# A Researcher's Guide to the Labour Force Survey: Its Evolution and the Choice of Public Use versus Master Files

Pierre Brochu

Department of Economics, University of Ottawa

June 2021

**Abstract**

To balance researchers' need for detailed information with respondents' confidentiality concerns, statistical agencies such as Statistics Canada commonly offer two versions of the same dataset: a public use file that is readily available and a master file with richer information but to which access is restricted. This article examines the choice of using public use versus master files of the Labour Force Survey (LFS). The article also provides researchers with a unified source of LFS information, including a thorough discussion of the structure of the LFS and its implication for research, such as the creation of mini-panels.

**Keywords:** household survey, microdata, data access, non-response

**Acknowledgements**

I thank Emmanuelle Bourbeau, Winnie Chan, Jonathan Créchet, Zechuan Deng, Stéphanie Lluis, Louis-Philippe Morin, René Morissette, Martin Pantel, Craig Riddell, and Mikal Skuterud for their helpful suggestions and Gabrielle Brochu and Richard Beard for their excellent research assistant work. I also thank the three anonymous referees for their comments and suggestions that helped improve the article. I am grateful to the librarians at Statistics Canada and at the University of Ottawa for helping me discover and track down historical records and publications. Finally, I thank David Green for suggesting that I write this article.

## 1 Introduction

Statistical agencies have come under increasing public scrutiny in recent years, with public debate centred on issues such as whether certain census or survey questions, for example those regarding citizenship, are too intrusive or politically motivated (United States; Williams 2019). This also includes whether the Census, or any survey, should be mandatory (Canada; Bouw and Chase 2010).[1] Of related concern is the worldwide rise in survey non-response (Barrett, Levell, and Milligan 2014; Green and Milligan 2010) and the public policy implications of that non-response ("Missing the Target" 2018).[2] At the core of these concerns are issues of privacy and confidentiality. In the latter case, individuals would be reticent to volunteer their personal information if there is a possibility that they can be identified in the data. Statistical agencies therefore face a dilemma in providing the detailed information that researchers need while still protecting respondents' identity. To address these countervailing concerns, statistical agencies (e.g., Statistics Canada) often provide two versions of the same dataset: a public use files version that is readily available but whose data have been modified (e.g., top-coding and collapsing responses into categories or intervals) and a master (confidential use) files version that provides more detail but to which access is restricted.

Researchers must go through an application process to gain access to Statistics Canada's master files data, and they must be able to show that these data are critical to the research at

[1] The Canadian long-form Census, which is held every five years, was discontinued in 2011 and replaced with the (non-mandatory) 2011 National Household Survey. The long-form Census was, however, restored as of 2016. See Green and Milligan (2010) and Veall (2010) for the merits of a mandatory long-form census.

[2] Statistical agencies will attempt to correct for non-response by adjusting the weights or using imputation methods. Having said this, high non-response rates put a lot on the shoulders of these correction techniques, and there is evidence that suggests that these correction methods are failing (e.g., Barrett et al. 2014). Given that governments rely on (household) survey data to set policy, poor-quality data will have public policy implications.

hand. Upon approval of the project, the data can be accessed only in secure, designated Research Data Centres (RDCs).[3] Although an extensive network of RDCs exists across Canada, not all researchers have easy access to these locations, especially researchers who are based outside of the country. Finally, a vetting process imposes restrictions on the type of results that can be taken out of these secure locations to protect confidentiality.[4]

In this article, I examine the choice of using public use versus confidential data in the context of the Labour Force Survey (LFS).[5] The LFS is one of Statistics Canada's flagship surveys. In addition to being the official source of employment and unemployment statistics for Canada, the LFS frame has historically been the starting point for many surveys, such as the Canadian Community Health Survey and the Survey of Financial Security (Statistics Canada 2008b). LFS data are also widely used by researchers both in economics (e.g., Green and Townsend 2010; Jones and Riddell 2019; Schirle 2015b) and in other social science disciplines (e.g., sociology [Fuller 2005; Hira-Friesen 2018] and political science [Bélanger and Gélineau 2010]).

The structure of the LFS survey—that it is conducted as a series of rolling panels in which households remain in the survey for six consecutive months—makes the public use versus master files comparison particularly relevant. That is because there is no way to follow

[3] As of September 2020, there were three federal RDCs in the nation's capital and 32 other RDCs at universities across the country. The United States also has its own RDC network (i.e., Federal Statistical Research Data Centers) and a similar number of branches across the United States where researchers can access restricted-use data.
[4] It should be noted that Statistics Canada also offers access to its master file data through the Real Time Remote Access system. There are two important limitations to following this route. First, researchers cannot work directly with the data but must instead submit programs through an online portal. Second, and most important, the vetting restrictions are very severe. Researchers can only ask for descriptive statistics; no regression analysis is permitted.
[5] The LFS is similar to the Current Population Survey (CPS), the United States' labour force survey, in terms of both survey design and questionnaire structure and content. Although there are small differences between the questionnaires that researchers can exploit (these are discussed later in this article), from a researcher's perspective, the key difference is access. There is only one version of the CPS, and it is readily available to researchers (i.e., a public use version). This explains why Donovan, Lu, and Schoellman (2020), for example, include the United States but not Canada when they explore cross-country differences in labour market flows.

individuals over time in the public use files. This matters because some labour market information, such as wages, is not refreshed beyond the individual's first interview (i.e., the first month of participation in the LFS) unless the individual changed employers (or is a new hire).[6] In the case of individuals aged 70 years and older, all labour market information is automatically carried forward into the next month. Given that one cannot identify when a household rotates into the survey in the public use files, using master files data may be more appropriate in research situations in which using stale data is problematic.

An additional aspect of the LFS of which many researchers may not be aware (or not know how to do), and for which I provide a detailed discussion in this article, is the possibility of creating mini-panels that are up to six months in length. Given that it is intended to be cross-sectional, there is no single person identifier to link individuals across consecutive months. The creation of mini-panels therefore requires a combination of variables that are only available in the master files.

A second and complementary goal of this article is to provide researchers with a unified source of LFS information. It is more common to observe comprehensive academic discussions of US datasets—particularly novel or more complex ones (e.g., Abowd, Haltiwanger, and Lane 2004; Abowd et al. 2009; Hamermesh, Frazis, and Stewart 2005)—than of Canadian datasets. The LFS may be a well-established dataset, but given that many researchers rely on many decades of LFS data (e.g., Baker and Drolet 2010; Brochu and Green 2013; Green and Townsend 2010; Heisz 2005; Jones and Riddell 2019; Schirle 2015b), it is important to understand how the survey has evolved through its periodic redesigns.

[6] More precisely, the LFS questionnaire indicates that the earnings-related questions are not asked in subsequent interviews if there is no change to questions 110, 114, 115, 116, 117, and 118 in the job description block. The union, job permanency, and establishment size questions are also bypassed in these circumstances.

The many *Guide to the Labour Force Survey* publications (although sporadic in the early years) are a good starting point for LFS information.[7] The *Methodology of the Canadian Labour Force Survey,* which is published every ten years or so,[8] provides more technical information and, in the later years, some of the additional historical considerations that researchers require. However, a variety of complementary sources, some of which are not readily available, are needed if one is to take full advantage of the many years of LFS data (especially the earlier years) and of the LFS design (e.g., using the LFS to construct six-month mini-panels).[9]

Given that the overarching goal of this article is to provide a researcher's perspective on the LFS, and it is directly intended for researchers interested in its data content, it is important to take a step back and provide a more general overview of the LFS—more precisely, to discuss its strengths and policy relevance and, in doing so, highlight where the LFS has been particularly valuable in providing insight and in which areas it is still underused.

A couple of features distinguish the LFS from other Canadian datasets. First, the LFS provides researchers with rich labour market information spanning multiple decades, allowing researchers to better understand the workings of the labour market and how it has evolved over time.[10] The LFS has been successfully used to look at employment and labour force participation

---

[7] The *Guide to the Labour Force Survey* was first published in 1997 and (typically) revised on a yearly basis as of 1999. It replaced the (now discontinued) *Guide to Labour Force Survey Data,* which was published at irregular intervals; two *Guide to Labour Force Survey Data* were published over the 1976–1996 interval (i.e., 1979 and 1992).

[8] Since 1976, Statistics Canada has produced five *Methodology of the Canadian Labour Force Survey* (i.e., Statistics Canada 1977a, 1990, 1998, 2008b; 2017b).

[9] Dominion Bureau of Statistics (1957) is a good source of information for the very early years of the LFS. Finally, Usalcas and Kinack (2017) is also a good starting point for those wanting to learn about the LFS and how it has changed over the 1945–2016 period. It is also a good source for more technical details (e.g., population rebasing).

[10] No other Canadian dataset has been able to match the LFS on these two fronts. Some provide detailed labour market information, but they do not have the span of the LFS (e.g., the Labour Market Activities Survey [1986–1990] and the Survey of Labour and Income Dynamics [SLID] [1993–2011]). Others, like the Canadian Census, provide time-consistent information over multiple decades, but they lack the rich labour market information found in the LFS.

(e.g., Jones and Riddell 2019; Schirle 2008), unemployment (e.g., Jones and Riddell 2006; Kroft et al. 2019), and job stability (e.g., Brochu 2013; Heisz 2005), to name a few.[11] It has also been relied on to evaluate a host of labour market policies, such as disability insurance (e.g., Milligan and Schirle 2019), employment insurance (EI; e.g., Lluis and McCall 2019), maternity leave (e.g., Baker and Milligan 2008), minimum wage (e.g., Brochu and Green 2013), and universal child care benefits (e.g., Schirle 2015a).

Second, the high frequency of the data (monthly) and its tight production deadline (less than one month from enumeration to its public release) makes the LFS a timely source of data. In particular, it has proven to be an invaluable resource during the coronavirus disease 2019 (COVID-19) pandemic. Many articles that have examined the initial impact of the pandemic in Canada, whether it be on employment and hours of work (e.g., Koebel and Pohler 2020; Lemieux et al. 2020), self-employed people (e.g., Béland, Fakorede, and Mikola 2020), gender employment differences among parents with young children (e.g., Qian and Fuller 2020), or early reopening dynamics (e.g., Jones et al. 2020), have relied on the public use files version of the LFS as their main source of data. Subsequent articles that have looked at employment dynamics from the onset of the pandemic through to the re-opening phase have relied on the panel feature of the LFS master files (e.g., Baylis et al. 2020; Brochu, Créchet, and Deng 2020). It should also be noted that as of April 2020, the LFS has been running a voluntary supplement

[11] It should be noted that the LFS has also been a valuable vehicle for many other supplements in the past. In some cases, they have proven to be an effective means of testing whether certain variables can be reliably measured and thus later added to the regular LFS survey. The 1984 Survey of Union Membership, for example, made it possible to test the addition of union membership and coverage, whereas the 1991 and 1995 Survey of Work Arrangements assessed various non-standard-work questions that later became part of the new LFS questionnaire. These supplements can be of great value to researchers not just because of the information they collect, but because they can be linked to subsequent months of the regular survey. Jones and Riddell (1999), for example, link the annual Survey of Job Opportunities to subsequent months of the LFS. As a result, they are able to compare transition rates of (a) job searchers, (b) those not searching for work but who reported wanting to work, and (c) those who neither searched for nor desired work. Many thanks to Craig Riddell for this historical insight regarding LFS supplements.

on the labour market impacts of COVID-19. The added questions cover such topics as whether respondents work from home and whether they have the capacity to meet their financial obligations, to name a few.

Although widely used within the discipline of economics, the LFS master files have been underused in other social sciences. Few non-economics papers (e.g., Smith et al. 2020) have taken advantage of the richness of the LFS master files, and to the best of my knowledge, none have taken advantage of its panel features. Areas in which the LFS has been underused include the immigration literature, which could rely more on the LFS. It may not have the size of the Census, but its unrivalled labour market data could help researchers better understand why current generations of immigrants are not faring as well as expected. Yet, as it stands, the evidence relying on the LFS is very limited (e.g., Hira-Friesen 2017; Skuterud and Su 2012). Second, more work can be done using the LFS to better understand the labour market implications of different family and household relationships. The LFS has very detailed information on the various labour market states, and using its master files, one can link the data of family and household members and even follow them over a six-month window. As such, one might, for example, see how an external shock to one family member affects the labour activities of other members over the short run (e.g., Lluis and McCall 2018). Finally, I am looking forward to seeing research using the recently added visible minority variables.

The rest of the article is organized as follows: in the next section, I provide a detailed historical overview of the LFS. I then compare and contrast the LFS's public use files and master files. The final section concludes.

## 2 Historical Overview of the Labour Force Survey

In this section, I provide a historical overview of the LFS from the perspective of an empirical researcher. The goal is to help researchers truly appreciate the subtleties of the LFS, which is critical to determining whether the LFS is appropriate for the research at hand. It also helps in determining how to best take advantage of the many years of LFS data, whether it be as repeated cross-sections or as mini-panels. Finally, I lay the foundation for the choice of whether to use the public use or master files. I start by focusing on the survey itself, then move on to interview and data collection methods, and finally provide a discussion of data quality as proxied by survey non-response rates.

### 2.1 The Survey

Since its inception as a quarterly survey in November 1945, and subsequently as a monthly survey (November 1952 onward), the LFS has been gathering labour market information on Canadian households. As of the 1960s, it became the official source of labour force statistics (e.g., unemployment, employment, and labour force participation rates), statistics that are barometers of labour market activity in Canada. These statistics have a direct bearing on the generosity of labour market programs. The magnitude of the regional unemployment rate, for example, helps determine eligibility requirements and benefit duration of the Canadian EI program.

The present-day target population of the LFS is individuals aged 15 years and older who live in Canada, are non-institutionalized, and who are not full-time members of the armed forces. Individuals living on reserves, and in very remote areas, are excluded from the survey.[12] The

[12] The initial surveys, which interviewed approximately 25,000 households, had similar restrictions compared with today's current practices, except that they were restricted to those aged 14 years and older and excluded individuals living in remote areas of provinces (Dominion Bureau of Statistics 1946). Coverage was extended to the remote areas of provinces starting in 1955. The lower age limit was increased to 15 years with the 1966 update.

LFS only started providing data for the Yukon, the Northwest Territories, and Nunavut in 1992, 2001, and 2004, respectively. Although not as large as the Census, the LFS is one of the larger surveys carried out by Statistics Canada; it interviews approximately 55,000 households per month, and it has been around this level since the mid- to late 1970s.[13]

### 2.1.1 1976–1989

Since 1976, the LFS has followed a rotating panel approach. The LFS interviews a household (i.e., individuals who live in the same dwelling irrespective of whether they are related) for six consecutive months, and every month one-sixth of the (sampled) households are replaced. It is important to recognize that the survey is anchored to dwellings; households are only part of the survey as long as they live at the sampled dwelling.[14] An individual who moves out before the end of the six-month window is, after the move, beyond the scope of the survey. Similarly, an individual who joins the dwelling after the first month is only asked labour market questions for the period after joining the surveyed household or dwelling. By tying the survey to dwellings (and surveying those who live in them) and replacing outgoing dwellings with new ones in the same or similar area, Statistics Canada ensures reasonable employment and unemployment estimates for geographical subregions.

During the household's "birth" interview, the LFS collects information on all household members, which involves both demographic (gender, age, education, and marital status) and labour market information. Subsequent interviews verify the list of household members, and if

[13] Following the 1976 sample ramp-up, the sample size of the LFS (1977–2017) ranged from 92,898 to 125,646 individual observations. This reflects changes, over time, in the number of households interviewed and also variation in the number of individuals per household.

[14] Formally, a dwelling is "any set of living quarters that is structurally separate and has a private entrance outside the building, or from a common hall or stairway inside the building" (Statistics Canada 2018, 10), whereas a household is formally defined as "any person or group of persons living in a dwelling. A household may consist of any combination of: one person living alone, one or more families or a group of people who are not related" (Statistics Canada 2018, 12).

no change has occurred, the interviewer carries forward the household demographic information and focuses on labour market questions. A birth interview may occur in the second (or later) month of the six-month window as a result of such circumstances as a complete change in household occupants or a dwelling vacancy in previous months. In these cases, the household is still considered a birth because it is in its first month of response. Labour market information would also be carried forward if the respondent's employer had not changed. The idea of carrying forward information so as to reduce respondent burden is not new to the LFS; it goes back to the pen-and-paper era in which the interviewer would enter the respondent's answers directly on the paper survey. Statistics Canada (1979) states that the interviewer would preprint the household information and attach it directly to the survey form. Some basic labour market information was also preprinted (Statistics Canada 1979). This included the following: "for whom did . . . work" (item 72); "when did . . . start working for this employer" (item 73); "what kind of business, industry or service was this" (item 74); "what kind of work was . . . doing" (item 75); "class of worker—main job" (item 76); and "class of worker—other job" (item 77). That way the interviewer could bypass these questions if the respondent's employer had not changed.

As briefly discussed in the introduction, the LFS now reuses all labour market information collected in the initial interview of those aged 70 years and older to reduce the response burden. This was first mentioned in the 1992 *Guide to Labour Force Survey Data* (and repeated in subsequent *Guide to the Labour Force Survey* publications [1997 onward]). They all state that the data are now carried forward, but not as of when. My own mini-panel analysis indicates that it was most probably introduced in the late 1980s, starting with the September 1988 incoming rotation group. As such, an individual aged 70 years and older who was, for

example, employed as of the first month will automatically be counted as employed for the next five months of the survey window (and his or her job tenure will be automatically updated).

Each month, the LFS collects the labour market information for a particular week (Sunday through Saturday), called the reference week. This reference week typically includes the 15th of the month. The first LFS, for example, asked respondents their labour market activities as of the 11–17 November 1945 reference week (Dominion Bureau of Statistics 1946). December is an exception. Statistics Canada moves up the reference week to ensure that enumeration is not too close to Christmas. For example, the reference week for the December 2018 LFS was 2–8 December. Enumeration occurs in the survey week, the week immediately after the reference week.[15] The fact that the LFS asks labour market questions about the previous week makes it less prone to recall error compared with other surveys (e.g., the SLID).[16] Given that the LFS is the official source of employment and unemployment data, providing timely release of such data is a priority. As such, the LFS has always had a tight deadline from enumeration to its public release. In 1952, it was a four-and-a-half-week process, whereas it now takes slightly less than four weeks. The monthly LFS statistics are typically released on the Friday of the fourth week from the start of enumeration. It should be noted that the release dates are publicly announced well in advance. This very tight production timeline makes the LFS a timely source of data that is critical when carrying out public policy debates, particularly in times of crisis (e.g., COVID-19

[15] The length of the enumeration period has varied over time. In its early days, the enumeration period lasted two weeks (Statistics Canada 1993). By the mid- to late 1970s, it had been reduced to a six-day period (i.e., Monday to Saturday). Starting in 1977, Statistics Canada began a post-survey-week follow-up in the week after the survey week (Monday and Tuesday) for the months of July and August, when temporary absences tended to be high (Gower 1979). Now, the enumeration is also carried out on Sundays; it starts the Sunday immediately after the reference week and ends on a Tuesday, nine days later. Statistics Canada will, on rare occasions, extend collection by one day (Statistics Canada 2008b, 2017b).

[16] The SLID, for example, asks labour market activity questions regarding the previous year. There is an extensive literature on the problems of recall error and its relationship to the length of the recall period (e.g., Bound, Brown, and Mathiowetz 2001).

pandemic; Milligan and Schirle 2020). To illustrate this point, statistics from the March 2020 LFS, which gathers labour market activities of Canadians for the 15–21 March reference week, were publicly released on 9 April 2020.

As with other long-running surveys, the LFS has gone through regular updates (typically post-censal). I provide next a short discussion of the changes, concentrating on those most relevant to empirical researchers. The focus is on the post-1976 period because of important changes introduced in 1976 that, for all intents and purposes, made for a new survey and because pre-1976 microdata are not available to researchers. The 1976 redesign introduced new elements such as rotation groups, Unemployment Insurance regions, and, most important, a revised questionnaire. Of note, the new LFS shied away from focusing on the main activity for determining labour force status, in effect relying on a more direct approach in which respondents are asked whether they worked during the reference week. The number of questions asked was substantially increased, with a distinction being made between hours at the main job versus other job(s), and respondents now being asked about usual hours worked and about overtime and extra hours. The educational attainment questions were also revamped, and school attendance questions were added. Another important area of change was the increased number of questions asked of those who are unemployed (e.g., what type of job they are seeking).[17] Finally, I should note that the sample size was also significantly increased, from approximately 35,000 households to 55,000 households.[18]

[17] See Usalcas and Kinack (2017) for further discussion. Statistics Canada (1977b) and MacDonald (1981) are also good sources (albeit more technical in nature), with the latter providing both pre- and post-1976 questionnaires.
[18] The survey ramp-up was not uniform across time and provinces. It started in urban areas (and Prince Edward Island). By February 1977, the sample size had reached 55,730 households. It should be recognized that the 1976 weights were adjusted to reflect these sample changes.

Although microdata are not available before 1976, Statistics Canada does provide historical labour force statistics, such as employment and unemployment counts and rates (overall and for select subgroups) that span 1976 (e.g., Statistics Canada 1991). To do so, they take advantage of the parallel run of the old and new questionnaires in 1975. They found, for example, that the new questionnaire measured more employment, particularly for women and youths, and also more unemployment, especially for women (Statistics Canada 2006). Statistics Canada then adjusted the pre-1976 series using what it learned from the parallel run, with the caveat that the adjustment that applies to 1975 also applies to earlier years. Of interest to researchers who rely on historical labour market statistics is the work carried out by Stephen Gordon through Project Link, in which Statistics Canada series are spliced together to produce a "rough-and-ready" time series (including seasonally adjusted ones) going as far back as 1945 (see Gordon n.d. for more details).[19]

The post-1981 and post-1986 censal revisions were minor in scope. They tended to be technical in nature and, for the most part, did not directly affect researchers.[20] A couple of changes, however, are worth noting. The first is with respect to occupation and industry classifications. In January 1984, the LFS moved from the 1971 census classification to the 1980 Standard Occupation Classification. At the same time, it also updated the industry classification, transitioning from the 1970 Standard Industry Classification (SIC) to the 1980 SIC. Since then, the LFS has gone through regular updates of both occupation and industry classifications (see Appendix A for more details). The LFS will provide, whenever possible, a historically consistent

[19] For the labour market statistics available through Project Link, the pre-1976 data come from the following three Statistics Canada publications: Dominion Bureau of Statistics (1958), Statistics Canada (1975), and Statistics Canada (1991).

[20] See Usalcas and Kinack (2017) for a detailed discussion of the technical changes across time. They also provide a detailed discussion of changes over time in geographical classification at the sub-provincial level.

occupation–industry variable bridging such changes. In cases in which the changes are considered too severe, however, the occupation–industry variable is discontinued, and a new variable is introduced on the basis of the updated classification. This problem is not unique to LFS users; it is one commonly faced by researchers who rely on multiple decades of data (e.g., the Census). The solution is to rely on crosswalks (concordance tables), such as those provided by Statistics Canada, that show the links between the two classifications, making it possible to create relatively consistent series at coarser occupation–industry groupings.

The second change that re-occurs at regular intervals, and is of relevance to researchers, is population rebasing. The LFS is meant to be reflective of its target population. The LFS relies on external sources—the most recent census and administrative data between censuses—to account for changes over time in the target population. The LFS will then update its estimated population counts when new census data are released. The weights provided in the LFS account for these population counts. As such, the 1984 rebasing that relies on the information from the 1981 Census (taking into account net undercoverage) led to a revision of LFS weights going back to 1976 (Statistics Canada 1984). Subsequent Censuses led to similar revisions. Given that researchers typically rely on weights when presenting summary statistics and carrying out regression analysis, these revisions could affect findings. In practice, however, such revisions have little impact. After every rebasing, Statistics Canada provides labour market statistics pre- and post-adjustment, and the differences between the series are always very minor. My own experience with the LFS (e.g., Brochu 2013; Brochu and Green 2013; Brochu, Gross, and Worswick 2020) suggests the same, as the historical revisions have very little impact on my

(regression) findings. In addition, my review of articles that that rely on the LFS does not find any evidence to suggest that findings are sensitive to population rebasing.[21]

### 2.1.2 1990s

The 1990s, however, saw important changes to the survey. It started in 1990, with significant modifications to the education questions. The focus changed from measuring years of education to measuring education attainment. Moreover, respondents are now asked whether they graduated from high school. The approach to measuring non-primary and non-secondary schooling also changed. Before the modifications, the LFS would ask whether the respondent had other education (other than high school), but the focus was on education that normally required high school graduation. This post-secondary criterion was, however, dropped with the advent of the new education questions. "Other education" now includes any education that counts toward a degree, certificate, or diploma from an educational institution. Finally, the LFS greatly expanded the set of non-high-school educational responses. As a result of these important changes, there are significant challenges to constructing time-consistent education groupings across the January 1990 threshold. See Appendix B for more detail.

Another dimension that matters to researchers is how the education measure in the LFS compares with that of other surveys. This is important when researchers need to use multiple sources of data, as is the case when examining the evolution of wages and earnings over time (e.g., Burbidge, Magee, and Robb 2002; Green and Townsend 2010). Burbidge et al. (2002)

[21] Given that seasonally adjusted series provided by Statistics Canada are based on raw (unadjusted) data, any adjustment to the raw data (such as population re-basing) will lead to adjustments to the seasonally adjusted data, too. I should mention that between censuses, the LFS revises its seasonal adjustment every year going back three years. Finally, the methods used for seasonally adjustment have been updated over the years (see the many editions of the *Statistics Canada Quality Guidelines*—e.g., Statistics Canada [2019]—for more information). As with the unadjusted labour market statistics, however, Statistics Canada provides readily available seasonally adjusted series from 1976 onward and has historical publications that provide series that span 1976 (e.g. Statistics Canada 1991).

compare the Survey of Consumer Finance (SCF), the SLID, and the LFS, and they show that one can define education categories in the LFS and SCF so that researchers who switch to the LFS for more recent data (because the SCF was discontinued) face less of a break in the education series than with the SLID.

The mid- to late 1990s brought about important revisions to the LFS. Although the changes were across the board (i.e., the questionnaire, the design of the sample, the collection technology, and the processing systems; Sunter et al. 1995), I focus on the new questionnaire, which was phased in by rotation group starting in September 1996.[22] The economy had gone through significant changes since the last major revision to the questionnaire (in 1976), and the LFS was "moving with the times" (Statistics Canada 1995). The labour force participation of women (and particularly young mothers) had dramatically increased. There was also a belief that globalization and the advent of the 1982 recession had altered the employer–employee relationship, thus polarizing labour market outcomes such as wages, hours worked, benefits, and job security (Sunter et al. 1997). As such, there was a need for new questions, and new responses to existing questions, to fill gaps in the data.[23] With this in mind, the LFS added a series of questions meant to improve the understanding of jobs held by Canadians (i.e., their characteristics and quality). The new content includes information on wages, hours worked, union membership and coverage, job permanency, and the size of the employing establishment.

The LFS now asks employees a series of wage and salary questions:

1. "Is . . . paid by the hour?"

[22] September: rotation three; October: rotations three and four; November: rotations three, four, and five; December: rotations three, four, five, and six; and January 1997: rotations three, four, five, six, one, and two.

[23] New labour market questions were also introduced to take advantage of the recently introduced computer-assisted interviewing that made it possible to better deal with the complex branching resulting from an expanded survey. Interviewing methods, including the use of computer-assisted interviewing, are discussed in more detail in the Interview Methods section.

2. “Does . . . usually receive tips or commissions?”

3. “[Including tips and commissions,] what is . . .’s hourly rate of pay?”

4. “What is the easiest way for you to tell us . . .’s wage or salary [including tips and commissions], before taxes and other deductions? Would it be yearly, monthly, weekly, or on some other basis?”

On the basis of the answers provided to these questions (and the usual-hours-of-work question), Statistics Canada derives hourly and weekly wage variables. A few points are worth stressing: first, the wage variables measure gross wages and salaries (i.e., before taxes or deductions) and include tips and commissions. Second, if individuals hold multiple jobs, the reported wage is for their main job (the job with the largest number of usual hours worked). Third, there are no wage data for the self-employed (or those working in a family business without pay) because one needs to be an employee to be asked wage and salary questions. Finally, these are usual wage measures. Respondents who are, for example, paid a fixed weekly salary will have both a weekly wage and an hourly wage. In the case of the latter, the LFS will simply divide the weekly wage by the usual number of hours worked per week.[24] It is also important to recognize that the wage questions are only asked at the first interview (unless the individual changes employer or is a new hire). For example, individuals who were employed in the first interview and remained with the same employer in subsequent months will see their wage carried forward from one month to the next (i.e., it will not be refreshed). As such, the wage data may not be up to date for all individuals. Only asking wage questions at the first interview was a decision long in the making. Statistics Canada started to work on the survey redesign in 1991 (Sunter et al. 1995). Statistics

[24] “Prior to January 1997, usual hours was the number of hours usually worked by the respondent in a typical week, regardless of whether they were paid. Beginning January 1997, usual hours for employees refers to their normal paid or contract hours, not counting any overtime. However, the definition of usual hours remains unchanged for the self-employed and unpaid family workers” (Statistics Canada 2018, 17).

Canada came to the conclusion that it was a just trade-off between providing researchers with monthly wage data and protecting confidentiality and reducing respondent burden. It was considered too invasive to consistently update wage information given that it could result in previously acquired information being shared with a different household member upon verification and updating (Sunter et al. 1995).[25]

Another key job characteristic now gathered in the revised questionnaire is union information. The new union status questions ask whether individuals are union members and, if not, whether they are covered by a union. It is common practice for researchers to treat as one both union members and those who are covered by a union but not members (e.g., Brochu, Gross, et al. 2020; Riddell 1993; Schirle 2015b). In both cases, they benefit from the union's presence. In practice, whether one focuses on coverage (which includes both members and those who are not members but are covered by a union) or unionization (just members) makes very little difference in Canada because the two series are very similar. This is not the case in many European countries in which collective agreements reached by a "representative" union and an employer association are extended to all workers in the sector (Card, Lemieux, and Riddell 2020).

The job permanency block of questions was introduced to get at the quality of the job, whether it was permanent, temporary, seasonal, or just casual in nature. To better understand the hours-of-work dimension, the LFS makes a distinction between paid and unpaid overtime hours, and it identifies whether the worker's hours tend to vary from week to week. As a final measure

[25] The American CPS is similarly mindful in its approach to the earnings data, because its earnings questions are only asked of those in the outgoing rotation group.

of quality, the LFS now provides information on establishment size (i.e., number of employees at the location of employment) and firm size (i.e., number of employees at all locations).

Of potential interest to many social science researchers, the LFS enriched the set of potential responses for a substantive number of questions. This was done to better account for the role of family responsibilities with respect to those who work (e.g., why they were absent from work, why they were working part time), those who are unemployed (e.g., why they left their jobs), or those out of the labour force (e.g., why they did not look for jobs). Similarly, there was an expansion of possible responses to account for economic conditions that were beyond the respondents' control (e.g., why they were working part time and why they lost a job). Finally, as of January 1997, the LFS started gathering dates of birth. That way, labour market information could be collected if the individual turned age 15 years during the six-month window (Statistics Canada 2018). It also meant that one could now observe a change of age in the data. If, say, an individual turned age 20 years halfway through the six-month window, one would observe an age of 19 years in the earlier months and 20 years in the later months.

When new questions are added, others may be dropped to mitigate the increased response burden. A noteworthy example is that up until 1996, the LFS had asked all employed workers whether they had looked for a new job in the past four weeks and, if so, what they had done to find one. Because of its confidential nature, this information was not included in the public use files. Interestingly, it was not included in the master files made available through the RDCs. This may be due to the fact that the information (variables) had been discontinued by the time the RDCs first opened in 2000.[26] The reduction of response burden may also be the reason why the

[26] Skuterud (2005) uses the on-the-job search information present in internal use LFS files to explore for changes in long-term employee loyalty. He finds that the proportion of employed workers looking for another job more than doubled over the 1976–1995 interval, which appears to be consistent with decreasing long-term search costs.

"previous separation" component of the LFS questionnaire (which asked of the employed why they had left their previous job and whether that job had been a part- or full-time job), which was introduced in 1996, was subsequently discontinued as of January 2006.

Starting in November 1999, the LFS expanded the marital code set from four to six: married and living in a common-law arrangement, which were previously grouped together, are now separate categories. Similarly, separated and divorced were also granted their own categories. Interestingly, the LFS has from very early on made a point of identifying those who are living in a common-law relationship (as opposed to being single) and including them with those who are married because their "intent is not to determine the legal marital status, but rather to supply family-related information that can be used in economic analysis" (Statistics Canada 1979, 21). It should be noted, however, that until very recently same-sex couples who reported being married were actually reclassified as common-law in the LFS. This changed with the 2021 revisions, when the LFS revised the marital status data retroactively back to 2006 to reflect the status reported by respondents (Statistics Canada 2021).

### 2.1.3 2000 Onward

The 2000s saw the inclusion of a set of immigrant and Indigenous identity questions.[27] Beginning in January 2006, the LFS asked respondents whether they were, or had ever been, a landed immigrant and, if so, the year and month when they first became one (Gilmore 2009).[28] The introduction of the Indigenous identity questions—that is, whether the respondents identified

[27] The Indigenous identity terminology is recent to the LFS. Indigenous identity questions were historically referred to as "Aboriginal status questions."

[28] It should be noted that temporary residents (TRs) are included in the LFS if they declare that they were contacted at their usual place of residence. An anonymous referee's own analysis, from having worked with both the census and the LFS, suggests that the LFS is increasingly capturing the TR population (especially foreign students). My own research on temporary foreign workers (TFWs) (i.e., Brochu, Gross, et al. 2020) suggests the same. Given that these TRs are not landed immigrants, they will fall under the non-immigrant category, and this holds true for both the public use and the master files.

as Inuit, Métis, or North American Indian—was staggered; it started in Alberta in January 2003 and was expanded to all western provinces in April 2004 and finally across Canada beginning in January 2007 (Pérusse 2008). There are two qualifiers. First, respondents are asked the Indigenous identity questions only if they were born in Canada or the United States; second, the added questions provide only an incomplete picture of Indigenous identity, because those living on reserves remained outside of the scope of the survey.

Although Census evidence at the time clearly documented the poor labour market performance of Indigenous identity (e.g., George and Kuhn 1994; Pendakur and Pendakur 1998, 2002), the Alberta government still saw a need to better understand its labour market outcomes when they requested that Indigenous identity questions be added to the LFS (Statistics Canada 2005a). A natural question is why the coverage of the LFS was not expanded at the same time those status questions were added in 2003. Maybe an even better question is why reserves were not covered in the first place. Both the 1979 and the 1992 *Guide to Labour Force Survey Data* state that the populations on Indian reserves are not covered by the LFS for conceptual (i.e., "general inapplicability of the survey concepts and definitions to the measurement of labour market conditions on reserves"; Statistics Canada 1979, 7), statistical (i.e., would require larger sampling rates to get reliable estimates of reserve population, resulting in problems of respondent burden; Statistics Canada 1979), and operational (i.e., "difficulties involved in carrying out monthly surveys in such area"; Statistics Canada 1979, 5) considerations. As of 2003, the *Guide to the Labour Force Survey* justified the exclusion of reserves by focusing on the operational dimensions (and their associated cost).

The debate on whether or not to include reserves in the LFS flared up to the national stage in the mid-2010s with the release of a Statistics Canada report on the Siksika Nation Pilot

LFS (i.e., Statistics Canada 2013) and the controversy surrounding the TFW programs (i.e., "Why Do We Know So Little" 2014). The pilot project was carried out on a reserve approximately one hour east of Calgary over the 2010–2011 period, and it found that the response rate was similar to that of other Albertans and the collection challenges were not unlike those experienced in other remote areas that are part of the LFS. As for the TFW controversy, the employment minister at the time, Jason Kenney, had announced that the low-end wage component of the program would be scaled back so as to force employers in high unemployment regions to not solely rely on the TFW program for its source of labour but to instead hire from underemployed groups that have difficulty getting a foothold in the labour market—which includes Indigenous peoples. It is easy to say that employers will not be allowed to hire TFWs in regions with unemployment higher than 6 percent, but if the government cannot get a true measure of regional unemployment (because a local reserve is not part of the count), this is not effective public policy.

Finally, the LFS added an ethnicity question in July 2020 that is meant to identify population groups designated as visible minorities (Statistics Canada 2020). The possible responses are the same as for the 2016 Census (i.e., White, South Asian, Chinese, Black, Filipino, Latin American, Arab, Southeast Asian, West Asian, Korean, and Japanese).

### 2.2 Interview Methods

Interviews were initially done in person, but as time went by phone interviews started to become more prevalent. As of 1972, Statistics Canada started using a "warm telephone procedure" in self-representing areas, that is cities of a certain size (Drew 1991). The initial interview was face to face, but subsequent interviews were conducted by phone, if the household

agreed. Sheridan, Drew, and Allard (1996) believe that using telephone interviews in subsequent interviews played a role in the drop in non-response rates from the 10–12 percent range in the early 1970s to the 5–6 percent range over the 1977–1996 period. This warm telephone procedure was extended to small urban centres and rural areas only with the post-1981 censal review (implemented over the 1984–1985 period). Statistics Canada delayed the introduction of telephone interviews in small urban centres and rural areas because of confidentiality issues, such as the presence of party lines (Statistics Canada 1979), which were phone lines shared by multiple households. The phones would ring in all households sharing the line, but a distinctive ring would identify the household for which the call was intended.

In 1990, the LFS started doing birth interviews over the phone for those living in high-rise apartments. This was done to deal with the security arrangements of those buildings and also with the difficulty with finding someone home (Sheridan et al. 1996). As of November 2004, first contact by phone was extended to all urban households, conditional on a telephone number being available (Statistics Canada 2008b).[29]

It should be noted that up to 1993, the LFS relied on pen-and-paper interviewing. In such a case, survey responses were recorded on a paper questionnaire irrespective of whether the interview was done in person or over the phone. With the introduction of computer-assisted interviewing, which was introduced between November 1993 and March 1994,[30] the interviewer would capture the responses on a laptop; if the interview is conducted in person, it is called computer-assisted personal interviewing (CAPI); if it is by telephone, it is called computer-

[29] When focusing on the incoming rotation (i.e., the initial interview), one observes that 98 percent of the interviews were done in person in 1979. The number dropped to 81 percent in 1990, to 73 percent in 1994, and finally, by 1999, to 66 percent (author's calculation).

[30] The phase-in was not done by rotation groups. One-third of the sample moved to computer-assisted interviewing in November 1993, the second third in December 1993, and the remaining third in March 1994.

assisted telephone interviewing (CATI). The CATI was initially decentralized (the interviewer would call from home), but over the June–September 2000 period, the LFS switched to a centralized system in which the interviewing was carried out from a call centre in each regional office (Statistics Canada 2008b). Mayda and Ford (2003) mention that the introduction of the Canadian Community Health Survey in 2000 required an important increase in the number of interviewers, so to meet capacity requirements they decided to adopt a centralized approach. The introduction of CATI made it possible to carry out (online) consistency edits and, as previously discussed, better deal with the complex branching resulting from the expanded 1996 questionnaire.

In addition to CAPI and CATI, there is now a third way of collecting the data. As of March 2015, the LFS started offering eligible respondents the option of answering online. As with most changes made by the LFS, its introduction was gradual. To start (i.e., over the March–October 2015 period), only half of each rotation group was given the option to answer online. It became available to all eligible respondents as of November 2015 and was phased in by rotation groups. By April 2016, it had been fully phased in (Statistics Canada 2017b). Only non-birth interviews are eligible for the electronic questionnaire (also referred to as the computer-assisted web interview), and it has to be carried out with the most knowledgeable person of the household, that is, the person who provided the information for all household members at the birth interview. If a household assigned to an electronic questionnaire does not respond or complete the survey within the first four days of collection, the case is transferred to CATI. It is expected that approximately one-fifth of subsequent interviews will be conducted by electronic questionnaire, with most of the remaining ones being done through CATI. See Statistics Canada (2017b) for a detailed discussion of current collection methods.

## 2.3 Non-Response

For household surveys such as the LFS, survey non-response is measured at the unit or household level. As such, the LFS unit non-response rate of 4.7 percent for January 1997, for example, implies that 4.7 percent of the households in the January 1997 LFS did not respond to the survey. Non-response could be due to the fact that Statistics Canada could not make first contact in the first month of the six-month window or simply because the household was temporarily away (e.g., on vacation) at one of the subsequent interviews.

Alternatively, one could be interested in the proportion of individuals who do not respond to the survey. Getting an accurate count of such a measure, however, is not possible because of the survey's structure. As discussed in the "Survey" section, the LFS is cross-sectional in nature. The January 2020 survey, for example, asks questions (including who lives in the dwelling) only for the reference week of January 2020. As such, there is no way of knowing how many individuals lived in a non-responding dwelling. One could worry that household non-response underestimates the true proportion of survey non-responders because it does not account for cases in which some (but not all) household members are not interviewed. Because of the option for proxy responses, however, this type of non-response (which Statistics Canada [2017b] calls person non-response) is of minor importance.

Household survey non-response has been on the rise since the mid- to late 1990s, and this is a worldwide phenomenon. For example, non-response rates for the American National Health Interview Survey were less than 10 percent in 1997 but rose to nearly 25 percent by 2013 (Meyer, Mok, and Sullivan 2015). Other household surveys were not exempt from this trend. Expenditures surveys in Commonwealth countries (Canada, Australia, and the United Kingdom) and the United States saw equally dramatic increases. By 2007, non-response for the United

Kingdom's Living Costs and Food Survey had almost reached 50 percent (Barrett et al. 2014). As a final illustration, the non-response rate for the Canadian General Social Survey, which had hovered around the 20 percent level for most of the 1985–1999 period, started to rise in the 2000s, averaging 43.7 percent over the 2005–2017 period. If survey non-response was random, it would not bias the estimators used in empirical work, and the cost of high non-response rates would simply be one of precision. There is much evidence, however, that non-response is in fact not random (e.g., Green and Milligan 2010; Meyer et al. 2015). Sabelhaus et al. (2014) find, for example, that non-response rates in the U.S. Consumer Expenditure Survey tend to be much higher at the top of the income distribution. Statistical agencies, such as Statistics Canada, will provide weights that will account for survey non-response, among others. But if non-response becomes very large, it is asking much of the weight variables. The *Journal of the American Medical Association* explicitly states that it restricts publication of research in which survey non-response is high (Davern 2013).[31]

Figure 1 shows the LFS quarterly unit non-response rates for the Quarter (Q) 2 1979–Q2 2020 period.[32] Before the COVID-19 pandemic, LFS non-response rates varied for reasons of seasonality (e.g., households have historically been more likely to be away on vacation during July), technical problems resulting from bad weather (e.g., the Quebec ice storm of January

[31] It should be noted that researchers also face problems of measurement error and item non-response, where the latter occurs when some (but not all) questionnaire items are missing. Although important, these issues are beyond the scope of this article. See Meyer, Mok and Sullivan (2015) for an overview of these concerns with respect to household surveys. For recent works, see, for example, Bollinger et al. (2019) and Meyer, Mittag, and George (2020).

[32] The non-response rates come from the following sources: *Labour Force Survey: Quality Report,* a monthly Statistics Canada publication that became semi-annual in the late 1980s (1979–1999); *Labour Force Survey: Annual Quality Report* (2000–2005), and, for 2006 onward, directly from Statistics Canada. It should be noted that the *Labour Force Survey: Quality Report* for the first half of 1999 was unavailable at the time of the writing of this article. The non-response rates for the first and second quarter of 1999 were read off of a figure (Figure 2.5 in the subsequent quality report) that showed how the non-response rate varied over the previous couple of years. Fortunately the non-response rates were very stable during that six-month period, and as such, the estimates are reasonable approximations.

1998), labour strife (e.g., the CATI interviewer strike in fall 2003), and the introduction of new technology (e.g., the introduction of a new computer-assisted interviewing system over the September 1999–November 1999 period). Yet, the quarterly rates (i.e., the unweighted three-month average) and yearly rates (i.e., the unweighted 12-month average) have been very low relative to other surveys and surprisingly stable. Over the Q2 1977–Quarter 3 1999 period, for example, the quarterly non-response rates only twice strayed outside of the 4–6 percent band.[33]

The early 2000s saw a creeping up of non-response. For example, the annual rates for 2003, 2004, and 2005 were 6.6 percent, 7.6 percent and 9.5 percent, respectively. The rise in non-response in 2003 can be explained by the combination of two events: Hurricane Juan making landfall in Atlantic Canada in late September, which resulted in significant technical difficulties at the Halifax regional office, and the LFS interviewer strike in September and October 2003 (Statistics Canada 2008a). Statistics Canada (2008a), however, believes that the introduction of telephone-first contact (starting in November 2004) can explain most of the rise observed in 2004 and 2005. The non-response rates then stabilized at (or slightly above) 10.0 percent in the 2010s. As of the last quarter of 2019, the non-response rate stood at 11.1 percent.[34]

[33] Although I do not have access to quarterly numbers before 1979 (Q2), the annual rate was 5.4 percent in both 1976 and 1977 (Gower 1979).

[34] LFS employment levels in the 2010s continued to track well with those of the Survey of Employment, Payrolls and Hours, an administrative dataset that is often compared with the LFS. When accounting for differences between the two surveys (e.g., Bender 2016; Statistics Canada 2018), the employment levels look very similar over the long run.

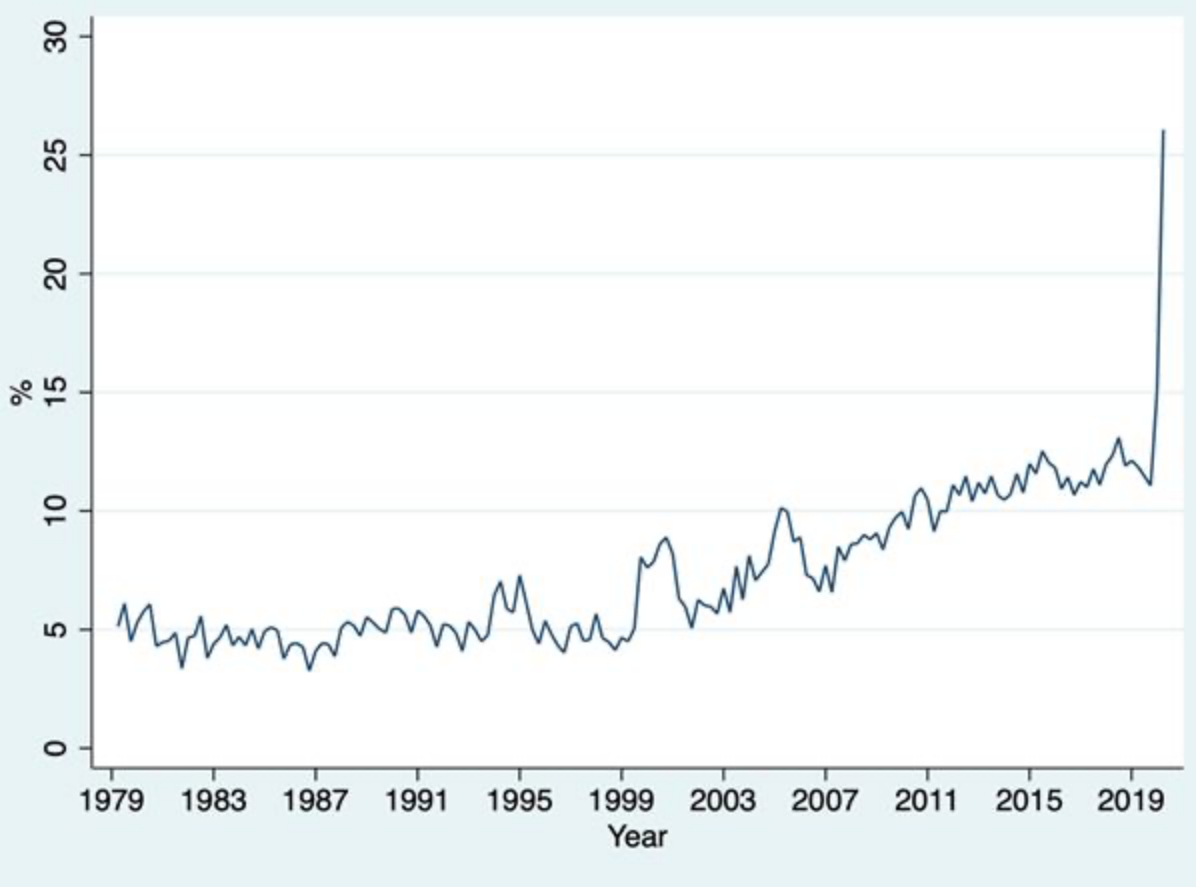


**Figure 1:** Labour Force Survey Quarterly Unit Non-Response Rates (Unweighted), Canada, Quarter 2 1979–Quarter 2 2020

Source: Statistics Canada.

The onset of COVID-19 severely hampered the data collection activities of statistical agencies, including those of Statistics Canada, leading to unusually high survey non-response rates in many countries. Figure 1 shows a rise that started in the first quarter of 2020 and continued into the second quarter. Focusing on monthly rates paints an even more dramatic picture. The non-response rate jumped to 22.1 percent in March 2020 (from 11.9 percent in February) and continued to rise in subsequent months, reaching 28.2 percent in June 2020. If one looks at these recent numbers through a historical lens, one must recognize that before the pandemic, the monthly non-response rate had surpassed the 13.0 percent threshold only twice

(13.0 percent in July 2018 and 13.4 percent in August 2018) since the introduction of the modern-day LFS in 1976.

Focusing on the years before the pandemic, one can observe that a non-response rate of 11.1 percent for the 2010s, albeit double what was observed in the earlier years of the survey, is still very low relative to other surveys.[35] The fact that the LFS is a mandatory survey probably has a role to play, but there is more to it.[36] The LFS is a flagship survey of Statistics Canada, and it is the official source of employment and unemployment numbers. As such, much thought and effort has been put into the LFS, which can to a great extent explain the historically low levels of survey non-response.

The approaches used to minimize non-response (and measurement error) include that important changes such as questionnaire revisions or technological innovations are typically years in the making, and they are thoroughly analyzed before, during, and even after their introduction. For example, work on the 1976 questionnaire redesign started in 1971. Given the significant changes to the survey, in terms of both its new design (e.g., introduction of rotation groups) and revamped questionnaire, it was, for all intents and purposes, a new survey, and Statistics Canada treated it as such. It was phased in over an eight-month period (starting in May 1974), but it only came into official use in January 1976. As such, one can think of the parallel run (over the full 1975 calendar year) as a calibration period. The 1996 questionnaire revision is another such example. Statistics Canada started working on the revision in 1991 (i.e., as a post-censal review); it involved extensive user consultation and field testing of the revised questionnaire (Sunter et al. 1995), and there was a five-month phase-in. The latter point is worth

[35] Historically, the CPS has had survey non-response rates similar to the LFS, with the exception that its rise from the 4–5 percent range started earlier, that is, in the mid-1990s (U.S. Census Bureau 2006, 2019).
[36] See Green and Milligan (2010) and Meyer et al. (2015) for detailed discussions of the mandatory versus voluntary issue.

stressing. Important changes are typically introduced in a gradual way, to minimize disruptions (and identify any potential problems in the process) that can affect non-response rates. Recall that the electronic questionnaire was introduced to half of the rotation groups over the March–October 2015 period, and it only began being phased in by rotation groups starting in November 2015.

Next, the LFS has developed well-defined (tried-and-true) procedural steps before, during, and after first contact to address non-response concerns. Statistics Canada, for example, has a long-standing practice of sending an introductory letter and an information brochure before the first contact. It also sends "refusal letters" to encourage reticent households to participate in the survey (Statistics Canada 2017b). Statistics Canada has shown a willingness to critically evaluate its past performances (e.g., by looking at historical non-response rates) and then make adjustments to its LFS procedures. This includes the introduction of the post–survey week follow-up pilot project (which later became a permanent feature) to address the fact that temporary absences were historically high in the month of July. Another such example is the introduction, in 1990, of phone birth interviews for those living in high-rise apartments.

Finally, training and validation programs have been important components of the LFS program. From very early on, there has been an emphasis on having trained employees (e.g., Gower 1979; Sheridan et al. 1996; Statistics Canada 2008b). Statistics Canada (2008b), for example, mentions a two-month training period for new interviewers that consists of both classroom and on-the-job training under the direct supervision of a senior interviewer. Interviewers were also trained in "doorstep diplomacy," that is, how to make initial contact and call backs (Gower 1979). Finally, Gower (1979) mentions that interviewers' performance is regularly monitored and reviewed—a point that more recent Statistics Canada publications also

stress (e.g., Statistics Canada 2008b). Sheridan et al. (1996) believe that the increased emphasis on interview training and on the importance of controlling the quality of an interviewer's work played a key role in the decrease in non-response over the late 1960s to early 1970s.[37]

Non-response rates in the 20–30 percent range, as observed since the start of the pandemic, are a source of concern. Yet, the comments made about the period when the non-response rate hovered around 10 percent also apply in current times. The recent numbers still compare favourably to other Statistics Canada surveys and to those of other statistical agencies—even if the benchmark was the pre-COVID period. It also means that dealing with non-response rates in the 20–30 percent range is not uncharted territory for the LFS. Finally, because it is a long-running survey, the LFS also has a long history of facing natural and manmade events (as discussed earlier) that have disrupted operations, at least regionally. Therefore, there is reason to be optimistic about the quality of the data provided by the LFS during these COVID times. See Brochu, Créchet, et al. (2020) for more details on how COVID-19 affected Statistics Canada operations and LFS non-response rates.

I wrap up this section with a discussion of the practical implications of the non-response rate patterns shown in Figure 1. As previously discussed, the weights provided by the LFS (and any other survey, for that matter) account for non-randomness in non-response. In the case of the LFS, however, they also reflect non-randomness in its design. The LFS follows a stratified sample design in which the country is divided into strata (i.e., subgroups) that are for the most part geographically defined. Having said this, special strata over the years have had socio-demographic dimensions (e.g., Indigenous strata in regions with a large prevalence of Indigenous

[37] Although now dated, interested readers may want to read Gower (1979) and Sheridan et al. (1996) because they provide insightful historical overviews of the many ways in which Statistics Canada has kept non-response low.

peoples). Within each stratum, the LFS defines clusters from which households are randomly chosen. Such a structure makes it possible to oversample a group of interest that may represent a small segment of the population. For example, to provide precise labour market indicators such as the unemployment rate at the sub-regional level (which are key to the generosity of the EI program), the LFS will oversample in less populated sub-regions and under-sample in more populated ones. Given that labour market outcomes can be very region specific, it is therefore important to apply weights when providing summary statistics. Having said this, I am not aware of any properly designed regression analysis for which the choice of whether to weight (or not) materially affected the findings.

## 3 Public Use versus Master Files

Statistics Canada offers both public use and master file versions of the LFS. An empirical researcher who wants to use the LFS faces the following trade-off: use readily available public use files in which information has been suppressed or rely on master files that provide more detailed data but to which access is restricted. I start this section by discussing how to create mini-panels. This can only be done using the master files. I then explore other differences between the public use and master files. The focus, as in the historical discussion of the survey, is on issues that directly affect researchers.

### 3.1 Mini-Panels

The LFS is not a panel per se because it follows dwellings, not individuals themselves. As such, if individuals leave the dwelling, they are beyond the scope of the survey. One can nevertheless take advantage of the fact that households are part of the LFS for six consecutive

months to construct mini-panels of up to six months in length (e.g., Brochu 2013; Brochu and Green 2013; Brochu, Gross, et al. 2020; Campolieti 2011; Chan, Morissette, and Frenette 2011; Heisz 2005; Jones and Riddell 1999, 2006, 2019; Lluis and McCall 2019; Lluis and Pan 2020; Skuterud and Su 2012). It should be noted that Statistics Canada has relied on the longitudinal features of the LFS to help with the redesign of the LFS survey (e.g., Kinack 1991).

Because the LFS is designed to provide cross-sectional samples, it does not have a single person identifier. However, one can identify individuals across the monthly files by using a combination of variables. Most of these are suppressed in the public use files. As such, a mini-panel analysis can only be carried out using the master files.

Up until very recently, the master file codebook of the LFS did not make explicit reference of how to create mini-panels. As such, researchers had to rely on the existing literature (see Brochu 2013; Brochu and Green 2013; Brochu, Gross, et al. 2020; Campolieti 2011; Chan, Morissette, and Frenette 2011; Heisz 2005; Jones and Riddell 1999, 2006, 2019; Lluis and McCall 2019; Lluis and Pan 2020; Skuterud and Su 2012) to get the necessary information. Yet, this information was often sparse and, even when available, would appear very opaque to a researcher because the stated variables (required to create the mini-panels) would vary depending on the period covered and on when the researcher accessed the data. In this article, therefore, I provide a detailed historical account of panel creation using the LFS and clear guidance on how to create mini-panels up to six months in length going back to 1976.

The record linkage approach, which I present next, was first posted on 12 September 2003 by Lee Grenon, who at that time was the Statistics Canada analyst at the University of British Columbia's RDC. This information, provided as a one-page document (see Grenon 2003), was subsequently made available to all researchers who use the LFS master files.

For the period from 1976 to 1995, one must rely on the survey year (SYEAR), the survey month (SMTH), the regional office (REGOFF), the unique household identifier within a regional office (DOCKET), and a sequential number that uniquely identifies a person within a household (LINE). The original document also recommended incorporating geographical variables (i.e., economic regions [ERTAB], census metropolitan areas and urban centres [CMATAB], and unemployment insurance region [UIRTAB]) to improve the match quality. However, these variables were only applicable for specific sub-periods (i.e., ERTAB and CMATAB for 1984–1986; ERTAB, CMATAB, and UIRTAB for 1987–1995). What complicates matters is that geographical boundaries are regularly updated. As such, some geographical variables have been discontinued and replaced by other variables, and others have been updated to reflect new boundaries with the changes being applied retroactively (whenever possible). At the time this article was written, the UIRTAB had been replaced by the EIERTAB (and is left blank before March 1995), and the CMATAB and ERTAB variables (although still in existence) reflect the 2001 Census geography boundaries for the 1987–1995 period. As such, these variables are no longer applicable for the matching process.

Also problematic when matching respondents across time is the closure of two regional offices, the St-John's regional office in April 1994 and the Winnipeg regional office in April 1995. The closure of the St-John's office, for example, causes matching difficulties for the March–April 1994 period for Newfoundland dwellings. One can, however, get around this problem by taking advantage of the fact that the unique DOCKET number of these dwellings remained unchanged upon their transfer to the Halifax regional office. It should also be noted that DOCKET numbers for Newfoundland just before the closure were larger than those found in the Maritimes (which was served by the Halifax office). In practice, this means that for

Newfoundland residents one must account for the fact that the REGOFF variable changed from 11 to 12, whereas the other matching variables remain unchanged. The Winnipeg regional office closed in April 1995. At the time, it served all of Manitoba and some of Saskatchewan (the rest of Saskatchewan was served by the Edmonton regional office). The fact that all cases were transferred to the Edmonton office and that there was no overlap in DOCKET numbers across regional offices makes matching possible. One needs only to account for the fact that their REGOFF variable changed from 16 to 17, whereas the other matching variables remain unchanged.

Finally, for 1996 onward, a different set of matching variables is needed because DOCKET was phased out with the survey redesign. Grenon (2003) recommends relying on SYEAR, SMTH, the one-digit province code (PROV1), the pseudo UIC regions (PSEUDOUI), the regional strata (FRAME), the super-stratum (STRAFRAM), the sample design type (TYPE), the first-stage sampling unit (CLUST), the rotation number (ROTATION), the number assigned to dwellings within a cluster (LISTLINE), the multiple dwelling code for structures that have more than one dwelling (MULT), and LINE.

Starting with the 2016 LFS update, the accompanying LFS codebook (i.e., the February 2016 implementation) introduced a new household identifier variable, HHLDID. In fact, it is a concatenation of 10 pre-existing variables: PROV, PROV1, PSEUDOUI, FRAME, STRAFRAM, TYPE, CLUST, ROTATION, LISTLINE, and MULT. What is novel is that it is the first time the master-file codebook recognizes that the combination of these variables (i.e., HHLDID) can uniquely identify a household. The codebook that accompanied the 2021 update (i.e., the January 2021 implementation) goes one step further and states that the HHLDID variable's universe is all respondents, which implies that it is a unique household identifier

dating back to 1976. The 2021 codebook does state that before 2006, single dwellings without additional dwellings had a blank value for the MULT variable (which is one variable that makes up the HHLDID variable). As such, it should be replaced with a zero when matching dwellings before and after 2006.

Although both the regional office approach in which one uses REGOFF, DOCKET, AND LINE (e.g., Brochu 2013; Brochu and Green 2013) and the approach in which one relies on HHLDID and LINE (e.g., Chan, Morissette, and Frenette 2011; Lluis and Pan 2020)[38] have proven to generate good match quality for the pre-1996 period, I recommend that researchers rely on the HHLDID and LINE variables to uniquely identify individuals across time, even for the pre-1996 period. This is for a couple of reasons: first, it is the approach presently recommended by Statistics Canada (as stated in the master-file codebook); second, it is much simpler and more straightforward to rely on one matching algorithm for all periods. One does not, for example, need to worry about matching difficulties at the period when one switches matching algorithms, which would be the case if one used the regional office approach for the pre-1996 period.

In theory, the matching variables should uniquely identify an individual over time. The fact that the LFS is not designed as a panel, however, may result in matches that do not represent the same individual. Madrian and Lefgren (1999) and Rivera Drew, Flood, and Warren (2014) provide interesting discussions of longitudinal matching in the CPS. Given that both the LFS and the CPS follow a rotating panel approach but are not panels per se, their content or advice is

[38] More precisely, they both use a close variant of the HHLDID and LINE approach for their pre-1996 mini-panels. Chan et al. (2011) use PROV1, PSEUDOUI, FRAME, STRAFRAM, TYPE, CLUST, ROTATION, LISTLINE, MULT, LINE, and three geographic variables (ERTAB, UIRTAB, and CMATAB). Lluis and Pan (2020) rely on PROV1, PSEUDOUI, FRAME, STRAFRAM, TYPE, CLUST, ROTATION, LISTLINE, MULT, and LINE. They also use the FAMID variable, but only because they are interested in identifying members of the same economic family within in a household.

equally relevant in the Canadian context. They suggest dropping observations when there are inconsistencies across months in socio-demographic characteristics, such as age and gender.[39] The LFS records the birthday of the respondent (i.e., day, month, and year), and as such, one would need to account for the possibility that a respondent becomes a year older within their six-month stay in the LFS.

Observations need to be dropped when constructing an LFS panel. The key culprit is the structure of the LFS itself—its rotating panel design. Researchers who want to construct, for example, a two-month panel will need to drop individuals who are part of the outgoing rotation group as of the first month of the panel. As such, one-sixth of the sample will be lost. Researchers who require longer balanced panels will need to drop more rotation groups. Brochu, Créchet, et al. (2020), for example, rely on a February 2020 through August 2020 panel for part of their analysis, to follow the path of individuals through the COVID-19 shutdown and advanced reopening phases.[40] This means, however, that they must drop five out of the six rotations.

Additional observations will be dropped for the following reasons: (a) a new household moves into the sampled dwelling, replacing the existing one; (b) individuals join or leave an existing household partway through the panel; (c) poor match quality; and (d) unusual circumstances (such as an individual refusing to provide any further information) that result in no information being recorded. One can identify new households entering the survey partway

[39] They would also include race, but this variable is less relevant in the Canadian context given that it was only just recently added to the LFS.
[40] This is not possible with the CPS, because its rotating panel design only allows for the creation of panels that are up to four months in length. See Brochu, Créchet, et al. (2020) for more detail.

through the six-month window using the BIRTH variable.[41] One can also identify some of the poor matching, the false positives. Brochu, Créchet, et al. (2020) use gender and age to identify inconsistencies across time (i.e., false positives) and find very few such cases in their 2020 data. One cannot, however, separate out the remaining reasons in the data (i.e., points b, c [false negatives], and d), but as a group, they tend to represent a small fraction of the sample. Brochu, Créchet, et al. (2020) find that for their two-consecutive-month panels, it represents only a couple of percentage points.

Finally, it should be mentioned that the LFS will impute the labour market information of individuals (but carry forward their socio-economic information) if they could not be reached in a particular month but there was information in a previous month. As such, changes in sample size will not fully reflect changes in non-response (such as those observed during the COVID times). See Brochu, Créchet, et al. (2020) for a more thorough discussion of this issue.

### 3.2 Modifications to Public Use Data

Given that public use data are readily available, Statistics Canada will suppress information to protect the confidentiality of respondents. That the data are essentially in the public domain (i.e., there is no controlled access, as with the master files) also has implications for accommodating changes in the public use files—whether it be the result of new questions being added to the survey or simply the desire to include data that had previously been suppressed in the public use files. Statistics Canada must account for the risk of residual disclosure associated with the comparison of different versions of the same data. As such,

[41] Until recently, the BIRTH variable was recorded at the individual level, and as such, different people within a household could have different values. It is now recorded at the household level. The historical values going back to 2006 were readjusted to make them consistent with the newer strategy.

changes to the public use files are not applied retroactively, as they are in the confidential use (master) files (Statistics Canada 2017c). This means that the differences between the public use and master files are period specific. It also means that it falls on the shoulders of the researcher to apply the appropriate concordance between variables over time for cases in which the series is broken through replacement of variables or re-categorization of responses.

### 3.2.1 Suppression of Variables

Trying to account for all the differences across the two versions of the survey and across time is beyond the scope of this article. Instead, I focus on key differences that would be most relevant to a broad class of researchers.

#### 3.2.1.1 Panel Creation Variables: A Further Discussion

The fact that one cannot follow individuals over time in the public use files has sample creation implications when using multiple periods of LFS data. Relying on consecutive months of data would require that the public use data be treated as repeated cross-sections, which is clearly not the case. Five-sixths of the sample in one month is re-interviewed in the next, which means that the observations are not independent. The researcher would thus face a problem of serial correlation, a confounding factor that cannot be accounted for when the researcher's unit of observation is the individual. Given the large sample sizes of the LFS, a common approach to abstract from the non-independence issue, is to rely on data that are six months apart (or more), ensuring that the same individuals only appear once in the data. It is common to focus on one particular month of each year (May LFS data: e.g., Baker and Drolet 2010; Morissette and Johnson 2005; November data: e.g., Green and Townsend 2010). For articles that focus on job

stability patterns in Canada and how (and why) they have evolved over time, it is common to rely on a spring and fall month for each year of data (e.g., Brochu 2013; Heisz 2005).

Not being able to follow individuals over time also means that one cannot identify new respondents, that is, members of the incoming rotation group. This matters when dealing with variables that are not refreshed in subsequent interviews, whether it be (a) socio-demographic variables that are only asked in the first interview (e.g., education and marital status); (b) labour market questions not asked in subsequent interviews unless the individual changed employer; or (c) labour market information of those aged 70 years and older being automatically passed on to the next month to reduce respondent burden. Estimating job stability illustrates this point. Job stability is commonly measured using the retention rate (survival rate) approach, which is the probability that a job as of period $t$ will continue into period $t + 1$ (e.g., Brochu 2013; Diebold, Neumark, and Polsky 1997).[42] Alternatively, some researchers in the minimum wage literature have focused on the separation rate (which is simply 1 – retention rate) or its various components (e.g., Brochu and Green 2013; Dube, Lester, and Reich 2016; Portugal and Cardoso 2006). If one were to rely on two-month mini-panels of the LFS, the retention rate of workers aged 70 years and older would be 100 percent—there would be no job instability whatsoever because job tenure is automatically updated by one month from its previous value—which is clearly an artifact of the survey design.

[42] Given that the LFS and CPS are long-running surveys with detailed labour market information, both have been relied upon when looking for long-term changes in job stability. The LFS has the advantage in that a job tenure question (that asks how long a worker has been working with their employer) has been part of the regular LFS questionnaire since 1976, whereas for the CPS, it is only available in the job tenure supplement. Given that this supplement was only asked at irregular intervals before 1996, the US job stability literature (e.g., Diebold, Neumark, and Polsky 1997) that relies on the CPS has had difficulty differentiating secular changes from cyclical ones. See Brochu (2013) for more details.

#### 3.2.1.2 Geographical Areas

Sub-regional geographic variables, other than province or state and large city identifiers, are commonly suppressed in public use files (e.g., Canadian long-form Census), and it is no different with the LFS.[43] The LFS public use files include only indicators for province of residence and, for more recent years, whether the household lives in one of the nine largest Census Metropolitan Areas (CMAs) and which one. For the earlier years, the CMA variable is binary and only identifies whether (or not) the respondent lives in one of the three largest CMAs. The use of coarse geographic controls is often sufficient, but there are situations in which sub-regional indicators are at the core of the identification strategy. For example, Brochu, Gross, et al. (2020), using the LFS master files, include place-of-residence indicators that identify cities (CMA) as well as smaller towns (i.e. Census Agglomeration (CA)) and rural areas of each province when comparing layoff rates of TFWs with those who are Canadian born. This is done to control for the possibility that observed differences in layoff rates could be driven by TFWs being employed in regions of the country in which layoffs may be less likely to occur.

#### 3.2.1.3 Ethnicity

Ethnicity information, as measured by Indigenous identity questions and those recently added that are designed to identify visible minority populations, is suppressed in the public use files. Researchers who want to take advantage of the rich labour market information of the LFS to better understand the outcomes of visible minorities must therefore rely on the master files. Recall that those living on reserve are not covered by the LFS, and as such, the master files can only be used to explore the labour market outcomes of those living off reserve (e.g., Lamb 2015).

[43] Sometimes even coarse geographical identifiers are not available. Brochu and Morin (2012) had to rely on the sensitive data (master files) of the American General Social Survey because no state identifiers are available in the public use version.

#### 3.2.1.4 Household Members

The variables that can be used to identify individuals across monthly surveys in the master files can also be used to link individuals (aged 15 years and older) who belong to the same household. By using the economic family identifier, one can go a step further and link individuals within the household who belong to the same economic family, where an economic family is defined as "a group of two or more persons who live in the same dwelling and who are related by blood, marriage (including common-law) or adoption" (Statistics Canada 2018). This can help in the understanding of how the socio-economic characteristics of household or economic family members relate to each other. It can, for example, help us understand how outside shocks can affect the labour market outcomes of households and economic families. It should be mentioned that the master files do explicitly provide some economic family and spousal information (more detailed in the spousal case), of which a small (basic) subset used to be provided in the public use file. This information was discontinued as of January 2017, when the public use files added immigrant variables, among others. See Statistics Canada (2017c) for a description of other changes to variables that occurred at this time. There exists an extensive literature on spousal labour supply and the "added worker effect." See, for instance, Stephens (2002), Blundell, Pistaferri, and Saporta-Eksten (2016), Hendren (2017), and Bredtmann, Otten, and Rulff (2018).

#### 3.2.1.5 Other Variables

There are additional variables that are only available in the master files that could interest social scientists. One can, for example, identify the country of birth and where the respondent completed their post-secondary education (if any) for a person who is, or ever was, a landed immigrant. Moreover, for those who have children, one can identify how many children the

respondents have in various age categories. Finally, for researchers whose outcome of interest is the wage, the master files also have information on whether the respondent is paid by the hour and whether they usually receive tips or commissions.

### 3.2.2 Top-Coding

Although there is top-coding in the public use files of the LFS, it is not typically a concern for researchers. Age, for example, is top-coded at 70, which is well beyond the usual retirement age. As such, researchers who look at labour market outcomes typically have an upper age restriction well below 70 years, and this holds true irrespective of whether they use public use or master files data (public use files: Green and Townsend 2010; Schirle 2015b; master files: Brochu and Green 2013; Chan et al. 2011). Hours of work (during the reference week) is another example in which top-coding takes place. Actual and usual hours, for both the main job and for all jobs combined, are capped at 99 hours in the public use files. This is a very high bar. To put this number in perspective, a worker who works all seven days of the reference week would need to work a little more than 14 hours per day to hit the cap.

It should be noted that the master files also top-code some variables, but the cap (if any) tends to be much higher than in the public use files. In the case of age, for example, it was 99 years in the pre-1996 questionnaire period, and it is 130 years in the new one. Hours variables were also capped in the pre-1996 period of the master files (i.e., at 99 hours), but the cap was subsequently removed.

Interestingly, the LFS does not top-code its hourly wage variable. Having said this, the LFS will carry out some risk analysis, and on the basis of these results may perform some

imputation to add uncertainty to the data (i.e., some key variables of the most unique records are replaced by less unique values).[44]

### 3.2.3 Collapsing of Responses into Intervals or Coarse Groups

Collapsing data into intervals or coarser groups, although necessary to protect confidentiality in the public use files, can be problematic for some researchers. I focus on age, immigrant status, and occupation variables to illustrate this point.

The five-year age intervals found in the public use files may be too broad for life-cycle researchers (e.g., Jung and Kuhn 2019).[45] Age intervals can be troublesome when the research question targets a very specific subpopulation. Morin (2015), for example, examines the effect of a labour supply shock on the earnings of high school graduates. More specifically, he takes advantage of the abolition of grade 13 in Ontario, which generated a large (double) cohort of high-school graduates simultaneously entering the labour market. His identification strategy critically rests on age being observed in years in the data. Finally, it should be noted that age categories can be problematic for researchers who rely on a synthetic cohort approach (e.g., the job stability literature: Brochu 2013; Heisz 2005).

The public use files do identify immigrant status (i.e., whether the person is or ever was a landed immigrant), but the information is very sparse. There are only two categories for years since migration (i.e., less than ten years since arrival or ten years or more), whereas the master files collect the year the person first became a landed immigrant and the month if it was in the

[44] For some of the LFS supplements before the introduction of the new 1997 questionnaire, the wage information is top-coded; for example, the hourly wage in the 1984 Survey of Union Membership is capped at $45, whereas for the 1991 and 1995 Survey of Work Arrangements, it is capped at $50 and $40, respectively (Card, Lemieux, and Riddell 2004).

[45] The public use files provide two age variables, AGE_12 and AGE_6. AGE_12 divides the full sample into 12 (five-year) age categories (i.e., 15–19, 20–24, . . . , 65–69, and 70 years and older). The AGE_6 variable, however, focuses on younger individuals and separates them into six categories (i.e., 15–16, 17–19, 20–21, 22–24, 25–26, and 27–29).

past five years. The coarseness of information in the public use files is therefore problematic when analyzing immigrant–native-born wage differentials. The Canadian evidence (e.g., Frenette and Morissette 2003; Green and Worswick 2012) suggests that the assimilation profile of immigrants is more complex than two immigration categories can accommodate, and it should be able to reflect a changing convergence pattern over time. It should also be mentioned that those who want to look at long-term outcomes of immigrants must, at present, rely on the master files; immigrant-related variables became available only as of 2017 in the public use files but as of the mid-2000s in the master files.

The coarser occupation groupings provided in the public use files are often more than adequate, but there are situations in which a more detailed breakdown is needed. TFW programs are meant to satisfy labour shortages that tend to be job specific, and as such, Brochu, Gross, et al. (2020) control for occupation at a very disaggregated level (413 occupation controls) when comparing labour market outcomes of TFWs with those of native-born workers. Although such fine occupation controls did not materially affect their findings, but in many cases resulted in larger effects, it was nevertheless important to include them if they were to credibly argue that they were in fact measuring a TFW effect.

## 4 Conclusion

In this article, I provide a unified source of information on the LFS. Given that researchers often use multiple decades of LFS data, it is important to know how it has evolved over time (through its periodic redesign). I also offer a detailed discussion of the structure of the LFS (e.g., that it is conducted as a series of rolling panels) and its implications for research, whether it be constructing mini-panels or using such variables as wage or tenure. Finally,

building on these considerations, in this article I compare and contrast the public use and master file versions of the LFS. Given the costs associated with accessing the master files, it is important to have a clear understanding of the subtle (and not-so-subtle) differences between the two versions, if one is to use the appropriate files advantageously.

## References


Abowd, J.M., J. Haltiwanger, and J. Lane. 2004. "Integrated Longitudinal Employer-Employee Data for the United States." American Economic Review (P&P) 94(2):224–9.

Abowd, J.M., B.E. Stephens, L. Vilhuber, F. Andersson, K.L. McKinney, M. Roemer, and S. Woodcock. 2009. "The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators." In Producer Dynamics: New Evidence from Micro Data, ed. T. Dunne, J.B. Jensen, and M.J. Roberts, 149–230. Cambridge, MA: National Bureau of Economic Research.

Baker, M., and M. Drolet. 2010. "A New View of the Male/Female Pay Gap." Canadian Public Policy/Analyse de politiques 36(4):429–64.

Baker, M., and K. Milligan. 2008. "How Does Job-Protected Maternity Leave Affect Mothers' Employment?" Journal of Labor Economics 26(4):655–91.

Barrett, G., P. Levell, and K. Milligan. 2014. "A Comparison of Micro and Macro Expenditure Measures across Countries Using Differing Survey Methods." In Improving the Measurement of Consumer Expenditures, ed. C.D. Carroll, T.F. Crossley, and J. Sabelhaus, 263–86. Cambridge, MA: National Bureau of Economic Research.

Baylis, P., P.-L. Beauregard, M. Connolly, N. Fortin, D.A. Green, P.G. Cubillos, S. Gyetvay, C. Haeck, T.L. Molnar, G. Simard-Duplain, et al. 2020. "The Distribution of Covid-19 Related Risks." Working Paper No. 27881, National Bureau of Economic Research, Cambridge, MA.

Béland, L.-P., O. Fakorede, and D. Mikola. 2020. "Short-Term Effect of COVID-19 on Self- Employed Workers in Canada." Canadian Public Policy/Analyse de politiques 46(Supplement 1):S66–81.

Bélanger, E., and F. Gélineau. 2010. "Does Perceived Competence Matter? Political Parties and Economic Voting in Canadian Federal Elections." Journal of Elections, Public Opinion and Parties 20(1):83–101.

Bender, R. 2016. Measuring Employment and Unemployment in Canada and the United States—A Comparison. Labour Statistics Technical Papers. Cat. No. 75-005-M. Ottawa: Statistics Canada.

Blundell, R., L. Pistaferri, and I. Saporta-Eksten. 2016. "Consumption Inequality and Family Labor Supply." American Economic Review 106(2):387–435.

Bollinger, C.R., B.T. Hirsch, C.M. Hakoyem, and J.P. Ziliak. 2019. "Trouble in the Tails? What We Know about Earnings Nonresponse 30 Years after Lillard, Smith, and Welch." Journal of Political Economy 127(5):2143–85.

Bound, J., C. Brown, and N. Mathiowetz. 2001. "Measurement Error in Survey Data." In Handbook of Econometrics, Vol. 5, ed. J.J. Heckman and E. Leamer, 3705–843. Amsterdam: Elsevier.

Bouw, B., and S. Chase. 2010. "In Census Fight, Statisticians Find Strength in Numbers." Globe and Mail, 3 August.

Bredtmann, J., S. Otten, and C. Rulff. 2018. "Husband's Unemployment and Wife's Labor Supply: The Added Worker Effect Across Europe." Industrial and Labor Relations Review 71(5):1201–31.

Brochu, P. 2013. "The Source of the New Canadian Job Stability Patterns." Canadian Journal of Economics/Revue canadienne d'économique 46(2):412–40.

Brochu, P., J. Créchet, and Z. Deng. 2020. “Labour Market Flows and Worker Trajectories in Canada During COVID-19.” Working Papers Series No. 32, Canadian Labour Economics Forum.

Brochu, P., and D.A. Green. 2013. “The Impact of Minimum Wages on Labour Market Transitions.” Economic Journal 123(573):1203–35.

Brochu, P., T. Gross, and C. Worswick. 2020. “Temporary Foreign Workers and Firms: Theory and Canadian Evidence.” Canadian Journal of Economics/Revue canadienne d’économique 53(3):871–915.

Brochu, P., and L.-P. Morin. 2012. “Union Membership and Perceived Job Security: Thirty Years of Evidence from the American General Social Survey.” Industrial and Labor Relations Review 65(2):263–85.

Burbidge, J.B., L. Magee, and A.L. Robb. 2002. “The Education Premium in Canada and the United States.” Canadian Public Policy/Analyse de politiques 28(2):203–17.

Campolieti, M. 2011. “The Ins and Outs of Unemployment in Canada, 1976–2008.” Canadian Journal of Economics/Revue canadienne d’économique 44(4):1331–49.

Card, D., T. Lemieux, and W.C. Riddell. 2004. “Unions and Wage Inequality.” Journal of Labor Research 25(4):519–59.

Card, D., T. Lemieux, and W.C. Riddell. 2020. “Unions and Wage Inequality: The Roles of Gender, Skill and Public Sector Employment.” Canadian Journal of Economics/Revue canadienne d’économique 53(1):140–73.

Chan, P.C.W., R. Morissette, and M. Frenette. 2011. Workers Laid-Off During the Last Three Recessions: Who Were They, and How Did They Fare? Analytical Studies Branch Research Paper Series. Cat. No. 11F0019M. Ottawa: Statistics Canada.

Davern, M. 2013. “Nonresponse Rates Are a Problematic Indicator of Nonresponse Bias in Survey Research.” Health Services Research 48(3):905–12.

Diebold, F.X., D. Neumark, and D. Polsky. 1997. “Job Stability in the United States.” Journal of Labor Economics 15(2):206–33.

Dominion Bureau of Statistics. 1946. Labour Force Survey Bulletin No. 1. Cat. No. 71-001. Ottawa: Dominion Bureau of Statistics.

Dominion Bureau of Statistics. 1957. Labour Force Survey Enumeration Manual. Ottawa: Dominion Bureau of Statistics.

Dominion Bureau of Statistics. 1958. The Labour Force, November 1945–July 1958. Reference Paper No. 58, 1958 Revision. Cat. No. 71-502. Ottawa: Dominion Bureau of Statistics.

Donovan, K., W.J. Lu, and T. Schoellman. 2020. Labor Market Dynamics and Development. Staff Report No. 596. Minneapolis: Research Division, Federal Reserve Bank of Minneapolis.

Drew, J.D. 1991. “Research and Testing of Telephone Survey Methods at Statistics Canada.” Survey Methodology 17(1):57–68.

Dube, A., T.W. Lester, and M. Reich. 2016. “Minimum Wage Shocks, Employment Flows, and Labor Market Frictions.” Journal of Labor Economics 34(3):663–704.

Frenette, M., and R. Morissette. 2003. Will They Ever Converge? Earnings of Immigrant and Canadian-Born Workers over the Last Two Decades. Analytical Studies Branch Research Paper Series. Cat. No. 11F0019MIE. Ottawa: Statistics Canada.

Friesen, J. 2015. “Ottawa Failing to Include First Nations in Key Employment Data.” Globe & Mail, 23 January.

Fuller, S. 2005. “Public Sector Employment and Gender Wage Inequalities in British Columbia: Assessing the Effects of a Shrinking Public Sector.” Canadian Journal of Sociology 30(4):405–39.

George, P., and P. Kuhn. 1994. “The Size and Structure of Native-White Wage Differentials in Canada.” Canadian Journal of Economics/Revue canadienne d’économique 27(1):20–42.

Gilmore, J. 2009. The 2008 Canadian Immigrant Labour Market: Analysis of Quality of Employment. Immigrant Labour Force Analysis Series. Cat. No. 71-606-X. Ottawa: Statistics Canada.

Gordon, S. n.d. “Worthwhile Canadian Initiative.” At https://worthwhile.typepad.com/worthwhile_canadian_initi/stephen-gordon/.

Gower, A.R. 1979. “Non-Response in the Canadian Labour Force Survey.” Survey Methodology 5(1):29–58.

Gower, D. 1993. The Impact of the 1990 Changes to the Education Questions on the Labour Force Survey Staff Report, Labour and Household Surveys Analysis Division. Ottawa: Statistics Canada.

Green, D.A., and K. Milligan. 2010. “The Importance of the Long Form Census to Canada.” Canadian Public Policy/Analyse de politiques 36(3):383–8.

Green, D.A., and J. Townsend. 2010. “Understanding the Wage Patterns of Canadian Less Skilled Workers: The Role of Implicit Contracts.” Canadian Journal of Economics/Revue canadienne d’économique 43(1):373–403.

Green, D.A., and C. Worswick. 2012. “Immigrant Earnings Profiles in the Presence of Human Capital Investment: Measuring Cohort and Macro Effects.” Labour Economics 19(2):241–59.

Grenon, L. 2013. “Record Unique Identifiers.” Statistics Canada, 12 September.

Hamermesh, D.S., H. Frazis, and J. Stewart. 2005. “Data Watch: The American Time Use Survey.” Journal of Economic Perspectives 19(1):221–32.

Heisz, A. 2005. “The Evolution of Job Stability in Canada: Trends and Comparisons with U.S. Results.” Canadian Journal of Economics/Revue canadienne d’économique 38(1):105–27.

Hendren, N. 2017. “Knowledge of Future Job Loss and Implications for Unemployment Insurance.” American Economic Review 107(7):1778–823.

Hira-Friesen, P. 2017. “Immigrants and Precarious Work in Canada: Trends, 2006–2012.” Journal of International Migration and Integration 19(1):35–57.

Hira-Friesen, P. 2018. “Does Employment in Precarious Work Lead to Wage Disparities for Canadian Immigrants?” Canadian Ethnic Studies 50(1):69–86.

Jones, S.R.G., F. Lange, W.C. Riddell, and C. Warman. 2020. “Waiting for Recovery: The Canadian Labour Market in June 2020.” Canadian Public Policy/Analyse de politique 46(Supplement 2):S102–18.

Jones, S.R.G., and W.C. Riddell. 1999. “The Measurement of Unemployment: An Empirical Approach.” Econometrica 67(1):147–62.

Jones, S.R.G., and W.C. Riddell. 2006. “Unemployment and Nonemployment: Heterogeneities in Labor Market States.” Review of Economics and Statistics 88(2):314–23.

Jones, S.R.G., and W.C. Riddell. 2019. “Unemployment, Marginal Attachment, and Labor Force Participation in Canada and the United States.” Journal of Labor Economics 37(Supplement 2):399–441.

Jung, P., and M. Kuhn. 2019. “Earnings Losses and Labor Mobility Over the Life Cycle.” Journal of the European Economic Association 17(3):678–724.

Kinack, M. 1991. “Measuring Data Quality with Longitudinal Data.” In JSM Proceedings, Survey Research Methods Section, 514–9. Alexandria, VA: American Statistical Association.

Koebel, K., and D. Pohler. 2020. “Labor Markets in Crisis: The Causal Impact of Canada’s COVID-19 Economic Shutdown on Hours Worked for Workers across the Earnings Distribution.” Working Papers Series No. 25, Canadian Labour Economics Forum.

Kroft, K., F. Lange, M.J. Notowidigdo, and M. Tudball. 2019. “Long Time Out: Unemployment and Joblessness in Canada and the United States.” Journal of Labor Economics 37(Supplement 2):355–97.

Lamb, D. 2015. “The Economic Impact of the Great Recession on Aboriginal People Living Off Reserve in Canada.” Industrial Relations 70(3):457–85.

Lavoie, H. 1990. Labour Force, Feature Article: Changes to Educational Attainment Questions. Cat. No. 71-001. Ottawa: Statistics Canada, Household Surveys Division.

Lemieux, T., K. Milligan, T. Schirle, and M. Skuterud. 2020. “Initial Impacts of the COVID-19 Pandemic on the Canadian Labour Market.” Canadian Public Policy/Analyse de politiques 46(Supplement 1):S55–65.

Lluis, S., and B. McCall. 2018. “Spousal Labour Supply Adjustments.” Working Papers No. 18-010, Department of Economics, University of Waterloo, Waterloo, ON.

Lluis, S., and B. McCall. 2019. “Employment and Job Search Implications of the Extended Weeks and Working while on Claim Pilot Initiatives.” Canadian Public Policy/Analyse de politiques 45(2):129–72.

Lluis, S., and Y.A. Pan. 2020. “Marital Property Laws and Women’s Labour Supply.” Canadian Public Policy/Analyse de politiques 46(3):340–68.

MacDonald, B. 1981. Conceptual, Definitional and Methodological Changes in the Labour Force Survey. Research Paper No. 3, Cat. No. 71-601. Ottawa: Statistics Canada, Labour Force Survey Division.

Madrian, B.C., and L.J. Lefgren. 1999. “A Note on Longitudinally Matching Current Population Survey (CPS) Respondents.” Technical Working Paper No. 247, National Bureau of Economic Research, Cambridge, MA.

Mayda, J., and P. Ford. 2003. “Integration of CAPI and CATI Infrastructures at Statistics Canada.” In Proceedings of the Eighth International Blaise Users Conference, 225–32. Copenhagen, 21–23 May.

Meyer, B.D., N. Mittag, and R.M. George. 2020. “Errors in Survey Reporting and Imputation and Their Effects on Estimates of Food Stamp Program Participation.” Journal of Human Resources. Published online before print, 10 August.

Meyer, B.D., W.K.C. Mok, and J.X. Sullivan. 2015. “Household Surveys in Crisis.” Journal of Economic Perspectives 29(4):199–226.

Milligan, K., and T. Schirle. 2019. “Push and Pull: Disability Insurance, Regional Labor Markets, and Benefit Generosity in Canada and the United States.” Journal of Labor Economics 37(Supplement 2):289–323.

Milligan, K., and T. Schirle. 2020. “Milligan, Schirle—A Labour Force Survey Pandemic Primer.” Intelligence Primer, 8 April. At https://www.cdhowe.org/intelligence-memos/milligan-schirle-%E2%80%93-labour-force-survey-pandemic-primer.

“Missing the Target—Plunging Response Rates to Household Surveys Worry Policymakers.” 2018. The Economist, 24 May.

Morin, L.-P. 2015. “Cohort Size and Youth Earnings: Evidence from a Quasi-Experiment.” Labour Economics 32(C):99–111.

Morissette, R., and A. Johnson. 2005. “Are Good Jobs Disappearing in Canada?” Economic Policy Review 11(1):23–56.

Pendakur, K., and R. Pendakur. 1998. “The Colour of Money: Earnings Differentials among Ethnic Groups in Canada.” Canadian Journal of Economics/Revue canadienne d’économique 31(3):518–48.

Pendakur, K., and R. Pendakur. 2002. “Colour My World: Have Earnings Gaps for Canadian-Born Ethnic Minorities Changed Over Time?” Canadian Public Policy/Revue canadienne d’économique 28(4):489–512.

Pérusse, D. 2008. Aboriginal People Living Off-Reserve and the Labour Market: Estimates from the Labour Force Survey, 2007. Aboriginal Labour Force Analysis Series, Cat. No. 71-588-X. Ottawa: Statistics Canada.

Portugal, P., and A.R. Cardoso. 2006. “Disentangling the Minimum Wage Puzzle: Accessions and Separations.” Journal of the European Economic Association 4(5):988–1013.

Qian, Y., and S. Fuller. 2020. “COVID-19 and the Gender Employment Gap among Parents of Young Children.” Canadian Public Policy/Analyse de politiques 46(Supplement 2):S89–101.

Riddell, W.C. 1993. “Unionization in Canada and the United States: A Tale of Two Countries.” In Small Differences That Matter: Labor Markets and Income Maintenance in Canada and the United States, ed. D. Card and R.B. Freeman, 109–48. Cambridge, MA: National Bureau of Economic Research.

Rivera Drew, J.A., S. Flood, and J.R. Warren. 2014. “Making Full Use of the Longitudinal Design of the Current Population Survey: Methods for Linking Records across 16 Months.” Journal of Economic and Social Measurement 39(3):121–44.

Sabelhaus, J., D. Johnson, S. Ash, D. Swanson, T.I. Garner, J. Greenlees, and S. Henderson. 2014. “Is the Consumer Expenditure Survey Representative by Income?” In Improving the Measurement of Consumer Expenditures, ed. C.D. Carroll, T.F. Crossley and J. Sabelhaus, 241–62. Cambridge, MA: National Bureau of Economic Research.

Schirle, T. 2008. “Why Have the Labor Force Participation Rates of Older Men Increased Since the Mid-1990s?” Journal of Labor Economics 26(4):549–94.

Schirle, T. 2015a. “The Effect of Universal Child Benefits on Labour Supply.” Canadian Journal of Economics/Revue canadienne d’économique 48(2):437–63.

Schirle, T. 2015b. “The Gender Wage Gap in the Canadian Provinces, 1997–2014.” Canadian Public Policy/Analyse de politiques 41(4):309–19.

Sheridan, M., D.J. Drew, and B. Allard. 1996. “Response Rate and the Canadian Labour Force Survey: Luck or Good Planning?” In Proceedings of Symposium 96, Nonresponse Errors, Statistics Canada, 67–75. Ottawa, 12-15 November.

Skuterud, M. 2005. Explaining the Increase in On-the-Job Search. Analytical Studies Branch Research Paper Series, Cat. No. 11F019MIE. Ottawa: Statistics Canada.

Skuterud, M., and M. Su. 2012. “Immigrants and the Dynamics of High-Wage Jobs.” Industrial and Labor Relations Review 65(2):377–97.

Smith, P.M., C. Cawley, A. Williams, and C. Mustard. 2020. “Male/Female Differences in the Impact of Caring for Elderly Relatives on Labor Market Attachment and Hours of Work: 1997–2015.” Journals of Gerontology, Series B: Psychological Sciences and Social Sciences 75(3):694–704.

Statistics Canada. 1975. Historical Labour Force Statistics—Actual Data, Seasonal Factors, Seasonally Adjusted Data, 1974. Cat. No. 71-201. Ottawa: Statistics Canada.

Statistics Canada. 1977a. Methodology of the Canadian Labour Force Survey 1976. Cat. No. 71-526. Ottawa: Statistics Canada.

Statistics Canada. 1977b. The Revised Labour Force Survey Documents. Cat. No. 71-601. Ottawa: Statistics Canada.

Statistics Canada. 1979. Guide to Labour Force Survey Data. Cat. No. 71-528. Ottawa: Statistics Canada.

Statistics Canada. 1984. Labour Force, Feature Article: Changes to the Labour Force Survey Effective January 1984. Cat. No. 71-001. Ottawa: Statistics Canada.

Statistics Canada. 1990. Methodology of the Canadian Labour Force Survey 1984–1990. Cat. No. 71-526. Ottawa: Statistics Canada.

Statistics Canada. 1991. Historical Labour Force Statistics, 1990. Statistics. Cat. No. 71-201. Ottawa: Statistics Canada.

Statistics Canada. 1993. 75 Years and Counting: A History of Statistics Canada. Ottawa: Statistics Canada.

Statistics Canada. 1995. Labour Force, Feature Article: Moving with the Times: Introducing Change to the LFS. Cat. No. 71-001. Ottawa: Statistics Canada.

Statistics Canada. 1998. Methodology of the Canadian Labour Force Survey. Cat. No. 71-526-XPB. Ottawa: Statistics Canada.

Statistics Canada. 1999. Commentary: Notes to Readers. Cat. No. 71-001-PPB. Ottawa: Statistics Canada.

Statistics Canada. 2005a. Aboriginal Peoples Living Off-Reserve in Western Canada: Estimates From the Labour Force Survey, April 2004–March 2005. Cat. No. 71-587-XIE. Ottawa: Statistics Canada.

Statistics Canada. 2005b. Improvements in 2005 to the Labour Force Survey (LFS). Cat. No. 71F0031XIE. Ottawa: Statistics Canada.

Statistics Canada. 2006. Labour Force Survey, July 2006. Cat. No. 11-001-XIE. Ottawa: Statistics Canada.

Statistics Canada. 2008a. Labour Force Survey: Annual Quality Report, January-December 2005. Cat. No. 71-0004. Ottawa: Statistics Canada.

Statistics Canada. 2008b. Methodology of the Canadian Labour Force Survey. Cat. No. 71-526-X. Ottawa: Statistics Canada.

Statistics Canada. 2011. Improvements to the Labour Force Survey (LFS), The 2011 Revisions of the Labour Force Survey (LFS). Cat. No. 71F0031X. Ottawa: Statistics Canada.

Statistics Canada. 2013. Siksika Nation Pilot Labour Force Survey: Collection Experience and Results 2010 to 2011. Labour Statistics Technical Paper. Cat. No. 75-005-M. Ottawa: Statistics Canada.

Statistics Canada. 2015. Improvements to the Labour Force Survey (LFS), The 2015 Revisions of the Labour Force Survey (LFS). Cat No. 71F0031X. Ottawa: Statistics Canada.

Statistics Canada. 2017a. Labour Force Survey, January 2017. Cat. No. 11-001-XIE. Ottawa: Statistics Canada.

Statistics Canada. 2017b. Methodology of the Canadian Labour Force Survey. Cat. No. 71-526-X. Ottawa: Statistics Canada.

Statistics Canada. 2017c. 2017 Changes to the Labour Force Survey Public Use Microdata. Ottawa: Statistics Canada.

Statistics Canada. 2018. Guide to the Labour Force Survey, 2018. Cat. No. 71-543-G. Ottawa: Statistics Canada.

Statistics Canada. 2019. Statistics Canada Quality Guidelines. Cat. No. 12-539-X. Ottawa: Statistics Canada.

Statistics Canada. 2020. Labour Force Survey, July 2020. Cat. No. 11-001-XIE. Ottawa: Statistics Canada.

Statistics Canada. 2021. Improvements to the Labour Force Survey (LFS): The 2021 Revisions of the Labour Force Survey (LFS). Cat No. 71F0031X. Ottawa: Statistics Canada.

Stephens Jr., M. 2002. "Worker Displacement and the Added Worker Effect." Journal of Labor Economics 20(3):504–37. https://doi.org/10.1086/339615.

Sunter, D., M. Kinack, E. Akyeampong, and D. Charrette. 1995. Redesigning the Canadian Labour Force Survey Questionnaire: Development and Testing. Cat. No. 71F0009XPE. Ottawa: Statistics Canada.

Sunter, D., M. Kinack, E. Akyeampong, and D. Charrette. 1997. The Labour Force Survey: Development of a New Questionnaire for 1997. Cat. No. 71F0022XPE. Ottawa: Statistics Canada.

Usalcas, J., and M. Kinack. 2017. History of the Canadian Labour Force Survey, 1945 to 2016. Labour Statistics Technical Papers. Cat. No. 75-005-M. Ottawa: Statistics Canada.

U.S. Census Bureau. 2006. Design and Methodology: Current Population Survey. Technical Paper 66. Washington, DC: U.S. Census Bureau.

U.S. Census Bureau. 2019. Design and Methodology: Current Population Survey–America's Source for Labor Force Data. Technical Paper 77. Washington, DC: U.S. Census Bureau.

Veall, M.R. 2010. "2B or not 2B? What Should Have Happened with the Canadian Long Form Census? What Should Happen Now?" Canadian Public Policy/Analyse de politiques 36(3):395–9.

"Why Do We Know So Little about Native Joblessness?" 2014. Globe & Mail, 20 October.

Williams, T. 2019. "What You Need to Know about the Census Citizenship Question." New York Times, 27 June.

## Appendix A: Occupation and Industry Classifications

In this appendix, I provide more detail on the occupation and industry updating.

**Occupation**

- As of January 1984, the Labour Force Survey (LFS) transitioned to the 1980 Standard Occupation Classification (SOC) from the 1971 census classification (Statistics Canada 1984). The impact of the change was minor, except for managerial occupations.
- As of January 1999, the LFS transitioned to the 1991 SOC from the 1980 SOC (Statistics Canada 1999). This was considered to be an important change, and it affected the provision of time-consistent series. See the General Comment below.
- As of January 2005, the LFS transitioned from the 1991 SOC to the National Occupational Classification for Statistics (NOC-S 2001) (Statistics Canada 2005b). Although some major groupings were affected, for the most part there were relatively few changes.
- As of January 2011, the LFS transitioned from the NOC-S 2001 to the NOC-S 2006 (Statistics Canada 2011). The changes were minor in scope.
- As of January 2015, the LFS transitioned from the NOC-S 2006 to the NOC 2011 (Statistics Canada 2015).
- As of January 2017, the LFS transitioned from the NOC 2011 to the NOC 2016 (Statistics Canada 2017a).
- As of January 2021, the LFS transitioned from the NOC 2016 to the NOC 2016 version 1.3 (Statistics Canada 2021).

General Comment: The LFS provides researchers who rely on CANSIM data with revised historical series going back in time. Researchers who rely on microdata are provided with occupation variables that are time consistent going back in time. In both cases, the series are discontinued if the changes in classification are considered too severe.

**Industry**

- As of January 1984, the LFS transitioned from the 1970 Standard Industry Classification (SIC) to the 1980 SIC (Statistics Canada 1984). The impact of the changes was minor because they were mainly confined to the major industry groups.
- As of January 1999, the LFS transitioned from the 1980 SIC to the 1997 North American Industry Classification System (NAICS) (Statistics Canada 1999). The changes were considered to be important and affected the provision of time-consistent series.
- As of January 2005, the LFS transitioned from the NAICS 1997 to the NAICS 2002 (Statistics Canada 2005b). The changes affected only the construction and information and cultural industries, as well as changes within other major groups.
- As of January 2011, the LFS transitioned from the NAICS 2002 to the NAICS 2007 (Statistics Canada 2011). The changes were minor and, as such, had little impact on the revised series.
- As of January 2015, the LFS transitioned from the NAICS 2007 to the NAICS 2012 (Statistics Canada 2015).
- As of January 2021, the LFS transitioned from the NAICS 2012 to the NAICS 2017 version 3.0 (Statistics Canada 2021).

General Comment: As with occupations, the LFS will discontinue an industry series if the changes in industry classification are too severe.

## Appendix B: Education

In this appendix, I provide more details on why it is problematic to construct consistent education series across the January 1990 threshold. I do so by providing the questions themselves and by highlighting the answers that trigger follow-up questions. I also provide a discussion of how an interviewer would record the answers because this provides additional insight into the approach to measuring education before and after the change.

Finally, I focus the discussion on the pen-and-paper era because it makes for a clean contrast. Recall that the computer-assisted approach to recording answers only started as of November 1993, approximately four years after the new education questions were introduced. This is because the Household Record Docket, the one-page form onto which the socio-demographic information was recorded, remained unchanged across the January 1990 threshold. The interviewer still only entered two one-digit codes for education in item 38 of the Household Record Docket. The one-digit code entered in column 1 (of item 38) still reflected primary and secondary education, whereas the one-digit code entered in column 2 (of item 38) still measured "other" education. What did change, however, was the meaning of the codes within each column.

### Pre-January 1990 Questionnaire

The pre-January 1990 (pen-and-paper) questionnaire has four education-related questions.

The first question asked, "How many years of primary and secondary education has . . . completed?" The interviewer would then enter the answer as a one-digit code in the first column of item 38 of the Household Record Docket: 0 = "no schooling," 1 = "1 to 8 years of primary and secondary education," 2 = "9 to 10 years of primary and secondary education," 3 = "11 years of primary and secondary education," 4 = "12 years of primary and secondary education," or 5 = "13 years of primary and secondary education."

The second question asked, "Has . . . taken any post-secondary education?" and if the respondent answers yes, they are then asked the follow-up question "Did this education normally require high school graduation?" Only if the respondent answered yes again were they asked the fourth, and final, question, "Did . . . receive, a degree, certificate or diploma?" The interviewer would then enter the post-secondary answers as a one-digit code in the second column of item 38 of the Household Record Docket: 1 = "some post-secondary education," 2 = "received a post-secondary certificate or diploma," or 3 = "received a university degree." It is important to recognize that for a person to have some post-secondary or higher education, they had to have received some education that normally required a high school degree, and this education had to be one that would provide (if completed) a degree, certificate, or diploma.

### January 1990 (and Onward) Questionnaire

The January 1990 (and onward) questionnaire asks five education-related questions. Most important, the focus is on educational attainment.

The first question asks, “What is the highest grade of elementary or high school (secondary school) . . . ever completed?” and if the respondent answers grades 11, 12, or 13, they are asked a follow-up question, “Did . . . graduate from high school (secondary school)?” The interviewer then enters the answer as a one-digit code in the first column of item 38 of the Household Record Docket: 0 = “grade 8 or lower (Quebec: Secondary II or lower),” 1 = “grade 9–10 (Quebec: Secondary III or IV),” 2 = “grade 11–13 (Quebec: Secondary V),” or 3 = “grade 11–13 (Quebec: Secondary V), graduated from high school.”

The third education-related question asks, “Has . . . received any other education?” If the respondent answers yes, they are asked “Could this education be counted towards a degree, certificate or diploma from an educational institution?” Only if the respondent answers yes again are they asked, “What is the highest degree, certificate or diploma . . . has obtained?” The interviewer then enters the post-secondary answers as a one-digit code in the second column of item 38 of the Household Record Docket: 1 = “no postsecondary degree, certificate or diploma,” 2 = “trade certificate or diploma from a vocational school or apprenticeship training,” 3 = “non-university certificate or diploma from a community college, CEGEP, school of nursing, etc.,” 4 = “university certificate below a bachelor’s level,” 5 = “bachelor’s degree,” or 6 = “university degree or certificate above bachelor’s level.” It is important to note that the new “other” education questions are not restricted to post-secondary education, although it still has to be education that counts toward a degree, certificate, or diploma from an educational institution.

As a final point, it should be noted that the LFS typically asks respondents about their education when they rotate into the survey (i.e., only during their first interview). This approach was followed both in the pre-January 1990 questionnaire period and the January (onward) 1990 questionnaire period. There was one exception to this rule. The January 1990 respondents who had been in the LFS in prior month(s) were asked the new education questions. As such, there is a clear break in the education questions as of January 1990.

Lavoie (1990) and Gower (1993) are the main sources of information for this appendix. See Gower (1993) for more details on how the changes to the education questions cause difficulties when it comes to constructing time-consistent education categories.