Basic Income: Characteristics Related to Presence in and Absence From the Tax System

David A. Green
Pablo Gutierrez
Kevin Milligan
Erik Snowberg
Vancouver School of Economics, University of British Columbia

Date: January 2021

Author Note

The authors can be contacted at david.green@ubc.ca, pgutiecu@gmail.com, Kevin.Milligan@ubc.ca, and Erik.Snowberg@ubc.ca.

Research paper commissioned by the Expert Panel on Basic Income, British Columbia. We gratefully acknowledge funding from the Government of British Columbia (spcs46008190052 and spsc46008190046) that helped support this research. All inferences, opinions, and conclusions drawn in this paper are those of the authors, and do not reflect the opinions or policies of the Data Innovation Program or the Province of British Columbia. Part of the research and analysis are based on data from Statistics Canada and the opinions expressed do not represent the views of Statistics Canada.

Abstract

One claim advanced for a Basic Income is that it can be administered through the tax system, making it more transparent, easier to access, and less costly to administer than other transfer approaches. Of course, these advantages are lessened to the extent that people do not file taxes. In this paper, we data from a combination of linked Census, tax and death records, data on opioid overdoses, data from BC's IA system, and data from a survey of residents of Single Room Occupancy hotels in Vancouver to paint a picture of the extent and nature of "falling through the cracks" in the tax and transfer system. We find that between 3 and 6.6% of the Canadian population are not known to the tax system at all. Another group of people are in the tax system records but did not file in a given year. Adding these two groups together, we find that between 11% and 15% of the population are either not in the tax system at all or do not file taxes in a year.

The group who are in the tax records but did not file a T1 in a year are disproportionately non-earners from low income households. Not filing taxes shows considerable persistence across years. Thus, the problems associated with not filing taxes could end up focused on a small core of continual non-filers. The people not in the tax records or Census files at all are very disproportionately likely to have died from a death of despair (suicide, drugs or alcohol related) and to have died in low income neighbourhoods and neighbourhoods with a higher proportion of people who have moved in the previous year. This suggests that this is a vulnerable population. Given that they are missing from both tax related and survey attempts by the government to contact them, they would likely be a particularly difficult group to reach with tax-based social benefits.

Introduction

One claim advanced for having a basic income is that it can be administered through the tax system, making it more transparent, easier to access, and less costly to administer than other income transfer approaches. For example, Van Parijs and Vanderborght, leading advocates for a basic income, state, "… in sufficiently formalized economies with tax systems that work reasonably well, the overall administrative cost of achieving any given rate of take-up among net beneficiaries can safely be expected to be less for a universal scheme than for a means-tested one" (Van Parijs & Vanderborght, 2017, p.18). Of course, these advantages are lessened when people do not file taxes. In that case, added systems would need to be put in place to find these non-filers and transfer benefits to them. Of particular concern is the possibility that it is the lowest income earners and most vulnerable parts of our population who do not file taxes.

In this paper, we use various sources to investigate two questions:

- How many people are not present in the tax and transfer system data in a given year?
- To the best extent possible, what do their characteristics seem to be?

The answers to both questions are complicated by the fact that we are trying to investigate people who do not exist in key datasets. We attempt to address this issue by using linked census and tax system data for 2015 and 2016, treating the census as the complete population.[1] Of course, some people may be missed by both the tax data and the census. We form estimates of the size and characteristics of that group by using death certificate data linked to the other two datasets. The core idea is that everyone ends up in the death records at some point (not counting emigrants), and therefore those records provide a measure of the complete population. Even this approach has issues, since the set of people who appear only in the death records may not represent all the people who are not in the tax or census data in a year, even after conditioning on age. We address those issues using a bounding approach. Not surprisingly, we find evidence of a somewhat sizable group of people who either do not file taxes in a year (corresponding to 7% of the people present in the census) or are not seen in tax data at all, even when we look backwards for 10 years (corresponding to 3% of people in the census).

Knowing that some people are not present in tax data raises issues beyond implications for implementing a basic income through the tax system. It also points to the possibility that there are people who have "fallen through the cracks" of the current transfer system—that is, people who qualify for benefits but, at least according to tax records, are not receiving those benefits. We investigate the extent of falling through the cracks, in part, by examining the characteristics of those who do not have tax records.

[1] This work was initiated by us but has since developed into an extended project into data quality being carried out by one of us (Pablo Gutierrez), and Andrew Heisz and Eric Olsen from Statistics Canada.

A key reason why people may fall through the cracks could be homelessness. People who are homeless may have difficulty establishing their existence and location when applying for benefits. We investigate the relationship of homelessness to benefit receipt, in part, by using B.C. Income Assistance (IA) data to count the number of people who are receiving benefits but are listed as having no fixed address. Interestingly, we find that there are approximately 7,000 such people in the data in 2016, which is very similar to the number in the homeless count for the province for that year. This may suggest that people who are homeless are not entirely out of contact with the system, though their contact may be sporadic and amount to receiving payments that are less than what they are qualified for.

Of relevance to this paper, Robson and Schwartz (2020) use data from the Canadian Survey of Financial Security to examine patterns of tax non-filing. In the context of their survey, Statistics Canada links the survey respondents to their tax records based on location of residence, age, gender, and some responses on income. The cases that Statistics Canada was not able to match are counted as non-filers and amount to approximately 12% of the sample (weighted). Robson and Schwartz (2020) also report numbers from Statistics Canada publications showing that the agency's estimates of non-filing based on census data is around 14%. They investigate the characteristics of non-filers, highlighting lower filing among men, younger people, and the self-employed but no strong relationship with imputed income. They then provide estimates of the government benefits foregone by the non-filers.

We are able to make advances relative to Robson and Schwartz's (2020) very useful work because of access to data from tax, census, and death records. The Canadian Survey of Financial Security has a response rate of only 70%, raising questions about sample selection and potential biases in the estimates. The death data, in particular, does not suffer from this concern. In addition, we are able to break non-filers down into those who are present in the tax data (either because there are tax forms in their name but no T1 was filed for the year or because there are no tax forms in the current year but tax filings in previous year) and those who are not present at all in the tax data. Based on our estimates, the latter group does not show a significant relationship with imputed income, but the former group shows a strong relationship.

Linking Census and Tax Data

In our first exercise, we treat the respondents to the 2016 census as the true population of B.C., linking them to tax records. In that census, respondents were *not* asked to report their income. Instead, they were automatically linked to tax records using the combination of full name, date of birth, sex, and address. Since the census was conducted in June 2016, the nearest tax records were those for the 2015 tax year (which could be filed as late as the end of April 2016). Within the tax records, we can define three states:

- **tax filer:** a person who filed a T1 tax form in 2016 for the 2015 tax year.

- **linked non-filer**: a person who did not file a T1 tax form in 2016 but who is in the tax system in the sense of there being tax forms for the person for the 2015 tax year, and/or the person filed taxes in previous years or at least has a social insurance number.
- **non-linked:** a person with no records in the tax system. We include only individuals age 20 and over since teenagers are not eligible for some tax credits, so face different incentives to show up in the tax system.

Because our base is the universe of people captured in the 2016 census, we have information for all three groups. In particular, for all census members, we have age, sex, family status, and location. In addition, for those who filled out the long-form census, we have information on education, immigration status, geographic mobility, and more.

Of course, there will be people who are not captured in the census and these may be disproportionately people who are homeless or in other vulnerable situations (though, in recent censuses, Statistics Canada has made an effort to include people who are homeless). We return to discussing that issue in the section on death certificate data. First, though, we ignore any such biases and focus on what we can learn by treating the census as capturing the whole population.

Imputing Earnings, Income, and Benefit Receipt

We are particularly interested in the question of whether low-income individuals and families are likely to be missing from tax data since one goal of a basic income is to better transfer resources to that group. The difficulty, of course, is that we do not have income, earnings, or government transfer benefit receipt information for non-tax filers. In previous censuses, respondents were asked to self-report their income, or at least they were given the option to do so—though, of course, some chose not to respond. However, in the 2016 census, respondents were explicitly told that there would be no income questions and that Statistics Canada would link their census records to their tax records.

To address this issue, we use Statistics Canada's imputed values. For linked non-filers, the imputation uses information from tax forms that were submitted to the Canada Revenue Agency (CRA) by other parties and/or information from tax filings for the previous tax year for the individual. Key among the tax forms (other than the T1) are the T4 (submitted by employers to their employees and showing employee earnings) and the T5007 (submitted by provincial governments and showing receipt and amount of IA benefits). Receipt of employment insurance from the federal government is also recorded on tax forms of this type.

These forms allow for filling in fields in a person's (non-filed) T1. If the relevant slips do not exist, then they use a nearest-neighbour matching algorithm that finds matches for specific T1 fields using information from other people (T1 tax filers) and from the individual's own tax filings in previous years. For non-linked individuals, there are no slips available and the nearest-neighbour matching is done using candidates from both T1 filers and from available slips from linked non-filers.

In the census data, Statistics Canada reports imputed income fields for non-tax filers, treating it as an estimate of the actual values for those fields. Alternatively (and particularly for those for whom there are no tax forms), one can view the imputed income less as a direct estimate of income for the individual than as an index that aggregates their education, age, and other characteristics, along with features of their neighbourhood, to construct a simple measure of socio-economic status.

Results

The first numeric row in Table 1 reports the distribution of census members across the three states, showing that: 89.1% filed taxes for the 2015 tax year; 7.4% did not file for that tax year but were in the CRA system; and 3.5% were not in the CRA system.

Demographic Characteristics

The remainder of Table 1 contains distributions across the three states for various demographic characteristics. In each case, the first set of three columns gives the proportion of individuals in our three states for a given characteristic value. The second set of three columns gives the proportion of people in one of our states with a given characteristic value. Thus, the first row of "Education" shows that 4.5% of the high school dropouts were not linked, 7.8% were linked non-filers, and 87.7% were filers. From the second set of columns, high school dropouts made up 15.9% of the not linked, 13% of linked non-filers, 12.1% of filers, and 12.3% of everyone in the census. Thus, dropouts are somewhat more likely to be not linked and linked non-filers than is the case in the overall population—an initial indication that people of a lower socio-economic status are more likely to be missing from tax data. This is borne out by the remaining education numbers, with the proportions in the not linked and linked non-filers falling almost monotonically with education level. The 4.5% of high school dropouts who are not in the CRA data is substantial, as is the 12.3% who did not file in the tax year (i.e., those in one of the first two states).

Simply delivering benefits through the tax system would miss a substantial number of people of concern in a group. Other mechanisms would have to be enacted to make a universal basic income truly universal. At that same time, the education gradient is not terribly steep: even of those in the highest education group (those with a university degree above a bachelor's), about 8% did not file taxes in 2016. Thus, delivering a demogrant through the tax system would disproportionately benefit the most educated, but not to a huge degree.

Age has the same feature of greater under-representation in tax data for the lower income (younger) group. In this case, the differences are substantial. Among 20 to 24 year olds, 5.3% are not linked, and over 16% did not file taxes. In comparison, only 6% of those 65 to 69 years old are either not linked or linked non-filers.

On the other side of the inequality ledger, females would be more likely to receive a tax delivered "universal" benefit since they are less likely to be either not linked or linked non-filers. It is worth noting that when we run linear probability models with the dependent variable equalling 1 either for the not linked or for the linked non-filers, and the covariates being all the variables in these tables, the age profile becomes essentially flat. That is, the high rate of non-filing among young people is accounted for by variables such as being in the lowest decile of the earnings distribution rather than age. We report the regression results below.

Immigrants who arrived before 2016 were more likely to file taxes than people born in Canada. Having even one immigrant parent tended to raise the likelihood of a person filing taxes. Much lower rates of filing are observed for non-permanent residents (such as temporary foreign workers) and for people who arrived in 2016 and so might not have any 2015 Canadian income to report to the CRA. Neither of the latter outcomes is surprising.

Visible minority status, overall, has very little impact on filing rates, though some specific groups have very high not linked and linked non-filing rates.

Not having children is associated with higher rates of linked non-filing but not of being not linked. This is probably related to the incentives for those with children to file to receive child-related benefits. Finally, single, childless individuals have high rates of being not linked (5.9%) and of being linked non-filing (11.4%). This finding is important since other work in the overall project indicates that this group tends to be underserved by transfer support in general. Trying to rectify that through benefits delivered through the tax system as it currently stands would be less than perfectly successful.

Earnings and Income

In Table 2, we present similar information for different income concepts. The first three-column panel contains the distribution of individual level employment earnings (all income received as wages, salaries, and commissions from paid employment, and net self-employment income from a farm or non-farm unincorporated business and/or professional practice). The second three-column panel shows individual after-tax and transfer total income from all sources. Recall that earnings and income values are imputed for the not linked and calculated using existing forms for the linked non-filers. Consequently, it is likely better to interpret the results for the not linked in particular in this table not so much as telling us directly about the income distribution, but as telling us whether people whose characteristics such as education and age that make them likely to have low earnings and income are more or less likely to be observed in the administrative data.

Individuals in the first decile of employment earnings are slightly more likely to be not linked (4.1% compared to 3.5% for the population as a whole) but much more likely to be linked non-filers (19.2% compared to 7.4% for the population as a whole). The same pattern holds for the disposable income distribution. This also shows up in the second set of columns, showing the proportions of each data group in each decile. The non-linked population is spread quite evenly across the distribution in terms of their predicted earnings. This implies that they are not disproportionately low educated, young, etc. On the other hand, 28% of linked non-filers have predicted income in the bottom decile and 12% in the second decile, with everyone else spread somewhat evenly across the remaining deciles.

Thus, what is associated with low predicted earnings and income is not so much being missing from the tax system altogether as not filing in a given tax year. Recall that to be in the linked non-filer state, a person can have current tax year forms from employers or governments while not filing their T1, and/or have filed in an earlier year or simply having a social insurance number but not filing in 2016. It is also worth noting that the people in the bottom earnings decile likely have a strong incentive to file since they would get the maximum GST credit as well as B.C.'s carbon tax credit. Recall that the table only includes people age 20, and so our results do not reflect a lack of access to the GST credit for teenagers. One possibility is that these are people who receive at least intermittent IA, but are without a fixed address and do not have enough knowledge of the tax system and its credits to file. We will investigate this group further, but there are three points of interest to note here: First, there is a significant group at the bottom of the distribution who would be missed by a tax-based distribution of a basic income—or, at least, would be missed in some years. Second, these people are not responding to the incentives of the available tax credits and might not file in response to incentives associated with a basic income. Third, a tax system in which a T1 form is automatically filed for anyone who has information that has generated any tax form in a year would pick up many of the non-filers and address the issues raised here.

In the remainder of the income distribution there is a slight U-shape, with those in the top income decile having the highest filing rates. Again, a "universal" benefit delivered through the tax system would go disproportionately to those higher up the distribution.

Benefit Receipt and the Tax System

In Table 3, we present similar results for transfer benefit receipt and filing taxes. The first panel shows that receipt of employment insurance benefits is not associated with a disproportionate amount of non-filing of either type. On the other hand, recipients of IA are almost twice as likely to be linked to non-filers as to the rest of the population and are somewhat more likely to be non-linked. Overall, nearly one in five IA recipients did not file taxes compared to about one in ten of non-IA recipients. This raises obvious concerns about the connection of the tax system to the vulnerable population serviced by IA. We investigate IA recipiency in more detail using provincial administrative data below.

The next panel in Table 3 shows that among linked non-filers in a year approximately half file taxes in the following year. This shows that non-filing is a relatively persistent state but also that people move out of that state at a sizable rate. Related to the previous point on transfer beneficiaries who do not file, it would be interesting to know the extent to which linked non-filers are in that state because they have filed in previous years versus never having filed but still being in the CRA system because of forms filed by employers or governments in some years.

Finally, changing residence in the previous year is associated with an approximate doubling in the rates of both being non-linked and being a linked non-filer. Migration, of course, could be associated with many underlying characteristics, including being a temporary foreign worker.

Linear Probability Models

The tables to this point (Tables 1, 2 and 3) show the associations of individual characteristics with the status of being present or absent from administrative data. However, those characteristics will be correlated, making it difficult to know whether, for example, age reflects the true effects of age or correlations with other factors. We address this by estimating two linear probability models. The first is estimated on the set of all census respondents and has as its dependent variable a dummy variable equalling 1 if the person is in the tax data (is linked) and zero otherwise. The second is estimated on the set of all people who are both in the census and in the tax data and has as its dependent variable a dummy variable equalling 1 if the person filed taxes in 2016 (i.e., for the 2015 tax year). For each dependent variable, we estimate three specifications:

- Model 1 includes only demographic variables (age, education, gender, presence of children, family status, ethnicity, immigrant status, and Aboriginal status).
- Model 2 has the same demographic variables plus controls for weeks worked, employment income, IA, and investment income.
- Model 3 adds family an income decile.

In the three models, Model 1 includes a complete set of Census Metropolitan Area (CMA) effects while Models 2 and 3 add in industry effects.

The estimated coefficients from the regressions for the three models with linked as the dependent variables are given in Table 4. We do not report standard errors since, with few exceptions, the sample sizes are such that all the estimated coefficients are statistically significantly different from zero. The asterisks next to the estimated coefficients mark levels of statistical significance. As with the simple tabulations, linked status has a weak association with age,[2] with small estimated coefficients and no discernable pattern across age groups. Single

[2] Weak in the size of the effect: because of the sample size, the coefficients for the different age groups all represent statistically significant differences relative to the base group of 40 to 44 year olds.

adults are slightly less likely to be linked to the tax system than are married couples (with an estimated coefficient implying a 1.9% lower rate of connection), while having children has very little effect.

As in the simple tabulations, non-permanent residents and those who arrived in the year of the census survey are less likely to be linked to tax data, but otherwise, immigration status has very little effect. Changing residence in the last year is also negatively related to being linked to tax data. This points to a further difficulty with using the tax system to deliver benefits: those who move around (something one would expect for people with few fixed assets such as houses) are less likely to be in tax data.

Finally, having only high school or lower education is associated with a lower probability of being in the tax system (about 1.5% lower for high school dropouts compared to people with a bachelor's degree), but there is little discernable pattern with education level beyond high school.

The patterns related to demographic variables change very little when we add in controls for weeks worked and industry. The latter controls point to a positive association between being in the tax system and working, with people who did not work in 2015 or 2016 being 2.1% less likely to be in the tax system. For workers, however, there is little association with the number of weeks worked in 2015. This fits with evidence from the simple tabulations and the education coefficients: there is some association with being not linked and the categories of lowest earnings status (i.e., high school dropouts and those who did not work at all), but that association is moderate in size and there is no connection with higher levels of education or weeks worked. The implication is that being not linked is a somewhat randomly distributed status rather than a strong marker for the most economically vulnerable.

The section on response mode includes controls for the method of contact for the census, showing that those who filled out the census in paper form were less likely to be in the tax system. Including the filing controls does not substantially alter the other estimated effects.

In Table 5, we present the estimates for the models with tax filing as the dependent variable. Here, the relationships to the various skill, earnings, and demographic variables are stronger. When we include weeks worked, individual earnings, and family income variables (as in Model 3), people with zero reported earnings are eight percentage points less likely to file taxes. If they are in a family with income in the first decile of the family income distribution, they are another eight percentage points less likely to file taxes. Thus, a person from a poor family who does not work during the year is 16 percentage points less likely to file taxes. This is a very large effect.

Interestingly, the effect of IA conditioning on earnings and income is positive. Thus, the strong negative relationship between IA and filing taxes that we saw in the raw data is really capturing the fact that IA recipients have low income and earnings. Being in contact with the IA system conditional on having low income and earnings, if anything, increases the chances a person will file taxes. While it’s pure speculation, it’s possible that the IA staff tell recipients about incentives to filing, such as receiving the GST credit, and that individuals who do not receive IA may not know about (or do not know how to get the credit).

One can see, also, the effects of controlling for other factors when looking at any particular effect. For age, for example, the estimated pattern is U-shaped with those in their 20s and those over age 60 showing stronger tendencies to file taxes than the 40 to 44 year old base group. In contrast, in the simple tabulations, people in their 20s were observed to be less likely to file taxes (unconditional on, for example, their education or earnings level). Having low education reduces the probability of tax filing, but here, too, the effects are reduced when we control for income and earnings.

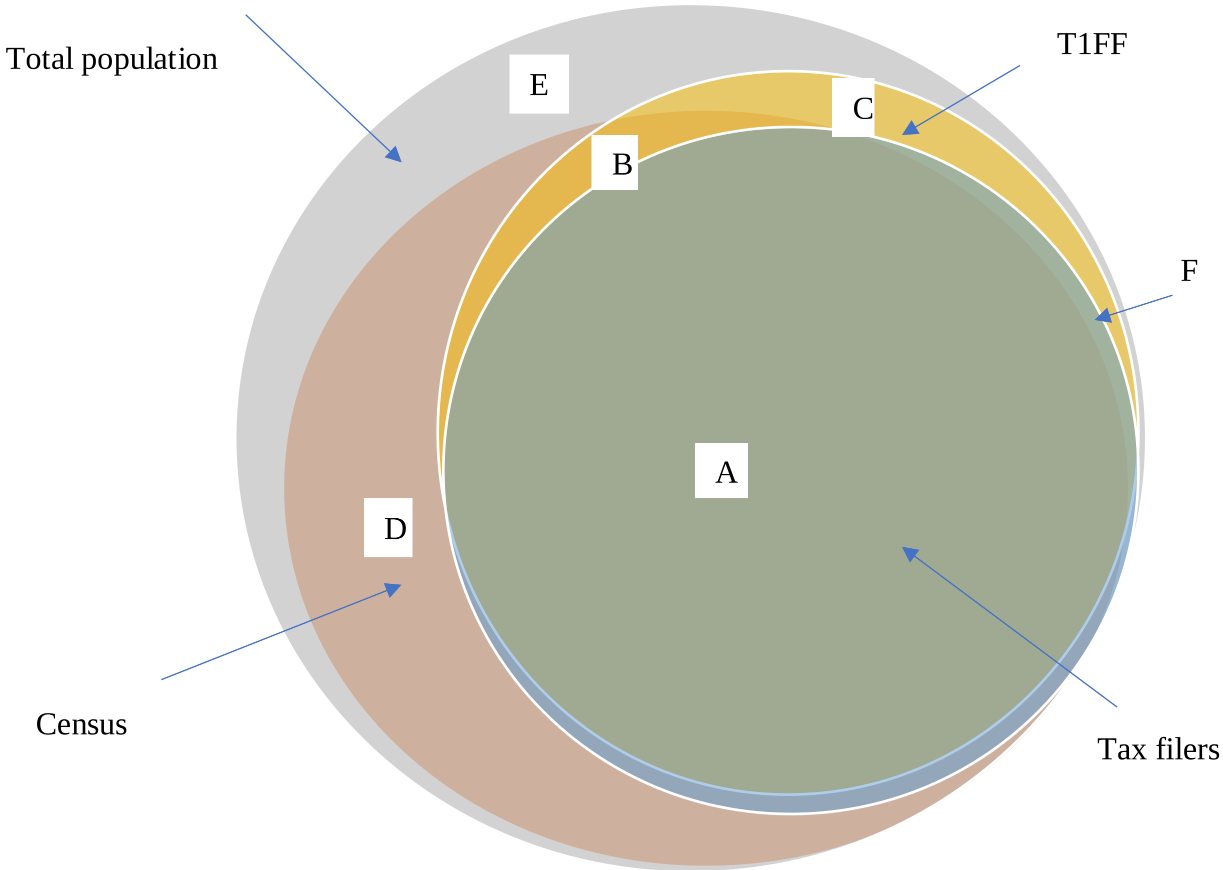
Together, these patterns reinforce the conclusion from the simple tabulations that not being in the tax system at all does not appear to have strong associations with characteristics that would place a person at the bottom of the income and earnings distribution. However, not filing taxes in a given year does have a strong association with having zero earnings and being in the first decile of the family income distribution. The concern, then, is less with people falling through the cracks altogether (in the sense of not interacting with the tax and transfer system at all) than with the system—at least as currently constituted—generating a systematic and substantial underfiling of taxes by the least well-off.

Going Beyond Census Data

In the exercise above, we treated the respondents to the 2016 census as if they were the complete population and then examined the characteristics of that sample who were in tax data. However, not everyone responds to the census, and we therefore need to consider the non-respondents both to get a sense of the size of the total population who are not in contact with the tax and transfer system, and to help in understanding any biases that might arise from treating the census as the complete population. It is helpful to start with a Venn diagram of the possible data states to set out the nature of the potential issues (Figure 1). Note that the different areas are drawn to emphasize visibility, not as an attempt to make them proportional to their actual relative sizes.

Figure 1

Venn Diagram of Dataset Overlaps



A: Tax filers[3] also in census

B: People in the tax system who do not file taxes who are in the census

C: People in the tax system who do not file taxes and who are not in the census

D: People in the census who are not in the tax system

E: People not in the census or tax data

F: People who file taxes but are not in the census

T1FF: A combined CRA file, processed over several months, made up of the individual T1 file, the T4 tax file, and a file pertaining to federal child benefits

The initial examination focused just on areas A, B, and D—all people in the census—to look at the percentages in areas A and B. In principle, we can observe the people in areas C and F in tax data and, therefore, can get a count of their number and observe the characteristics that are present in tax data (age, gender, location, and actual or calculated income). Doing this

---

[3] Tax filing refers to filing at T1 tax form by the end of April, 2016 for the 2015 tax year. Thus, this corresponds to the act of tax filing near to the Census survey date in June of 2016.

requires the union of tax and census data while we currently only have census data linked (or not linked) to tax data. We can obtain an estimate of the size of area F using a linkage of the Longitudinal Administrative Databank (LAD) to the census. The LAD is a 20% random sample of T1 tax filers in a given year. From its linkage to the census, we observe that approximately 10% of T1 tax filers in the 2016 tax filing year were not present in the census.

Data on the union of people in the tax system and the census still misses those in area E (people not in the census or tax data). To get at the size and composition of people in that group, we need a dataset that covers the population. Death certificate data provides a possible way to measure the entire population. Death certificate data from the Canadian Vital Statistics Death Database (CVSD) includes a record for everyone who dies in Canada in a given year. Since everyone dies eventually, everyone ends up in this dataset. The obvious exception is people who live in Canada for a period but then leave. Given that people who plan to emigrate are exposed to the same risks of death as those who do not, the death certificate data still includes a sample of people who would otherwise have left Canada at some point. However, the sampling rate for that population will be lower than for everyone else because they have less time in Canada in which they might die.

The CVSD is built on data sent to Statistics Canada from the provinces for each death. We work with the data aggregated to the annual level. The database includes information on age, sex, location of residence and of death (at the forward sorting code level), place of birth of the deceased, place of birth of the deceased's parents, whether the individual was married, cause of death, and locality of death (e.g., at home, in a park). Using the location data, we match the place of residence to census tract–level data to link in data on income distribution, education levels, proportion of immigrants, and proportion of Aboriginal people.

The census-based data stands in for individual data on education, income, immigrant status, etc., which are not present in the CVSD. The locality of death may also be useful in that dying at a locality that is not listed as the residence ("home") or a place of work may identify the person as being homeless. Similarly, cause of death (e.g., overdose) could provide some insight into whether the person was from a vulnerable population.

Using the death data to get at the selectivity of being in the census is complicated for a number of reasons. First, presence in the death data is heavily age conditioned. In the extreme, imagine that people died only of old age. In that case, the death data could be a representative draw of people over, say, age 70 but would not be representative of people younger than that. If, alternatively, death arrives randomly among the younger population, then the death data is a useful, if small, sampling of the younger population. However, death is likely non-random in the sense that poorer people are more likely to die prematurely.

The second issue is that we only see people in area E if they are dead (we have no other source of data to catch them). That means if we did something like pool the dead people with those observed in census or tax data, and then estimate the proportion of people in area E, we would obviously have a distorted picture of the latter proportion.

To understand the response to these issues, consider the questions of what proportion of the population is in area E and what factors determine being observed there. To start, consider estimating the unconditional probability, $P(E=1)$, where E is a dummy variable equal to 1 if the person is in group E. We can write this probability as:

$P(E=1) = P(E=1|Dth=1)*P(Dth=1) + P(E=1|Dth=0)*(1 – P(Dth=1))$ (1), where, Dth is a dummy variable corresponding to death.

We can observe $P(E=1|Dth=1)$ from the death data but nothing else on the right-hand side of (1).

It seems reasonable to bound $P(E=1)$ based on an assumption that the people who have not died are less likely to be in area E than those who have died. (To see this, look at the Bayesian formula, $P(E=1|Dth=1) = P(Dth=1|E=1)*[P(E=1)|P(Dth=1)]$, and note that being in area E makes it more likely that you die in any given year.) In that case, the upper bound on $P(E=1|Dth=0)$ is $P(E=1|Dth=1)$ and the lower bound is 0. So, the two bounds on $P(E=1)$ are $P(E=1|Dth=1)$ (upper) and $P(E=1|Dth=1)*P(Dth=1)$ (lower). The lower bound still requires an estimate for the unconditional probability of death, $P(Dth=1)$. Given that area E is likely to be small in size, it is not unreasonable to simply use the death rate in the rest of our data. Even if the death rate in area E is much higher, it would probably only raise $P(Dth=1)$ by a percentage point at most.

From data linking the death records for 2017, the 2016 census, and the 2015 and 2016 T1FF, 3.6% of deaths in 2017 were found in neither the census nor the T1FF. This is our estimate of $P(E=1|Dth=1)$ and is, therefore, our upper bound estimate of $Pr(E=1)$. Our lower bound estimate for the size of area E would be 3.6% times the overall death rate. The death rate in Canada is about 7.6 per 1000 people or 0.007. Thus, the lower bound is very close to zero.

We also know from the linked data that 86.75% of deaths in 2017 had an associated census record. Recall from our examination earlier that 3.5% of people in the census were not linked to tax data at all, and a further 7.4% were in the tax data but did not file taxes in the current year. If census respondents make up 87% of the population (attributing the death linkage rates to the whole population), then people who are in the census but not in the tax data account for $(0.035*0.87)*100 = 3\%$ of the total population. Adding this to our bounds on the estimate of the size of area E implies that between 3% and 6.6% of the total population is not in the tax records at all. From our census analysis, there is not a strong correlation between being not linked and predicted income, so these numbers do not represent who might need assistance but has fallen through the cracks. If we think that the bottom 20% of the income distribution need assistance, then the proportion of the population that is both in need of help and has fully fallen through the cracks is between about 0.6% and 1.2%. Some of these people are likely temporary foreign workers, but the numbers in Table 1 suggest that the group consists of others as well.

We can get some picture of who is in area E through two other (indirect) routes. The first is to examine deaths broken down by cause. In particular, work by Case and Deaton in the U.S. has highlighted what they call “deaths of despair”: deaths due to suicide, drug overdose, or alcoholism. In 2017, deaths of despair made up 4.5% of all deaths in Canada. Notably, 8% of all deaths had no associated T1 record for the 2015 tax year, compared to 27% of those who died a death of despair. Moreover, 14% of all deaths of despair fell into group E (i.e., were not present in either census or tax records). Deaths of despair make up 17% of deaths in group E—roughly four times their share in deaths in all the census and tax areas combined.

We also ran regressions of an indicator of being missing from both census and tax records on indicators for deaths of despair, age, and the proportion of people in the census tract where a person died who were Aboriginal, had income below the poverty line, moved in the last year, moved in the last five years, and immigrants. Fitting with our simple tabulations, deaths of despair are positively and statistically significantly related to being missing from both census and tax data. Among the demographic characteristics, the percentages of Aboriginal people, those with low income, and immigrants were all positive and statistically significant. The effects of both the percentage of variables for Aboriginal people and immigrants could reflect both low-income effects and definitional issues, since Aboriginal people are not in tax records due to treaty rights, and some immigrants may not be permanent residents. The percentage low-income variable has smaller effects than the other two, but the fact that its effects are positive and significant point to people in area E (those who are not in any government data) living in poor neighbourhoods and possibly being poor themselves.

We can use similar logic to our arguments examining the people in area E to uncover more about who might be present in census data but do not file taxes. Based on that logic, people who are in the tax data and census data but do not file taxes account for (0.074*0.86)*100 = 6.4% of the total population. Recall from both the simple tabulations and the regression analysis that non-filers disproportionately have characteristics that fit with being low income (41% of non-filers in the simple tabulations have imputed income that would put them in the bottom two deciles of the income distribution). Thus, there is a sizable group who likely need assistance but would not get it if they did not change their filing behaviour and transfers were heavily shifted to being delivered through the tax system. In addition, this group is not getting benefits such as the GST and carbon tax credits under the current system.

As mentioned earlier, 27% of people who die from deaths of despair did not file taxes in the year before their death. In a regression of a dummy variable for not filing taxes in 2015 or 2016 on the same variables used in the regression for not being present in either census or tax data, deaths of despair, percentage Aboriginal people, and percentage low income again enter positively and significantly.

Finally, it is worth highlighting that approximately 13% of people who died in Canada in 2017 were not in the 2016 census. Some of those missing may be immigrants who arrived after the 2016 census, but as immigration inflows in 2016 and 2017 amounted to roughly 0.8% of the total population this is unlikely to explain much of the result. We ran the same type of regression exercise using the linked death records data as for the non-filers specification but with the dependent variable equalling 1 for people who were missing only from census data. That regression shows strong relationship to a death from drugs or alcohol and to the proportion of the population in the area that moved in the previous year as well as a positive and significant relationship with the proportion who were low income. Thus, the census seems likely to be systematically missing vulnerable populations who are low income, living with addiction, and moving around a lot. In some ways this is not a surprise, but the size of the missing group is striking. Since these vulnerable populations almost certainly have higher death rates, they are more likely to be in our death data, and so the 13% missing is an upper bound on the percentage missing from the census in the population at large. But we can express the result in the other direction: 45% of people who died from drug or alcohol overdoses in 2017 were not linked to the 2016 census. To some degree, this may represent a problem with linkages for people in this population—that is, some of these people may be in the census but we aren't able to identify them. Of course, this type of linkage problem is what we are trying to highlight, and the size of the non-matches suggests there is likely a substantial portion of this population missing from the census even after accounting for matching problems. Dying from these causes is a marker of extreme vulnerability, which raises important issues for how we characterize society using census and other survey data.

Data From Opioid Overdoses

We commissioned a set of special tables based on a Statistics Canada project examining the opioid epidemic. In particular, they identified non-fatal opioid overdose cases using administrative records from the BC Emergency Health Services, Medical Service Plan records, the Discharge Abstract Database, and the National Ambulatory Care Reporting System. The analysis includes all confirmed non-fatal overdose cases that occurred between January 1, 2014 and December 31, 2016.[4] To avoid double counting, overdose events occurring in a 24-hour period to the same individual were treated as a single event. In 2014–2016, a total of 11,843 individuals (7,515 males and 4,318 females) experienced a non-fatal overdose, and there were 17,249 non-fatal overdose events (MacDougall et al., 2019).

[4] Cases of non-fatal overdose were identified based on the protocols detailed in MacDougall et al. (2019).

The data on non-fatal overdose cases was linked to tax data from which we can observe: whether they have employment earnings and, if they do, how much, whether they filed taxes, and whether they received IA.

A key result is that people who have suffered an opioid overdose in a year tend to be low tax filers. Of those people who had at least one opioid overdose in 2014, 74% of males and 81% of females filed taxes. Both percentages correspond to slight declines from the year before the overdose. Note that these are percentages of people for whom there is a link to tax data. The comparable number from our census data was 92.6% of the overall population. The overdose population was also much more likely to file their taxes late, with 21% of males and 14% of females filing late in the year they had their overdose. The rate of late filing in the general population is 5%.

The data also shows a decline in the proportion of people who have earned income at the time of overdose (note that the data shows an earlier jump down in the 2008 recession—the decline related to an opioid event is similar in magnitude, though stretched out of time more, compared to the recession). The proportion in receipt of IA also declines after the event—slightly, but in contrast to the trend beforehand. Together, these imply that people have substantially less access to earnings or IA transfers after an event. However, at the moment data on disability transfers is not included, so the decline in other earnings sources might be offset by increased categorization as disabled.

Overall, this data indicates that a vulnerable population of considerable concern—those who are suffering opioid overdoses—suffer from low income near the time of an overdose event and have particularly weak connections to the tax system. Responses to their low-income situation that rely on delivery of benefits through the tax system face the obvious problem of their low connection to the system.

Income Assistance Data-based Estimates

One way to read our results on the size of the not linked and non-filers group is that the size of the population who are both in need and are fully falling through the cracks in the sense of not being connected to the tax and transfer system at all is not large (about 1% of the total population). However, even this size is a concern and needs to be addressed, but it does suggest that much of the vulnerable population is in contact with the transfer system that is meant to help them. We can get a rough gauge of whether that is true by examining provincial data on the issuance of IA benefits. The results below were developed as part of the Guaranteed Basic Income project, commissioned by the Ministry of Social Development and Poverty Reduction, Province of British Columbia. The Income Assistance data is from British Columbia Ministry of Social Development and Poverty Reduction(2019).

The case files for IA system include a location code, with "no fixed address"(NFA) being a possible entry for that code. One could interpret the NFA recipients as corresponding to individuals experiencing homelessness. In November 2017, just over 8,000 IA cases had this code. This is strikingly similar to the 7,655 people who were recorded as being homeless in the spring 2018 homelessness count. A potential implication is that many people without homes are, in fact, in contact with the IA system. Whether their contact is sporadic is unclear, and that is something we intend to examine.

Data From an SRO Hotel

The final data source we employ is from the Hotel Study, a research project in which a sample of occupants of single-room occupancy (SRO) hotels in Vancouver's Downtown Eastside were followed for a median of two years beginning in 2009 (Vila-Rodriguez et al., 2013). The initial sample size was 491. Much of the focus of the surveys and analysis was on health outcomes, but there is information on income sources from the baseline survey. This sample set is of interest because they are marginally housed[5] and are candidates for being missed by the census and/or the tax and transfer system.

The baseline survey at the time people entered the study included questions on primary and secondary sources of income. Respondents were asked about wage and salary income, income from legal self-employment, a list of public transfers (including employment insurance, workers' compensation, IA, etc.), and "other." Of those under age 60 at the time of entry into the study, 89% listed their primary source of income as either IA or disability benefits. Another 7.7% listed their primary source as either paid or self-employed earnings. The remaining 3% mainly said either "none" or "other." Thus, the vast majority of those in this marginal housing are in receipt of IA or disability benefits. That there are only 2%–3% who do not report either earnings or transfer income may fit with the 3% of the population (as captured in death records) who are not present in tax data. This does not mean that the people living in the SROs were getting all the transfers to which they were entitled, much less that they were getting adequate and consistent income, but it does potentially fit with the message that the percentage who are completely falling through the cracks in the sense of having no contact with the tax and transfer system could be quite small.

---

[5] The SRO units were described by the study as follows: "The single-room occupancy hotels typically comprise single rooms of 80 to 120 square feet (8–12m^2), with a sink and possibly a hotplate. Toilet and shower facilities, located at the end of hallways, are shared by 10 to 15 tenants. All single-room occupancy hotels housing study participants were over 75 years old and had evidence of bedbug, cockroach, and mouse infestation" (Vila-Rodriguez et al., 2013, p. 1414).

Conclusion

We have used data from a combination of linked census, tax and death records, data on opioid overdoses, data from B.C.'s IA system, and data from a survey of residents of SRO hotels in Vancouver to paint a picture of the extent and nature of falling through the cracks in the tax and transfer system. A sizable portion of people in the 2016 census—approximately one in nine—did not file taxes in that year. Of those, 7.6% were present in the tax system but did not file a T1 for the 2015 tax year. Another 3% were not present in the tax system records at all. The latter group seems potentially worrying and are candidates for individuals fully falling through the cracks. Estimates using tax data, however, indicate that this group is only slightly disproportionately low income and low educated. The people in this group seem, by and large, to be spread somewhat evenly across the population. The group who are in the tax records but did not file a T1 in a year, however, are disproportionately non-earners from low-income households. Importantly, not filing taxes shows considerable (though far from perfect) persistence across years. Thus, the problems associated with not filing taxes could end up focused on a small core of continual non-filers. Interestingly, being in receipt of IA benefits is negatively correlated with filing taxes when examined on its own, but once we control for earnings and household income, it is actually a positive indicator for filing taxes. People from disadvantaged situations who are in contact with the IA system are less likely to fall through the cracks in the tax system than those who are not.

The results based on census data suffer from the concern that the census itself does not capture the entire population. Based on a linkage to death records, we know that just over 13% of people who died in 2017 in Canada were not in the 2016 census. As much as 3.6% of people were in neither tax system records nor the census. These people are very disproportionately likely to have died from a death of despair (suicide, drugs, or alcohol related) and to have died in low-income neighbourhoods and neighbourhoods with a higher proportion of people who have moved in the previous year. This suggests that the people missing from both survey and tax data are more likely to be from vulnerable populations than the people who are in the census but not linked to tax system data. Given that they are missing from both tax-related and survey attempts by the government to contact them, they would likely be a particularly difficult group to reach with tax-based social benefits. The high proportion of them who died from deaths of despair implies that this omission is particularly troubling.

If we add our estimate of the people in neither tax data nor the census to the people present in the census who are not in the tax system, between 3% and 6.6% of the Canadian population are not known to the tax system at all. Added to the people who do not file taxes in a year, this means between 11% and 15% of the population are not in the tax system at all or do not file taxes in a year.

These findings complement those in Robson and Schultz (2020), who find that approximately 12% of respondents to a Statistics Canada survey did not file taxes in the survey year. Our data supports the magnitude of their finding but also allows us to break the non-filers down further—in particular into those who are in neither the tax system nor the census. As a side point, it also allows us to isolate people who died in 2017 who were not in the 2016 census. Strikingly, the group missing from the census makes up 13% of people who died in 2017. Even more strikingly, 45% of the people who died from a drug or alcohol overdose in 2017 could not be matched to the 2016 census. This suggests either (or likely both) an issue with making linkages for this population in government data (which, of course, is the problem we are trying to highlight) or that the under-representation in the census of a very vulnerable group (those dying from overdoses) is sizable, and implies that our characterization of the population from the census and surveys is imperfect in an important dimension.

Taken together, these results imply that delivering a universal benefit through the current tax system will miss a considerable number of people, particularly among the lowest income and most vulnerable populations. Indeed, a universal benefit would go disproportionately to those with characteristics that put them at the top of the income and resource distributions. Of course, the offering of a sizable basic income would provide an incentive for more people to file taxes. The evidence on the responsiveness to such incentives in this data is mixed. The larger filing rates of people with children in our data fits with incentives to do so to get the Canada Child Benefit (though this may also reflect recent actions by the federal government to ensure use of those benefits), but the low rates of filing among those in the bottom earnings decile point in the opposite direction since those people have an incentive to file in order to get GST and carbon tax credits.

References

Anne Case, & Angus Deaton (2015). Rising Morbidity and Mortality in Midlife Among White Non-Hispanic Americans in the 21st Century. *Proceedings of the National Academy of Sciences of the United States of America,* 49, 15078-15083.

British Columbia Ministry of Social Development and Poverty Reduction [creator] (2020): BC Employment and Assistance (BCEA) V02. Data Innovation Program, Province of British Columbia [publisher].2019.

MacDougall, L., Semolina, K., Otterstatter, M., Zhao, B., Chong, M., Godfrey, D., Mussavi-Rizi, Al, Sutherland, J., Kuo, M., & Kendall, P. (2019). Development and characteristics of the provincial overdose cohort in British Columbia, Canada. *PLoS One, 14*(1).

Robson, J., & Schwartz, S. (2020). Who doesn't file a tax return? A portrait of nonfilers. *Canadian Public Policy, 46,*3, 323–339.

Van Parijs, P., & Vanderborght, Y. (2017). *Basic income: A radical proposal for a free society and a sane economy.* Harvard University Press.

Vila-Rodriguez, F., J. Panenka, W. J , Long, D. J., Thornton, A. E., Vertinsky, T., Wong, H., Barr, A. M., Procyshyn, R. M., Sidhu, J. J., Smith, G. N., Buchanan, T., Kraiden, M., Krausz, M., Montaner, J., MacEwan, G. W., & Honer, W. G. (2013). The hotel study: Multimorbidity in a community sample living in marginal housing. *American Journal of Psychiatry, 170*(12), 1413–1422.

Table 1

Tax Filing Status of 2016 Census Respondents in B.C.: Demographic Characteristics)

		Row percentage (total)				Column percentage (level)			
		Not linked	Non-filers	Filers		Not linked	Non-filers	Filers	Total
	Total	0.035	0.074	0.891		1.000	1.000	1.000	1.000
Variables									
Education	No high school	0.045	0.078	0.877		0.159	0.130	0.121	0.123
	High school grad	0.042	0.085	0.873		0.353	0.332	0.283	0.289
	Less than bachelor's degree	0.031	0.074	0.895		0.292	0.324	0.325	0.324
	Bachelor's degree	0.027	0.062	0.912		0.132	0.143	0.176	0.172
	University above bachelor's degree	0.028	0.055	0.917		0.016	0.014	0.020	0.019
Sex	Female	0.029	0.064	0.907		0.436	0.443	0.525	0.516
	Male	0.040	0.085	0.875		0.564	0.557	0.475	0.484
Age group	20–24	0.053	0.112	0.834		0.121	0.120	0.074	0.079
	25–29	0.041	0.104	0.856		0.098	0.117	0.080	0.083
	35–39	0.027	0.090	0.884		0.062	0.098	0.080	0.081
	45–49	0.041	0.094	0.865		0.105	0.112	0.086	0.088
	55–59	0.038	0.076	0.886		0.106	0.100	0.096	0.097
	65–69	0.030	0.030	0.940		0.067	0.031	0.083	0.078
	75+	0.022	0.014	0.964		0.056	0.016	0.094	0.087
Immigrant status	Non-permanent resident	0.184	0.224	0.591		0.118	0.067	0.015	0.022
	Born in Canada, at least one parent born outside			0.901			0.194	0.203	0.201
	Born in Canada and both parents born in Canada	0.037	0.088	0.876		0.462	0.520	0.430	0.438

	Immigrant arrived in 2016	0.071	0.402	0.528		0.007	0.018	0.002	0.003
	Immigrant arrived before 2016	0.026	0.043	0.931		0.250	0.193	0.346	0.331

		Row percentage (total)			**Column percentage (level)**			
Variables		**Not linked**	**Non-filers**	**Filers**	**Not linked**	**Non-filers**	**Filers**	Total
		0.035	0.074	0.891	1.000	1.000	1.000	1.000
Visible minority	South Asian	0.040	0.057	0.904	0.085	0.057	0.076	0.075
	Chinese	0.034	0.052	0.914	0.111	0.078	0.114	0.112
	Filipino	0.028	0.047	0.924	0.024	0.019	0.031	0.030
	Latin American	0.056	0.114	0.830	0.016	0.015	0.009	0.010
	Korean	0.059	0.062	0.879	0.021	0.011	0.012	0.013
	Not visible minority	0.031	0.075	0.894	0.600	0.671	0.665	0.663
Number of children	No children	0.037	0.080	0.883	0.811	0.809	0.744	0.751
	1	0.028	0.063	0.909	0.089	0.093	0.112	0.110
	2	0.024	0.052	0.924	0.072	0.074	0.108	0.104
	3	0.029	0.052	0.920	0.022	0.019	0.028	0.027
Family type	Spouse	0.027	0.054	0.919	0.487	0.448	0.635	0.616
	Lone parent	0.035	0.074	0.891	0.056	0.055	0.054	0.055
	Never married Son/Daughter	0.018	0.105	0.877	0.049	0.131	0.091	0.093
	Not in a census family	0.059	0.114	0.826	0.405	0.364	0.218	0.236

Table 2

Tax Filing Status of 2016Census Respondents in B.C.: Imputed Income and Employment Deciles

Variable		Row percentage (total)			Column percentage (level)			
		Not linked	Non-filers	Filers	Not linked	Non-filers	Filers	Total
		0.035	0.074	0.891	1.000	1.000	1.000	1.000
Employment decile	1	0.041	0.192	0.767	0.119	0.260	0.086	0.100
	2	0.032	0.049	0.919	0.093	0.066	0.103	0.100
	3	0.027	0.036	0.936	0.079	0.049	0.105	0.100
	4	0.029	0.050	0.921	0.085	0.068	0.103	0.100
	5	0.037	0.081	0.882	0.107	0.109	0.099	0.100
	6	0.037	0.071	0.891	0.108	0.096	0.100	0.100
	7	0.037	0.076	0.887	0.107	0.102	0.100	0.100
	8	0.035	0.069	0.896	0.101	0.094	0.100	0.100
	9	0.034	0.064	0.902	0.099	0.086	0.101	0.100
	10	0.035	0.051	0.914	0.103	0.069	0.103	0.100
Income decile	1	0.042	0.212	0.746	0.122	0.286	0.084	0.100
	2	0.040	0.090	0.869	0.117	0.122	0.098	0.100
	3	0.033	0.051	0.916	0.096	0.069	0.103	0.100
	4	0.033	0.054	0.912	0.096	0.074	0.102	0.100
	5	0.034	0.060	0.906	0.097	0.081	0.102	0.100
	6	0.033	0.060	0.907	0.095	0.081	0.102	0.100
	7	0.032	0.058	0.910	0.092	0.079	0.102	0.100
	8	0.032	0.057	0.911	0.093	0.077	0.102	0.100
	9	0.033	0.054	0.914	0.094	0.072	0.103	0.100
	10	0.034	0.044	0.922	0.098	0.060	0.103	0.100

Table 3

Tax Filing Status of Census Respondents in B.C.: Tax and Transfer-related Variables

Variables		Row percentage (total)			Column percentage (level)			
		Not linked	Non-filers	Filers	Not linked	Non-filers	Filers	Total
		0.035	0.074	0.891	1.000	1.000	1.000	1.000
Employment insurance	Yes	0.035	0.076	0.890	0.076	0.078	0.076	0.076
	No	0.035	0.074	0.892	0.924	0.922	0.924	0.924
Income Assistance	Yes	0.055	0.138	0.807	0.054	0.063	0.031	0.034
	No	0.034	0.072	0.894	0.946	0.937	0.969	0.966
Declared taxes previous year	Yes	0.000	0.032	0.968	0.000	0.380	0.957	0.881
	No	0.000	0.543	0.457	0.000	0.620	0.043	0.085
	Not linked	1.000	0.000	0.000	1.000	0.000	0.000	0.035
Change of residence 1 year before	Yes	0.077	0.166	0.757	0.110	0.110	0.042	0.049
	No	0.032	0.069	0.898	0.890	0.890	0.958	0.951

Table 4

Linear Probability Model Estimates for Being Linked to the Tax System

Variable	Model 1	See note	Model 2	See note	Model 3	See note
Age						
15	-0.008	***	-0.001		-0.003	***
16	-0.009	***	-0.005	***	-0.008	***
17	-0.010	***	-0.008	***	-0.010	***
18	-0.012	***	-0.013	***	-0.013	***
19	-0.011	***	-0.012	***	-0.012	***
20-24	-0.004	***	-0.004	***	-0.004	***
25-29	0.010	***	0.010	***	0.010	***
30-34	0.015	***	0.015	***	0.015	***
35-39	0.013	***	0.013	***	0.013	***
45-49	-0.003	***	-0.003	***	-0.003	***
50-54	0.004	***	0.004	***	0.004	***
55-59	0.004	***	0.006	***	0.006	***
60-64	0.006	***	0.012	***	0.011	***
65-69	0.011	***	0.021	***	0.020	***
70-74	0.018	***	0.031	***	0.031	***
75+	0.020	***	0.035	***	0.036	***
40-44	ref		0.000		ref	
Female	0.009	***	0.009	***	0.009	***
Census family status						
Lone parent	-0.005	***	-0.004	***	-0.002	***
Never married son/ daughter	0.025	***	0.026	***	0.026	***
Grandchild in a census family with no parent of grandchild present	-0.017	***	-0.013	***	-0.011	***
Person not in a census family: foster child	-0.118	***	-0.111	***	-0.111	***
Person not in a census family: all other	-0.019	***	-0.018	***	-0.015	***
Spouse or common-law partner	ref		ref		ref	
Presence of children						
One	0.008	***	0.007	***	0.007	***
Two	0.007	***	0.008	***	0.008	**

Three	0.008	***	0.009	***	0.009	***
Four	0.007	***	0.010	***	0.011	***
Five or more	0.007	***	0.011	***	0.012	***
No children	ref		ref		ref	
Visible minority						
South Asian	-0.001	*	0.000		0.002	***
Chinese	-0.007	***	-0.005	***	-0.008	***
Black	-0.014	***	-0.013	***	-0.008	***
Filipino	0.004	***	0.002	***	0.003	***
Latin American	-0.009	***	-0.009	***	-0.006	***
Arab	-0.008	***	-0.006	***	-0.004	**
Southeast Asian	-0.011	***	-0.010	***	-0.009	***
West Asian	-0.003	***	-0.001		0.001	**
Korean	-0.014	***	-0.012	***	-0.013	***
Japanese	-0.016	***	-0.016	***	-0.017	***
Other visible minority	-0.005	**	-0.005	**	-0.003	***
Multiple visible minorities	-0.002	**	-0.002	**	-0.003	**
Aboriginal	-0.003		-0.001		0.000	
Not visible minority	ref		ref		ref	
Aboriginal identity						
North American Indian	-0.015	***	-0.014	***	-0.010	***
Métis	0.000		-0.001		-0.001	
Inuit	-0.012	***	-0.013	***	-0.003	
Other Aboriginal identity	-0.016	***	-0.015	***	-0.013	***
Non-Aboriginal identity	ref		ref		ref	
Immigration status						
Non-permanent resident	-0.141	***	-0.138	***	-0.136	***
Born in Canada and at least one parent born outside	0.000		0.000		0.000	
Immigrant arrived in 2016	-0.027	***	-0.026	***	-0.025	***
Immigrant arrived before 2016	0.003	***	0.003	***	0.003	***
Born in Canada and both parents born in Canada	ref		ref		ref	

Residential variables						
Lives on a reserve	-0.047	***	-0.044	***	-0.036	***
Changed CA/CMA residence code last year	-0.025	***	-0.022	***	-0.021	***
Changed CA/CMA residence code last five years	-0.002	***	0.000		0.000	
Highest education achieved						
No high school	-0.015	***	-0.011	***	-0.008	***
High school graduation certificate or equivalency certificate	-0.011	***	-0.009	***	-0.007	***
Less than bachelor's degree	-0.006	***	-0.004	***	-0.002	***
University certificate or diploma above bachelor's level	-0.002	***	-0.002	***	-0.002	**
Degree in medicine, dentistry, veterinary medicine, or optometry	0.002	**	0.000		-0.001	
Master's degree	0.005	***	0.004	***	0.003	***
Doctorate degree	0.003	***	0.001	***	0.001	***
Bachelor's degree	ref		ref		ref	
Census labour responses						
Did not work in 2015 or 2016	-		-0.021	***	-0.016	***
Did not work in 2015 worked in 2016	-		-0.013	***	-0.010	***
Worked 1–13 weeks part time	-		0.007	***	0.007	***
Worked 14–26 weeks part time			0.005	***	0.005	***
Worked 27–39 weeks part time	-		0.004	***	0.004	***
Worked 40–48 weeks part time	-		0.002	***	0.002	***
Worked 49–52 weeks part time	-		0.001	***	0.001	***
Worked 1–13 weeks full time	-		0.008	***	0.008	***
Worked 14–26 weeks full time	-		0.009	***	0.009	***

Worked 27–39 weeks full time	-		0.008	***	0.008	***
Worked 40-–weeks full time	-		0.003	***	0.003	***
Worked 49-–weeks full time	-		ref		ref	
Commuter worker	-		-0.031	***	-0.031	***
Response mode						
Internet self-response	-		-		0.017	***
Paper self-response	-		-		-0.035	***
Canvasser interview	-		-		-0.007	***
Phone interview	-		-		ref	
Intercept	0.974	***	0.977	***	0.961	***
Control for CA/CMA(136)	Yes		Yes		Yes	***
Control for industry sector (22)	No		Yes		Yes	***
R-Square	0.029		0.037		0.043	
N	7,139,225		7,139,225		7,139,225	

Note: Asterisks refer to average marginal effects of covariates on the probability of being linked between the census and the tax data files: * = significant at 10%; ** = significant at 5%; *** = significant at 1%.

Table 5

Linear Probability Model Estimates for Filing a T1 Tax Form

Variable	M1	See note	M2	See note	M3	See note
Age group						
15	-0.762	***	-0.711	***	-0.712	***
16	-0.646	***	-0.613	***	-0.614	***
17	-0.461	***	-0.452	***	-0.453	***
18	-0.156	***	-0.16	***	-0.159	***
19	0.002	**	-0.01	***	-0.007	***
20–24	0.042	***	0.03	***	0.034	***
25–29	0.034	***	0.03	***	0.032	***
30–34	0.018	***	0.017	***	0.018	***
35–39	0.007	***	0.007	***	0.007	***
45–49	0.003	***	0.004	***	0.003	***
50–54	0.019	***	0.021	***	0.02	***
55–59	0.038	***	0.043	***	0.042	***
60–64	0.058	***	0.075	***	0.071	***
65–69	0.081	***	0.116	***	0.106	***
70–74	0.092	***	0.139	***	0.124	***
75+	0.106	***	0.156	***	0.139	***
40–44	Ref		Ref		Ref	
Female	0.019	***	0.02	***	0.017	***
Census family status						
Lone parent	-0.017	***	-0.02	***	-0.008	***
Never married son/daughter	-0.028	***	-0.027	***	-0.038	***
Grandchild in a census family with no parent of grandchild present	-0.071	***	-0.068	***	-0.072	***
Person not in a census family : foster child	-0.099	***	-0.086	***	-0.089	***
Person not in a census family : all other	-0.043	***	-0.043	***	-0.028	***
Spouse or common-law partner	Ref		Ref		Ref	
Presence of children						
One	0.036	***	0.038	***	0.041	***
Two	0.047	***	0.05	***	0.055	***
Three	0.053	***	0.06	***	0.066	***
Four	0.061	***	0.074	***	0.081	***
Five or more	0.073	***	0.093	***	0.1	***

No children	Ref		Ref		Ref	
Visible minority status						
South Asian	0.010	***	0.013	***	0.014	***
Chinese	0.022	***	0.024	***	0.031	***
Black	-0.045	***	-0.044	***	-0.041	***
Filipino	0.028	***	0.022	***	0.021	***
Latin American	-0.018	***	-0.014	***	-0.012	***
Arab	-0.012	***	0		0.008	***
Southeast Asian	0.013	***	0.014	***	0.016	***
West Asian	-0.001		0.011	***	0.021	***
Korean	0.017	***	0.027	***	0.037	***
Japanese	-0.007	***	-0.007	***	-0.007	***
Other visible minority	-0.013	***	-0.012	***	-0.01	***
Multiple visible minorities	-0.010	***	-0.009	***	-0.006	***
Aboriginal	-0.056	***	-0.049	***	-0.045	***
Not visible minority	Ref		Ref		Ref	
Aboriginal identity						
North American Indian	-0.010	**	-0.004		-0.003	
Métis	0.030	***	0.025	***	0.023	***
Inuit	0.034	***	0.027	***	0.025	***
Other Aboriginal identity	-0.018	***	-0.014	***	-0.013	***
Non-Aboriginal identity	Ref		Ref		Ref	
Immigration status						
Non-permanent resident	-0.142	***	-0.122	***	-0.107	***
Born in Canada and at least one parent born outside	-0.009	***	-0.008	***	-0.008	***
Immigrant arrived in 2016	-0.411	***	-0.371	***	-0.358	***
Immigrant arrived before 2016	0.014	***	0.017	***	0.019	***
Born in Canada and both parents born in Canada	Ref		Ref		Ref	
Residential variables						
Lives on a reserve	-0.054	***	-0.022	***	-0.015	***
Changed CA/CMA residence code last year	-0.066	***	-0.059	***	-0.053	***
Changed CA/CMA residence code last five years	-0.008	***	-0.004	***	-0.001	**
Highest education achieved						
No high school	-0.038	***	-0.024	***	-0.018	***

High school graduation certificate or equivalency certificate	-0.023	***	-0.016	***	-0.013	***
Less than bachelor’s degree	-0.012	***	-0.009	***	-0.006	***
University certificate or diploma above bachelor level	-0.004	***	-0.003	***	-0.003	***
Degree in medicine, dentistry, veterinary medicine, or optometry	0.005	***	0.014	***	0.007	***
Master’s degree	0.004	***	0.002	***	0.001	**
Doctorate degree	0.000		-0.004	***	-0.006	***
Bachelor’s degree	Ref		Ref		Ref	
Census labour responses						
Did not work in 2015 or 2016	-		0.014	***	0.02	***
Did not work in 2015 worked in 2016	-		-0.024	***	-0.019	***
Worked 1–13 weeks full time	-		0.015	***	0.018	***
Worked 14–26 weeks full time	-		0.009	***	0.01	***
Worked 27–39 weeks full time	-		0.007	***	0.008	***
Worked 40-–weeks full time	-		0.006	***	0.007	***
Worked 1-–weeks part time	-		0.026	***	0.028	***
Worked 14–26 weeks part time	-		0.026	***	0.028	***
Worked 27–39 weeks part time	-		0.017	***	0.02	***
Worked 40-–weeks part time	-		0.017	***	0.02	***
Worked 49-–weeks part time	-		0.021	***	0.023	***
Worked 49–52 weeks full time	-		Ref		Ref	
Inter CA/CMA commuter worker	-		-0.004	***	-0.006	***
Wages and salaries reported on T4						
0	-		-0.098	***	-0.078	***
$1–3,000	-		-0.029	***	-0.009	***
$3,000–11,327	-		0.018	***	0.035	***
$11,328–23000	-		0.002	***	0.013	***
$23,001–40,000	-		-0.003	***	0.003	***
$100,000+	-		0		-0.007	***
$40,001–100,000	-		Ref	***	Ref	
Investment income declared on T3 or T5 slips						
$1–999	-		0.029	***	0.027	***
$1,000–9,999	-		0.037	***	0.032	***
$10,001–99,999	-		0.046	***	0.034	***
$100,000+	-		0.034	***	0.019	***

0	-		Ref		Ref	
Benefits reported on T5007 Slips						
Social assistance						
Some social assistance benefits	-		0.036	***	0.054	***
Some workers' compensation benefits	-		0.008	***	0.004	***
Economic family income decile						
Decile 1	-		-		-0.082	***
Decile 2	-		-		-0.016	***
Decile 3	-		-		-0.016	***
Decile 4	-		-		-0.011	***
Decile 5	-		-		-0.007	***
Decile 6	-		-		-0.004	***
Decile 8	-		-		0.004	***
Decile 9	-		-		0.009	***
Decile 10	-		-		0.017	***
Decile 7	-		-		Ref	
Intercept	0.883	***	0.886	***	0.877	***
Control for CA/CMA	Yes		Yes		Yes	
Control for industry	No		Yes		Yes	
R-Square	0.274		0.291		0.295	
N	6,929,860		6,929,860		6,929,860	

Note: Asterisks refer to average marginal effects of covariates on the probability of being linked between the census and the tax data files: * = significant at 10%; ** = significant at 5%; *** = significant at 1%.