Microdata User Guide

Employment Insurance Coverage Survey

2010

Statistics Canada Statistique Canada

Canada

Table of Contents

1.0 Introduction

The Employment Insurance Coverage Survey (EICS) was conducted by Statistics Canada with the cooperation and support of Human Resources and Skills 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
Special Surveys Division
Telephone: 613-951-3321 or call toll-free 1-800-461-9050
Fax: 613-951-4527
E-mail: ssd@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.

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 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 mothers of an infant who received maternity and 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, 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. In 2010, 44% of the unemployed were potentially eligible for Employment Insurance. Of those who were potentially eligible, 83.9% could meet the entrance requirements of the program and were very likely to receive benefits during their unemployment spell, if they claimed. The remaining 16% did not have enough.

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.

[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.

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 past 12 months. The industry coding corresponds to the North American Industry Classification System 2007 (NAICS 2007). Occupation codes are based on the National Occupational Classification for Statistics 2006 (NOC-S 2006), January 1987 to present.

For the EICS the industry coding corresponds to the North American Industry Classification System 1997 (NAICS 1997). Occupation codes are based on the Standard Occupational Classification 1991 (SOC 1991) and the National Occupational Classification for Statistics 2001 (NOC-S 2001).

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. The interviews are 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 then 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 distinct subpopulations of interest called “Type”. Type is defined as follows:

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 three months,
5) mothers of infants less than one year old working during the reference week.

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

Mothers
In this survey, the term “mother” refers to mothers (by birth or adoption) of an infant aged less than one year old during the LFS reference week. Many mothers were not part of the survey sample prior to 2000. In particular, mothers working full-time and mothers 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 mother of an infant during the survey reference week (see definition of Mothers above).

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

The original survey types consisted of:

- Type 1 - same as current;

- Type 2 - including part-time mothers;
- Type 3 - excluding mothers who have not worked in two years; and
- Type 4 - including mothers 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 three 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 “mothers”, 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 5.4.[3] Sections 5.5 and 5.6 describe how the EICS departed from the basic LFS design.

5.1 Population Coverage

The 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 Aboriginal settlements in the provinces; full-time members of the Canadian Armed Forces and the institutionalized population. 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

The LFS has undergone an extensive redesign, culminating in the phasing in of the new design in November 2004. The LFS sample is based upon a stratified, multi-stage design employing probability sampling at all stages of the design. The design principles are the same for each province.

5.2.1 Primary Stratification

Provinces are divided into economic regions (ER) and employment insurance economic regions (EIER). ERs are geographic areas of more or less homogeneous economic structure formed on the basis of federal-provincial agreements. They are relatively stable over time. EIERs are also geographic areas, and are roughly the same size and number as ERs, but they do not share the same definitions. Labour force estimates are produced for the EIERs for the use of Human Resources and Skills Development Canada.

The intersections of the two types of regions form the first level of stratification for the LFS. These ER/EIER intersections are treated as primary strata and further stratification is carried out within them (see Section 5.2.3). Note that a third set of regions, census metropolitan areas (CMA), is also respected by stratification in the current LFS design, since each CMA is also an EIER.

5.2.2 Types of Areas

The primary strata (ER/EIER intersections) are further disaggregated into three types of areas: rural, urban and remote areas. Urban and rural areas are loosely based on the Census definitions of urban and rural, with some exceptions to allow for the formation of strata in some areas. Urban areas include the largest CMAs down to the smallest villages categorized by the 1991 Census as urban (1,000 people or more), while rural areas are made up of areas not designated as urban or remote.

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

All urban areas are further subdivided into two types: those using an apartment list frame and an area frame, as well as those using only an area frame.
Approximately 1% of the LFS population is found in remote areas of provinces which are less accessible to LFS interviewers than other areas. For administrative purposes, this portion of the population is sampled separately through the remote area frame. Some populations, not congregated in places of 25 or more people, are excluded from the sampling frame.

5.2.3 Secondary Stratification

In urban areas with sufficiently large numbers of apartment buildings, the strata are subdivided into apartment frames and area frames. The apartment list frame is a register maintained for the 18 largest cities across Canada. The purpose of this is to ensure better representation of apartment dwellers in the sample as well as to minimize the effect of growth in clusters, due to construction of new apartment buildings. In the major cities, the apartment strata are further stratified into low income strata and regular strata.

Where it is possible and/or necessary, the urban area frame is further stratified into regular strata, high income strata, and low population density strata. Most urban areas fall into the regular urban strata, which, in fact, cover the majority of Canada's population. High income strata are found in major urban areas, while low density urban strata consist of small towns that are geographically scattered.

In rural areas, the population density can vary greatly from relatively high population density areas to low population density areas, resulting in the formation of strata that reflect these variations. The different stratification strategies for rural areas were based not only on concentration of population, but also on cost-efficiency and interviewer constraints.

In each province, remote settlements are sampled proportional to the number of dwellings in the settlement, with no further stratification taking place. Dwellings are selected using systematic sampling in each of the places sampled.

5.2.4 Cluster Delineation and Selection

Households in final strata are not selected directly. Instead, each stratum is divided into clusters, and then a sample of clusters is selected within the stratum. Dwellings are then sampled from selected clusters. Different methods are used to define the clusters, depending on the type of stratum.

Within each urban stratum in the urban area frame, a number of geographically contiguous groups of dwellings, or clusters, are formed based upon 2001 Census counts. These clusters are generally a set of one or more city blocks or block-faces. The selection of a sample of clusters (always six or a multiple of six clusters) from each of these secondary strata represents the first stage of sampling in most urban areas. In some other urban areas, census enumeration areas (EA) are used as clusters. In the low density urban strata, a three stage design is followed. Under this design, two towns within a stratum are sampled, and then 6 or 24 clusters within each town are sampled.

For urban apartment strata, instead of defining clusters, the apartment building is the primary sampling unit. Apartment buildings are sampled from the list frame with probability proportional to the number of units in each building.

Within each of the secondary strata in rural areas, where necessary, further stratification is carried out in order to reflect the differences among a number of socio-economic

characteristics within each stratum. Within each rural stratum, six EAs or two or three groups of EAs are sampled as clusters.

5.2.5 Dwelling Selection

In all three types of areas (urban, rural and remote areas) selected clusters are first visited by enumerators in the field and a listing of all private dwellings in the cluster is prepared. From the listing, a sample of dwellings is then selected. The sample yield depends on the type of stratum. For example, in the urban area frame, sample yields are either six or eight dwellings, depending on the size of the city. In the urban apartment frame, each cluster yields five dwellings, while in the rural areas and EA parts of cities, each cluster yields 10 dwellings. In all clusters, dwellings are sampled systematically. This represents the final stage of sampling.

5.2.6 Person Selection

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 the elderly (age 70 and over) by carrying forward their responses for the initial interview to the subsequent five months in the survey.

5.3 Sample Size

The sample size of eligible persons in the LFS is determined so as to meet the statistical precision requirements for various labour force characteristics at the provincial and sub-provincial level, to meet the requirement of federal, provincial and municipal governments as well as a host of other data users.

The monthly LFS sample consists of approximately 60,000 dwellings. After excluding dwellings found to be vacant, dwellings demolished or converted to non-residential uses, dwellings containing only ineligible persons, dwellings under construction, and seasonal dwellings, about 54,000 dwellings remain which are occupied by one or more eligible persons. From these dwellings, LFS information is obtained for approximately 100,000 civilians aged 15 or over.

5.4 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.

Because of the rotation group feature, it is possible to readily conduct supplementary surveys using the LFS design but employing less than the full size sample.

5.5 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 mothers of infants.

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

5.5.1 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 three months;
5) mothers of infants less than one year old 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.

5.5.2 Type 4: A Special Case

Respondents sampled with Type = 4 are not all targeted by the survey. Only those who have experienced an interruption in work in the two months prior to the survey reference week need to be interviewed. This information was not available from the LFS sample frame. Therefore, all full-time workers with short job tenure at their current job were selected. The question on work interruption is asked in the EICS and respondents who worked continually over the two month period prior to the reference week are 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.

5.5.3 Sub-sampling

At the initial stage, sub-sampling was done to arrive at the target sample of 3,600 and to balance the representation of groups according to the relevance of the Employment Insurance program to them. The sub-sampling criteria are summarized by focus type, as follows:

Type 1
All persons were included.

Type 2

a) full-time students were sub-sampled at the rate of 70%,
b) persons working 20 to 24 hours during the reference week,
c) persons working 25 to 29 hours during the reference week were sub-sampled at the rate of approximately 50%,
d) the remaining cases were all included in the sample.

Type 3
a) full-time students who left their last job because of school and full-time students who did not leave their job more than one year ago were sub-sampled at the rate of 50%,
b) the remaining persons were all included.

Type 4
All persons were included.

Type 5
All persons were included.

5.5.4 Other Exclusions

At the second stage of sub-sampling, when three or more persons targeted by the EICS lived in the same household, only two persons were selected into the survey, unless they were all unemployed. In this case, a maximum of three persons were kept in the sample. This was done to reduce the response burden within the household.

Some persons did not respond to the LFS interview (they had imputed data) or gave no permission to LFS personnel to conduct telephone interviews with them. These were also excluded from the EICS.

5.6 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**
Newfoundland and Labrador	619
Prince Edward Island	449
Nova Scotia	857
New Brunswick	780
Quebec	2,698
Ontario	3,769
Manitoba	1,289
Saskatchewan	1,001
Alberta	1,571
British Columbia	1,639
Canada	**14,672**

6.0 Data Collection

Data collection for the Labour Force Survey (LFS) is carried out each month during the week following the LFS reference week. The reference week is normally the week containing the 15th day of the month.

6.1 Interviewing for the Labour Force Survey

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. Each interviewer contacts approximately 75 dwellings per month.

Dwellings new to the sample in urban areas are contacted by telephone if the telephone number is available from administrative files otherwise the dwelling is contacted through a personal visit using the computer-assisted personal interview (CAPI). 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 armed forces. Provided there is a telephone in the dwelling and permission has been granted, subsequent interviews are conducted by telephone. This is done out of a centralized computer-assisted telephone interviewing (CATI) unit where cases are assigned randomly to interviewers. As a result, approximately 85% of all households are interviewed by telephone. In these 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.

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.

At the conclusion of the LFS monthly interviews, interviewers introduce the supplementary survey, if any, to be administered to some or all household members that month.

6.2 Supervision and Quality Control

All LFS interviewers are under the supervision of a staff of senior interviewers who are responsible for ensuring that interviewers are familiar with the concepts and procedures of the LFS and it’s many supplementary surveys, and also for periodically monitoring their interviewers and reviewing their completed documents. The senior interviewers are, in turn, under the supervision of the LFS program managers, located in each of the Statistics Canada regional offices.

6.3 Non-response to the Labour Force Survey

Interviewers are instructed to make all reasonable attempts to obtain LFS 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 cooperation. 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. For households non-responding to the LFS and for which LFS information was obtained in the previous month, this information is brought forward and used as the current month's LFS information. No supplementary survey information is collected for these households.

6.4 Data Collection Modifications for the Employment Insurance Coverage Survey

Household members selected for the Employment Insurance Coverage Survey (EICS) are contacted three to seven weeks after their last LFS interview. All interviews are conducted over the telephone and 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 at first contact are 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 interviewer at the time of the interview using a computerized questionnaire. The computerized questionnaire reduces processing time and costs associated with data entry, transcription errors and data transmission. The response data are encrypted to ensure confidentiality and sent via modem to the appropriate Statistics Canada Regional Office. From there they are transmitted over a secure line to Ottawa 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 interviewer is prompted, through message screens on the computer, to modify the information. However, for some questions 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 5.5.2). 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 7.5.5).

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 consistency (working individuals reporting zero earnings).

7.3 Coding of Open-ended Questions

A few data items on the questionnaire were recorded by interviewers 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 1997), the Standard Occupational Classification (SOC-1991) and the National Occupational Classification for Statistics (NOC-S 2001)) 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, reason for interrupting work, job search method used, reason why spouse did not claim benefits or why both parents did.

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 codebooks.

There was no imputation done for the 2010 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 public use 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 codebook available for the public use microdata file (refer to Chapter 13.0) includes a note that identifies all questionnaire items used to create each derived variable.

7.5.1 Grouping of Continuous Data Items

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), the number of weeks worked during the reference year (WEEKSCAT), and the number of weeks of advance notice from the employer before a job loss (NOTICE_W).

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 OCC6), the most important job search method during the reference week (JOBSRCH), help needed in finding a job (HELPFIND), planned or current childcare arrangement (CHLDCARE), household income in the month before the birth/adoption (M_HHINC - for mothers only), type of economic family (EFAMILY) and a few others.

7.5.2 Combining Identical Questions

Before 2004, the EICS questionnaire (refer to Chapter 12.0) contained a number of questions which were duplicated two or three times with slightly different wording. This practice was frequent for questions related to claiming and receiving EI benefits. Different wording was used for mothers to reflect a different time reference, the birth or adoption of a child, and also between working and non-working respondents to make the questions more relevant to the respondent's situation. In most cases, the derived variable was created by simply combining answers from two questions. This is the case for respondents currently working at a job or business (WORKNOW), the type of benefits received in the reference week or reference month (BENTYP), the number of weeks of EI benefits received since last applied (BENWEEKS), the amount of EI benefits received (BENAMNT), receive any advance formal notice from the employer before a work interruption (NOTICE), took a break from working during pregnancy or since birth/adoption (BREAKWRK), and parental benefits claimed by the spouse (SPCLAIM).

Similarly, questions regarding employment after birth or adoption (EMPAGREE, SAMEMP, and WORKCOND) and on childcare arrangements (CHLDCARE) were asked differently for mothers on leave than for mothers who had already returned to work.

7.5.3 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.

7.5.4 Combining Two or More Different Questions

Variables such as union status (UNIONCA), type of work arrangement (WRKTYP), reason stopped working at job (RSWORK), 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, insurance or other benefits (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 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.

7.5.5 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 Human Resources and Skills Development Canada (HRSDC). 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 telephone survey.

The following is a description of the logic of the taxonomy used by HRSDC 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 not potentially eligible for EI.

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 or RSWORK).
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 persons eligible for EI 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

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, detailed industry and occupation, job tenure, number of months since last

worked, age of the baby in months (mothers only) and several other detailed variables are only available on the survey master file.

Users requiring access to information excluded from the microdata files may purchase custom tabulations. 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	In-scope Sample	Response	Response Rate (%)
Newfoundland and Labrador	582	477	82
Prince Edward Island	419	359	86
Nova Scotia	800	676	85
New Brunswick	731	629	86
Quebec	2,526	2,181	86
Ontario	3,509	2,966	85
Manitoba	1,179	1,018	86
Saskatchewan	902	784	87
Alberta	1,383	1,157	84
British Columbia	1,502	1,265	84
Canada	**13,533**	**11,512**	**85**

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 5.5.2 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 EICS sample consisted of one rotation group from the LFS for the "regular" EICS population and of two rotation groups for "mothers".

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 January, April, July and November LFS collection period. Collection lasted five weeks for each EICS cycle.

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. All derived variables were specified in decision tables. For each variable, the process generates a summary table documenting the rules applied and rule counts. The distribution for each derived variable was compared to that of the questionnaire items used in creating it. The derived variables were also cross-classified with other related variables to ensure internal consistency and limit the risk of errors in the derivation rules. A comparison of the distribution over the 2009 and 2010 period was also conducted to ensure historical comparability of the information included on the public use microdata files.

8.2.4 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. It was consistently more pronounced among the full-time employed (Type = 4 or 5) over the years and also marginally for men, but there is no marked difference across broad age groups.

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.

8.2.5 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.

For example, suppose that based upon the 2003 EICS results, one estimates that 81% of individuals are eligible for Employment Insurance among the potentially eligible and this

estimate is found to have a standard error of 0.03. Then the coefficient of variation of the estimate is calculated as:

$$\left(\frac{0.03}{0.81} \right) X \ 100 \ \% \ = \ 3.7\%$$

There is more information on the calculation of coefficients of variation in Chapter 10.0.

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 mothers for whom the information is available (returned to work already or know plans) and its denominator is the number of mothers 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 mothers 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 MOTHER = 1 and WORKNOW = 1 (mother of an infant less than one year old 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 mothers,

a) estimate the total number of weeks ($\hat{X}$) as described above,
b) estimate the number of mothers currently working ($\hat{Y}$) in this category by summing the final weights of all records with MOTHER = 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. Approximate variances for simple estimates such as totals, proportions and ratios (for qualitative variables) can be derived using the accompanying Approximate Sampling Variability Tables.

For other analysis techniques (for example linear regression, logistic regression and analysis of variance), a method exists which can make the variances calculated by the standard packages

more meaningful, by incorporating the unequal probabilities of selection. The method rescales the weights so that there is an average weight of 1.

For example, suppose that analysis of all male respondents is required. The steps to rescale the weights are as follows:

1) select all respondents from the file who reported SEX = men;
2) calculate the AVERAGE weight for these records by summing the original person weights from the microdata file for these records and then dividing by the number of respondents who reported SEX = men;
3) for each of these respondents, calculate a RESCALED weight equal to the original person weight divided by the AVERAGE weight;
4) perform the analysis for these respondents using the RESCALED weight.

However, because the stratification and clustering of the sample's design are still not taken into account, the variance estimates calculated in this way are likely to be under-estimates.

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.

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 30, the weighted estimate should be considered to be of unacceptable quality.

For weighted estimates based on sample sizes of 30 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 30 or more, and low coefficients of variation in the range of 0.0% to 16.5%. No warning is required.
2) Marginal	Estimates have a sample size of 30 or more, and high coefficients of variation in the range of 16.6% to 33.3%. Estimates should be flagged with the letter E (or some similar identifier). They should be accompanied by a warning to caution subsequent users about the high levels of error, associated with the estimates.
3) Unacceptable	Estimates have a sample size of less than 30, or very high coefficients of variation in excess of 33.3%. Statistics Canada recommends not to release estimates of unacceptable quality. However, if the user chooses to do so then estimates should be flagged with the letter F (or some similar identifier) and the following warning should accompany the estimates: "Please be warned that these estimates [flagged with the letter F] do not meet Statistics Canada's quality standards. Conclusions based on these data will be unreliable, and most likely invalid."

9.6 Release Cut-off's for the Employment Insurance Coverage Survey

The following table provides an indication of the precision of population estimates as it shows the release cut-offs associated with each of the three quality levels presented in the previous section. These cut-offs are derived from the coefficient of variation (CV) tables discussed in Chapter 10.0.

For example, the table shows that the quality of a weighted estimate of 5,000 people with Type 1 possessing a given characteristic in the Atlantic Provinces is marginal.

Note that these cut-offs apply to estimates of total number of persons possessing a characteristic. To estimate ratios, users should not use the numerator value (nor the denominator) in order to find the corresponding quality level. Rule 4 in Section 10.1 and Example 4 in Section 10.1.1 explain the correct procedure to be used for ratios.

Province and Region for TYPE = 1	**Acceptable CV 0.0% to 16.5%**	**Marginal CV 16.6% to 33.3%**	**Unacceptable CV > 33.3%**
Atlantic Provinces	12,000 & over	3,200 to < 12,000	under 3,200
Quebec	39,000 & over	10,500 to < 39,000	under 10,500
Ontario	47,500 & over	12,400 to < 47,500	under 12,400
Manitoba and Saskatchewan	20,500 & over	6,700 to < 20,500	under 6,700
Alberta	34,500 & over	10,500 to < 34,500	under 10,500
British Columbia	48,500 & over	15,000 to < 48,500	under 15,000
Western Provinces	32,900 & over	8,700 to < 32,900	under 8,700
Canada	**40,700 & over**	**10,200 to < 40,700**	**under 10,200**

Province and Region for MOTHER = 1	**Acceptable CV 0.0% to 16.5%**	**Marginal CV 16.6% to 33.3%**	**Unacceptable CV > 33.3%**
Atlantic Provinces	22,200 & over	17,100 to < 22,200	under 17,100
Quebec	70,500 & over	35,100 to < 70,500	under 35,100
Ontario	48,100 & over	15,800 to < 48,100	under 15,800
Manitoba and Saskatchewan	8,400 & over	2,500 to < 8,400	under 2,500
Alberta	20,000 & over	6,800 to < 20,000	under 6,800
British Columbia	26,100 & over	15,200 to < 26,100	under 15,200
Western Provinces	35,400 & over	11,100 to < 35,400	under 11,100
Canada	**76,600 & over**	**22,000 to < 76,600**	**under 22,000**

Province and Region for MOTHER = 0 and TYPE = 3	**Acceptable CV 0.0% to 16.5%**	**Marginal CV 16.6% to 33.3%**	**Unacceptable CV > 33.3%**
Atlantic Provinces	27,100 & over	7,600 to < 27,100	under 7,600
Quebec	79,700 & over	22,500 to < 79,700	under 22,500
Ontario	122,700 & over	34,200 to < 122,700	under 34,200
Manitoba and Saskatchewan	14,800 & over	4,000 to < 14,800	under 4,000
Alberta	48,100 & over	14,300 to < 48,100	under 14,300
British Columbia	65,200 & over	19,300 to < 65,200	under 19,300
Western Provinces	55,100 & over	14,500 to < 55,100	under 14,500
Canada	**75,900 & over**	**19,200 to < 75,900**	**under 19,200**

Province and Region for MOTHER = 0 and TYPE = 2 or 4	**Acceptable CV 0.0% to 16.5%**	**Marginal CV 16.6% to 33.3%**	**Unacceptable CV > 33.3%**
Atlantic Provinces	23,200 & over	6,300 to < 23,200	under 6,300
Quebec	62,500 & over	16,400 to < 62,500	under 16,400
Ontario	63,500 & over	16,200 to < 63,500	under 16,200
Manitoba and Saskatchewan	17,600 & over	4,600 to < 17,600	under 4,600
Alberta	50,000 & over	13,800 to < 50,000	under 13,800
British Columbia	43,200 & over	11,400 to < 43,200	under 11,400
Western Provinces	43,100 & over	10,900 to < 43,100	under 10,900
Canada	**41,600 & over**	**10,300 to < 41,600**	**under 10,300**

10.0 Approximate Sampling Variability Tables

In order to supply coefficients of variation (CV) which would be applicable to a wide variety of categorical estimates produced from this microdata file and which could be readily accessed by the user, a set of Approximate Sampling Variability Tables has been produced. These CV tables allow the user to obtain an approximate coefficient of variation based on the size of the estimate calculated from the survey data.

The coefficients of variation are derived using the variance formula for simple random sampling and incorporating a factor which reflects the multi-stage, clustered nature of the sample design. This factor, known as the design effect, was determined by first calculating design effects for a wide range of characteristics and then choosing from among these a conservative value to be used in the CV tables which would then apply to the entire set of characteristics.

The table below shows the conservative value of the design effects as well as sample sizes and population counts by province for survey type and mother status, which were used to produce the Approximate Sampling Variability Tables for the Employment Insurance Coverage Survey (EICS).

Province and Region for TYPE = 1	**Design Effect**	**Sample Size**	**Population**
Atlantic Provinces	1.64	557	123,925
Quebec	1.64	439	323,391
Ontario	1.52	635	688,838
Manitoba and Saskatchewan	3.47	261	62,735
Alberta	1.73	186	135,612
British Columbia	2.38	229	175,653
Western Provinces	1.78	676	374,000
Canada	**1.87**	**2,307**	**1,410,154**

Province and Region for MOTHER = 1	**Design Effect**	**Sample Size**	**Population**
Atlantic Provinces	45.68	179	24,591
Quebec	12.56	226	105,162
Ontario	3.58	264	144,932
Manitoba and Saskatchewan	2.04	221	33,154
Alberta	2.34	152	55,601
British Columbia	8.49	94	34,002
Western Provinces	5.17	467	122,757
Canada	**7.40**	**1,136**	**397,442**

Province and Region for MOTHER = 0 and TYPE = 3	Design Effect	Sample Size	Population
Atlantic Provinces	3.29	603	162,892
Quebec	3.20	562	461,430
Ontario	3.65	707	770,410
Manitoba and Saskatchewan	1.70	430	117,479
Alberta	2.16	267	210,272
British Columbia	2.39	303	290,439
Western Provinces	2.67	1,000	618,190
Canada	**3.07**	**2,872**	**2,012,923**

Province and Region for MOTHER = 0 and TYPE = 2 or 4	Design Effect	Sample Size	Population
Atlantic Provinces	3.02	806	192,840
Quebec	2.37	960	753,158
Ontario	2.02	1,368	1,236,284
Manitoba and Saskatchewan	2.29	897	205,530
Alberta	2.55	556	347,366
British Columbia	1.73	642	479,931
Western Provinces	2.49	2,095	1,032,826
Canada	**1.87**	**5,229**	**3,215,109**

All coefficients of variation in the Approximate Sampling Variability Tables are approximate and, therefore, unofficial. Estimates of actual variance for specific variables may be obtained from Statistics Canada on a cost-recovery basis. Since the approximate CV is conservative, the use of actual variance estimates may cause the estimate to be switched from one quality level to another. For instance a *marginal* estimate could become *acceptable* based on the exact CV calculation.

Remember: If the number of observations on which an estimate is based is less than 30, the weighted estimate is most likely unacceptable and Statistics Canada recommends not to release such an estimate, regardless of the value of the coefficient of variation.

10.1 How to Use the Coefficient of Variation Tables for Categorical Estimates

The following rules should enable the user to determine the approximate coefficients of variation from the Approximate Sampling Variability Tables for estimates of the number, proportion or percentage of the surveyed population possessing a certain characteristic and for ratios and differences between such estimates.

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

The coefficient of variation depends only on the size of the estimate itself. On the Approximate Sampling Variability Table for the appropriate geographic area, locate the estimated number in the left-most column of the table (headed “Numerator of Percentage”) and follow the asterisks (if any) across to the first figure encountered. This figure is the approximate coefficient of variation.

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

The coefficient of variation of an estimated proportion or percentage depends on both the size of the proportion or percentage and the size of the total upon which the proportion or percentage is based. Estimated proportions or percentages are relatively more reliable than the corresponding estimates of the numerator of the proportion or percentage, when the proportion or percentage is based upon a sub-group of the population. For example, the proportion of unemployed receiving regular Employment Insurance (EI) benefits during the reference week is more reliable than the estimated number of unemployed receiving regular EI benefits during the reference week. (Note that in the tables the coefficients of variation decline in value reading from left to right).

When the proportion or percentage is based upon the total population of the geographic area covered by the table, the CV of the proportion or percentage is the same as the CV of the numerator of the proportion or percentage. In this case, Rule 1 can be used.

When the proportion or percentage is based upon a subset of the total population (e.g. those in a particular sex or age group) reference should be made to the proportion or percentage (across the top of the table) and to the numerator of the proportion or percentage (down the left side of the table). The intersection of the appropriate row and column gives the coefficient of variation.

Rule 3: Estimates of Differences Between Aggregates or Percentages

The standard error of a difference between two estimates is approximately equal to the square root of the sum of squares of each standard error considered separately. That is, the standard error of a difference $\left(\hat{d} = \hat{X}_1 - \hat{X}_2\right)$ is:

$$\sigma_{\hat{d}} \sqrt{\left(\hat{X}_1 \alpha_1\right)^2 + \left(\hat{X}_2 \alpha_2\right)^2}$$

where $\hat{X}_1$ is estimate 1, $\hat{X}_2$ is estimate 2, and α_1 and α_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively. The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d}$. This formula is accurate for the difference between separate and uncorrelated characteristics, but is only approximate otherwise.

Rule 4: Estimates of Ratios

In the case where the numerator is a subset of the denominator, the ratio should be converted to a percentage and Rule 2 applied. This would apply, for example, to the case where 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 standard error of the ratio of the estimates is approximately equal to the square root of the sum of squares of each coefficient of variation considered separately multiplied by $\hat{R}$. That is, the standard error of a ratio $\left(\hat{R} = \hat{X}_1 / \hat{X}_2\right)$ is:

$$\sigma_{\hat{R}} = \hat{R} \sqrt{\alpha_1^{\,2} + \alpha_2^{\,2}}$$

where α_1 and α_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively. The coefficient of variation of $\hat{R}$ is given by $\sigma_{\hat{R}} / \hat{R}$. The formula will tend to overstate the error if $\hat{X}_1$ and $\hat{X}_2$ are positively correlated and understate the error if $\hat{X}_1$ and $\hat{X}_2$ are negatively correlated.

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.

10.1.1 Examples of Using the Coefficient of Variation Tables for Categorical Estimates

The following examples based on the EICS 2003 are included to assist users in applying the foregoing rules. Please note that the data for these examples are different than the results obtained from the current survey and are only to be used as a guide.

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

Suppose that a user estimates that 400,393 unemployed individuals received regular EI benefits during the reference week. How does the user determine the coefficient of variation of this estimate?

1) Refer to the coefficient of variation table for TYPE = 1 CANADA.

Employment Insurance Coverage Survey, 2000 to 2003

Approximate Sampling Variability Tables - TYPE = 1 Canada

NUMERATOR OF PERCENTAGE ('000)	ESTIMATED PERCENTAGE													
	0.1%	1.0%	2.0%	5.0%	10.0%	15.0%	20.0%	25.0%	30.0%	35.0%	40.0%	50.0%	70.0%	90.0%
1	98.9	98.4	97.9	96.4	93.9	91.2	88.5	85.7	82.8	79.8	76.6	70.0	54.2	31.3
2	******	69.6	69.3	68.2	66.4	64.5	62.6	60.6	58.5	56.4	54.2	49.5	38.3	22.1
3	******	56.8	56.6	55.7	54.2	52.7	51.1	49.5	47.8	46.1	44.2	40.4	31.3	18.1
4	******	49.2	49.0	48.2	46.9	45.6	44.2	42.8	41.4	39.9	38.3	35.0	27.1	15.6
5	******	44.0	43.8	43.1	42.0	40.8	39.6	38.3	37.0	35.7	34.3	31.3	24.2	14.0
6	******	40.2	40.0	39.4	38.3	37.2	36.1	35.0	33.8	32.6	31.3	28.6	22.1	12.8
7	******	37.2	37.0	36.5	35.5	34.5	33.4	32.4	31.3	30.2	29.0	26.4	20.5	11.8
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
250	******	******	******	******	******	******	******	5.4	5.2	5.0	4.8	4.4	3.4	2.0
300	******	******	******	******	******	******	******	4.9	4.8	4.6	4.4	4.0	3.1	1.8
350	******	******	******	******	******	******	******	******	4.4	4.3	4.1	3.7	2.9	1.7
400	******	******	******	******	******	******	******	******	******	**4.0**	3.8	3.5	2.7	1.6
450	******	******	******	******	******	******	******	******	******	******	3.6	3.3	2.6	1.5
500	******	******	******	******	******	******	******	******	******	******	******	3.1	2.4	**1.4**
750	******	******	******	******	******	******	******	******	******	******	******	******	2.0	1.1
1,000	******	******	******	******	******	******	******	******	******	******	******	******	******	1.0

NOTE: For correct usage of these tables, please refer to the microdata documentation.

2) The estimated aggregate (400,393) does not appear in the left-hand column (the "Numerator of Percentage" column), so it is necessary to use the figure closest to it, namely 400,000.

3) The coefficient of variation for an estimated aggregate is found by referring to the first non-asterisk entry on that row, namely, 4.0%.

4) So the approximate coefficient of variation of the estimate is 4.0%. The finding that there were 400,393 (to be rounded according to the rounding guidelines in Section 9.1) unemployed individuals received regular EI benefits during the reference week is publishable with no qualifications.

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

Suppose that the user estimates that 605,777 / 740,586 = 81.8% 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) Refer to the coefficient of variation table for TYPE = 1 CANADA.

2) Because the estimate is a percentage which is based on a subset of the total population (i.e., unemployed individuals potentially eligible to receive EI benefits), it is necessary to use both the percentage (81.8%) and the numerator portion of the percentage (605,777) in determining the coefficient of variation.

3) The numerator, 605,777, does not appear in the left-hand column (the "Numerator of Percentage" column) so it is necessary to use the figure closest to it, namely 500,000. Similarly, the percentage estimate does not appear as any of the column headings, so it is necessary to use the percentage closest to it, 90.0%.

4) The figure at the intersection of the row and column used, namely 1.4% is the coefficient of variation to be used.

5) So the approximate coefficient of variation of the estimate is 1.4%. The finding that 81.8% 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 543,846 / 718,300 = 75.7% of the regular employed population in Quebec contributed to EI, while 775,530 / 1,172,069 = 66.2% of the regular population in Ontario contributed to EI. How does the user determine the coefficient of variation of the difference between these two estimates?

1) Using the MOTHER = 0 and TYPE = 2 or 4 QUEBEC coefficient of variation table and the MOTHER = 0 and TYPE = 2 or 4 ONTARIO coefficient of variation table in the same manner as described in Example 2 gives the CV of the estimate for Quebec as 2.5%, and the CV of the estimate for Ontario as 2.5%.

Employment Insurance Coverage Survey, 2000 to 2003

Approximate Sampling Variability Tables - MOTHER = 0 and TYPE = 2 or 4 Quebec

NUMERATOR OF PERCENTAGE	ESTIMATED PERCENTAGE													
('000)	0.1%	1.0%	2.0%	5.0%	10.0%	15.0%	20.0%	25.0%	30.0%	35.0%	40.0%	50.0%	70.0%	90.0%
1	******	103.2	102.7	101.1	98.4	95.7	92.8	89.9	86.8	83.6	80.4	73.4	56.8	32.8
2	******	73.0	72.6	71.5	69.6	67.6	65.6	63.5	61.4	59.1	56.8	51.9	40.2	23.2
3	******	59.6	59.3	58.4	56.8	55.2	53.6	51.9	50.1	48.3	46.4	42.4	32.8	18.9
4	******	51.6	51.4	50.6	49.2	47.8	46.4	44.9	43.4	41.8	40.2	36.7	28.4	16.4
5	******	46.2	45.9	45.2	44.0	42.8	41.5	40.2	38.8	37.4	35.9	32.8	25.4	14.7
6	******	42.1	41.9	41.3	40.2	39.1	37.9	36.7	35.4	34.1	32.8	30.0	23.2	13.4
7	******	39.0	38.8	38.2	37.2	36.2	35.1	34.0	32.8	31.6	30.4	27.7	21.5	12.4
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
150	******	******	******	******	******	******	******	7.3	7.1	6.8	6.6	6.0	4.6	2.7
200	******	******	******	******	******	******	******	******	1.0	5.9	5.7	5.2	4.0	2.3
250	******	******	******	******	******	******	******	******	******	******	5.1	4.6	3.6	2.1
300	******	******	******	******	******	******	******	******	******	******	******	4.2	3.3	1.9
350	******	******	******	******	******	******	******	******	******	******	******	3.9	3.0	1.8
400	******	******	******	******	******	******	******	******	******	******	******	******	2.8	1.6
450	******	******	******	******	******	******	******	******	******	******	******	******	2.7	1.5
500	******	******	******	******	******	******	******	******	******	******	******	******	**2.5**	1.5

NOTE: For correct usage of these tables, please refer to the microdata documentation.

Employment Insurance Coverage Survey, 2000 to 2003

Approximate Sampling Variability Tables - MOTHER = 0 and TYPE = 2 or 4 Ontario

NUMERATOR OF PERCENTAGE	ESTIMATED PERCENTAGE													
('000)	0.1%	1.0%	2.0%	5.0%	10.0%	15.0%	20.0%	25.0%	30.0%	35.0%	40.0%	50.0%	70.0%	90.0%
1	122.9	122.4	121.8	119.9	116.7	113.4	110.0	106.5	102.9	99.2	95.3	87.0	67.4	38.9
2	******	86.5	86.1	84.8	82.5	80.2	77.8	75.3	72.8	70.1	67.4	61.5	47.6	27.5
3	******	70.7	70.3	69.2	67.4	65.5	63.5	61.5	59.4	57.2	55.0	50.2	38.9	22.5
4	******	61.2	60.9	59.9	58.3	56.7	55.0	53.3	51.4	49.6	47.6	43.5	33.7	19.4
5	******	54.7	54.4	53.6	52.2	50.7	49.2	47.6	46.0	44.3	42.6	38.9	30.1	17.4
6	******	50.0	49.7	48.9	47.6	46.3	44.9	43.5	42.0	40.5	38.9	35.5	27.5	15.9
7	******	46.3	46.0	45.3	44.1	42.9	41.6	40.3	38.9	37.5	36.0	32.9	25.5	14.7
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
350	******	******	******	******	******	******	******	******	5.5	5.3	5.1	4.6	3.6	2.1
400	******	******	******	******	******	******	******	******	******	5.0	4.8	4.3	3.4	1.9
450	******	******	******	******	******	******	******	******	******	******	4.5	4.1	3.2	1.8
500	******	******	******	******	******	******	******	******	******	******	******	3.9	3.0	1.7
750	******	******	******	******	******	******	******	******	******	******	******	******	**2.5**	1.4
1,000	******	******	******	******	******	******	******	******	******	******	******	******	******	1.2

NOTE: For correct usage of these tables, please refer to the microdata documentation.

2) Using Rule 3, the standard error of a difference $\left(\hat{d} = \hat{X}_1 - \hat{X}_2\right)$ is:

$$\sigma_{\hat{d}} = \sqrt{\left(\hat{X}_1\alpha_1\right)^2 + \left(\hat{X}_2\alpha_2\right)^2}$$

where $\hat{X}_1$ is estimate 1 (Quebec), $\hat{X}_2$ is estimate 2 (Ontario), and α_1 and α_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively.

That is, the standard error of the difference $\hat{d} = 0.757 - 0.662 = 0.095$ is:

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{\left[(0.757)(0.025)\right]^2 + \left[(0.662)(0.025)\right]^2} \\ &= \sqrt{(0.0003581) + (0.0002739)} \\ &= 0.025\end{aligned}$$

3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d} = 0.025 / 0.095 = 0.263$

4) So the approximate coefficient of variation of the difference between the estimates is 26.3%. The difference between the estimates is considered marginal and Statistics Canada recommends that this estimate be flagged with the letter E (or some similar identifier) and be accompanied by a warning to caution subsequent users about the high levels of error associated with the estimate.

Example 4: Estimates of Ratios

Suppose that the user estimates that 543,846 of the regular employed population in Quebec contributed to EI, while 775,530 of the regular 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 employed individuals in Quebec who contributed to EI. The denominator of the estimate ($\hat{X}_2$) is the number of employed individuals in Ontario who contributed to EI.

2) Refer to the coefficient of variation tables for MOTHER = 0 and TYPE = 2 or 4 QUEBEC and MOTHER = 0 and TYPE = 2 or 4 ONTARIO.

3) The numerator of this ratio estimate is 543,846. The figure closest to it is 500,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 2.5%.

4) The denominator of this ratio estimate is 775,530. The figure closest to it is 750,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 2.5%

5) So the approximate coefficient of variation of the ratio estimate is given by Rule 4, which is:

$$\alpha_{\hat{R}} = \sqrt{\alpha_1^{\,2} + \alpha_2^{\,2}}$$

where α_1 and α_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively. That is:

$$\begin{aligned}\alpha_{\hat{R}} &= \sqrt{(0.025)^2 + (0.025)^2} \\ &= \sqrt{0.000625 + 0.000625} \\ &= 0.035\end{aligned}$$

6) The obtained ratio of Quebec versus Ontario individuals in the regular employed population contributing to EI is 543,846 / 775,530 which is 0.70 (to be rounded according to the rounding guidelines in Section 9.1). The coefficient of variation of this estimate is 3.5%, 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 employed population who contributed to EI, to individuals aged 25 to 44 years in the regular employed population who contributed to EI is 1.24 for Manitoba and Saskatchewan, while it is 1.21 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 approximate coefficient of variation for the Manitoba and Saskatchewan ratio ($\hat{R}_1$) and the Alberta ratio ($\hat{R}_2$) as in Example 4. The approximate CV for the Manitoba and Saskatchewan ratio is 12.8% and 11.3% for Alberta.

2) Using Rule 3, the standard error of a difference ($\hat{d} = \hat{R}_1 - \hat{R}_2$) is:

$$\sigma_{\hat{d}} = \sqrt{\left(\hat{R}_1\alpha_1\right)^2 + \left(\hat{R}_2\alpha_2\right)^2}$$

where α_1 and α_2 are the coefficients of variation of $\hat{R}_1$ and $\hat{R}_2$ respectively. That is, the standard error of the difference $\hat{d}$ = 1.24 – 1.21 = 0.03 is:

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{\left[(1.24)(0.128)\right]^2 + \left[(1.21)(0.113)\right]^2} \\ &= \sqrt{(0.025192) + (0.018695)} \\ &= 0.209\end{aligned}$$

3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d}$ = 0.209 / 0.03 = 6.967.

4) So the approximate coefficient of variation of the difference between the estimates is 696.7%. 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 Coefficient of Variation Tables 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 $\left(\hat{X} - k, \hat{X} + k\right)$ 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 from the Approximate Sampling Variability Tables by first determining from the appropriate table the coefficient of variation of the estimate $\hat{X}$, and then using the following formula to convert to a confidence interval ($CI_{\hat{x}}$):

$$CI_{\hat{x}} = \left(\hat{X} - t\hat{X}\alpha_{\hat{x}}, \hat{X} + t\hat{X}\alpha_{\hat{x}}\right)$$

where $\alpha_{\hat{x}}$ is the determined coefficient of variation of $\hat{X}$, and

$t = 1$ if a 68% confidence interval is desired;
$t = 1.6$ if a 90% confidence interval is desired;
$t = 2$ if a 95% confidence interval is desired;
$t = 2.6$ if a 99% confidence interval is desired.

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 Coefficient of Variation Tables to Obtain Confidence Limits

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:

$\hat{X}$ = 81.8% (or expressed as a proportion 0.818)

t = 2

$\alpha_{\hat{x}}$ = 1.4% (0.014 expressed as a proportion) is the coefficient of variation of this estimate as determined from the tables.

$CI_{\hat{x}}$ = {0.818 – (2) (0.818) (0.014), 0.818 + (2) (0.818) (0.014)}

$CI_{\hat{x}}$ = {0.818 – 0.023, 0.818 + 0.023}

$CI_{\hat{x}}$ = {0.795, 0.841}

With 95% confidence it can be said that between 79.5% and 84.1% of unemployed individuals who were potentially eligible to receive EI benefits were eligible to receive EI benefits.

10.3 How to Use the Coefficient of Variation Tables 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 on the difference $\hat{X}_1 - \hat{X}_2$ be $\sigma_{\hat{d}}$.

If $t = \dfrac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}}$ 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 Coefficient of Variation Tables 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 0.025. Hence,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0.757 - 0.662}{0.025} = \frac{0.095}{0.025} = 3.80$$

Since t = 3.80 is greater than 2, it must be concluded that there is a significant difference between the two estimates at the 0.05 level of significance.

10.4 Coefficients of Variation for Quantitative Estimates

For quantitative estimates, special tables would have to be produced to determine their sampling error. Since most of the variables for the EICS are primarily categorical in nature, this has not been done.

As a general rule, however, the coefficient of variation of a quantitative total will be larger than the coefficient of variation of the corresponding category estimate (i.e., the estimate of the number of persons contributing to the quantitative estimate). If the corresponding category estimate is not releasable, the quantitative estimate will not be either. For example, the coefficient of variation of the total number of unemployed receiving regular EI benefits would be greater than the coefficient of variation of the corresponding proportion of unemployed receiving regular EI benefits. Hence, if the coefficient of variation of the proportion is unacceptable (making the proportion not releasable), then the coefficient of variation of the corresponding quantitative estimate will also be unacceptable (making the quantitative estimate not releasable).

Coefficients of variation of such estimates can be derived as required for a specific estimate using a technique known as pseudo replication. This involves dividing the records on the microdata files into subgroups (or replicates) and determining the variation in the estimate from replicate to replicate. Users wishing to derive coefficients of variation for quantitative estimates may contact Statistics Canada for advice on the allocation of records to appropriate replicates and the formulae to be used in these calculations.

10.5 Coefficient of Variation Tables

Refer to EICS2010_CVTabsE.doc for the coefficient of variation tables.

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.

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 mothers, the fraction is two-sixths.

2) An adjustment to account for the EICS sub-sampling (refer to Section 5.5.1).

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. Human Resources and Skills 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 mothers of infants less than one year old, is excluded from this calibration.

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