

# **Microdata User Guide**

## **ONTARIO MATERIAL DEPRIVATION SURVEY**

2009



Statistics  
Canada

Statistique  
Canada

Canada



## Table of Contents

<b>1.0</b>	<b>Introduction</b>	<b>5</b>
<b>2.0</b>	<b>Background</b>	<b>7</b>
<b>3.0</b>	<b>Objectives</b>	<b>9</b>
<b>4.0</b>	<b>Concepts and Definitions</b>	<b>11</b>
4.1	Labour Force Survey Concepts and Definitions	11
4.2	Ontario Material Deprivation Survey Concepts and Definitions	12
<b>5.0</b>	<b>Survey Methodology</b>	<b>13</b>
5.1	Population Coverage	13
5.2	Sample Design	13
5.2.1	Primary Stratification	13
5.2.2	Types of Areas	13
5.2.3	Secondary Stratification	14
5.2.4	Cluster Delineation and Selection	14
5.2.5	Dwelling Selection	15
5.2.6	Person Selection	15
5.3	Sample Size	15
5.4	Sample Rotation	15
5.5	Modifications to the Labour Force Survey Design for the Ontario Material Deprivation Survey	15
5.6	Sample Size for the Ontario Material Deprivation Survey	16
<b>6.0</b>	<b>Data Collection</b>	<b>17</b>
6.1	Interviewing for the Labour Force Survey	17
6.2	Supervision and Quality Control	17
6.3	Non-response to the Labour Force Survey	17
6.4	Data Collection Modifications for the Ontario Material Deprivation Survey	18
6.5	Non-response to the Ontario Material Deprivation Survey	18
<b>7.0</b>	<b>Data Processing</b>	<b>19</b>
7.1	Data Capture	19
7.2	Editing	19
7.3	Coding of Open-ended Questions	19
7.4	Imputation	19
7.5	Creation of Derived Variables	20
7.6	Weighting	20
<b>8.0</b>	<b>Data Quality</b>	<b>23</b>
8.1	Response Rates	23
8.2	Survey Errors	23
8.2.1	The Frame	23
8.2.2	Data Collection	24
8.2.3	Data Processing	24
8.2.4	Non-response	24
8.2.5	Measurement of Sampling Error	26

<b>9.0</b>	<b>Guidelines for Tabulation, Analysis and Release.....</b>	<b>29</b>
9.1	Rounding Guidelines.....	29
9.2	Sample Weighting Guidelines for Tabulation.....	29
9.3	Definitions of Types of Estimates: Categorical and Quantitative.....	30
9.3.1	Categorical Estimates.....	30
9.3.2	Quantitative Estimates.....	30
9.3.3	Tabulation of Categorical Estimates.....	30
9.3.4	Tabulation of Quantitative Estimates.....	31
9.4	Guidelines for Statistical Analysis.....	31
9.5	Coefficient of Variation Release Guidelines.....	32
<b>10.0</b>	<b>Bootstrap Method for Variance.....</b>	<b>35</b>
10.1	Statistical Package for Variance Estimation.....	35
<b>11.0</b>	<b>Weighting.....</b>	<b>37</b>
11.1	Weighting Procedures for the Labour Force Survey.....	37
11.2	Weighting Procedures for the Ontario Material Deprivation Survey.....	38
<b>12.0</b>	<b>Questionnaires.....</b>	<b>41</b>
12.1	The Labour Force Survey Questionnaire.....	41
12.2	The Ontario Material Deprivation Survey Questionnaire.....	41
<b>13.0</b>	<b>Record Layout with Univariate Frequencies.....</b>	<b>43</b>

## **1.0 Introduction**

The Ontario Material Deprivation Survey (OMDS) was conducted by Statistics Canada in March and April 2009 with the support of the Cabinet Office of the Government of Ontario. This manual has been produced to facilitate the manipulation of the microdata file based on the survey results.

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

### Statistics Canada

Andrew Heisz  
Income Statistics Division  
Telephone: 613-951-3748  
Fax: 613-951-0080  
E-mail: [Andrew.Heisz@statcan.gc.ca](mailto:Andrew.Heisz@statcan.gc.ca)  
Client Services  
Income Statistics Division  
Telephone: 613-951-7355 or call toll-free 1-888-297-7355  
Fax: 613-951-3012  
E-mail: [income@statcan.gc.ca](mailto:income@statcan.gc.ca)

### Cabinet Office of the Government of Ontario

Bruce Baldwin, Executive Coordinator  
Health, Social, Education and Children's Policy  
Cabinet Office  
99 Wellesley St. W, Rm, 4340  
Toronto, Ontario  
M7A 1A1  
Telephone: 416-325-7801  
E-mail: [Bruce.Baldwin@ontario.ca](mailto:Bruce.Baldwin@ontario.ca)



## **2.0 Background**

The Ontario Material Deprivation Survey (OMDS) was conducted on behalf of the Ontario Government. The aim is to gauge the ability of families to satisfy basic material needs such as food, clothing, housing as well as social needs of participation and leisure. The survey includes ten items deemed to be necessities, due to lack of resources. The list of ten items comes from a 2008 study by the Daily Bread Food Bank of Toronto. It also includes questions on income. The OMDS was fielded in March and April 2009 as a supplement to the Labour Force Survey.





### **3.0 Objectives**

The survey was designed to produce estimates of the incidence of missing two or more items (out of ten) for Ontario and various geographic subgroups. Estimates were to be produced at the individual level. It was also designed to provide contextual data on demographic and socioeconomic characteristics of Ontario families and individuals.



## 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 Ontario Material Deprivation Survey (OMDS) are given in Section 4.2. Users are referred to Chapter 12.0 of this document for a copy of the actual survey questionnaire(s) used.

### 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<sup>1</sup> 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<sup>2</sup>; 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.

---

<sup>1</sup> 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.

<sup>2</sup> 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. Since 1997, these statistics have been based on the North American Industry Classification System (NAICS) and the Standard Occupational Classification (SOC-91). Prior to 1997, the 1980 Standard Industrial Classification and the 1980 Standard Occupational Classification were used.

### **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 15<sup>th</sup> 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 than 30 hours per week at their main or only job.

### **Economic Family**

Economic family refers to a group of two or more persons who live in the same dwelling and are related to each other by blood, marriage, common-law or adoption.

## **4.2 Ontario Material Deprivation Survey Concepts and Definitions**

### **Material Deprivation**

Material deprivation refers to a state where a person or family is without an item or denied an activity deemed to be a necessity because of lack of resources.

### **Low Income Cut-off (LICO)**

The LICO is an income threshold below which a family is expected to spend 20 percentage points more than average on food, shelter, and clothing. The LICOs are calculated separately for different family sizes (1, 2, 3, 4, 5, 6, and 7 or more members) residing in areas of different population size (rural, urban below 30,000, 30,000 to 99,000, 100,000 to 499,999, and 500,000 and over). Statistics Canada produces before- and after- tax LICOs for each calendar year. For the actual values, please see the Statistics Canada's publication *Low income cut-offs for 2008 and low income measures for 2007*, Catalogue no. 75F0002M-No. 002

### **Area size**

The area size variable is based on the urban/rural status of the enumeration area (defined by Statistics Canada) in which the majority of the postal codes fall. Urban areas have minimum population concentrations of 1,000 people and a population density of at least 400 people per square kilometre based on the 2001 Census population counts. All the territory outside the urban areas is considered rural.

## **5.0 Survey Methodology**

The Ontario Material Deprivation Survey (OMDS) was administered in March and April 2009 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.<sup>3</sup> Sections 5.5 and 5.6 describe how the OMDS departed from the basic LFS design in March and April 2009.

### **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 in Canada's 10 provinces. Specifically excluded from the survey's coverage are residents of the Yukon, Northwest Territories and Nunavut, persons living on Indian Reserves, full-time members of the Canadian Armed Forces and inmates of institutions. These groups together represent an exclusion of approximately 2% of the population aged 15 or over.

### **5.2 Sample Design**

The LFS has undergone an extensive redesign, culminating in the introduction of the new design at the end of 1994. 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.

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.

---

<sup>3</sup> A detailed description of the LFS design is available in the Statistics Canada publication entitled *Methodology of the Canadian Labour Force Survey*, Catalogue no. 71-526-XPB.

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 1991 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. LFS 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 102,000 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 Ontario Material Deprivation Survey**

The OMDS used two of the six rotation groups in the March LFS sample and three rotation groups in the April LFS sample. For the OMDS, coverage was at the economic family level. Unlike the LFS where information is collected for all eligible household members, the OMDS collected information about an economic family and its members. If the LFS household consisted

of more than one economic family, one of them was randomly selected to participate in the OMDS.

While the LFS carries forward information obtained during the initial interviews, the OMDS collected information from and about persons 70 years of age and over, as well as members of the armed forces not living on military bases.

## **5.6 Sample Size for the Ontario Material Deprivation Survey**

The following table shows the number of economic families as well as persons 15 years of age and older in the LFS sampled rotations who were eligible for the OMDS supplement. This table includes households which were non-respondents to the LFS.

<b>Sample Size</b>		
<b>Ontario households</b>	<b>Ontario economic families</b>	<b>Economic family members aged 15 and over</b>
12,787	12,787	25,418



## **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 15<sup>th</sup> 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 are usually 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 its 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 Ontario Material Deprivation Survey**

Upon completion of the Labour Force Survey interview, the interviewer asked to speak to the most knowledgeable person to participate in the OMDS interview. If this person was not available, the interviewer arranged for a convenient time to call back. The collection period was extended by one week following the LFS week to allow interviewers time to contact these individuals. The OMDS respondent was asked a set of material deprivation questions, as well as questions on sources of family income and personal income of family members aged 15 and over. In order to reduce response burden, income information was collected for up to and including four persons only within one economic family.

#### **6.5 Non-response to the Ontario Material Deprivation Survey**

The next stage of data collection for households responding to the LFS was to administer the OMDS. In the event that the supplementary survey was met with a refusal, a senior interviewer made another call to try to gain participation.

## **7.0 Data Processing**

The main output of the Ontario Material Deprivation Survey (OMDS) 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 Editing**

The first stage of survey processing undertaken at head office was the replacement of any “out-of-range” values on the data file with blanks. This process was designed to make further editing easier.

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.

The second type of error treated involved a lack of information in questions which should have been answered. For this type of error, a non-response or “not-stated” code was assigned to the item.

### **7.3 Coding of Open-ended Questions**

There were no open-ended questions in the OMDS.

### **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.

In the case of the OMDS, donor imputation was used to fill in missing data for non-response to the question asking for the total personal income in 2008. Some respondents did not want or were not able to report salary or personal income other than by selecting a range. In those cases a conversion to dollar values took place. A randomly selected value was imputed in place of a range. For the highest open range of “\$100,000 or more” a value of \$100,999 was used. The data file includes flags indicating where the original response was given as a range or was missing and later imputed. Further information on the imputation process is given in Chapter 8.0 (Data Quality).

## **7.5 Creation of Derived Variables**

A number of data items on the microdata file have been derived by combining items on the questionnaire in order to facilitate data analysis.

A measure of material deprivation – number of types of material deprivation experienced due to the lack of affordability - was derived from the set of material deprivation questions.

Total family income before tax was derived by adding incomes of all family members.

The LICO status was derived by comparison of the total family income with the low income cut-off values published by Statistics Canada.

## **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 Ontario families experiencing one or more types of material deprivation is to be estimated, it would be done by selecting the records referring to those families 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 reduction of the number of variables, grouping values into wider categories, and coding specific values into the “not stated” category. 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.

### **Differences between the OMDS master file and the public use file**

#### File structure

Unlike the master file, the OMDS public use file is designed as a family level file. The variables that originate from the OMDS supplement (material deprivation and family income) are at the family level on both files, while some of the LFS variables in order to be included on the OMDS PUMF had to be derived as family, not person, characteristics. These are: number of family members in each of three broad age groups, the highest education level of the major income earner, and immigrant status of the major income earner.

Several LFS variables describing the specifics of employment are not included on the public use file.

#### Grouped variables

The number of categories was reduced in the following variables: area size, dwelling type, type of economic family, immigrant status, family sources of income, and main source of income. The total family income, as well as the income of the major income earner, were converted to ranges.

#### Capped variables

The following variables were capped: family size (capped at 6), number of employed persons in the economic family, number of unemployed persons in the economic family, and number of family members in each of the three age groups. The top income range was capped at \$100,000 and the total weekly number of hours worked by all family members was capped at 80 hours.

#### Suppressions of values on selected records

Dwelling type was suppressed and replaced with “not stated” on 599 records, area size on 309 records, and the ownership of dwelling variable on 177 records. The suppressions of other variables (main source of income, number of unemployed, number of employed, highest level of education) affected a much lower number of records.

#### Recoding of variables

A small number of records had a demographic variable recoded to avoid potential identification of respondents resulting from an unusual combination of characteristics.



## 8.0 Data Quality

### 8.1 Response Rates

The following table summarizes the response rates to the Labour Force Survey (LFS) and to the Ontario Material Deprivation Survey (OMDS).

Province	LFS Selected Households	LFS Responding Households	LFS Response Rate* (%)	OMDS Responding Households	OMDS Response Rate** (%)	OMDS Overall Response Rate*** (%)
Ontario	12,787	12,058	94.3	10,703	88.8	83.7

**Note:** The OMDS overall response rate is based on all LFS household records. The LFS responding households include respondents carried forward from the previous month.

\* The LFS response rate is the number of LFS responding households as a percentage of the number of LFS selected households.

\*\* The OMDS response rate is the number of OMDS responding households as a percentage of the number of LFS responding households.

\*\*\* The OMDS overall response rate is the number of OMDS responding households as a percentage of the number of LFS selected households.

### 8.2 Survey Errors

The estimates derived from this survey are based on a sample of households. 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

Since the OMDS was a supplement to the LFS, the frame used was the LFS frame. Any non-response to the LFS had an impact on the OMDS frame. The quality of the sampling

variables in the frame was very high. The OMDS sample consisted of Ontario economic families in five LFS rotation groups. The criteria used for the OMDS selection (such as a rotation group) were not missing for any LFS records.

Note that the LFS frame excludes about 2% of all households in the 10 provinces of Canada. Therefore, the OMDS frame also excludes a similar proportion of Ontario households. It is unlikely that this exclusion introduces any significant bias into the survey data.

### **8.2.2 Data Collection**

Interviewer training consisted of reading the OMDS Interviewer's Manual, practicing with the OMDS 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 collected the OMDS information after the LFS information was collected. The collection period ran from March 22 to April 6, 2009 and from April 19 to May 4, 2009. In general, interviews went smoothly and quickly with an average interview time of slightly less than 5 minutes.

### **8.2.3 Data Processing**

Data processing of the OMDS was done in a number of steps including clean-up, pre-edit, verification, coding, flow editing, imputation and estimation. At each step a picture of the output files was taken and an easy verification was made comparing files at the current and previous step. This greatly improved the data processing stage.

During processing of the data, 10 OMDS household/economic family level records, as well as 4 records at the person level, did not match to corresponding records in the LFS. Thus they were coded as out-of-scope and were dropped from further processing. When supplementary survey records do not match to host survey records they must be dropped since a weight cannot be derived for them.

Responses to income questions were examined for consistency between reporting of wage/salary as a source of family income and the actual reporting of wage/salary values for any member of the family. Also, cases in which families reported no income for 2008 were analyzed using socio-demographic and labour information. Due to a low number of cases with possible discrepancies, the original responses were retained. Very low as well as very high incomes were also examined. The impact of these values on key estimates, such as the median household/economic family income, was determined to be minimal, so no corrective action was taken.

### **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, no member of the household was able to provide the information, or the respondent refused to participate in the survey. Total non-response was handled by adjusting the weight of households that responded to the survey to compensate for those that did not respond.

In most cases, partial 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.



The following table presents response levels to the set of deprivation questions. The item non-response was very low, varying from 0 to 0.8%. Although there were some differences between those families who responded to all the deprivation questions and those who did not answer one or more deprivation questions, given the fact that the item non-response is so low, the impact on the final estimates should be negligible.

The item response rate for the deprivation questions combined with the OMDS response rate was between 88.0% and 88.8%. The overall item response rate (accounting also for the LFS non-response) ranged from 83.0% and 83.7%.

Items - Type of Deprivation	Deprivation item - responding households	Item non-response rate (%)	Item response rate accounting for OMDS non-response* (%)	Overall item response rate accounting for OMDS and LFS non-response** (%)
Eat fresh fruit and vegetables every day	10,703	0	88.8	83.7
Get dental care if needed	10,673	0.3	88.5	83.5
Eat meat, fish or a vegetarian equivalent at least every <u>other</u> day	10,703	0	88.8	83.7
Able to replace or repair broken or damaged appliances such as a vacuum or a toaster	10,661	0.4	88.4	83.4
Have appropriate clothes for job interviews	10,614	0.8	88.0	83.0
Able to get around your community, either by having a car or by taking the bus or an equivalent mode of transportation	10,691	0.1	88.7	83.6
Able to have friends or family over for a meal at least once a month	10,674	0.3	88.5	83.5
House or apartment free of pests, such as cockroaches	10,670	0.3	88.5	83.4
Able to buy some small gifts for family or friends at least once a year	10,678	0.2	88.6	83.5
Have a hobby or leisure activity	10,684	0.2	88.6	83.6

\* The OMDS item response rate combined with the OMDS response rate of 88.8%

\*\* The OMDS item response rate combined with the OMDS response rate of 88.8% and the LFS response rate of 94.3%

All imputations involved donors that were selected using a score function. For each item non-response or partial non-response records (also called recipient records), we compared certain characteristics to characteristics from all the donors. When the characteristics were the same between a donor and the recipient, a value was added to the score of that donor. The donor with the highest score was deemed the “closest”

donor and was chosen to fill in missing pieces of information of the non-respondents. If there was more than one donor with the highest score, a random selection occurred. The pool of donors was made up in such a way that the imputed value assigned to the recipient, in conjunction with other non-imputed items from the recipient would still pass the edit.

Donor imputation was simultaneously conducted for both personal income and family income. Family income was derived by summing the total personal income of all members of a given family.

There were 341 high income families who were excluded from the donor pool in the donor imputation module for family income. Their reported income examined in the context of demographic and labour force characteristics was quite unique, so they did not meet the criteria of a good donor. Among these families, there were 56 with income much higher than expected. Their reported income could not be retained and had to be imputed.

The cases of imputation of personal income include records where it was imputed for 54 individuals who belonged to families with 5 or more adults. Their income information was missing because to reduce response burden personal income was collected for a maximum of 4 family members aged 15 and over.

	<b>Step 1 Personal Income</b>	<b>Step 2 Family Income</b>
Imputed	2,747	1,679
Total	21,529	10,703
Rate (%)	12.8	15.7

The OMDS imputation process worked well. A flag indicating which records had imputed income values is included on the Master File.

### **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 uses 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 survey results, one estimates that 5.9% of Ontario families reported that they can not afford to eat fresh fruit and vegetables every

day and this estimate is found to have a standard error of 0.003. Then the coefficient of variation of the estimate is calculated as:

$$\left( \frac{0.003}{0.059} \right) \times 100 \% = 5.1 \%$$

More information on the calculation of coefficients of variation can be found 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 Ontario Material Deprivation Survey (OMDS) 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 OMDS 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 OMDS.

#### **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 Ontario families who experienced at least one material deprivation or the percentage of children below age 15 who lived in Ontario families experiencing at least one deprivation 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.

##### Example of a Categorical Question:

Q: Do you and your family eat fresh fruit and vegetables every day?

R: Yes / No

Q: Is this because you cannot afford it, or for some other reason?

R: Cannot afford it / Some other reason

#### **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 income of Ontario families who experienced two or more material deprivations.

##### Example of a Quantitative Question:

Q: What is your best estimate of your total income from wages and salaries from all jobs, before deductions, including tips and commissions, during the year ending December 31, 2008?

R: |\_|\_|\_|\_|\_|\_| dollars

#### **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 the total number of deprivations experienced by Ontario families multiply the value reported in derived variable MDNUM (number of types of deprivations experienced) by the final weight for the record, then sum this value over all records with MDNUM > 0 (at least one deprivation reported).

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 deprivations experienced by Ontario families who experienced at least one deprivation,

- a) estimate the total number of deprivations ( $\hat{X}$ ) as described above,
- b) estimate the number of families ( $\hat{Y}$ ) in this category by summing the final weights of all records with MDNUM > 0, then
- c) divide estimate a) by estimate b) ( $\hat{X} / \hat{Y}$ ).

## 9.4 Guidelines for Statistical Analysis

The OMDS 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.

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

## **9.5 Coefficient of Variation Release Guidelines**

Before releasing and/or publishing any estimates from the OMDS, 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**

<b>Quality Level of Estimate</b>	<b>Guidelines</b>
1) Acceptable	<p>Estimates have a sample size of 30 or more, and low coefficients of variation in the range of 0.0% to 16.5%.</p> <p>No warning is required.</p>
2) Marginal	<p>Estimates have a sample size of 30 or more, and high coefficients of variation in the range of 16.6% to 33.3%.</p> <p>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.</p>
Unacceptable	<p>Estimates have a sample size of less than 30, or very high coefficients of variation in excess of 33.3%.</p> <p>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:</p> <p>“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.”</p>



## 10.0 Bootstrap Method for Variance

In order to determine the quality of the estimate and to calculate the coefficient of variation (CV), the standard deviation must be calculated. Confidence intervals also require the standard deviation of the estimate. The OMDS uses a multi-stage survey design and calibration, which means that there is no simple formula that can be used to calculate variance estimates. Therefore, an approximate method was needed. The Rao-Wu bootstrap method is used because the sample design and calibration needs to be taken into account when calculating variance estimates. The Rao-Wu bootstrap method does this, and with the use of the Bootvar program, discussed in the next section, is a method that is fairly easy for users.

The OMDS uses the Rao-Wu bootstrap method described in the 1987 paper: Resampling inference with complex survey data. *Journal of the American Statistical Association*, 83, 231-241.

Independently, in each stratum, a simple random sample of  $(n-1)$  of the  $n$  units in the sample is selected with replacement. Note that since the selection is with replacement, a unit may be chosen more than once. This step is repeated  $R$  times to form  $R$  bootstrap samples. For each of the  $R$  bootstrap samples, bootstrap weights are calculated for each unit in the bootstrap sample (units not selected in a given bootstrap sample are assigned a weight of zero). These bootstrap weights are based on the initial sample design weight, the number of times a given unit has been selected and the initial sample size as well as the bootstrap sample size. These weights are then adjusted according to the same weighting process as the regular weights: non-response adjustment, calibration, etc. The entire process (selecting simple random samples, recalculating weights for each stratum) is repeated several times, yielding  $R$  different bootstrap weights for each unit in the original sample. The OMDS uses  $R=1,000$ , to produce 1,000 bootstrap samples with 1,000 potential different weights for each unit.

The end result is 1,000 final bootstrap weights for each unit in the sample. The variation among the 1,000 possible estimates based on the 1,000 bootstrap weights are related to the variance of the estimator based on the regular weights and can be used to estimate it.

There are a number of reasons why a user may need to calculate the coefficient of variation of estimates with the bootstrap method. For example, although some standard statistical packages allow sampling weights to be incorporated in the analyses, the variances that are produced often do not properly take into account the design and/or calibration of the weights, whereas the Bootstrap variance program does. Also, for estimates of quantitative variables, separate tables are required to determine their sampling error.

### 10.1 Statistical Package for Variance Estimation

Statistics Canada has developed a program that can perform bootstrap variance estimation: the Bootvar program.

The Bootvar program is available in SAS or SPSS format. It is made up of macros that compute variances for totals, ratios, differences between ratios and for linear and logistic regression.

Bootvar may be downloaded from Statistics Canada's Research Data Centre (RDC) website. Users must accept the Bootvar Click-Wrap Licence before they can read the files. There is a document on the site explaining how to adapt the system to meet users' needs.

SAS: [http://www.statcan.gc.ca/rdc-cdr/bootvar\\_sas-eng.htm](http://www.statcan.gc.ca/rdc-cdr/bootvar_sas-eng.htm)

SPSS: [http://www.statcan.gc.ca/rdc-cdr/bootvar\\_spss-eng.htm](http://www.statcan.gc.ca/rdc-cdr/bootvar_spss-eng.htm)



## 11.0 Weighting

Since the Ontario Material Deprivation Survey (OMDS) 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 Ontario Material Deprivation Survey**

The principles behind the calculation of the weights for the OMDS 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 OMDS microdata file.

- 1) An adjustment to account for the use of a 5/6 (five-sixth) sub-sample, instead of the full LFS sample
- 2) An adjustment to account for the additional non-response to the supplementary survey i.e., non-response to the OMDS households that did respond to the LFS. The procedure is similar to the LFS non-response weight adjustment, but groupings are based on different variables.

At this stage the weight is comprised of two components: the inverse of the sampling rate and the non-response adjustment. A third component, the family weighting adjustment described below, was added to improve accuracy of estimates.

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. Using a linear regression model, auxiliary information is used to arrive at the final weight. The regression is set up to ensure that the final weights it produces sum to the census projections for the auxiliary variables, namely various age-sex groups, economic regions and census metropolitan areas. This improves the reliability of estimates that can be produced by the OMDS.

At the same time as ensuring consistency with external census counts, the family weighting procedure also ensures that every member of the economic family is assigned the same weight.

The Master file has been created at the person level and, consequently, the weights on the file are “person weights”. To use a family weight, one has to select the weight of any family member. For example, when looking for an estimate of the number of families with no material deprivation, one can look at the weighted derived variable MDNUM = 0 where PRSNID = 1 (every family on the file includes a person identified as “1”).





## **12.0 Questionnaires**

### **12.1 The Labour Force Survey Questionnaire**

The Labour Force Survey questionnaire (LFS\_QuestE.pdf) is used to collect information on the current and most recent labour market activity of all household members 15 years of age or older. It includes questions on hours of work, job tenure, type of work, reason for hours lost or absent, job search undertaken, availability for work, and school attendance.

### **12.2 The Ontario Material Deprivation Survey Questionnaire**

The Ontario Material Deprivation Survey (OMDS) questionnaire was used in March and April 2009 to collect the information for the supplementary survey. The file OMDS2009\_QuestE.pdf contains the English questionnaire.



## **13.0 Record Layout with Univariate Frequencies**

See OMDS2009\_PUMF\_CdBk.pdf for the record layout with univariate frequencies.