Microdata User Guide

Employment Insurance Coverage Survey

2020

Statistics Canada Statistique Canada

Canada

Centre for Labour Market Information

Contents

1.0 Introduction

The Employment Insurance Coverage Survey (EICS) was conducted by Statistics Canada with the cooperation and support of Employment and Social Development Canada. This manual has been produced to facilitate the use of the microdata and the interpretation of the survey results.

Any question about the data set or its use should be directed to:

Statistics Canada
Client Services
Centre for Labour Market Information
Telephone: 1-866-873-8788
E-mail: statcan.labour-travail.statcan@statcan.gc.ca

2.0 Background

The Employment Insurance Coverage Survey (EICS) was launched in 1997, primarily in response to a need to better understand the relationship between the number of persons in receipt of Employment Insurance (EI) benefits and the number of unemployed as reported by the Labour Force Survey.

The EI administrative data is limited with respect to the population covered and the variables available: information is available on accepted claims but not for disallowed claims or for non-claimants. The administrative data also lacks demographic and household information which is necessary for social analysis.

The survey results fill several of these data gaps and allow users to draw a comprehensive profile of the unemployed and other persons who may have been entitled to EI benefits due to a recent break in employment or a situation of underemployment.

The scope of the survey was broadened in 2000 to cover the access to maternity and parental benefits. These changes were implemented one year before the expansion of the parental benefits program in January 2001.

Prior to 2020, the target population for parents was mothers of infants aged one year or less during the reference week. Starting in 2020, the target population for parents was expanded to parents of infants aged 18 months or less to include fathers in households without a female parent and parents who have chosen to receive extended parental benefits.

In 2020, as a result of changes to the Employment Insurance program during the COVID-19 pandemic, only parents of infants aged 18 months or less were sampled.

3.0 Objectives

The primary objective of the Employment Insurance Coverage Survey (EICS) is to track the performance of the Employment Insurance (EI) program, by finding out how many people are covered by EI, what proportion of people receive benefits and which groups of people who may need EI do not get access to Employment Insurance. The Employment Insurance Coverage Survey also covers access to maternity and parental benefits.

The data are used to measure the coverage of the Canadian population by Employment Insurance and the role EI benefits play in contributing to personal and household income during periods of unemployment or underemployment. The unemployed as well as working individuals (e.g. beneficiaries with earnings) and those categorized as not in the labour force by the Labour Force Survey (LFS) are the objects of analysis under this topic. The latter two groups also receive Employment Insurance benefits in significant numbers.

The factors cited most frequently to explain variations in EI coverage are: not qualifying for EI, exhausting benefits, serving a waiting period after job separation, or not claiming EI. The magnitude of these and other factors and their correlation to personal characteristics, seasonal and business cycles, and regions of Canada can be investigated using this survey to improve our understanding of the reasons why some unemployed do not receive EI benefits.

Through the survey data, analysts will also be able to observe the characteristics and situation of people not covered by EI and of those who exhausted EI benefits, the job search intensity of the unemployed, expectation of recall to a job, and alternate sources of income and funds.

Survey data pertaining to maternity and parental benefits answer questions on the proportion of parents of an infant who received maternity or parental benefits, the reason why they don't and about sharing parental benefits with their spouse. The survey also allows looking at the timing and circumstances related to the return to work, the income adequacy of households with young children and more.

The Employment Insurance Coverage Survey

The survey was designed to produce a series of precise measures of the unemployed population in order to identify groups with low probability of receiving benefits. Such groups include:

- the long-term jobless;
- labour market entrants and students;
- people becoming unemployed after uninsured employment;
- people who have left jobs voluntarily; and
- individuals who are eligible, given their employment history, but do not claim or otherwise receive benefits

Employment Insurance coverage of the unemployed

The survey data were used to classify individuals as either "potentially eligible"' by EI or "not potentially eligible", based on information provided by respondents about their claiming and receiving of benefits, their perceived reasons for not receiving benefits or for not claiming, and their recent labour market history. The term "potentially eligible for Employment Insurance" is used here to describe unemployed people who, during the reference week, received EI benefits or were in a position to receive them because of their recent insurable employment and subsequent job loss. The term "not potentially eligible" describes the situation of those who did not receive benefits and could not have received them even if they had claimed them, as determined from the reported information.

The EICS provides an insight into the composition of the unemployed, particularly those not receiving Employment Insurance benefits during the period of a reference week. It provides a more meaningful picture of who does or does not have access to EI benefits than do beneficiary/unemployed (B/U) ratio indicators. The beneficiary/unemployed (B/U) ratio is calculated for a given week by dividing the number of regular EI beneficiaries by the total number of unemployed people.

4.0 Concepts and Definitions

This chapter outlines concepts and definitions of interest to the users. The concepts and definitions used in the Labour Force Survey (LFS) are described in Section 4.1 while those specific to the Employment Insurance Coverage Survey (EICS) are given in Section 4.2.

4.1 Labour Force Survey Concepts and Definitions

Labour Force Status

Designates the status of the respondent vis-à-vis the labour market: a member of the non-institutional population 15 years of age and over is either employed, unemployed or not in the labour force.

Employment

Employed persons are those who, during the reference week:

a) did any work[1] at all at a job or business; or
b) had a job but were not at work due to factors such as own illness or disability, personal or family responsibilities, vacation, labour dispute or other reasons (excluding persons on layoff, between casual jobs, and those with a job to start at a future date).

Unemployment

Unemployed persons are those who, during the reference week:

a) were on temporary layoff during the reference week with the expectation of recall and were available for work; or
b) were without work, had actively looked for work in the past four weeks, and were available for work[2] ; or
c) had a new job to start within four weeks from the reference week, and were available for work.

Not in the Labour Force

Persons not in the labour force are those who, during the reference week, were unwilling or unable to offer or supply labour services under conditions existing in their labour markets, that is, they were neither employed nor unemployed.

Industry and Occupation

The Labour Force Survey provides information about the occupation and industry attachment of employed and unemployed persons, and of persons not in the labour force who have held a job in the

[1] Work includes any work for pay or profit, that is, paid work in the context of an employer-employee relationship, or self-employment. It also includes unpaid family work, which is defined as unpaid work contributing directly to the operation of a farm, business or professional practice owned and operated by a related member of the same household. Such activities may include keeping books, selling products, waiting on tables, and so on. Tasks such as housework or maintenance of the home are not considered unpaid family work.

[2] Persons are regarded as available for work if they:

i. reported that they could have worked in the reference week if a suitable job had been offered; or if the reason they could not take a job was of a temporary nature such as: because of own illness or disability, personal or family responsibilities, because they already have a job to start in the near future, or because of vacation (prior to 1997, those on vacation were not considered available).
ii. were full-time students seeking part-time work who also met condition i) above. Full-time students currently attending school and looking for full-time work are not considered to be available for work during the reference week.

past 12 months. The industry coding corresponds to the North American Industry Classification System 2012 (NAICS 2012). Occupation codes are based on the National Occupational Classification (NOC) 2016.

For the EICS the industry coding corresponds to the North American Industry Classification System 2012 (NAICS 2012). Occupation codes are based on the National Occupational Classification (NOC) 2016.

Reference Week
The entire calendar week (from Sunday to Saturday) covered by the Labour Force Survey each month. It is usually the week containing the 15th day of the month. Data collection is conducted during the following week, called the Survey Week, and the labour force status determined is that of the reference week.

Full-time Employment
Full-time employment consists of persons who usually work 30 hours or more per week at their main or only job.

Part-Time Employment
Part-time employment consists of persons who usually work less than 30 hours per week at their main or only job.

4.2 Employment Insurance Coverage Survey Concepts and Definitions

Type
The EICS sample represents five types of distinct subpopulations of interest to researchers:

1) persons who were unemployed during the reference week;
2) persons employed part-time during the reference week;
3) persons not in the labour force during the reference week;
4) persons employed full-time during the reference week who started their current job during the previous two months;
5) parents of infants aged 18 months or less working during the reference week.

The type often determines which questions are asked in the survey.

Parents
In this survey, the term “parent” refers to mothers or fathers (by birth or adoption) of an infant aged 18 months or less during the LFS reference week. Many parents were not part of the survey sample prior to 2000. In particular, parents working full-time and parents not in the labour force and who have not worked in the past two years (or ever) were not included in the survey prior to 2000.

“Regular” population
Not the parent of an infant during the survey reference week (see definition of Parents above).

Original sample
Refers to the population targeted by the EICS before it was expanded to include all parents of an infant.

The original survey targeted people belonging in:

- Type 1 - same as current;
- Type 2 - including part-time parents;
- Type 3 - excluding parents who have not worked in two years; and
- Type 4 - including parents with a recent break in employment.

It is important to note that only the definition of Type 1 (the unemployed) has not changed since 1997.

Reference week
The sample used for this survey is selected from persons who have completed their participation in the LFS. Although interviews are done up to seven weeks after the LFS interviews, the reference week for the survey is the same as for the LFS.

Reference month
The reference month refers to the month which contains the reference week. This is the reference period for questions related to income.

Reference year
For "parents", the reference year is the 12-months prior to the birth or adoption of their child.
For the "regular" EICS population, the reference year is the 12-month period ending with the reference month.

Working during the reference week
Working during reference week refers to any work of an hour or longer duration performed for pay or profit.

Full-time/part-time employment
Full-time employment in this survey means that the persons usually work 30 hours or more per week in their job or jobs. Part-time employment consists of all other persons, that is, those who usually work less than 30 hours per week.

The LFS defines part-time work differently for multiple job holders: it applies the 30 hour criterion only to the main job.

Insurable employment
Refers to work that is insured by the Employment Insurance (EI) program against an interruption of earnings. Self-employment and some other types of employment are excluded. The survey identifies insurable employment based on the person having EI premiums deducted from their pay and the class of worker.

Employment Insurance Claimant
A claimant is a person who submitted an EI claim during a specified period.

Employment Insurance Beneficiary
A beneficiary is someone who, upon claiming EI benefits, qualifies and receives benefits for a particular period (for instance, the reference week, the reference month or since the last work interruption).

Potentially eligible for Employment Insurance
Term used in analysis to describe unemployed people who, during the reference week, received EI benefits or were in a position to receive them because of their recent insurable employment and subsequent job loss. This includes all unemployed persons with some insurable employment in the last 12 months who did not quit their job without cause or in order to return to school.

Eligible for Employment Insurance

This is a subset of the potentially eligible population. It includes people who received or expect to receive EI benefits in their current unemployment spell and individuals who have worked in a paid job in the year prior to losing or leaving their last job and likely accumulated enough hours to qualify for EI benefits.

Not potentially eligible for Employment Insurance
This group includes unemployed persons without insurable employment in the last 12 months and also persons who quit their job without cause or in order to return to school.

5.0 Survey Methodology

The Employment Insurance Coverage Survey (EICS) has been administered since 1997 to a sub-sample of the dwellings in the Labour Force Survey (LFS) sample, and therefore its sample design is closely tied to that of the LFS. The LFS design is briefly described in the Sections 5.1 to 5.5[3]. Sections 5.6 and 5.7 describe how the EICS departed from the basic LFS design.

5.1 Population Coverage

The Labour Force Survey (LFS) is a monthly household survey of a sample of individuals who are representative of the civilian, non-institutionalized population 15 years of age or older. It is conducted nationwide, in both the provinces and the territories.

Excluded from the survey's coverage are: persons living on reserves and other Indigenous settlements in the provinces, full-time members of the Canadian Armed Forces, the institutionalized population, and households in extremely remote areas with very low population density. These groups together represent an exclusion of approximately 2% of the population aged 15 and over.

National Labour Force Survey estimates are derived using the results of the LFS in the provinces. Territorial LFS results are not included in the national estimates, but are published separately.

5.2 Sample Design and Sample Size

Every ten years, after the decennial population census, the LFS undergoes a sample redesign to reflect changes in population characteristics and new definitions of geographical boundaries. From 2005 to 2014, the LFS sample design was based on information from the 2001 Census and reflected the population size, provincial distribution and the sub-provincial boundaries as of 2001.

Since that time, there has been significant population growth, change in population and labour market characteristics, as well as a realignment of municipal and census metropolitan area (CMA) boundaries. In January 2015, a new sample design was introduced, based on the 2011 Census information and geographical boundaries.

The LFS design strata, which are a way to divide the population in order to make sampling more efficient, are set out to be homogeneous with respect to some key labour market variables. An added benefit of stratification is to stabilize the sample size by areas. However, the strata become less efficient the further the design is from the source year (i.e., 2001 for the design in place from 2005 to 2014) and as the population and labour market characteristics shift over time.

The 2015 design defines new strata based on the most recent census information (2011), the National Household Survey (2011), as well as administrative data. The large majority of strata are constructed to improve the design's efficiency. A few strata are set aside to target specific sub-populations. High-income strata are constructed in most CMAs to group together areas with higher prevalence of high-income households. During LFS sampling, these special strata are not treated differently, but they are available for other surveys, such as the Survey of Household Spending. Similarly, Indigenous strata are constructed in regions with a higher prevalence of Indigenous persons in Saskatchewan, Alberta and British Columbia. These strata were constructed based on the requirements of external clients to have funded additional sample in the LFS to target special and Indigenous populations in these provinces.

[3] For comprehensive information on the LFS methodology see the publication Methodology of the Canadian Labour Force Survey, catalogue no. 71-526-X.

The sample is allocated to provinces and to strata within provinces in the way that best meets the need for reliable estimates at various geographic levels. Preliminary studies showed that, given the current sample size, the reliability guidelines used to allocate the sample in 2005 were no longer achievable. The guidelines for the 2015 design are therefore slightly different from those used in 2005.

5.3 Selection of Dwellings

A two-stage approach is used for all provinces except Prince Edward Island. Within each stratum, 'clusters' containing approximately 230 households are defined as small areas. During the first stage of sampling, a number of clusters, typically six, is selected from each stratum. For each selected cluster, a list of its dwellings is created. For most clusters, the list is extracted from the Dwelling Universe Frame created and maintained by Statistics Canada. For the remaining clusters, using a map of the cluster, an interviewer goes to the cluster and lists all dwellings. During the second stage of sampling, a sample of dwellings is selected from these lists.

Given the specific characteristics of Prince Edward Island, sampling is done in a single stage by using a complete list of addresses for all strata and selecting dwellings from this list.

5.4 Selection of Household Members

Demographic information is obtained for all persons in a household for whom the selected dwelling is the usual place of residence. Labour force information is obtained for all civilian household members 15 years of age or older. Respondent burden is minimized for persons aged 70 years and over by carrying forward their responses from the initial interview to the subsequent five months in survey.

5.5 Sample Rotation

The LFS follows a rotating panel sample design, in which households remain in the sample for six consecutive months. The total sample consists of six representative sub-samples or panels, and each month a panel is replaced after completing its six-month stay in the survey. Outgoing households are replaced by households in the same or a similar area. This results in a five-sixths month-to-month sample overlap, which makes the design efficient for estimating month-to-month changes. The rotation after six months prevents undue respondent burden for households that are selected for the survey.

5.6 Modifications to the Labour Force Survey Design for the Employment Insurance Coverage Survey

The EICS is collected in four cycles each year. For each cycle, the EICS uses the rotation group that has just completed its six months in the LFS. The EICS collection follows the LFS collection for the months of March, June, October and December. This sample is augmented by a second rotation for each cycle for parents of infants.

The survey estimates are produced for the reference year by averaging over the four cycles covered by the survey.

Note: Due to changes to Statistics Canada’s operations in response to COVID-19 public health measures, the April collection cycle was delayed to September for the 2020 reference year and the sample was limited to parents of infants aged 18 months or less.

5.6.1 Target Population

The target population for this survey is a subpopulation of the LFS and focuses on five groups (or types) of persons who are potential employment insurance recipients:

1) persons who were unemployed during the reference week;
2) persons employed part-time during the reference week;

3) persons not in the labour force during the reference week;
4) persons employed full-time during the reference week who started their current job during the previous two months;
5) parents of infants aged 18 months or less working during the reference week.

Of most relevance are the unemployed and the jobless, but part-time workers can also receive benefits, e.g. if they recently had an interruption in earnings and are entitled to retain Employment Insurance (EI) benefits while working due to small employment earnings.

One rotation group from the LFS typically includes approximately 5,500 individuals falling in one of the five target groups (out of a total sample of approximately 22,000 individuals aged 15 and over). Full-time employed and those not in the labour force during the reference week who have not worked for two years were the principal exclusions.

5.6.2 Type 4: A Special Case

Only the full-time employed (Type 4) who have experienced an interruption in work in the two months prior to the survey reference week need to be interviewed. Since this information was not available from the LFS interview, all full-time workers with short job tenure at their current job were selected. The question on work interruption was asked in the EICS and respondents who worked continually over the two months prior to the reference week were not asked further questions. They are out-of-scope for the survey and their records are dropped in processing (refer to Section 7.2). In a year, roughly 40% of those selected with Type = 4 are dropped for this reason.

5.6.3 Other Exclusions

In order to reduce the response burden within the household, a maximum of two persons per household are selected, with the exception of cases where three selected members of the household are all unemployed (the limit is then three).

A small number of records are excluded before sample selection begins: respondents with invalid names, phone numbers or employer names, imputed records from LFS, records denying Statistics Canada permission to call them, and respondents aged 70 and older.

5.7 Sample Size by Province for the Employment Insurance Coverage Survey

The following table shows the number of persons in the LFS sampled rotations that were selected in the EICS sample.

Province	**Sample Size**
	2020
Newfoundland and Labrador	31
Prince Edward Island	41
Nova Scotia	65
New Brunswick	52
Quebec	289
Ontario	440
Manitoba	176
Saskatchewan	140
Alberta	218
British Columbia	152
Canada	**1604**

6.0 Data Collection

6.1 Interviewing for the Labour Force Survey

Data collection for the LFS is carried out each month over the ten days following the LFS reference week. The reference week is normally the week containing the 15th day of the month.

Statistics Canada interviewers are employees hired and trained to carry out the LFS and other household surveys. Each month, they contact the sampled dwellings to obtain the required labour force information.

LFS interviews are conducted by telephone interviewers working out of regional office CATI (Computer Assisted Telephone Interview) sites. Personal visits from a field interviewer were suspended after March 2020 due to the COVID-19 Pandemic. The interviewer first obtains socio-demographic information for each household member and then obtains labour force information for all members aged 15 and over who are not members of the regular Canadian Armed Forces. Subsequent interviews are conducted by telephone. In subsequent monthly interviews, the interviewer confirms the socio-demographic information collected in the first month and collects the labour force information for the current month. Starting in 2015, respondents also have the option of completing the survey on-line for subsequent interviews.

In each dwelling, information about all household members is usually obtained from one knowledgeable household member. Such 'proxy' reporting, which accounts for approximately 65% of the information collected, is used to avoid the high cost and extended time requirements that would be involved in repeat visits or calls necessary to obtain information directly from each respondent.

If, during the course of the six months that a dwelling normally remains in the sample, an entire household moves out and is replaced by a new household, information is obtained about the new household for the remainder of the six-month period.

6.2 Supervision and Quality Control

All LFS interviewers are under the supervision of senior interviewers who are responsible for ensuring that their staff are familiar with the concepts and procedures of the LFS and its many supplementary surveys, as well as periodically monitoring their interviews. The senior interviewers are, in turn, under the supervision of the LFS program managers.

6.3 Non-response to the Labour Force Survey

Non-response to the LFS tends to average about 10% of eligible households. Interviewers are instructed to make all reasonable attempts to obtain interviews with members of eligible households. For individuals who at first refuse to participate in the LFS, a letter is sent from the Regional Office to the dwelling address stressing the importance of the survey and the household's co-operation. This is followed by a second call (or visit) from the interviewer. For cases in which the timing of the interviewer's call (or visit) is inconvenient, an appointment is arranged to call back at a more convenient time. For cases in which there is no one home, numerous call backs are made. Under no circumstances are sampled dwellings replaced by other dwellings for reasons of non-response.

Each month, after all attempts to obtain interviews have been made, a small number of non-responding households remain. A weight adjustment is applied to account for non-responding households.

6.4 Data Collection Modifications for the Employment Insurance Coverage Survey

Household members selected for the Employment Insurance Coverage Survey (EICS) are contacted within seven weeks of their last LFS interview. Households are contacted by letter, and are given the

option of completing the questionnaire online. Non-response follow up by telephone interviewers begins two weeks after online collection starts. Proxy response is not allowed in the EICS. There may be more than one person selected in each household, but never more than three.

6.5 Non-response to the Employment Insurance Coverage Survey

Similar to the LFS, the interviewers are asked to make all reasonable efforts to obtain the EICS interview. Refusals may be followed up by a senior interviewer. However, contrary to the LFS, no letters are sent to help obtain the respondent's cooperation.

7.0 Data Processing

The main output of the Employment Insurance Coverage Survey (EICS) is a "clean" microdata file. This chapter presents a brief summary of the processing steps involved in producing this file.

7.1 Data Capture

Responses to survey questions are captured directly by the questionnaire application at the time of the interview which reduces processing time and costs associated with data entry, transcription errors and data transmission. The response data are encrypted to ensure confidentiality and transferred over a secure network for further processing.

Some editing is done directly at the time of the interview. Where the information entered is out of range (too large or small) of expected values, or inconsistent with the previous entries, the respondent or interviewer is prompted, through message screens on the computer, to verify or modify the information. However, for most questions respondents or interviewers have the option of bypassing the edits, and of skipping questions if the respondent does not know the answer or refuses to answer. Therefore, the response data are subjected to further edit and imputation processes once they arrive in head office.

7.2 Verification and Editing

Electronic text files containing the daily transmissions of completed cases are combined to create the "raw" survey file. At the end of collection, this file should contain one record for each sampled individual. Before further processing, verification is performed to identify and eliminate potential duplicate records and to drop non-response and out-of-scope records.

There are a number of circumstances where respondents may be found out-of-scope of the EICS. By far, the majority of out-of-scope sampled cases are found among Type 4 respondents (refer to Section 5.6.2). A small number of other records are dropped after verifying the accuracy of the information used in sampling. Finally, a very small percentage of the sample is no longer in-scope of the EICS at time of the interview due to death, moving to an institution or moving outside of the country.

A criterion is defined for dropping non-response records. In the EICS, the respondent must have at least responded to the items required to derive the Employment Insurance (EI) coverage variable COV (refer to Section 7.5.4).

Editing consists in modifying the data at the individual variable level. The first step in editing is to determine which items from the survey output need to be kept on the survey master file. Subsequently, invalid characters are deleted and the data items are formatted appropriately. Text fields are stripped off the main files and written to a separate file for coding.

The first type of error treated was errors in questionnaire flow, where questions which did not apply to the respondent (and should therefore not have been answered) were found to contain answers. In this case a computer edit automatically eliminated superfluous data by following the flow of the questionnaire implied by answers to previous, and in some cases, subsequent questions. For skips based on answered questions, all skipped questions are set to "Valid skip" (6, 96, 996, etc.). For skips based on "Don't know" or "Refusal", all skipped questions are set to "Not stated" (9, 99, 999, etc.). The remaining empty items are filled with a numeric value (9, 99, 999, etc. depending on variable length). These codes are reserved for processing purposes and mean that the item was "Not stated".

There was no other type of editing or imputation done on questionnaire items. Therefore, some internal inconsistency may become apparent when conducting analysis. One notable example is the item on

hourly earnings (HRLYEARN) which does include a small percentage of outliers and internal inconsistency (working individuals reporting zero earnings).

7.3 Coding of Open-ended Questions

A few data items on the questionnaire were recorded in an open-ended format. In the EICS, the coding process assigns standard codes to the industry and occupation descriptions provided by the respondents (North American Industry Classification System (NAICS 2017), and the National Occupational Classification (NOC 2016)) and to the country of birth. Also, "Other, specify" fields with a significant number of text answers were examined and coded to existing categories. In some occasions, new categories were created to facilitate the analyses of the textual information. These were items relating to reasons for not claiming or receiving benefits, industry, occupation, or reason for not returning to work.

7.4 Imputation

Imputation is the process that supplies valid values for those variables that have been identified for a change either because of invalid information or because of missing information. The new values are supplied in such a way as to preserve the underlying structure of the data and to ensure that the resulting records will pass all required edits. In other words, the objective is not to reproduce the true microdata values, but rather to establish internally consistent data records that yield good aggregate estimates.

We can distinguish between three types of non-response. Complete non-response is when the respondent does not provide the minimum set of answers. These records are dropped and accounted for in the weighting process (see Chapter 11.0). Item non-response is when the respondent does not provide an answer to one question, but goes on to the next question. These are usually handled using the "not stated" code or are imputed. Finally, partial non-response is when the respondent provides the minimum set of answers but does not finish the interview. These records can be handled like either complete non-response or multiple item non-response.

Imputation was used to eliminate or reduce missing information caused by application problems in 2000 and 2001. This procedure was not repeated in subsequent years. Users will find item specific information in the notes included in the survey master file data dictionary.

There was no imputation done for the 2020 Employment Insurance Coverage Survey.

7.5 Creation of Derived Variables

A large number of data items on the microdata file have been derived by combining items on the questionnaire in order to facilitate data analysis. All items on the microdata file were given a short name that abbreviates the variable description (in English).

There are several types of derived variables on the data file. This section provides general information about each type of derived variables. The data dictionary available for the microdata file includes a note that identifies all questionnaire items used to create each derived variable.

7.5.1 Grouping of Continuous Data Item

Most data items collected as continuous variables are only included on the public use microdata file as grouped variables. Examples of such items are the age of the respondent (AGECAT), job tenure (TENURE_G), and the number of weeks worked during the reference year (WEEKSCAT).

In other situations, categorical response items were regrouped to create meaningful categories or to reduce the risk of identifying individuals with unique sets of answers. This is the case for highest level of

educational attainment (EDUC), industry and occupation (NAICS6 and NOC6), planned or current childcare arrangement (CHLDCARE), parental benefits plan or option chosen (BENPLAN – parents only), type of economic family (EFAMILY) and a few others.

7.5.2 Combining Data from the Labour Force Survey and the Employment Insurance Coverage Survey

Questions related to the employer and employment conditions were only asked in the EICS if the information was not available from the Labour Force Survey (LFS). In the LFS, these questions relate to the current job, or, for some items, to the previous job if held in the previous year. The EICS is looking for this information for all respondents who worked in the previous two years. Generally, the variable name used in the LFS microdata file was used (FTPT, HRLYEARN). Many of these employment related variables were grouped for the EICS public use microdata file.

7.5.3 Combining Two or More Different Questions

Variables such as union status (UNIONCA), type of work arrangement (WRKTYP), made a claim for EI in the last 12 months or since the month last worked (CLAIM), received EI benefits (BENEFIT), reason did not receive or claim EI benefits for the reference week or since birth/adoption (RNBENRW), received additional payments from employer (ADDPAYM), and looking for work within community or province (LOOKOUT) are derived using more than one questionnaire item.

In these cases, the algorithm used to create the new variable is usually fairly intuitive. For instance, the variable on type of work arrangements is created by combining full-time or part-time status, permanent or temporary employment status and reason for temporary employment and class of worker as follows:

Full-time or part-time status (FTPT)
Coverage: Paid employees at last or current job
1 Full-time
2 Part-time

Permanent or temporary job status (PERMTEMP) (only available on survey Master file)
Coverage: Paid employees at last or current job
1 Permanent
2 Not permanent, seasonal job
3 Not permanent, temporary, term or contract job
4 Not permanent, casual job
5 Not permanent, work done through a temporary help agency
6 Not permanent, other

Class of worker at main job (COW)
Coverage: Respondents who ever worked
1 Public or private employee
2 Self-employed incorporated/unincorporated (with/without employees)
3 Private, unpaid family worker

Type of work arrangement (WRKTYP) (derived variable)
Coverage: Respondents who ever worked
01 Permanent, full-time worker (FTPT = 1 and PERMTEMP = 1)
02 Permanent, part-time worker (FTPT = 2 and PERMTEMP = 1)
03 Permanent, work hours unknown (FTPT = 9 and PERMTEMP = 1)

04 Not permanent, seasonal worker (PERMTEMP = 2)
05 Not permanent, other (PERMTEMP = 3, 4 or 5)
06 Self-employed (COW = 2)

Other derived variables are created using more complex rules. This is the case of COV, a derived variable created to establish coverage of the EI program.

7.5.4 Taxonomy of Employment Insurance Coverage: the COV Variable

The EICS provides information on the situation of non-working individuals relative to EI benefits. It is a survey and not an administrative data source. The EI administrative data represents the actual decisions of Employment Insurance agents about benefit claims received by Employment and Social Development Canada (ESDC). On the other hand, in the EICS, estimates of the degree of coverage of the Canadian population by the EI program are made on the basis of behaviours, events and perceptions reported by respondents in a household survey.

The following is a description of the logic of the taxonomy used by ESDC in reporting EI coverage of the unemployed. The categories of coverage were determined in a hierarchical order described below.

The first four categories are mutually exclusive and regroup all respondents who have received benefits since they last worked or expected to receive benefits for the reference week when interviewed. Some respondents in these four groups have left their job, returned to school, were self-employed in their last job or without work for more than one year. Despite these circumstances, the fact that they have received EI benefits in the past year clearly establishes their eligibility.

COV = 1	Respondent received regular EI benefits in the reference week (using BENEFIT and BENTYP).
COV = 2	Respondent received special EI benefits in the reference week (using BENEFIT and BENTYP).
COV = 3	Respondent did not receive benefits during the reference week but expects to receive benefits in the non-working period (using BENEFIT and RNBENRW). Persons are considered to be in a position of receiving benefits when they indicate that they claimed EI benefits and say that they did not receive EI benefits during the reference week but are: still expecting benefit payments for that week, or are serving a waiting period, or benefits are being withheld due to severance or other payments or other reasons.
COV = 4	Respondent did not receive benefits for the reference week but received some EI benefits since he/she last worked in the last 12 months.

The taxonomy of the EI coverage then goes on to identify respondents who did not contribute to EI and therefore are <u>not potentially eligible for EI</u>.

COV = 12	Respondent has never worked.
COV = 11	Respondent last worked more than 12 months ago.
COV = 10	Respondent was not a paid employee in their last job or stated that they did not contribute to EI in their last job (using WRKTYP and RNBENRW).

The classification continues with the remaining respondents who contributed to EI but are not potentially eligible because of their reason for leaving their last job.

COV = 9	Respondent reported not claiming or receiving benefits because they went to school or gave their reason for leaving their last job as going to school (using RNBENRW).
COV = 8	Respondent reported not claiming or receiving benefits because they quit their last job voluntarily and other respondents who indicated that they quit their last job.

For the remaining respondents (about one in seven unemployed individuals) the main task was to determine EI eligibility based on hours worked in the year preceding the interruption of work.

The last three categories in the taxonomy of COV rest largely (but not exclusively) on a survey based estimate of insurable hours worked in the previous year. This estimate takes into consideration the number of weeks worked in that year, the number of weekly hours worked on average when working full-time and hours worked on average when working part-time. Usual hours worked in the most recent job or average hours worked for all part-timers and full-timers are used in case of non-response. The entrance criterion is set at 700 hours for all, the highest entrance criteria across the country.

COV = 7	Respondent reported not claiming or receiving EI benefits because of a lack of sufficient hours of insurable work or because they had no recent work (using RNBENRW). Respondents whose tenure at the last job was less than or equal to three months since no information is available on the insurability of the hours worked at previous jobs within the year (could have been self-employment or other uninsured employment) (using TENURE_G). Survey estimate of insurable hours is less than 700 hours.
COV = 5	Survey estimate of insurable hours is 700 or greater but respondent did not claim EI benefits.
COV = 6	Survey estimate of insurable hours is 700 or greater and respondent claimed EI benefits (did not receive).

This concludes the definition of COV. The derived variable ELIGIBLE summarises COV as follows:

1 Potentially eligible, eligible (COV = 1 to 6)

2 Potentially eligible, not eligible (COV = 7)

3 Not potentially eligible (COV = 8 to 12).

The main measure of EI coverage published from this survey expresses the estimate of eligible (ELIGIBLE = 1) as a percentage of potentially eligible (ELIGIBLE = 1 or 2).

7.6 Weighting

The principle behind estimation in a probability sample such as the LFS is that each person in the sample "represents", besides himself or herself, several other persons not in the sample. For example, in a simple random 2% sample of the population, each person in the sample represents 50 persons in the population.

The weighting phase is a step which calculates, for each record, what this number is. This weight appears on the microdata file, and must be used to derive meaningful estimates from the survey. For example if the number of parents receiving maternity or parental benefits is to be estimated, it is done by selecting the records referring to those individuals in the sample with that characteristic and summing the weights entered on those records.

Details of the method used to calculate these weights are presented in Chapter 11.0.

7.7 Suppression of Confidential Information

Statistics Canada is prohibited by law from releasing any data which would divulge information obtained under the Statistics Act that relates to any identifiable person, business or organization without the prior knowledge or the consent, in writing, of that person, business or organization. Various confidentiality rules are applied to all data that are released or published to prevent the publication or disclosure of any information deemed confidential. If necessary, data are suppressed to prevent direct or residual disclosure of identifiable data. It should be noted that the "Public Use" Microdata Files (PUMF) may differ from the survey "master" files held by Statistics Canada. These differences usually are the result of actions taken to protect the anonymity of individual survey respondents. The most common actions are the suppression of data items and grouping values into wider categories. For certain variables that are susceptible to identifying individuals, the PUMF may have been treated with local suppression, that is, some of the values in the master file may have been coded as "not stated" on the PUMF.

The survey master file includes geographic identifiers for the 10 provinces and for the EI economic regions. The PUMF does not contain any geographic identifiers below the provincial level and some provinces were grouped (i.e., Atlantic region and Manitoba with Saskatchewan). Grouping of provinces was done to avoid excessive data suppression of useful variables.

The survey master file includes the respondent's precise age while the PUMF contains age groups only. Similarly, ungrouped variables such as industry and occupation, job tenure, number of months since last worked, and age of the baby in months (parents only) are only available on the survey master file.

Users of the PUMF requiring access to information excluded from the microdata files, custom tabulations may be purchased. Estimates generated will be released to the user, subject to meeting the guidelines for analysis and release outlined in Chapter 9.0 of this document.

8.0 Data Quality

8.1 Response Rates

The following tables summarize the number of in-scope persons, number of respondents and resulting response rate to the Employment Insurance Coverage Survey (EICS).

Province	2020		
	In-scope Sample	**Response**	**Response Rate (%)**
Newfoundland and Labrador	31	18	58
Prince Edward Island	40	20	50
Nova Scotia	64	44	69
New Brunswick	51	29	57
Quebec	286	193	67
Ontario	426	288	68
Manitoba	167	112	67
Saskatchewan	135	96	71
Alberta	214	148	69
British Columbia	144	92	64
Canada	**1558**	**1040**	**67**

Note: The EICS response rate is the number of EICS responding individuals as a percentage of the number of EICS selected individuals in-scope (refer to Sections 5.6.2 and 7.2).

8.2 Survey Errors

The estimates derived from this survey are based on a sub-sample of individuals from the Labour Force Survey. Somewhat different estimates might have been obtained if a complete census had been taken using the same questionnaire, interviewers, supervisors, processing methods, etc. as those actually used in the survey. The difference between the estimates obtained from the sample and those resulting from a complete count taken under similar conditions, is called the sampling error of the estimate.

Errors which are not related to sampling may occur at almost every phase of a survey operation. Interviewers may misunderstand instructions, respondents may make errors in answering questions, the answers may be incorrectly entered on the questionnaire and errors may be introduced in the processing and tabulation of the data. These are all examples of non-sampling errors.

Over a large number of observations, randomly occurring errors will have little effect on estimates derived from the survey. However, errors occurring systematically will contribute to biases in the survey estimates. Considerable time and effort were taken to reduce non-sampling errors in the survey. Quality assurance measures were implemented at each step of the data collection and processing cycle to monitor the quality of the data. These measures include the use of highly skilled interviewers, extensive training of interviewers with respect to the survey procedures and questionnaire, observation of interviewers to detect problems of questionnaire design or misunderstanding of instructions, procedures to ensure that data capture errors were minimized, and coding and edit quality checks to verify the processing logic.

8.2.1 The Frame

Because the EICS was a supplement to the Labour Force Survey (LFS), the frame used was the LFS sample. Any non-response to the LFS had an impact on the EICS frame. The quality of the sampling variables in the frame was very high. The 2020 EICS sample consisted of two rotation groups from the LFS for the "parent" population.

Note that the LFS frame excludes about 2% of all households in the 10 provinces of Canada. Therefore, the EICS frame also excludes the same proportion of households in the same geographical area. It is unlikely that this exclusion introduces any significant bias into the survey data. The EICS frame also excludes full non-response to the LFS and item non-response to variables used in the selection criteria.

The variables on the EICS frame were quite up-to-date since they were collected from the LFS at most three weeks before the beginning of the EICS collection.

8.2.2 Data Collection

Interviewer training consisted of reading the EICS Interviewer's Manual, practicing with the EICS training cases on the computer, and discussing any questions with senior interviewers before the start of the survey. A description of the background and objectives of the survey was provided, as well as a glossary of terms and a set of questions and answers. Interviewers started collecting the EICS information two weeks after the end of the March, June, October, and December LFS collection periods, in July, September, November, and January respectively. Collection lasted approximately five weeks for each EICS cycle.

In 2020, due to COVID-19 related shutdowns in March and April, the EICS used the rotation groups that completed their six months in the LFS in June, August, October and December.

8.2.3 Data Processing

Data processing of the EICS was done in a number of steps including verification, coding, editing, estimation, confidentiality, etc. At each step a picture of the output files is taken and a report showing changes to each variable from one step to the other is created. The verification of these processing reports greatly reduces the risk of introducing errors in the data at the processing stage.

Verification

Electronic text files containing the daily transmissions of completed cases are combined to create the "raw" survey file. All EICS records could be matched to their corresponding record from the LFS and no records were lost or dropped.

Duplicate records are sometimes created due to transmission problems. When this happens, one of two identical records is dropped or, if the duplicates are not absolutely identical, the record with the most information is kept. In the EICS, duplicates were rarely found.

Editing

Editing consists of modifying the data at the individual variable level. The main type of editing carried out for the EICS data is called "flow" edits (refer to Section 7.2). The reports produced by the flow edit system were thoroughly examined to detect potential errors introduced in processing. This examination focussed on items with high incidence of "Not stated" answers and items where a valid answer was changed to a "Valid skip" or "Not stated". Very few situations could not be explained. The verification process however revealed a number of response errors (refer to Section 8.2.4).

Coding

Industry and occupation were coded by a specially trained group of people, which helped reduce the risk of coding errors. Items unique to this survey are likely more subject to coding errors or inconsistent coding from year to year. No specific measure of coding errors is available.

Derived Variables
A large number of derived variables were created from the EICS collected data. The distribution for each derived variable was compared to that of the questionnaire items used in creating it. A comparison of the distribution between the 2019 and 2020 period was also conducted to ensure historical comparability.

As a result of the changes in the survey population and questionnaire (Section 5.6), caution should be used when making comparisons with data from previous survey years.

8.2.4 Response error for question BK_Q05A

Response issues to question BK_Q05A ("Have you **taken a break** from working of a week or more **[since the birth or adoption of your child]**?") were identified for parents with children aged 12 to 18 months. Some of these parents who indicated that they had not taken a break from work also reported that they had received maternity and/or parental benefits and had not worked while receiving these benefits. As BK_Q05A is used to help determine duration of leave for parents and as a flow condition for the "Work after birth (WA)" block, there was an artificial increase in the number of parents with a duration of leave of 0 weeks and decrease in the number of responses received for the WA block.

To maintain consistency in the survey data, responses and duration of leave derived variables (ALLEAVE, LEAVECAT) were not adjusted; however, it is recommended to use the number of weeks of benefits received as a proxy for duration of leave for these respondents.

In SAS, the following code can be used to create a derived variable for the duration of leave (DURLEAVE) on the master file:

```
/* If respondent is a parent with a child aged 12 months or older
      who did not take a break and did not work while receiving benefits,
      duration of leave is the number of weeks of benefits received.*/
if PARENT = 1 and CHLDMTHS >= 12 and BK_05A = 2 and BE_75 = 2
      then DURLEAVE = BENWEEKS;
else DURLEAVE = ALLEAVE;
```

8.2.5 Non-response

A major source of non-sampling errors in surveys is the effect of non-response on the survey results. The extent of non-response varies from partial non-response (failure to answer just one or some questions) to total non-response.

Total non-response occurred because the interviewer was either unable to contact the respondent or the respondent refused to participate in the survey.
Total non-response was handled by adjusting the weight of individuals who responded to the survey to compensate for those who did not respond.

In most cases, partial (item) non-response to the survey occurred when the respondent did not understand or misinterpreted a question, refused to answer a question, or could not recall the requested information.

There was no imputation of data to compensate for total or item non-response in the EICS.

8.2.6 Measurement of Sampling Error

Since it is an unavoidable fact that estimates from a sample survey are subject to sampling error, sound statistical practice calls for researchers to provide users with some indication of the magnitude of this sampling error. This section of the documentation outlines the measures of sampling error which Statistics Canada commonly used and which it urges users producing estimates from this microdata file to use also.

The basis for measuring the potential size of sampling errors is the standard error of the estimates derived from survey results.

However, because of the large variety of estimates that can be produced from a survey, the standard error of an estimate is usually expressed relative to the estimate to which it pertains. This resulting measure, known as the coefficient of variation (CV) of an estimate, is obtained by dividing the standard error of the estimate by the estimate itself and is expressed as a percentage of the estimate.

9.0 Guidelines for Tabulation, Analysis and Release

This chapter of the documentation outlines the guidelines to be adhered to by users tabulating, analyzing, publishing or otherwise releasing any data derived from the survey microdata files. With the aid of these guidelines, users of microdata should be able to produce the same figures as those produced by Statistics Canada and, at the same time, will be able to develop currently unpublished figures in a manner consistent with these established guidelines.

9.1 Rounding Guidelines

In order that estimates for publication or other release derived from these microdata files correspond to those produced by Statistics Canada, users are urged to adhere to the following guidelines regarding the rounding of such estimates:

a) Estimates in the main body of a statistical table are to be rounded to the nearest hundred units using the normal rounding technique. In normal rounding, if the first or only digit to be dropped is 0 to 4, the last digit to be retained is not changed. If the first or only digit to be dropped is 5 to 9, the last digit to be retained is raised by one. For example, in normal rounding to the nearest 100, if the last two digits are between 00 and 49, they are changed to 00 and the preceding digit (the hundreds digit) is left unchanged. If the last digits are between 50 and 99 they are changed to 00 and the preceding digit is incremented by 1.

b) Marginal sub-totals and totals in statistical tables are to be derived from their corresponding unrounded components and then are to be rounded themselves to the nearest 100 units using normal rounding.

c) Averages, proportions, rates and percentages are to be computed from unrounded components (i.e. numerators and/or denominators) and then are to be rounded themselves to one decimal using normal rounding. In normal rounding to a single digit, if the final or only digit to be dropped is 0 to 4, the last digit to be retained is not changed. If the first or only digit to be dropped is 5 to 9, the last digit to be retained is increased by 1.

d) Sums and differences of aggregates (or ratios) are to be derived from their corresponding unrounded components and then are to be rounded themselves to the nearest 100 units (or the nearest one decimal) using normal rounding.

e) In instances where, due to technical or other limitations, a rounding technique other than normal rounding is used resulting in estimates to be published or otherwise released which differ from corresponding estimates published by Statistics Canada, users are urged to note the reason for such differences in the publication or release document(s).

f) Under no circumstances are unrounded estimates to be published or otherwise released by users. Unrounded estimates imply greater precision than actually exists.

9.2 Sample Weighting Guidelines for Tabulation

The sample design used for the Employment Insurance Coverage Survey (EICS) was not self-weighting. When producing simple estimates including the production of ordinary statistical tables, users must apply the proper survey weights.

If proper weights are not used, the estimates derived from the microdata files cannot be considered to be representative of the survey population, and will not correspond to those produced by Statistics Canada.

Users should also note that some software packages may not allow the generation of estimates that exactly match those available from Statistics Canada, because of their treatment of the weight field.

9.3 Definitions of Types of Estimates: Categorical and Quantitative

Before discussing how the EICS data can be tabulated and analyzed, it is useful to describe the two main types of point estimates of population characteristics which can be generated from the microdata file for the EICS.

9.3.1 Categorical Estimates

Categorical estimates are estimates of the number, or percentage of the surveyed population possessing certain characteristics or falling into some defined category. The number of unemployed who received Employment Insurance (EI) benefits during the reference week or the proportion of the unemployed eligible for EI benefits are examples of such estimates. An estimate of the number of persons possessing a certain characteristic may also be referred to as an estimate of an aggregate.

Examples of Categorical Questions:

Q: Were Employment Insurance premiums deducted from your wages or salary at that job with (employer name)?
R: Yes / No

Q: What type of benefits did you receive that week?
R: Training / Regular / Maternity (only if female) / Parental / Sickness / Fishing / Other

9.3.2 Quantitative Estimates

Quantitative estimates are estimates of totals or of means, medians and other measures of central tendency of quantities based upon some or all of the members of the surveyed population. They also specifically involve estimates of the form $\hat{X}/\hat{Y}$ where $\hat{X}$ is an estimate of surveyed population

quantity total and $\hat{Y}$ is an estimate of the number of persons in the surveyed population contributing to that total quantity.

An example of a quantitative estimate is the average number of months of leave taken from work after the birth or adoption of a child. The numerator is an estimate of the total number of months of leave taken by all parents for whom the information is available (returned to work already or know plans) and its denominator is the number of parents taking leave of a known duration.

Examples of Quantitative Questions:

Q: How long was this break from working, in terms of months?
R: |_|_| months

Q: During the weeks that you worked full-time, how many hours on average did you work per week?
R: |_|_|_| hours

9.3.3 Tabulation of Categorical Estimates

Estimates of the number of people with a certain characteristic can be obtained from the microdata file by summing the final weights of all records possessing the characteristic(s) of interest. Proportions and ratios of the form $\hat{X}/\hat{Y}$ are obtained by:

a) summing the final weights of records having the characteristic of interest for the numerator ($\hat{X}$),
b) summing the final weights of records having the characteristic of interest for the denominator ($\hat{Y}$), then
c) dividing estimate a) by estimate b) ($\hat{X}/\hat{Y}$).

9.3.4 Tabulation of Quantitative Estimates

Estimates of quantities can be obtained from the microdata file by multiplying the value of the variable of interest by the final weight for each record, then summing this quantity over all records of interest. For example, to obtain an estimate of total number of weeks of Employment Insurance (EI) received by parents of an infant who have already returned to work, multiply the value reported in derived variable BENWEEKS (weeks received EI) by the final weight for the record, then sum this value over all records with PARENT = 1 and WORKNOW = 1 (parents of an infants aged 18 months or less who are currently working).

To obtain a weighted average of the form $\hat{X}/\hat{Y}$, the numerator ($\hat{X}$) is calculated as for a quantitative estimate and the denominator ($\hat{Y}$) is calculated as for a categorical estimate. For example, to estimate the average number of weeks EI was received by parents,

a) estimate the total number of weeks ($\hat{X}$) as described above,
b) estimate the number of parents currently working ($\hat{Y}$) in this category by summing the final weights of all records with PARENT = 1 and WORKNOW = 1, then
c) divide estimate a) by estimate b) ($\hat{X}/\hat{Y}$).

9.4 Guidelines for Statistical Analysis

The EICS is based upon a complex sample design, with stratification, multiple stages of selection, and unequal probabilities of selection of respondents. Using data from such complex surveys presents problems to analysts because the survey design and the selection probabilities affect the estimation and variance calculation procedures that should be used. In order for survey estimates and analyses to be free from bias, the survey weights must be used.

While many analysis procedures found in statistical packages allow weights to be used, the meaning or definition of the weight in these procedures may differ from that which is appropriate in a sample survey framework, with the result that while in many cases the estimates produced by the packages are correct, the variances that are calculated are poor.

The calculation of more precise variance estimates requires detailed knowledge of the design of the survey. Such detail cannot be given in this microdata file because of confidentiality. Variances that take the complete sample design into account can be calculated for many statistics by Statistics Canada on a cost-recovery basis. The method available to approximate the true variance is to use a replication method, namely the bootstrap method. This method is known to correctly approximate the true value of the variance. A file containing 1,000 bootstrap weights is available. Variance calculation using 1,000 bootstrap weights involves calculating the estimates with each of these 1,000 weights and then, calculating the variance of these 1,000 estimates.

The user of the EICS PUMF may also compute their own variance estimates using an accompanying set of bootstrap weights that is generated in a way that is conducive to maintaining confidentiality[4]. These PUMF bootstrap weights can be used to compute variance estimates for totals, proportions, and ratios, as well as parameters used in other analysis techniques such as linear regression, logistic regression, and analysis of variance. See Section 10.0 for a detailed description and examples.

Due to the reduction in sample size in 2020, the bootstrap weights used to produce official estimates and the bootstrap weights accompanying the PUMF are both generated using the generalized bootstrap method, which only requires a random adjustment generated from a probability distribution. Traditionally the bootstrap weights used to produce EICS official estimates have been generated using the Rao Wu Yue re-scaling method.

9.5 Coefficient of Variation Release Guidelines

Before releasing and/or publishing any estimates from the EICS, users should first determine the quality level of the estimate. The quality levels are acceptable, marginal and unacceptable. Data quality is affected by both sampling and non-sampling errors as discussed in Chapter 8.0. However for this purpose, the quality level of an estimate will be determined only on the basis of sampling error as reflected by the coefficient of variation as shown in the table below. Nonetheless users should be sure to read Chapter 8.0 to be more fully aware of the quality characteristics of these data.

First, the number of respondents who contribute to the calculation of the estimate should be determined. If this number is less than 5, the weighted estimate should be considered to be of unacceptable quality.

For weighted estimates based on sample sizes of 5 or more, users should determine the coefficient of variation of the estimate and follow the guidelines below. These quality level guidelines should be applied to rounded weighted estimates.

All estimates can be considered releasable. However, those of marginal or unacceptable quality level must be accompanied by a warning to caution subsequent users.

Quality Level Guidelines

Quality Level of Estimate	**Guidelines**
1) Acceptable	Estimates have a sample size of 5 or more, and low coefficients of variation in the range of 0.0% to 15%. No warning is required.
2) Marginal	Estimates have a sample size of 5 or more, and high coefficients of variation in the range of 15% to 35%. The following warning should accompany the estimates: "Use with caution"
3) Unacceptable	Estimates have a sample size of less than 5, or very high coefficients of variation in excess of 35%. Statistics Canada recommends not to release estimates of unacceptable quality. The following warning should accompany the estimates: "Too unreliable to be published"

10.0 Bootstrap Weights

The bootstrap weights file provided with the EICS analytical file contains 1,000 sets of bootstrap weights stored in columns. The columns are named WRPMb, where $b = 1, \ldots, 1000$. One set of bootstrap weights is called a replicate, so each column represents one replicate. In general, variance estimation

using bootstrap weights is done by computing the estimate of a parameter using each set of bootstrap weights and then computing the variance of the resulting estimates. This method is valid not only for totals, means, and ratios, but also for non-linear parameters such as medians and other quantiles. These bootstrap weights can be used to compute a variety of variance estimates based on EICS data. In this section, some examples are provided to illustrate the potential uses of the bootstrap weights.

Note that the PUMF bootstrap weights were generated in such a way to ensure confidentiality. Therefore, while the results of computations using the PUMF bootstrap weights will be comparable to the results using the official bootstrap weights, they will not be the same. Estimates of variance for specific variables using the official bootstrap weights may be obtained from Statistics Canada on a cost-recovery basis.

10.1 How to Use the Bootstrap Weights for Categorical Estimates

Generally, to use bootstrap weights to estimate the variance of a parameter estimate, one starts with an analysis file containing the variables of interest and any domain variables in addition to the bootstrap weights. Such a file can be derived, for example, by merging the analytical file and the bootstrap weights file. Once the analysis file is created, the following steps can be used to compute the bootstrap variance:

1. Compute the survey estimate, $\hat{X}$, using the variable of interest and the final survey weights.
2. Compute a bootstrap estimate, $\hat{X}^{*(b)}$, for each bootstrap replicate b, by using the variable of interest and the bootstrap weights for each replicate. Since the EICS provides 1,000 replicates, this step should result in the 1,000 bootstrap estimates, $\hat{X}^{*(1)}, \ldots, \hat{X}^{*(1000)}$.
3. Compute the bootstrap variance of the 1,000 bootstrap estimates using the formula:

$$\widehat{\mathrm{var}}(\hat{X}) = \sum_{b=1}^{1000} \left(\hat{X}^{*(b)} - \hat{X}\right)^2 / 1000, \tag{1}$$

where $\hat{X}$ is the parameter estimate computed using the final weights and $\hat{X}^{*(b)}$ is the bootstrap estimate for replicate b. If estimates are needed by domain, such as province or age group, the above steps would be implemented separately for each level of the desired domain.

4. Compute the estimated coefficient of variation $\widehat{\mathrm{cv}}$ as

$$\widehat{\mathrm{cv}}(\hat{X}) = \frac{\sqrt{\widehat{var}(\hat{X})}}{\hat{X}} \tag{2}$$

The following rules should enable the user to estimate the coefficients of variation using the bootstrap weights for estimates of the number, proportion or percentage of the surveyed population possessing a certain characteristic and for ratios and differences between such estimates. In all the rules, the following notation is used:

n The sample size of the domain or subpopulation being analyzed.

b The index of the bootstrap replicates. In the EICS, $b = 1, \ldots, 1000$.

$\hat{X}$ The parameter estimate computed using the final weights.

$\hat{X}^{*(b)}$ The b-th parameter estimate computed using the b-th replicate.

w_i The final weight for the i-th individual.

$w_i^{*(b)}$ The bootstrap weight for the i-th individual for the b-th replicate.

I_i The indicator function for the i-th individual, where $I_i = 1$ if the individual possesses the characteristic of interest, and 0 otherwise.

Rule 1: Estimates of Numbers of Persons Possessing a Characteristic (Aggregates)
The estimate of the number of persons possessing a characteristic is computed as

$$\hat{X} = \sum_{i=1}^{n} w_i I_i. \tag{3}$$

The bootstrap estimate of the number of persons possessing a characteristic is computed as

$$\hat{X}^{*(b)} = \sum_{i=1}^{n} w_i^{*(b)} I_i, \tag{4}$$

for $b = 1, \ldots, 1000$. The variance of $\hat{X}$ is then computed as in Equation (1).

Rule 2: Estimates of Proportions or Percentages of Persons Possessing a Characteristic

The estimate of the proportion of persons possessing a characteristic is computed as

$$\hat{X} = \frac{\sum_{i=1}^{n} w_i I_i}{\sum_{i=1}^{n} w_i}. \tag{5}$$

The bootstrap estimate of the number of persons possessing a characteristic is computed as

$$\hat{X}^{*(b)} = \frac{\sum_{i=1}^{n} w_i^{*(b)} I_i}{\sum_{i=1}^{n} w_i^{*(b)}}, \tag{6}$$

for $b = 1, \ldots, 1000$. The variance of $\hat{\theta}$ is then computed as in Equation (1).

Rule 3: Estimates of Differences Between Aggregates or Percentages

The estimate of the difference between two totals is computed as

$$\hat{d} = \hat{X}_1 - \hat{X}_2,$$

where $\hat{X}_1$ and $\hat{X}_2$ are computed as in Equations (3) or (5). The bootstrap estimate of the difference between two totals is computed as

$$\hat{d}^{*(b)} = \hat{X}_1^{*(b)} - \hat{X}_2^{*(b)},$$

where $\hat{X}_1^{*(b)}$ and $\hat{X}_2^{*(b)}$ are computed as in Equations (4) or (6). The variance of $\hat{d}$ is then computed as in Equation (1).

Rule 4: Estimates of Ratios

This rule is a generalization of Rule 2 where the numerator and denominator can be arbitrary quantities. In the case where the numerator is a subset of the denominator, Rule 2 can be applied. For example, if the denominator is the number of unemployed potentially eligible for EI and the numerator is the number of unemployed eligible for EI.

In the case where the numerator is not a subset of the denominator, as for example, the ratio of the number of unemployed in receipt of regular EI benefits as compared to the number of unemployed in receipt of any other type of benefits, the estimate of the ratio can be computed as

$$\hat{R} = \frac{\hat{X}_1}{\hat{X}_2}, \tag{7}$$

where $\hat{X}_1$ and $\hat{X}_2$ are computed as in Equations (3) or (5).

The bootstrap estimate of the ratio is computed as

$$\hat{R}^{*(b)} = \frac{\hat{X}_1^{*(b)}}{\hat{X}_2^{*(b)}}, \tag{8}$$

where $\hat{X}_1^{*(b)}$ and $\hat{X}_2^{*(b)}$ are computed as in Equations (4) or (6). The variance of $\hat{R}$ is then computed as in Equation (1).

Rule 5: Estimates of Differences of Ratios

In this case, Rules 3 and 4 are combined. The CVs for the two ratios are first determined using Rule 4, and then the CV of their difference is found using Rule 3.

The estimate of a difference of ratios is computed as

$$\hat{\delta} = \hat{R}_1 - \hat{R}_2, \tag{8}$$

where $\hat{R}_1$ and $\hat{R}_2$ are computed as in Equation (7).

The bootstrap estimates of the difference of ratios is computed as

$$\hat{\delta}^{*(b)} = \hat{R}_1^{*(b)} - \hat{R}_2^{*(b)}, \tag{9}$$

where $\hat{R}_1^{*(b)}$ and $\hat{R}_2^{*(b)}$ are computed as in Equation (8). The variance of $\hat{\delta}$ is then computed as Equation (1).

10.1.1 Examples of Using the Bootstrap Weights for Categorical Estimates

Computing variance estimates using bootstrap weights is greatly facilitated by the use of a statistical software package, especially one that can handle replication methods. For example, the packages SAS and Stata have the capability to estimate variances using bootstrap weights (Gagné et al., 2014)[5]. In addition, Statistics Canada provides a set of SAS macros called Bootvar that can compute variances for many types of statistics. Bootvar can be requested free of charge from Statistics Canada. The following

[5] Gagné, C., Roberts, G., Keown, L. Weighted estimation and bootstrap variance estimation for analyzing survey data: How to implement in selected software. *Research Data Centres, Information and Technical Bulletin*, Vol. 6 no. 1, 5-70.

examples are included to assist users in applying the foregoing rules. The calculations were done using SAS 9.3.
It should be noted that these examples refer to estimates associated with the typical target population of the survey and are not directly relatable to the 2020 target population. The methods however, remain applicable.

Example 1: Estimates of Numbers of Persons Possessing a Characteristic (Aggregates)

Suppose that a user estimates that 353,000 unemployed individuals received regular EI benefits during the reference week. How does the user determine the coefficient of variation of this estimate?

Using the bootstrap weights, a user would compute 1,000 estimates of individuals who received regular EI benefits using the 1,000 sets of bootstrap weights. Applying Equations (1) and (2) to these 1,000 estimates gives an estimated CV of 0.0223.

Example 2: Estimates of Proportions or Percentages of Persons Possessing a Characteristic

Suppose that the user estimates that 511,700 / 606,700 = 84.3% of unemployed individuals potentially eligible to receive EI benefits were eligible to receive EI benefits. How does the user determine the coefficient of variation of this estimate?

1) Compute 1,000 estimates of unemployed individuals who eligible to receive EI benefits and 1,000 estimates of unemployed individuals potentially eligible to receive EI benefits using the bootstrap weights.

2) Using the estimates from the previous step, compute 1,000 estimates of the proportion of unemployed individuals potentially eligible to receive EI benefits who were eligible to receive EI benefits.

3) Applying Equations (1) and (2) to these 1,000 estimates gives an estimated CV of 0.0206.

4) So the estimated coefficient of variation of the estimate is 2.1%. The finding that 84.3% of unemployed individuals potentially eligible to receive EI benefits were eligible to receive EI benefits can be published with no qualifications.

Example 3: Estimates of Differences Between Aggregates or Percentages

Suppose that a user estimates that 163,900 / 253,300 = 64.7% of the regular unemployed population in Quebec contributed to EI, while 250,100 / 434,700 = 57.5% of the regular unemployed population in Ontario contributed to EI. How does the user determine the coefficient of variation of the difference between these two estimates?

1) For each province, two sets of 1,000 estimates must be computed using the bootstrap weights: (i) the estimate of the number of regular unemployed population and (ii) the estimate of the number of regular unemployed population who contributed to EI.

2) For each province, compute the 1,000 estimates of the proportion of the regular unemployed population who contributed to EI using the estimates from Step (1).

3) Using the two sets of 1,000 estimates from Step (2), compute 1,000 estimates of the difference between the proportion of regular unemployed population in Quebec who contributed to EI and the proportion of regular unemployed population in Ontario who contributed to EI.

4) Applying Equations (1) and (2) to these 1,000 difference estimates, the estimated CV is computed to be 0.634.

5) So the estimated coefficient of variation of the difference between the estimates is 63.4%. The difference between the estimates is considered unacceptable, but may be the result of the small estimated difference of proportions of 0.072. In this situation, further analysis such as computing confidence intervals may be helpful.

Example 4: Estimates of Ratios

Suppose that the user estimates that 163,900 of the regular unemployed population in Quebec contributed to EI, while 253,100 of the regular unemployed population in Ontario contributed to EI. The user is interested in comparing the estimate of Quebec versus that of Ontario in the form of a ratio. How does the user determine the coefficient of variation of this estimate?

1) First of all, this estimate is a ratio estimate, where the numerator of the estimate ($\hat{X}_1$) is the number of unemployed individuals in Quebec who contributed to EI. The denominator of the estimate ($\hat{X}_2$) is the number of unemployed individuals in Ontario who contributed to EI.

2) For each province, compute 1,000 estimates of the number of the regular unemployed population who contributed to EI.

3) Using the results of Step (2), compute the 1,000 estimates of the ratio of the estimated number of Quebec to the estimated number in Ontario.

4) Applying Equations (1) and (2) to these 1,000 estimates gives an estimated CV of 0.075.

So the estimated coefficient of variation of the ratio estimate is 7.5%, which makes the estimate releasable with no qualifications.

Example 5: Estimates of Differences of Ratios

Suppose that the user estimates that the ratio of individuals aged 15 to 24 years in the regular unemployed population who contributed to EI, to individuals aged 25 to 44 years in the regular unemployed population who contributed to EI is 0.50 for Manitoba and Saskatchewan, while it is 0.62 for Alberta. The user is interested in comparing the two ratios to see if there is a statistical difference between them. How does the user determine the coefficient of variation of the difference?

1) First, calculate the Manitoba and Saskatchewan ratio ($\hat{R}_1$) and the Alberta ratio ($\hat{R}_2$) using Equation (7) and the final weights, as well as the 1,000 ratios for each region ($\hat{R}_1^{*(b)}$ and $\hat{R}_2^{*(b)}$, respectively, for $b = 1, \ldots, 1000$) using Equation (8) and the bootstrap weights.

2) Using the results of Step (1), calculate the estimated differences $\hat{d} = \hat{R}_1 - \hat{R}_2$ and $\hat{d}^{*(b)} = \hat{R}_1^{*(b)} - \hat{R}_2^{*(b)}$, for $b = 1, \ldots, 1000$.

3) Using the results of Step (2) and applying Equations (1) and (2) to these 1,000 estimates gives an estimated coefficient of variation of 2.409.

4) So the estimated coefficient of variation of the difference between the estimates is 240.9%. The difference between the estimates is considered unacceptable and Statistics Canada recommends this estimate not be released. However, should the user choose to do so, the estimate should be flagged with the letter F (or some similar identifier) and be accompanied by a warning to caution subsequent users about the high levels of error, associated with the estimate.

10.2 How to Use the Bootstrap Weights to Obtain Confidence Limits

Although coefficients of variation are widely used, a more intuitively meaningful measure of sampling error is the confidence interval of an estimate. A confidence interval constitutes a statement on the level of confidence that the true value for the population lies within a specified range of values. For example a 95% confidence interval can be described as follows:

> If sampling of the population is repeated indefinitely, each sample leading to a new confidence interval for an estimate, then in 95% of the samples the interval will cover the true population value.

Using the standard error of an estimate, confidence intervals for estimates may be obtained under the assumption that under repeated sampling of the population, the various estimates obtained for a population characteristic are normally distributed about the true population value. Under this assumption, the chances are about 68 out of 100 that the difference between a sample estimate and the true population value would be less than one standard error, about 95 out of 100 that the difference would be less than two standard errors, and about 99 out of 100 that the difference would be less than three standard errors. These different degrees of confidence are referred to as the confidence levels.

Confidence intervals for an estimate, $\hat{X}$, are generally expressed as two numbers, one below the estimate and one above the estimate, as $(\hat{X} - k, \hat{X} + k)$ where k is determined depending upon the level of confidence desired and the sampling error of the estimate.

Confidence intervals for an estimate can be calculated directly by first determining the estimated standard error, $\hat{se}(\hat{X}) = \sqrt{\hat{var}(\hat{X})}$, of the estimate $\hat{X}$, and then using the following formula to convert to a confidence interval $CI(\hat{X})$:

$$CI(\hat{X}) = (\hat{X} - t \cdot \hat{se}(\hat{X}), \hat{X} + t \cdot \hat{se}(\hat{X})$$

where t is the approximate endpoint of the normal distribution. Depending on the level of confidence, the following table can be used:

t	Level
1.0	68% confidence interval is desired
1.6	90% confidence interval is desired
2.0	95% confidence interval is desired
2.6	99% confidence interval is desired

Another way to compute confidence intervals using the bootstrap weights is to determine the percentiles of the distribution of estimates corresponding to the desired level of confidence. This is done by first sorting the estimates from the 1,000 bootstrap replicates in ascending order. Let $\hat{X}_{(p)}$ be the estimated p-th percentile of X. Then, $\hat{X}_{(p)}$ is the value of corresponding to the p-th estimate of the sorted data. To obtain the confidence limits using this method, one would simply compute $\hat{X}_{(p)}$ and $\hat{X}_{(100-p)}$, the p-th and $(100-p)$-th percentiles of the bootstrap replicates. The resulting confidence interval is

$$CI(\hat{X}) = (\hat{X}_{(p)}, \hat{X}_{(100-p)}).$$

Note: Release guidelines which apply to the estimate also apply to the confidence interval. For example, if the estimate is not releasable, then the confidence interval is not releasable either.

10.2.1 Example of Using the Standard Error to Obtain Confidence Limits

Using the table above, a 95% confidence interval for the estimated proportion of unemployed individuals who were potentially eligible to receive EI benefits were eligible to receive EI benefits (from Example 2, Section 10.1.1) would be calculated as follows:

$$\begin{aligned} \hat{X} &= 0.843 \\ t &= 2 \\ \hat{se}(\hat{X}) &= 0.0173 \\ CI(\hat{X}) &= (0.843 - 2 \cdot 0.0173, 0.843 + 2 \cdot 0.0173) \\ &= (0.8084, 0.8776) \end{aligned}$$

With 95% confidence it can be said that between 80.8% and 87.8% of unemployed individuals who were potentially eligible to receive EI benefits were eligible to receive EI benefits.

10.2.2 Example of Using the Bootstrap Weights to Obtain Confidence Limits

Using the bootstrap percentiles, the 95% confidence interval would be calculated as follows:

$$\begin{aligned} \hat{X}_{(2.5)} &= 0.8111 \\ \hat{X}_{(97.5)} &= 0.8794 \\ CI(\hat{X}) &= (0.8111, 0.8794) \end{aligned}$$

With 95% confidence it can be said that between 81.1% and 87.9% of unemployed individuals who were potentially eligible to receive EI benefits were eligible to receive EI benefits.

10.3 How to Use the Bootstrap Standard Error to do a t-test

Standard errors may also be used to perform hypothesis testing, a procedure for distinguishing between population parameters using sample estimates. The sample estimates can be numbers, averages, percentages, ratios, etc. Tests may be performed at various levels of significance, where a level of significance is the probability of concluding that the characteristics are different when, in fact, they are identical.

Let $\hat{X}_1$ and $\hat{X}_2$ be sample estimates for two characteristics of interest. Let the standard error of the difference $\hat{d} = \hat{X}_1 - \hat{X}_2$ be $\text{se}(\hat{d})$.

If $t = \frac{\hat{X}_1 - \hat{X}_2}{se(\hat{d})}$ is between -2 and 2, then no conclusion about the difference between the characteristics is justified at the 5% level of significance. If however, this ratio is smaller than -2 or larger than +2, the observed difference is significant at the 0.05 level. That is to say that the difference between the estimates is significant.

10.3.1 Example of Using the Bootstrap Standard Error to do a t-test

Let us suppose that the user wishes to test, at 5% level of significance, the hypothesis that there is no difference between the proportion of the regular employed population in Quebec who contributed to EI and the proportion of the regular employed population in Ontario who contributed to EI. From Example 3, Section 10.1.1, the standard error of the difference between these two estimates was found to be 0.0455. Hence,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{se(\hat{d})} = \frac{0.647 - 0.575}{0.0455} = \frac{0.072}{0.0455} = 1.58$$

Since t = 1.58 is less than 2, it must be concluded that there is no significant difference between the two estimates at the 0.05 level of significance.

10.4 Coefficients of Variation for Quantitative Estimates

For quantitative estimates, the rules in 10.1 can be adapted to determine their sampling error. For example, to determine the coefficient of variation of the average of usual hourly earnings, one can adapt Rule 2 by replacing the indicator variable I_i with a variable representing usual hourly earnings. For example, if z_i = hourly earnings, then adapting Equation (5), we would obtain

$$\hat{X} = \frac{\sum_{i=1}^{n} w_i z_i}{\sum_{i=1}^{n} w_i},$$

an estimate of the average usual hourly earnings. See also Section 9.3.4.

11.0 Weighting

Since the Employment Insurance Coverage Survey (EICS) used a sub-sample of the Labour Force Survey (LFS) sample, the derivation of weights for the survey records is clearly tied to the weighting procedure used for the LFS. The LFS weighting procedure is briefly described below.

11.1 Weighting Procedures for the Labour Force Survey

In the LFS, the final weight attached to each record is the product of the following factors: the basic weight, the cluster sub-weight, the stabilization weight, the balancing factor for non-response, and the province-age-sex and sub-provincial area ratio adjustment factor. Each is described below.

Basic Weight

In a probability sample, the sample design itself determines weights which must be used to produce unbiased estimates of population. Each record must be weighted by the inverse of the probability of selecting the person to whom the record refers. In the example of a 2% simple random sample, this probability would be 0.02 for each person and the records must be weighted by 1 / 0.02 = 50. Due to the complex LFS design, dwellings in different regions will have different basic weights. Because all eligible individuals in a dwelling are interviewed (directly or by proxy), this probability is essentially the same as the probability with which the dwelling is selected.

Cluster Sub-weight

The cluster delineation is such that the number of dwellings in the sample increases very slightly with moderate growth in the housing stock. Substantial growth can be tolerated in an isolated cluster before the additional sample represents a field collection problem. However, if growth takes place in more than one cluster in an interviewer assignment, the cumulative effect of all increases may create a workload problem. In clusters where substantial growth has taken place, sub-sampling is used as a means of keeping interviewer assignments manageable. The cluster sub-weight represents the inverse of this sub-sampling ratio in clusters where sub-sampling has occurred.

Stabilization Weight

Sample stabilization is also used to address problems with sample size growth. Cluster sub-sampling addressed isolated growth in relatively small areas whereas sample stabilization accommodates the slow sample growth over time that is the result of a fixed sampling rate along with a general increase in the size of the population. Sample stabilization is the random dropping of dwellings from the sample in order to maintain the sample size at its desired level. The basic weight is adjusted by the ratio of the sample size, based on the fixed sampling rate, to the desired sample size. This adjustment factor is known as the stabilization weight. The adjustment is done within stabilization areas defined as dwellings belonging to the same employment insurance economic region and the same rotation group.

Non-response

For certain types of non-response (i.e. household temporarily absent, refusal), data from a previous month's interview with the household if any, is brought forward and used as the current month's data for the household. Respondents that refused the data linkage question were considered non-response.

In other cases, non-response is compensated for by proportionally increasing the weights of responding households. The weight of each responding record is increased by the ratio of the number of households that should have been interviewed, divided by the number that were actually interviewed. This adjustment is done separately for non-response areas, which are defined by employment insurance economic region, type of area, and rotation group. It is based on the assumption that the households

that have been interviewed represent the characteristics of those that should have been interviewed within a non-response area.

Labour Force Survey Sub-weight
The product of the previously described weighting factors is called the LFS sub-weight. All members of the same sampled dwelling have the same sub-weight.

Sub-provincial and Province-Age-Sex Adjustments
The sub-weight can be used to derive a valid estimate of any characteristic for which information is collected by the LFS. However, these estimates will be based on a frame that contains some information that may be several years out of date and therefore not representative of the current population. Through the use of more up-to-date auxiliary information about the target population, the sample weights are adjusted to improve both the precision of the estimates and the sample's representation of the current population.

Independent estimates are available monthly for various age and sex groups by province. These are population projections based on the most recent census data, records of births and deaths, and estimates of migration. In the final step, this auxiliary information is used to transform the sub-weight into the final weight. This is done using a calibration method. This method ensures that the final weights it produces sum to the census projections for the auxiliary variables, namely totals for various age-sex groups, economic regions, census metropolitan areas, rotation groups, household and economic family size. Weights are also adjusted so that estimates of the previous month's industry and labour status estimates derived from the present month's sample, sum up to the corresponding estimates from the previous month's sample. This is called composite estimation. The entire adjustment is applied using the generalized regression technique.

This final weight is normally not used in the weighting for a supplement to the LFS. Instead, it is the sub-weight which is used, as explained in the following paragraphs.

11.2 Weighting Procedures for the Employment Insurance Coverage Survey

The principles behind the calculation of the weights for the EICS are identical to those for the LFS. However, further adjustments are made to the LFS sub-weights in order to derive a final weight for the individual records on the EICS microdata file.

1) An adjustment to account for the use of a one-sixth sub-sample, instead of the full LFS sample. In the case of the parents, the fraction is two-sixths.

2) An adjustment to account for the EICS sub-sampling (refer to Section 5.6.1).

3) An adjustment to account for the additional non-response to the supplementary survey i.e., non-response to the EICS for individuals who did respond to the LFS or for which previous month's LFS data was brought forward. The procedure is similar to the LFS non-response weight adjustment, but groupings are based on different variables.
These variables are the province, type of respondent, sex and a grouping of employment insurance regions.

4) A final adjustment is done using two external non-overlapping independent sources. Employment and Social Development Canada provides estimated counts for regular beneficiaries with

and without earnings. The other source is LFS data which provides estimated counts for unemployment (not seasonally adjusted). The adjustment is done within a calibration process which ensures that the estimates produced with the EICS data match the counts from the external sources. The final calibrated weight is equal to the weight before the calibration multiplied by the factor necessary to calibrate to the applicable independent source. The extended part of the EICS survey population, comprised of the parents of infants less than eighteen months old, is excluded from this calibration. Due to the change in target population in 2020, this step was not performed on the 2020 weights.

The resulting weight WTPM is the final weight which appears on the EICS microdata file.