	0.250	(0.202, 0.298)
	Constant	8.006	(7.418, 8.595)	8.558	(8.030, 9.086)
	Number	3484	...	4414	...
Confirmed parent	Parental income	0.305	(0.246, 0.365)	0.273	(0.220, 0.325)
	Constant	7.681	(7.032, 8.330)	8.451	(7.867, 9.034)
	Number	3035	...	4414	...
Sample with COP	Parental income	0.263	(0.207, 0.319)	0.232	(0.182, 0.282)
	Constant	8.137	(7.530, 8.745)	8.747	(8.200, 9.295)
	Number	3517	...	4394	...

... not applicable

Note: Confidence intervals are normal-approximation confidence intervals with 95% coverage.
Source: Statistics Canada, Longitudinal and International Study of Adults, 2014, and T1FF, 1982 to 2013.

References

Atkinson, A. B., Maynard, A. K., & Trinder, C. G. (1983). *Parents and Children: Incomes in Two Generations*. London: Heineman Educational Books.

Becker, Gary S., and Nigel Tomes. 1979. "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility." *Journal of Political Economy* 87(6): 1153 – 89.

Becker, G. S., & Tomes, N. (1986). Human capital and the rise and fall of families. *Journal of Labour Economics, 4*(3, Part 2), S1-S39.

Blanden, J. (2005). Love and money: Intergenerational mobility and marital matching. *Analytical Studies Branch Research Paper Series, Family and Labour Studies Division, Ottawa*.

Bloome, D. (2017). Childhood family structure and intergenerational income mobility in the United States. *Demography, 54*(2), 541-569.

Chen, W. H., Ostrovsky, Y., & Piraino, P. (2017). Lifecycle variation, errors-in-variables bias and nonlinearities in intergenerational income transmission: new evidence from Canada. *Labour Economics, 44*, 1-12.

Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014a). Where is the land of opportunity? The geography of intergenerational mobility in the United States. *Quarterly Journal of Economics, 129*(4), 1553-1623.

Connolly, Marie, Catherine Haeck and David Lapierre (2018), "Social mobility trends in Canada: Going up the Great Gatsby Curve." Working paper.

Cook, K., & Demnati, A. (2000). Weighting the Intergenerational Income Data File. *Social Survey Methods Division, Statistics Canada, mimeo*.

Corak, M., & Heisz, A. (1995). The intergenerational income mobility of Canadian men. *Canadian Business Economics*, 59-69.

Corak, M., & Heisz, A. (1999). The intergenerational earnings and income mobility of Canadian men: Evidence from longitudinal income tax data. *The Journal of Human Resources 34*(3), 504-533.

Corak, M. (2017). "Divided Landscapes of Economic Opportunity: The Canadian Geography of Intergenerational Income Mobility", *Human Capital and Economic Opportunity Working Paper* No. 2017-043.

Frenette, M., Green, D. A., & Milligan, K. (2007). The tale of the tails: Canadian income inequality in the 1980s and 1990s. *Canadian Journal of Economics/Revue canadienne d'économique*, 40(3), 734-764.

Frenette, M., Green, D., & Picot, G. (2004). *Rising income inequality in the 1990s: An exploration of three data sources*. Statistics Canada, Analytical Studies Branch.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, 153-161.

Hemeon, J. (2016). *Historical Data Linkage Quality: The Longitudinal and International Study of Adults, and Tax Records on Labour and Income*. (Statistics Canada, Catalogue no. 89-648-X). Retrieved from http://www.statcan.gc.ca/pub/89-648-x/89-648-x2016001-eng.htm

Messacar, D. (2017). Big tax data and economic analysis: Effects of personal income tax reassessments and delayed tax filing. *Canadian Public Policy, 43*(3), 261-283.

Oreopoulos, P. (2003). The long-run consequences of living in a poor neighbourhood. *Quarterly Journal of Economics, 118*(4), 1533-1575.

Oreopoulos, Philip, Marianne Page and Ann Huff Stevens (2008), "The intergenerational effects of worker displacement," *Journal of Labor Economics* 26(3): 455-483.

Solon, Gary (1992). Intergenerational income mobility in the United States. *American Economic Review, 82*(3), 393-408.

Statistics Canada (2018). Longitudinal and International Study of Adults 2016 (Wave 3) User Guide. Ottawa: Statistics Canada.

Zimmerman, D. J. (1992). Regression toward mediocrity in economic stature. *American Economic Review, 82*(3), 409-429.