

Methodology Branch

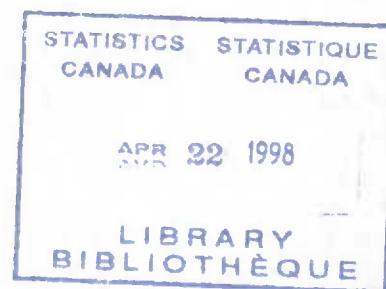
Household Survey  
Methods Division

Direction de la méthodologie

Division des méthodes  
d'enquêtes des ménages

**WORKING PAPER**

**METHODOLOGY BRANCH**



**EVALUATING LABOUR FORCE SURVEY DATA  
USING THE 1991 CENSUS OF POPULATION**

**HSMD-98-003 E**

**Benoit Allard, Isabelle Lévesque and Johane Dufour**

**Household Survey Methods Division**

**Statistics Canada**

**March 1998**

## **EVALUATING LABOUR FORCE SURVEY DATA USING THE 1991 CENSUS OF POPULATION**

Benoit Allard, Isabelle Lévesque and Johane Dufour  
Household Survey Methods Division

### **ABSTRACT**

A project to match data from the Labour Force Survey (LFS) with data from the 1991 Census of the population was undertaken at Statistics Canada. The main goal of this project was to conduct a number of studies on the quality of LFS data, using 1991 Census information on the population. The increasing use of the LFS as a general vehicle for household surveys, the significant number of supplements added each month to the LFS, and the implementation of new surveys at Statistics Canada are reasons that led to the realization of this large-scale project. Three studies have been carried out: i) a comparison of the profiles of respondents and nonrespondents, ii) a coverage study, and iii) a comparison of LFS and Census responses. For the first study, the linkage between both surveys was done at the household level and a database was built containing the characteristics of LFS nonrespondent households which were successfully matched to the Census. For the other two studies, the linkage process was pursued at the person level and a database containing LFS and Census information on matched and unmatched individuals was built.

## **ÉVALUER LES DONNÉES DE L'ENQUÊTE SUR LA POPULATION ACTIVE PAR L'ENTREMISE DU RECENSEMENT DE LA POPULATION DE 1991**

Benoit Allard, Isabelle Lévesque et Johane Dufour  
Division des méthodes d'enquêtes des ménages

### **RÉSUMÉ**

Un projet d'appariement entre les données de l'Enquête sur la population active (EPA) et celle du Recensement de la population de 1991 a été entrepris à Statistique Canada. Le but principal de ce projet consistait en la réalisation d'un certain nombre d'études sur la qualité des données de l'EPA en utilisant l'information provenant du Recensement de la population de 1991. L'utilisation grandissante de l'EPA comme véhicule général pour les enquêtes-ménages, le nombre important d'enquêtes supplémentaires qui se greffent à tous les mois à l'EPA ainsi que la réalisation de nouvelles enquêtes à Statistique Canada s'inscrivent au nombre des raisons qui ont motivé la réalisation de ce projet d'envergure. Trois études ont été réalisées: i) une comparaison du profil des répondants et des non-répondants, ii) une étude de la couverture et iii) une comparaison des réponses recueillies par l'EPA et par le Recensement. Pour la première étude, l'appariement entre les deux enquêtes a été effectué au niveau des ménages et une base de données, contenant les caractéristiques des ménages non-répondants de l'EPA qui étaient appariés au fichier du Recensement, a été construite. Pour ce qui est des deux autres études, la procédure d'appariement des données a été appliquée au niveau des personnes et une base de données contenant l'information provenant de l'EPA et du recensement a été créée pour les individus appariés et non appariés.

**TABLE OF CONTENTS**

<b>INTRODUCTION .....</b>	<b>1</b>
<b>1. OVERVIEW OF THE LABOUR FORCE SURVEY .....</b>	<b>3</b>
<b>2. METHODOLOGY OF THE STUDY .....</b>	<b>3</b>
2.1. Household Linkage .....	3
2.2. Person Linkage .....	5
<b>3. FIRST STUDY: CHARACTERISTICS OF LFS NONRESPONDENTS .....</b>	<b>6</b>
3.1. Household Size .....	7
3.2. Age and Sex .....	8
3.3. Labour Force Status .....	10
3.4. Household Income .....	12
3.5. Summary of Results .....	13
<b>4. SECOND STUDY: COVERAGE .....</b>	<b>14</b>
4.1. Dwellings Erroneously Identified as Vacant .....	15
4.2. Missed Persons .....	17
4.2.1. Labour Force Status .....	18
4.2.2. Age and Sex .....	18
4.2.3. Marital Status .....	20
4.3. Summary of Results .....	20
<b>5. THIRD STUDY: COMPARISON OF LFS AND CENSUS RESPONSES .....</b>	<b>21</b>
5.1. Comparison of Responses at the Dwelling Level .....	21
5.2. Comparison of Responses at the Individual Level .....	22
5.2.1. Comparison of Labour Force Status .....	22
5.2.2. Comparison of Age and Sex .....	24
5.2.3. Comparison of Marital Status .....	25
5.2.4. Comparison of Education Level .....	26
5.3. Summary of Results .....	28
<b>6. CONCLUSION AND FUTURE PLANS .....</b>	<b>29</b>
<b>ACKNOWLEDGMENTS .....</b>	<b>31</b>
<b>REFERENCES .....</b>	<b>33</b>
<b>APPENDIX A: FLOW CHART OF THE LINKAGE PROCESSES .....</b>	<b>35</b>
<b>APPENDIX B: MATCH RATES AT THE CMA/CA LEVEL .....</b>	<b>37</b>
<b>APPENDIX C: PERSON LINKAGE PROCESS .....</b>	<b>39</b>
<b>APPENDIX D: NEW CATEGORIZATION OF THE CENSUS EDUCATION VARIABLE ....</b>	<b>49</b>

---





## INTRODUCTION

The Labour Force Survey (LFS), like any survey, is subject to nonsampling errors coming from various sources. The LFS estimates are based primarily on the individuals and households selected for the survey who actually answer the questions presented by the LFS interviewers. Biases can be introduced when not all of the selected dwellings answer the survey, when the sample does not adequately represent the target population, or when some measurement error is introduced by the questionnaire and collection method.

To investigate these sources of bias, a project to link a sample of LFS dwellings to the 1991 Census database was suggested. Information from the Census could then be obtained for these dwellings. In order to access the Census data, a proposal to link LFS samples to the 1991 Census was submitted to the Data Access and Control Services Division late in 1994, since at Statistics Canada all projects involving record linkage and confidentiality issues require their approval by them. The subsequent acceptance of the project has generated the work presented in this paper.

Creating the linked databases required a great deal of work by the LFS Data Quality Unit. The resulting databases (individual and household levels) were used to perform three studies on the quality of LFS data. The first study undertaken is a comparison between respondent and nonrespondent households in the LFS using the Census information. The second is a coverage study that was carried out to learn more about the characteristics of the dwellings and the individuals who present coverage problems for the LFS. The third study is a comparison of LFS and Census responses to specific questions. The latter was done to evaluate the extent to which two different questionnaires and collection methods can produce different results, using the same sample. The linked databases can also be used for further studies on the LFS or its supplements.

This paper is divided into six sections. The following section briefly describes the Canadian Labour Force Survey. The second section presents the methodology of the study; the linkage process is then explained. In the next section, respondent and nonrespondent households in the LFS are compared. The paper then presents the results of the coverage study. The fifth section compares individuals' responses to specific questions in both surveys, the LFS and the 1991 Census. Finally, recommendations for future work are outlined in the last section of the paper.

---





## **1. OVERVIEW OF THE LABOUR FORCE SURVEY**

The LFS is the largest ongoing household survey conducted by Statistics Canada. This survey produces monthly estimates of labour force characteristics of the Canadian population, such as employment and unemployment at national and provincial levels, as well as by industry and occupation. The LFS uses a stratified multi-stage sampling plan with the dwelling as the final sampling unit. The sample is split into six representative sub-samples or panels, and each month the dwellings from one of the panels (one-sixth of the sample) are replaced. Thus, selected dwellings remain in the survey for six consecutive months. Approximately 59,000 households, representing 125,000 individuals aged 15 and over, are in the sample each month (Singh, Drew, Gambino, Mayda, 1990).

Statistics Canada's Regional Offices (RO) employ roughly 1,000 interviewers to conduct the LFS interviews. The first (or birth) interview with the household in the dwelling is conducted in person by an interviewer. The five subsequent interviews are mostly conducted by telephone.

## **2. METHODOLOGY OF THE STUDY**

In view of realizing the three studies on the quality of LFS data, it was necessary to obtain detailed information about households in the sample. The 1991 Census was the best candidate source for providing such information for several reasons. Firstly, the Census is a cost-effective means of obtaining information without imposing further burden on the respondent. Secondly, all dwellings that make up the population for the LFS are listed, and occupied dwellings are contacted by the Census enumerator. Moreover, Census nonresponse is low because individuals are legally bound to complete the questionnaire and because of the census follow-up procedure. Finally, the Census is appropriate since it is already used as a benchmark for the LFS. Consequently, the 1991 Census was retained as the best candidate to provide the information needed.

This section presents, in more detail, the process that permitted creation of the linked databases containing both LFS and Census information at the household and individual levels. The household match was the first step undertaken in this major linkage process and is described in the first subsection. Characteristics of nonrespondent households were then available and a comparison of nonrespondent and respondent households in the LFS could then be tackled. A person-to-person match was then pursued to study the characteristics of persons who present coverage problems for the LFS and to compare the responses of matched individuals to LFS and Census questions. The second subsection details the person linkage, which is the second step of the linkage process.

A flow chart of the household and person linkage processes is provided in Appendix A to help understand and visualize the steps followed while matching the LFS and the 1991 Census data.

### **2.1. Household Linkage**

The task of linking two huge sources of data, such as the LFS sample and the 1991 Census, was not an easy one. The first issue to resolve was the difference in the reference periods of these two surveys. The 1991 Census was held on June 4<sup>th</sup>, just in the middle of two LFS occurrences. The LFS is usually conducted during the week following the one which contains the fifteenth day of the month. Accordingly, it was decided that using the May and June 1991 LFS samples would provide a reasonable approximation. Moreover, to ensure

---

sufficient data for which to draw inferences on LFS nonrespondents, since the LFS nonresponse rate is approximately 5% each month, it was necessary to incorporate other months of the 1991 LFS. Households that rotated out between January and April 1991 and that were nonrespondents throughout their stay in the LFS sample were added to the LFS database. At this stage, the LFS file for the study contained 85,143 dwellings.

The LFS file and the Census file did not have a unique identifier that could be used to directly link the two files since, to ensure confidentiality, names and addresses are not captured by the Census, and the LFS captures only addresses. An intermediate file had to be used to allow the matching of the two files. The Address Register (AR) was the only existing vehicle at Statistics Canada that could be used for this task. The AR is a list of residential addresses covered by Geography Division's Area Master File at Statistics Canada (see Swain et al., 1991), which contains a unique matching key with the Census file. Since the LFS files contain addresses, the AR files would be initially linked to the LFS files using addresses, and, after successful completion of this linkage, the merged file would then be matched to the Census file. However, the AR only covers large urban areas - those with a population greater than 50,000; this represents about 50% of the Enumeration Areas (EAs) in the 1991 Census. Accordingly, using the AR imposed supplementary constraints to the study: only large urban areas could be studied. The initial LFS sample contained 56,250 urban dwellings (approximately 60% of the total). This number of dwellings was reduced to 33,190 because only urban areas with a population greater than 50,000 were covered by the AR. Although about 50% of all EAs were covered by the AR, less than 50% of the initial sample was covered (33,190 out of 85,143) because the LFS oversampled rural areas.

The matching key between the AR and the LFS file is the address. In order to perform this link, the LFS addresses had to be parsed to standardize street names, civic number, street designation, etc. (AR addresses are already standardized). The software PAAS (Postal Address Analysis System) was used to do the standardization. The use of PAAS allowed the automatic standardization of 92% of the LFS addresses. After manual editing, this success rate increased by another 7% to 99%. This left 32,861 addresses which had crossed the standardization process with success.

The next task consisted of matching the 32,861 LFS dwellings with a standardized address to the AR to obtain the key "Province / Federal District / Enumeration Area / Household Number" (PROV/FED/EA/HHLD number) that would permit the extraction of Census data. Several combinations of matching variables (street number, street name, street designator, apartment number, apartment designator, municipality, postal code) were used for automated matching. To be accepted, a match had to be exact and unique. After matching 78.8% of the addresses automatically, an additional 13.8% were matched manually. Most of these had not been matched automatically because of: 1) spelling errors in the LFS listing address; or 2) incorrect address standardization. As a result of this exercise, it was possible to obtain the Census key for 92.6% of the LFS addresses (30,422 dwellings).

The following step allowed the extraction of the Census data, but not without problems. When the AR was created, the data capture step was not subject to a 100% validation check. As a result, even if the AR gave a Census key for 30,422 LFS dwellings, only 29,696 (97.6%) were found in the Census file. Consequently, it is possible that 2% of Census extracted data are not the desired ones, if the data capture error rate is uniform through all AR files.

Another 71 dwellings had to be excluded from the matched sample because they were matched to a collective dwelling in the Census. Some dwellings were enumerated as collective dwellings by the Census, but as several private dwellings by the LFS.

---



The match rates at the Census Metropolitan Area / Census Agglomeration (CMA/CA) level ranged between 54.3% (in Halifax, Nova Scotia) and 94.4% (in Victoria, British Columbia) of the initial sample. Most of the match rates were between 80% and 90% (see Appendix B for more details).

*It is important to note that: to account for the uneven match rates, the initial LFS weights (which are the weights determined by the sample design) were adjusted by a factor equal to the inverse match rate, at the CMA/CA level. The initial LFS weights were used since a weight was required for nonresponding households as well as for vacant dwellings. These adjusted weights were used for both LFS and Census data.*

At this stage, another 851 matched dwellings were excluded from the analysis because they were not inside a CMA or CA even though they were matched using the Address Register. This left 28,774 dwellings for the LFS-Census file. For 5,677 of these dwellings (18.7%), Statistics Canada was in possession of long Census questionnaire (2B) information.

The resulting database (household level) of the household linkage step was used to conduct the first quality study i.e. the characteristics of nonrespondent households.

## 2.2. Person Linkage

In order to match individuals, vacant dwellings were removed from the LFS-Census database. Nonrespondent households in the LFS were also eliminated since person-to-person matches could not be performed. Table 2.1 displays the number of vacant and occupied dwellings found in both surveys after household linkage. The number of nonresponding households in the LFS (after household linkage) is also given.

**Table 2.1: Vacant and Occupied Dwellings in the LFS and the Census**

	Census		Total
	Vacant	Occupied	
<b>L</b>			
<b>F</b>			
<b>S</b>			
<b>Vacant</b>	580	813	1,393
<b>Nonresponse</b>	39	1,759	1,798
<b>Response</b>	134	25,449	25,583
<b>Total</b>	753	28,021	28,774

From Table 2.1, it can be seen that there was a total of 1,393 vacant dwellings in the LFS out of the 28,774 matched dwellings (4.8%). Among these, 813 were occupied by one or more persons in the Census while 580 were also vacant. These 813 dwellings represent undercoverage on the part of the Labour Force Survey. In total, 753 of the matched dwellings were vacant during Census day. However, 134 of them were occupied by one or more persons in the LFS; this can imply overcoverage for the LFS. Also, there were 1,798 LFS nonresponding households after household linkage.

This left 25,449 households (i.e. dwellings occupied in both surveys) for which the person-to-person matches had to be reconciled. Among these: 1) 22,790 dwellings had the same household size for the LFS as for the Census; 2) 1,818 dwellings had a smaller household size in LFS than in the Census; and 3) 841 dwellings had a greater household size in LFS than in the Census.

To perform the person-to-person matches, all of the demographic variables available for both surveys were used (age, sex, marital status and education level). Characteristics of the dwelling were also of use; namely the type of dwelling and the owned/rented flag. The person linkage was then realized by comparing all of these variables for the two surveys. While most of the Census variables have both imputed and unimputed values, the LFS variables used for matching do not. The values for each LFS variable are either the original ones or the imputed values (which happens rarely for the demographic variables) with no indication whether the value was imputed. To take account of this and to consider the fact that unimputed values must be used as often as possible to avoid the contamination of the values by the Census imputation methods, new variables combining the two types of values (imputed and unimputed) were created. These new variables were derived as follows: if the unimputed value was blank or invalid, the new variable took the same value as the imputed one; otherwise, the new variable would get the unimputed value. Only the age could not be handled that way because only the imputed age was available. But, since the unimputed month of birth (a flag indicating whether the date of birth is before or after June 3) and the unimputed decade and year of birth were available, an unimputed age could then be calculated. According to the results, 98.8% (65,734) of the persons in the Census database had equal values for the imputed and unimputed ages while 0.2% (101) had unequal values. For the remaining 1.0% (708), the unimputed age could not be determined.

The 101 cases with unequal imputed and unimputed values were scrutinized at more closely. For one of these cases, it was decided to use the unimputed age since it seemed to be the most realistic given the responses to other variables. For all other cases, both ages (imputed and unimputed) were considered in the linkage. At first, the person-to-person match was performed with the unimputed age and then, if the result was unsatisfactory, the match was done with the imputed age.

The person linkage was performed on the 25,449 occupied dwellings common to both surveys. These households contained 65,090 persons in the LFS and 66,543 persons in the Census. A total of 24,193 (95.1%) households were matched automatically and 30 (0.1%) more were matched manually. Altogether that translates to 24,223 (95.2%) matched households and 61,132 persons. Among these, 12,262 (20.1%) individuals had completed the long Census questionnaire (2B form). At the household level, 4,844 (20.0%) of the matched households had the 2B information (20% of dwellings in the Census received the 2B form).

After completion of the person linkage process, there were 1,226 (4.8%) unmatched households, 3,958 (6.1%) unmatched LFS persons and 5,411 (8.1%) unmatched Census persons. There were more unmatched persons in the Census compared to the LFS since the Census data contained more persons originally. For the unmatched households, the conclusion made after the person linkage process was to assume that the LFS and Census households which were linked together via the Address Register were not the same household. Perhaps the household linkage process led to the wrong Census key; or perhaps the household moved between the time the LFS was conducted and Census Day.

More details concerning the person linkage and its performance are given in Appendix C. The resulting database (individual level) of the person linkage step was used to perform second and third studies, i.e. the coverage study and the comparison between LFS and Census responses.

---



### **3. FIRST STUDY: CHARACTERISTICS OF LFS NONRESPONDENTS**

The LFS nonresponse rate usually fluctuates between 4% and 6%, and is usually higher in LFS supplements. Two techniques are used by the LFS to compensate for total nonresponse. The first one is applied to nonresponding households that responded the previous month. The labour force information is then simply carried forward from the previous month to the current month. However, this procedure is not applied for two consecutive months. The second technique corrects total nonresponse when there is no information from the previous month. Nonresponse is then compensated for by inflating the weight of responding households which belong to the same rotation group, employment insurance region and area (urban, rural, remote) with a factor equivalent to the inverse of the response rate. For that, the assumption is made that nonrespondents do not differ from the respondents, which, in practice, is not always the case (see Allard and Dufour, 1994). More information is then needed on the characteristics of nonrespondents. Moreover, since January 1997, the new LFS questionnaire which contains questions on earnings has been in place, which can have an impact on the nonresponse rate. Studying the income profile of both respondents and nonrespondents takes on greater importance. With information on nonrespondents taken from the Census, it is possible to study the differences in characteristics between respondents and nonrespondents, and thus evaluate the potential extent of nonresponse bias in the LFS.

With the approval of the linkage project between the LFS and the 1991 Census, a comparison between respondent and nonrespondent households in the LFS could then be performed. Using the first database (household level) built with household matches only (prior to person linkage, so as to avoid reducing the sample further), three different profiles were built: the nonrespondent profile, drawn from the matched Census data, and two respondent profiles: 1) from the matched Census data; and 2) from the original LFS data, using the full sample (matched and unmatched). Comparing these three profiles was the goal of this study and this section presents the results. In the first subsection, percentage distributions of the nonrespondent and respondent households are compared by household size. The second and third subsections present percentage distributions of individuals for the three categories of households (nonrespondent ones, respondent ones from the Census and respondent ones from the LFS) by age and sex and by labour force activity respectively. The fourth subsection relates to household income. The comparison is only made between nonrespondent and respondent LFS households using Census data since questions on earnings were not yet implemented in the LFS. The last subsection gives a summary of the study on the characteristics of nonrespondents.

#### **3.1. Household Size**

Weighted percentage distributions of nonrespondent and respondent households, broken down by household size, are presented in Table 3.1. The two respondent profiles - the one drawn from the Census data and the one drawn from the LFS data - are close to each other, indicating that the matching process did not introduce considerable bias in the estimates. However, the nonrespondent profile is considerably different from the respondent profiles. Single-person households represent about 2 out of every 5 nonrespondent households; for respondent households, only 1 in 4 is a single person. On average, nonrespondent households are smaller than respondent households by about 0.5 person (2.1 vs. 2.6 persons).

---

**Table 3.1: Weighted Distributions of Nonrespondent and Respondent Households by Household Size (%)**

Household size	Nonrespondent (Census data)	Respondent (Census data)	Respondent (LFS data)
1	39.5	24.0	24.7
2	32.8	31.5	31.7
3	14.2	17.6	17.2
4	7.8	17.1	17.1
5	3.4	6.9	6.8
6	1.6	2.0	1.6
7+	0.6	0.9	0.8
Total	100.0	100.0	100.0
Average hhld. size	2.10 persons	2.62 persons	2.60 persons

This finding - that larger households are more likely to respond - is consistent with previous studies. The more persons in a household, the easier it is for the interviewer to contact at least one member in that household. The current LFS weighting system does not account for the difference in household size between respondents and nonrespondents when adjusting for nonresponse. Some research is being done to see if household size distributions taken from external sources (such as the Census) could be used for post-stratification.

### 3.2. Age and Sex

Table 3.2 shows the weighted distributions of the individuals in nonrespondent and respondent households by sex and age group (0-14, 15-19, 20-24, 25-34, 35-44, 45-64 and 65+). Once again, the two respondent profiles (Census and LFS) are similar. Nonrespondents are close to respondents for sex over all ages (about 48.5% male and 51.5% female).

Looking at the age distribution for both sexes, it can be seen that nonrespondents are older on average (38.5 years) than respondents (about 35 years). Accordingly, there are fewer children (0-14) and teenagers (15-19) in nonrespondent households, as well as more middle-aged persons (45-64) and seniors (65+). This is consistent with nonrespondent households being smaller in size than respondent households (as seen in Table 3.1). Since children and teenagers usually live in larger households (i.e. with parents and siblings) and seniors tend to live in smaller households (i.e. without children), it is not surprising to observe that nonrespondents, who live in smaller households than respondents, are also slightly older.

Although the male and female populations have somewhat different age distributions, the differences between nonrespondents and respondents are similar for males and females. There is a higher percentage of persons aged over 45 among nonrespondents, and a lower percentage of persons under 20, for both males and females.

**Table 3.2: Weighted Distribution of Individuals  
in Nonrespondent and Respondent Households by Age and Sex (%)**

Sex	Age group	Nonrespondent (Census data)	Respondent (Census data)	Respondent (LFS data)
Male	All	48.8	48.5	48.4
Female		51.2	51.5	51.6
Total		100.0	100.0	100.0
Male	0-14	13.6	20.5	21.6
	15-19	5.4	6.8	6.7
	20-24	8.4	8.2	7.8
	25-34	21.1	18.4	18.3
	35-44	17.5	16.0	16.4
	45-64	22.8	20.6	20.2
	65+	11.1	9.5	9.1
	Total	100.0	100.0	100.0
Female	0-14	14.5	18.9	19.6
	15-19	4.3	6.1	6.2
	20-24	8.2	7.6	7.4
	25-34	16.8	17.6	17.8
	35-44	17.2	16.6	16.7
	45-64	23.5	20.3	19.8
	65+	15.6	12.8	12.6
	Total	100.0	100.0	100.0
Both	0-14	14.1	19.7	20.5
	15-19	4.8	6.4	6.4
	20-24	8.3	7.9	7.6
	25-34	18.9	18.0	18.1
	35-44	17.4	16.3	16.5
	45-64	23.2	20.4	20.0
	65+	13.4	11.2	10.9
	Total	100.0	100.0	100.0
Both	Average age	38.5 years	35.1 years	34.7 years

In Table 3.3, the weighted age distributions of nonrespondents and respondents are broken down by household size. For size 1 households, there is a large difference in the proportion of seniors: only 22% of nonrespondent single persons are over 65, compared to 35% for respondents. Also, the proportion of persons aged between 20 and 44 is larger for nonrespondents in single-person households than for respondents. Single adults between 20 and 44 years of age are usually more difficult to contact because they tend to be away from home more often; on the other hand, seniors can be reached more easily and are thus more likely to respond.



Interestingly, for any fixed household size larger than 1, the age distribution of nonrespondents is very similar to that of respondents. It seems, from the results of Table 3.3, that age does not have a major influence on nonresponse except in the case of single-person households, where nonresponse is higher for younger individuals.

**Table 3.3: Weighted Distribution of Individuals in Nonrespondent and Respondent Households by Age and Household Size (%)**

	Age group	Household size					Total
		1	2	3	4	5+	
<b>Nonrespondent (Census data)</b>	0-14	-	2.0	18.6	31.4	33.4	14.1
	15-19	0.4	1.3	7.2	8.5	11.2	4.8
	20-24	7.9	8.2	10.5	6.4	7.9	8.3
	25-34	23.7	19.7	21.2	16.1	10.7	18.9
	35-44	18.2	12.8	18.0	22.0	20.3	17.4
	45-64	27.5	31.6	21.3	13.8	12.1	23.2
	65+	22.3	24.4	3.3	1.9	4.5	13.4
	Total	100.0	100.0	100.0	100.0	100.0	100.0
<b>Respondent (Census data)</b>	0-14	-	3.3	18.6	30.7	34.9	19.7
	15-19	0.6	2.2	5.8	9.6	10.5	6.4
	20-24	5.1	8.0	9.0	7.4	8.5	7.9
	25-34	18.9	21.0	22.1	15.4	13.4	18.0
	35-44	14.9	12.1	15.8	20.9	16.7	16.3
	45-64	25.0	29.7	23.1	14.3	12.7	20.4
	65+	35.5	23.7	5.6	1.7	3.3	11.2
	Total	100.0	100.0	100.0	100.0	100.0	100.0
<b>Respondent (LFS data)</b>	0-14	-	3.4	19.7	32.1	37.0	20.5
	15-19	0.8	2.2	6.5	9.4	10.3	6.4
	20-24	5.2	8.4	8.9	6.9	7.3	7.6
	25-34	20.3	21.2	21.3	15.9	12.7	18.1
	35-44	15.8	12.0	16.2	21.2	16.6	16.5
	45-64	24.6	30.4	22.5	13.2	11.5	20.0
	65+	33.4	22.4	4.9	1.3	4.6	10.9
	Total	100.0	100.0	100.0	100.0	100.0	100.0

### 3.3. Labour Force Status

In Table 3.4, the weighted distribution of individuals in nonrespondent and respondent households is examined by labour force status and household size. Out-of-scope persons (children under 15 and full-time members of the armed forces) are not included in this table. Because only the long Census questionnaire has questions

on work, only about one-fifth of the sample could be used for profiles drawn from the Census. These results should be interpreted with caution as the nonrespondent sample is very small (292 households).

**Table 3.4: Weighted Distribution of Individuals in Nonrespondent and Respondent Households by Labour Force Status and Household Size (%)**

	Labour Force Status	Household Size			Total
		1	2	3+	
<b>Nonrespondent (Census data)</b>	Employed	58.3	58.4	67.8	62.2
	Unemployed	4.6	4.3	6.5	5.2
	Not in Labour Force	37.1	37.3	25.7	32.6
	Total	100.0	100.0	100.0	100.0
	Unemployment rate	7.3	6.9	8.7	7.8
<b>Respondent (Census data)</b>	Employed	49.1	56.5	68.6	62.9
	Unemployed	4.5	5.1	7.2	6.3
	Not in Labour Force	46.4	38.4	24.2	30.8
	Total	100.0	100.0	100.0	100.0
	Unemployment rate	8.4	8.3	9.5	9.1
<b>Respondent (LFS data)</b>	Employed	50.6	56.3	68.0	62.4
	Unemployed	5.2	5.4	7.6	6.7
	Not in Labour Force	44.2	38.3	24.4	30.9
	Total	100.0	100.0	100.0	100.0
	Unemployment rate	9.3	8.7	10.1	9.6

The distribution of persons by labour force characteristic (Employed, Unemployed, Not in Labour Force), and the Unemployment rate, are similar for the two respondent profiles. This suggests that the Census characteristics are close to what the LFS would have measured if they had responded. As for the nonrespondents (all household sizes combined), the proportion of Not in Labour Force is about 2 percentage point higher than for respondents and at least 1 percentage point lower for Unemployed and, consequently, the Unemployment rate is about 1.5 percentage point lower.

For households of size 2 and 3+, respondents and nonrespondents have similar labour force status distributions. However, for households of size 1 there are large differences. The proportion of Employed persons among nonrespondents is about 8 percentage points higher than for respondents, and the proportion of inactive persons (Not in Labour Force) is about 8 points lower. This is not surprising, since a single employed person is often away from home and is difficult to contact, while a person who is not in labour force is likely to spend more time at home.

As for the unemployment rates, they are lower for nonrespondents than for respondents for all household sizes. In households of size 2 and 3+, there are fewer unemployed persons among nonrespondents than among

respondents. As a result, the weight adjustment procedure used to compensate for LFS nonresponse may be introducing a bias in the labour force estimates.

From Table 3.1, it was noticed that the average size of nonrespondent households is less than that of respondent households by about 0.5. Table 3.5 was produced to see how this difference is distributed over the four labour force characteristics (Employed, Unemployed, Not in Labour Force and Out-of-scope). The table shows the average number of persons per household by labour force status. Again, the results for respondent census data compare closely to respondent LFS data. However, for each labour force status, the average number of persons in nonrespondent households is less than for respondent households. No particular status seems to carry the lion's share of the difference in average household size between nonrespondents and respondents. Nonrespondents have slightly lower participation and unemployment rates than respondents, but their employment-to-population ratio is about the same.

**Table 3.5: Average Number of Persons per Household by Labour Force Status**

Labour Force Status	Nonrespondent (Census data)	Respondent (Census data)	Respondent (LFS data)
Employed	1.09	1.32	1.28
Unemployed	0.09	0.13	0.14
Not in Labour Force	0.57	0.65	0.64
Out-of-scope	0.36	0.52	0.54
Total	2.10	2.62	2.60
Total in-scope	1.75	2.10	2.06
Participation rate (%)	67.4	69.2	69.1
Unemployment rate (%)	7.8	9.1	9.6
Employment-population ratio (%)	62.2	62.9	62.4

### 3.4. Household Income

It is interesting to examine the impact of household income on nonresponse, especially since new questions on earnings were added to the LFS questionnaire in January 1997. Table 3.6 shows the weighted distribution of total household income (for 1990, the reference year used in the 1991 Census) by household size for nonrespondent and respondent households. Since income was not measured by the LFS at the time of the study, only the Census profile is available for the respondents.

For any fixed household size, the percentage of households with an income less than \$20,000 is higher among nonrespondents than respondents. Also, the median income of nonrespondent households is somewhat less than the median income of respondent households. When all household sizes are combined, the difference between nonrespondent and respondent households is very large - a difference of more than \$10,000 in median income - mostly because of the smaller size of nonrespondent households.

Although these results are interesting, they do not fully reflect the situation of the new LFS questionnaire in that the new earnings questions are expected to affect nonresponse in a very different way. These results do



not tell how much the new earnings questions will influence a respondent's willingness to respond to future surveys, nor how this influence will depend on the respondent's income. Table 3.6 suggests that, without income questions in the questionnaire to influence response patterns, nonrespondents tend to have a lower income than respondents. With earnings questions added to the LFS questionnaire, the results may differ considerably.

**Table 3.6: Weighted Distribution of Nonrespondent and Respondent Households by Total Income and Household Size (%)**

		Household size			
		1	2	3+	Total
<b>Nonrespondent (Census data)</b>	<b>Total income</b>				
	\$0 - \$19,999	58.4	25.8	13.5	35.3
	\$20,000 - \$49,999	31.1	39.5	37.3	35.6
	\$50,000 +	10.5	34.7	49.2	29.1
	Total	100.0	100.0	100.0	100.0
	Median income	\$16,225	\$34,923	\$45,722	\$29,647
<b>Respondent (Census data)</b>	\$0 - \$19,999	51.4	17.9	9.0	22.0
	\$20,000 - \$49,999	39.9	44.3	31.5	37.5
	\$50,000 +	8.7	37.8	59.6	40.5
	Total	100.0	100.0	100.0	100.0
	Median income	\$17,595	\$38,179	\$54,377	\$40,000

### 3.5. Summary of Results

The main findings of the study on the characteristics of nonrespondents are that:

1. Smaller households are more prone to nonresponse than larger households;
2. In households of size one, nonrespondents are younger, more likely to be employed and less likely to be out of the Labour Force than respondents;
3. In households of two or more persons, respondents and nonrespondents have similar characteristics, except that nonrespondents have a slightly lower unemployment rate and a slightly lower household income.

The Allard and Dufour study of 1994 led to the same results, with the exception that nonrespondents had a slightly *higher* unemployment rate than respondents. (The LFS - Census 1991 linkage study applies only to areas covered by the AR while the Allard and Dufour study applied to the entire LFS sample.) However, the two studies do show consistent results in that single-person nonrespondents are more likely to be employed and less likely to be out of the Labour Force than respondents. Thus, while it is not clear what kind of bias (upward or downward) is being introduced to the unemployment rate with the current nonresponse adjustment

procedure, both studies agree that some upward bias is perhaps being introduced into the estimates of employment and of the participation rate.

#### 4. SECOND STUDY: COVERAGE

Coverage errors occur whenever the target population is not adequately represented by the sample. These errors can be introduced at several stages of the survey and generally result, for the LFS, in an undercoverage of the population. In the LFS-Census linkage project, undercoverage (usually called *slippage* in the LFS) is represented by individuals in the Census that have not been paired with an individual from the LFS, either because a satisfactory match could not be found within the Census household that was successfully linked to their LFS household, or because they were linked to a dwelling that was vacant in the LFS. Conversely, LFS overcoverage is represented in our sample by those LFS individuals who could not be linked to a Census individual.

The breakdown of the sample, for the purposes of this study, is as follows: in addition to the 61,132 LFS-Census persons successfully matched,

- i) 2,723 Census persons whose household was successfully linked to an occupied LFS household, but who could not be linked to an LFS individual within that household (undercoverage);
- ii) 1,783 Census persons whose household was linked to a vacant dwelling in the LFS (undercoverage);
- iii) 1,432 LFS persons whose household was successfully linked to an occupied Census household, but who could not be linked to a Census individual within that household (overcoverage);
- iv) 280 LFS persons whose household was linked to an unoccupied dwelling in the Census (overcoverage).

This study estimated an LFS undercoverage rate of 7.53% and an LFS overcoverage rate of 3.01%. These are comparable to the usual slippage rate and the Census undercoverage rate, respectively. Table 4.1 displays the breakdown of these rates.

**Table 4.1: LFS Undercoverage and Overcoverage Rates**

	Undercoverage rate	Overcoverage rate
<b>Persons not matched within an occupied household</b>	4.34%	2.52%
<b>Persons whose household was matched to a vacant dwelling</b>	3.19%	0.49%
<b>Total</b>	<b>7.53%</b>	<b>3.01%</b>

The breakdown of undercoverage identifies two of the main causes of slippage in the LFS: persons missed, and dwellings erroneously identified as vacant. (The other main cause of slippage is listing errors, which cannot be studied here because they do not show up in the LFS sample.) The coverage study was undertaken to learn more about the characteristics of the dwellings and the persons who present coverage problems for the LFS. This task was accomplished using the second database resulting from the person linkage process. The

comparison of the characteristics of unmatched dwellings (dwellings erroneously identified as vacant) with those of matched dwellings is presented in the first subsection. Three characteristics are studied: the tenure component (whether the dwelling is owned or rented), the type of dwelling and the household size of the missed households. The second part of the coverage study is done at the person level, i.e. the characteristics of unmatched persons (under- and over-covered) and those of matched persons are compared. The unmatched persons comprise both the persons missed within matched households (after person linkage), and the persons missed because their dwelling was erroneously identified as vacant. The second subsection presents the results of this study for the following characteristics: labour force status, age and sex, and marital status.

#### 4.1. Dwellings Erroneously Identified as Vacant

Table 4.2 displays the weighted frequency distribution of matched and unmatched dwellings by tenure, for the LFS and the Census as well. LFS dwellings erroneously identified as vacant are represented by the unmatched dwellings on the side of the Census; they are a cause of slippage in the LFS. The unmatched dwellings on the LFS side represent overcoverage on the part of the LFS. Comparing the unmatched and the matched distributions, it can be noticed that more than 75% of the unmatched dwellings were rented while almost 60% of the matched dwellings were owned.

**Table 4.2: Weighted Distribution of Matched and Unmatched Dwellings by Tenure**

Tenure	LFS		CENSUS	
	Matched (%)	Unmatched (%)	Matched (%)	Unmatched (%)
<b>Owned</b>	58.80	21.13	58.46	22.76
<b>Rented</b>	41.20	78.87	41.54	77.24

Another interesting characteristic to analyse is the type of dwelling in which people live. Since the LFS and the Census did not use the same categories for the type of dwelling, the Census categories had to be modified to correspond to the LFS ones. In Table 4.3, all of the categories are presented for both the LFS and the Census. It can be seen from Table 4.3 that there were three categories in the LFS ("Double", "Institution", "Hotel Rooming or Lodge, Logging Camp, Construc. etc") that were not used by the Census. Moreover, the Census category "Semi-detached" was not used by the LFS. Also, dwellings that could not be included in a specific category were classified as "Other" in the LFS while they were categorized as "Other Single Attached" or "Other movable" in the Census. Thus, the simplest way of making the LFS and the Census categories comparable was to combine the Census "Other Single Attached", "Other movable" and "Semi-detached" into one category that will correspond to the "Other" category in the LFS.

It should be noted that before November 1991, the "Low-Rise Apart/Flat" and "High-Rise Apart" LFS categories were combined into one category. Therefore, at the time of the 1991 May and June Labour Force Surveys, these breakdowns were not available; the analysis will then be done with the combined category named "Apart/Flat".



Table 4.3: Type of Dwelling Categories for the LFS and the Census

LFS	CENSUS
Single Detached	Single Detached
Double	-----
Row or Terrace	Row House
Duplex	Duplex
Low-Rise Apart/Flat	Apt. (<5 stories)
High-Rise Apart	Apt. (>=5 stories)
Institution	-----
Hotel Rooming or Lodge, Logging Camp, Construc. etc	-----
Mobile Home	Mobile Home
Other	-----
-----	Other Single Attached
-----	Other movable
-----	Semi-detached

The weighted frequency distributions of matched and unmatched dwellings by type of dwelling are presented in Table 4.4 for both surveys. The results show that the unmatched dwellings differ from the matched ones. The dwellings present the most coverage problems are apartments/flats, with proportions around 70% in the unmatched portion, compared to 36% in the matched portion. Single detached households, on the other hand, represent nearly one-half of the matched portion, but only 11.21% and 18.52% (LFS and Census, respectively) of the unmatched portion. These results are not surprising, since it is generally easier to determine whether a dwelling is vacant in the case of a single detached house than for an apartment. Some of the reasons may be: i) in a single detached dwelling, there are more accessible windows by which an interviewer can look to verify if the dwelling is vacant; or ii) in apartment blocks, there is often no sign on the door to indicate if the apartment is for rent or for sale, while for a single detached, these signs are easily noticeable. These results explain why rented dwellings are more prone to coverage problems than owned ones (detached houses are more likely to be owned than apartments).



**Table 4.4: Weighted Distribution of Matched and Unmatched Dwellings by Type of Dwelling**

Type of dwelling	LFS		CENSUS	
	Matched (%)	Unmatched (%)	Matched (%)	Unmatched (%)
Single Detached	48.96	11.21	48.73	18.52
Double	4.82	2.21	-----	-----
Row or Terrace	5.87	9.06	5.64	3.30
Duplex	2.88	2.72	3.02	5.12
Apart/Flat	36.62	74.15	36.31	66.55
Institution	0.04	-----	-----	-----
Hotel/Camp/Constr.	0.01	-----	-----	-----
Mobile Home	0.77	0.65	0.61	0.03
Other	0.02	-----	5.70	6.48

Table 4.5 gives the average household size of matched and unmatched dwellings for both the LFS and the Census. The table shows that missed households (i.e. erroneously identified as vacant) are smaller in size than the ones reached by the survey. In Section 3 on the nonresponse study, it was shown that smaller households are more likely to be nonrespondents - this study indicates that they are also more likely to be marked as vacant.

**Table 4.5: Average Household Size for Matched and Unmatched Dwellings**

	LFS		CENSUS	
	Matched (%)	Unmatched (%)	Matched (%)	Unmatched (%)
Hhld size	2.52	2.04	2.52	2.31

#### 4.2. Missed Persons

To learn more about missed persons (persons missed within matched households and persons missed because their dwelling was erroneously identified as vacant), the characteristics of matched and unmatched individuals are compared. Variables such as the labour force status, the age and sex groups and the marital status are studied.

#### 4.2.1. Labour Force Status

Table 4.6 shows the weighted frequency distribution of matched and unmatched individuals by labour force status. Unmatched individuals on the Census side represent LFS undercoverage, while unmatched individuals on the LFS side represent LFS overcoverage. The overcoverage and undercoverage rates are also presented in Table 4.6. It can be seen from the table that LFS missed persons are more likely to be unemployed and less likely to be out of scope than matched individuals. As well, the persons that represent LFS overcoverage are more prone to be unemployed and less prone to be out of scope than the persons of the covered population. This results in a considerable difference in unemployment rates between the matched and the unmatched individuals, as shown in Table 4.7. The unemployment rate for Census persons not matched with the LFS (LFS undercoverage) is roughly twice that of matched individuals. Also, the employment-population ratio is lower for unmatched individuals (both over- and under-covered). It is interesting to point out that unmatched individuals on both sides have similar labour force profiles. However, they are very different from the profiles of matched individuals.

**Table 4.6: Weighted Distribution of Matched and Unmatched Individuals by Labour Force Status**

LF Status	LFS			CENSUS		
	Matched (%)	Unmatched (%)	Overcoverage rate (%)	Matched (%)	Unmatched (%)	Undercoverage rate (%)
Employed	49.70	48.33	2.91	50.44	46.15	5.85
Unemployed	5.23	8.90	4.98	4.78	10.47	12.96
NLF	24.36	27.66	3.38	24.57	28.02	7.19
Out of scope	20.70	15.12	2.20	20.21	15.36	4.91

**Table 4.7: Labour Force Rates for Matched and Unmatched Individuals**

Rate	LFS		CENSUS	
	Matched	Unmatched	Matched	Unmatched
Unemployment rate	9.52%	15.55%	8.65%	18.50%
Participation rate	69.28%	67.42%	69.21%	66.89%
Empl.-pop. ratio	62.68%	56.94%	63.22%	54.52%

#### 4.2.2. Age and Sex

Table 4.8 shows the weighted frequency distribution of matched and unmatched individuals by age and sex. Again, unmatched individuals on the Census side represent LFS undercoverage, while unmatched individuals

on the LFS side represent LFS overcoverage. From Table 4.8, it can be seen that there is a smaller proportion of unmatched males and females under 15 years old and a greater proportion of unmatched males and females 20 to 29 years old, for LFS and the Census as well. The difference in proportions is roughly two-to-one for the 20-24 age groups of both sexes. The 20-24 and 25-29 groups display large undercoverage and overcoverage rates, especially for males. For the Census, males 30-39 years old have also a high undercoverage rate since they are overrepresented in the unmatched population comparatively to the matched one. Also, males over 40 and females over 30 are overrepresented in the matched population compared to the unmatched population, in both surveys; they have relatively low undercoverage and overcoverage rates. For both surveys, the proportion of males is higher for unmatched individuals than for matched ones, i.e. males are associated with more coverage problems than females. It is interesting to note that females make the majority of matched individuals, but males are in majority among unmatched individuals for both surveys. These results are consistent with those of the LFS itself, which usually has high slippage rates for persons between 20-29, especially for males.

**Table 4.8: Weighted Distribution of Matched and Unmatched Individuals by Age and Sex**

Age-Sex group		LFS			CENSUS		
		Matched (%)	Unmatched (%)	Overcoverage rate (%)	Matched (%)	Unmatched (%)	Undercoverage rate (%)
Males	0-14	10.38	7.90	2.29	10.27	6.97	5.21
	15-19	3.26	3.13	3.04	3.28	3.43	7.82
	20-24	3.73	7.12	5.56	3.73	7.89	14.64
	25-29	4.39	6.88	4.61	4.36	8.53	13.67
	30-39	8.04	8.78	3.26	8.14	10.82	9.72
	40-54	9.54	9.17	2.88	9.57	7.47	5.95
	55 +	9.01	7.08	2.37	9.03	6.82	5.77
Males		48.35	50.25	3.10	48.39	51.95	8.05
Females	0-14	10.15	6.89	2.05	10.04	7.45	5.68
	15-19	3.15	3.19	3.02	3.13	3.01	7.22
	20-24	3.69	7.50	5.90	3.74	7.23	13.53
	25-29	4.55	5.96	3.88	4.48	6.15	10.01
	30-39	9.04	7.70	2.56	9.04	7.04	5.93
	40-54	9.89	9.81	2.97	9.93	7.42	5.71
	55+	11.17	8.69	2.34	11.26	9.78	6.58
Females		51.65	49.75	2.88	51.61	48.05	7.05

### 4.2.3. Marital Status

Table 4.9 displays the weighted frequency distribution of matched and unmatched individuals by marital status. First, it should be noted that the differences observed between both LFS and Census distributions of matched individuals can be attributed in part to the different wordings of the two questionnaires (these will be explained further in Section 5). Comparing the unmatched and matched individuals, one can notice that the unmatched persons are more likely to be single, separated or divorced than matched persons. So, the proportion of married persons is higher among matched individuals. Single persons usually present greater coverage problems than married persons because of their mobility.

**Table 4.9: Weighted Distribution of Matched and Unmatched Individuals by Marital Status**

Marital status	LFS			CENSUS		
	Matched (%)	Unmatched (%)	Overcoverage rate (%)	Matched (%)	Unmatched (%)	Undercoverage rate (%)
<b>Married</b>	48.49	39.88	2.47	43.21	29.58	5.28
<b>Single</b>	41.70	48.28	3.45	44.69	55.79	9.23
<b>Widow/ widower</b>	4.54	4.76	3.13	4.65	4.45	7.22
<b>Separated/ divorced</b>	5.28	7.08	3.97	7.44	10.19	10.03

### 4.3. Summary of Results

The results of the coverage study suggest that LFS dwellings erroneously identified as vacant are widely different from those of the covered population. Dwellings which present coverage problems are more often apartments/flats. They are also associated with smaller household size. The persons not covered by the LFS are also very different from those of the covered population. Persons who present coverage problems are much more likely to be unemployed than those who do not. They are also more likely to be young adults, male and unmarried.

Coverage problems can introduce substantial bias into the LFS estimates if there are large differences between the covered population and individuals who are not covered. This study suggests that the portion of the target population which is not covered may have an unemployment rate larger than that of the covered portion of the population. As a result, the LFS may be underestimating the unemployment rate. Therefore, it is very important to keep slippage rates at a minimum for all domains of estimation.



## 5. THIRD STUDY: COMPARISON OF LFS AND CENSUS RESPONSES

A third goal of the LFS-Census linkage project was to compare LFS and Census responses at the dwelling and person levels. The database resulting from the person-to-person matches (Section 2.2) was used to achieve this task. Using the households and the persons who were successfully matched, two profiles could be drawn: one from the LFS data, and one from the Census data. These two profiles were then compared for several variables.

The first part of this study compares responses at the dwelling level (subsection 5.1). This means that only information on matched households was kept to compare the two profiles. Two characteristics related to the dwelling were studied: 1) tenure (whether the dwelling is owned or rented), and 2) type of dwelling. For the second part of the study (subsection 5.2), responses were compared at the person level. This time, the LFS and the Census profiles were analysed using the linked individuals. The following variables were studied: labour force status, age, sex, marital status and education level. A summary of the results of parts one and two is provided in subsection 5.3.

### 5.1. Comparison of Responses at the Dwelling Level

Table 5.5 shows the weighted percentage distribution of matched households by tenure for the LFS as well as for the Census. Since a few missing values were found in the LFS data, the results are based on 24,069 households instead of 24,223 households. It can be noticed from Table 5.5 that the total distributions are quite similar: 58.80% of owned dwellings and 41.20% of rented ones for the LFS compared to 58.46% of owned dwellings and 41.54% of rented ones for the Census. In fact, 98.20% of the dwellings had the same answer for the tenure component. This high rate is not surprising since people are less reluctant to answer this type of question than more confidential questions such as age or income. Unequal responses between the LFS and the Census are not frequent (1.80%), and it should be noted that 3.43% of the Census values that differ from the LFS values are imputed ones.

**Table 5.5: Weighted Percentage Distribution of Matched Households by Tenure (%)**

		CENSUS		
		Owned	Rented	Total
L F S	Owned	57.73	1.07	58.80
	Rented	0.73	40.47	41.20
Total		58.46	41.54	100.00

Table 5.6 displays the weighted percentage distribution of matched households by type of dwelling. Again, the Census categories for that characteristic were modified as in Section 4.1, so that they could correspond to the LFS ones. A few missing values were found, so the results are based on 24,144 households out of 24,223 linked households. The distributions are very similar except for the "Other" category, which is understandable

since the definition of this category is not exactly the same in the LFS and in the Census. Besides, if the LFS' "Double", "Institution" and "Hotel Rooming or Lodge, Logging Camp, Construc. etc" were combined with the "Other" category, the percentage would rise to 4.89, which is relatively close to 5.70%. Combining them would also increase the percentage of equal LFS and Census responses from 88.02% to 91.77%. The 8.23% of unequal values could be attributed in part to imputation (10.42% of unequal values were imputed for the Census) and to the fact that the original LFS and Census' categorizations were not the same.

**Table 5.6: Weighted Percentage Distribution of Matched Households by Type of Dwelling (%)**

	Single Detached	Double	Row or Terrace	Duplex	Apart / Flat	Institution	Hotel / Camp / Constr.	Mobile Home	Other
LFS	48.96	4.82	5.87	2.88	36.62	0.04	0.01	0.77	0.02
CENSUS	48.73	-----	5.64	3.02	36.31	-----	-----	0.61	5.70

## 5.2. Comparison of Responses at the Individual Level

The differences between LFS and Census responses are examined for the linked individuals coming from the database resulting from the person linkage step. Variables such as labour force activity, age, sex, marital status and education level are studied.

### 5.2.1. Comparison of Labour Force Status

Table 5.7 displays the weighted cross tabulation of LFS-Census matched individuals by labour force status. Since all questions on labour force activity are part of the Census long questionnaire (2B), and as one fifth of the population receives the Census long questionnaire, the results are based on only 12,262 individuals rather than 61,132, which represents 20.1% of the matched sample.

The total distributions of the labour force status for the LFS and the Census are slightly different. The greatest differences are observed for the employed and unemployed individuals. The LFS registers 49.48% of employed individuals while in the Census, the rate is almost 1 percentage point higher (50.44%). For all other categories, the LFS records a slightly higher percentage than the Census. The percentage of unemployed individuals is about 0.50% higher in the LFS (5.23% in the LFS compared to 4.78% in the Census). As for the percentages of people Not in Labour Force or Out of Scope, the rates are quite similar, although they are a little higher for the LFS. There are 24.84% of individuals Not in Labour Force for the LFS and 24.57% for the Census, and there are 20.45% of individuals Out of Scope for the LFS and 20.21% for the Census.

The differences in labour force data for the employed and unemployed individuals result in a difference of about 1 percentage point in the unemployment rate (see Table 5.8). The participation rates for the LFS and the Census are similar: less than 0.1% of a percentage point separates the two values. Also, the Census records a slightly higher employment-population ratio (63.22% for the Census compared to 62.68% for the LFS).

**Table 5.7: Weighted Distribution of Matched Individuals by Labour Force Status (%)**

	CENSUS				Total
	Employed	Unempl.	NLF	Out of Scope	
<b>Employed</b>	46.80	1.00	1.63	0.04	49.48
<b>L Unemployed</b>	1.73	2.56	0.92	0.01	5.23
<b>F NLF</b>	1.83	1.20	21.76	0.05	24.84
<b>S Out of Scope</b>	0.07	0.01	0.25	20.11	20.45
<b>Total</b>	50.44	4.78	24.57	20.21	100.00

**Table 5.8: Weighted Labour Force Rates for Matched Individuals**

	LFS	CENSUS
<b>Unemployment rate</b>	9.52%	8.65%
<b>Participation rate</b>	69.28%	69.21%
<b>Employment-Population ratio</b>	62.68%	63.22%

Among the 12,262 individuals who have non-missing labour force values for both the LFS and the Census, 91.23% have the same labour force status, which implicitly means that 8.77% of the individuals have unequal values. It can be seen from Table 5.7 that most of the unequal values are mainly detected in the Employed, Unemployed and NLF categories. To account for the size of each category, discrepancy rates (i.e. percentage of persons with a given labour force status in LFS that have a different labour force status in the Census) were calculated to see if, for a given category, the proportion of unequal values was significant. Table 5.9 shows the weighted discrepancy rates by labour force status in the LFS. It can be noticed from that table that the discrepancy rates are low for employed and out-of-scope persons (5.41% and 1.66% respectively) - these two LF statuses are relatively easy to measure in a survey questionnaire. Discrepancy rates are significantly higher for unemployed and NLF people. Persons who are unemployed or inactive go through a more complicated set of questions, with greater chance of measurement error. Part of the very high discrepancy rate for unemployed persons can perhaps be explained by the difference between the two questionnaires: only five questions are used in the Census questionnaire to determine labour force status *and* number of hours worked. Also, the Census is self-administered while the LFS is done by interview.

There is also a difference in the reference dates of the two surveys. The LFS database includes information from the May 1991 and June 1991 surveys, using reference weeks in the middle of the month, while the Census was conducted June 4<sup>th</sup>, using the preceding week as the reference period. Selecting the matched individuals (from the database resulting from the person linkage step) that were both in May and June of 1991



LFS surveys, it was observed that 5% of them had changed labour force status between the two months. So, differences in labour force status can also be attributed in part to the time difference in data collection.

**Table 5.9: Weighted Discrepancy Rates by Labour Force Status in LFS**

<b>Labour Force Status in LFS</b>	<b>Discrepancy Rate</b>
<b>Employed</b>	5.41%
<b>Unemployed</b>	51.03%
<b>NLF</b>	12.41%
<b>Out of Scope</b>	1.66%

### **5.2.2. Comparison of Age and Sex**

Age and sex are important variables, in both the LFS and the Census, and they should not be left blank since these variables are used to edit and impute other variables, or as post-stratification variables. As some people are reluctant to give their age, missing values can sometimes be found and are compensated for by imputation. In this subsection, the LFS and Census' age and sex characteristics are compared for all of the matched individuals (i.e. 61,132 persons).

Table 5.10 presents the weighted percentage distributions of matched individuals by age and sex groups. The total percentages of males and females for the LFS and the Census are quite similar: 48.35% males and 51.65% females in the LFS compared to 48.39% males and 51.61% females in the Census. Also, the distributions by age/sex groups are comparable for both surveys. This is understandable since the age and sex characteristics were the main variables used to match persons (person linkage process). Besides, 97.35% of the matched individuals have no more than 1 year of difference between the LFS age and the Census age. The remaining individuals (2.65%) have 2 or 3 years of difference between the two ages, which is in part due to imputation of missing values (3.33% of these Census value are imputed). The difference between LFS and Census ages can also be attributed to a difference in the LFS and the Census questionnaires: the Census asked the date of birth while the LFS asked the age. For proxy interviews, it is obviously easier for respondents to give the age of a person than the date of birth. Comparing sexes independently, it was noticed that 1.04% of the 61,132 matched individuals have different values in the LFS and in the Census. But, of these individuals, 21.30% have an imputed Census sex.

Table 5.10: Weighted Percentage Distribution of Matched Individuals by Age and Sex (%)

		Age group							Total
		0-14	15-19	20-24	25-29	30-39	40-54	55+	
Males	LFS	10.38	3.26	3.73	4.39	8.04	9.54	9.01	48.35
	CENSUS	10.27	3.28	3.73	4.36	8.14	9.57	9.03	48.39
Females	LFS	10.15	3.15	3.69	4.55	9.04	9.89	11.16	51.65
	CENSUS	10.04	3.13	3.74	4.48	9.04	9.93	11.25	51.61

### 5.2.3. Comparison of Marital Status

The weighted cross tabulation of the LFS and the Census percentage distributions of the marital status is displayed in Table 5.11. According to this Table, the total distributions for the LFS and the Census are somewhat different. The only category with similar percentages is the widow/widower category. The LFS identifies more persons as married and fewer as single, separated or divorced than does the Census. The discrepancies found between both surveys can be explained by the difference between the two questionnaires: the wording for the marital status categories is not exactly the same for the LFS as for the Census. The 1991 LFS had four categories worded as 1) *Now Married or Living Common-Law*, 2) *Single (Never Married)*, 3) *Widow or Widower* and 4) *Separated or Divorced*, while the 1991 Census had five of them: 1) *Legally Married (and Not Separated)*, 2) *Legally Married and Separated*, 3) *Divorced*, 4) *Widowed* and 5) *Never Married (Single)*. The widowed category is the only category worded the same, which explains why similar percentages were observed for the LFS and the Census.

The discrepancy rate (i.e. the percentage of individuals that have different marital statuses in the two surveys) is 6.91%, with a very low Census imputation rate (only 2.25% for persons with a discrepancy). The great majority of discrepancies are of one of the three following types. Firstly, the type of discrepancy most often met is: *Now Married or Living Common-Law* in the LFS and *Single* in the Census (3.50% of all persons). The Census questionnaire words the married category as *Legally Married (and Not Separated)*, which reduces the likelihood of persons in a common-law relationship to select that response. Secondly, another type of discrepancy often observed is: *Married* in the LFS and *Separated/Divorced* in the Census (1.82%). As mentioned earlier, the Census category for separated is worded *Legally Married and Separated*, this makes it clearer that *Separated* should be the response instead of *Married*. The third type of discrepancy most often met is *Single* in the LFS and *Separated/Divorced* in the Census with a rate of 0.63%.

Table 5.11: Weighted Percentage Distribution of Matched Individuals by Marital Status (%)

	CENSUS				Total
	Legally Married (and Not Separated)	Never Married (Single)	Widowed	Separated / Divorced <sup>1</sup>	
<b>Now Married or Living Common-Law</b>	42.93	3.50	0.24	1.82	48.49
<b>L F S Single (Never Married)</b>	0.13	40.89	0.04	0.63	41.70
<b>Widow / Widower</b>	0.03	0.11	4.16	0.24	4.54
<b>Separated / Divorced</b>	0.12	0.19	0.22	4.75	5.28
<b>Total</b>	43.21	44.69	4.65	7.44	100.00

<sup>1</sup> The "Separated / Divorced" category includes two Census categories: i) Legally Married and Separated and ii) Divorced

#### 5.2.4. Comparison of Education Level

Comparison of education level was the hardest to achieve since the LFS and the Census use very different sets of questions, which resulted in different categories for the derived variable. The LFS uses five questions to derive two education variables: 1) the highest grade of elementary / secondary school completed and 2) the highest degree / certificate / diploma obtained from an educational institution above elementary / secondary school. For the Census, one variable, the highest level of schooling, was derived from four questions. To be able to compare responses between the two surveys, the Census variable was recategorized as in the LFS. You will find more details on the new categorization of the Census education variable in Appendix D.

Table 5.12 presents the weighted percentage distributions of matched individuals for the first education variable (highest grade of elementary / secondary school). Since Census education questions figure only on the long questionnaire, only 9,743 people out of 61,132 had non-missing values for both surveys; which represents 15.9% of the matched individuals. It can be seen from the table that small differences are observed between the LFS and the Census responses. This can be explained by the difference in the wording of the LFS and Census' questions. The Census asks the respondent to write down the highest grade (or year) of secondary (high school) or elementary school *ever attended* while the LFS asks for the highest grade of elementary or high school (secondary school) *ever completed*. As the highest grade attended can sometimes be higher than the highest grade completed, the Census will tend to have more individuals in the highest categories. As a matter of fact, the LFS records higher percentages of people with a grade 8 and lower or a grade 9 and 10, and lower percentages of people with a grade 11 - 13 that did or did not graduate high school.

A different way of comparing education levels between both surveys is to look at the number of equal values. It was found that 75.1% of the matched individuals have the same LFS and Census responses. This percentage of equal values is lower than those for the labour force status, the age and sex and the marital status because the education level was the variable for which it was the hardest to find a correspondence between the LFS



and the Census variables. The great majority of the discrepancies are of the three following types: i) *Graduate from High School* for the LFS and *Grade 11 - 13* for the Census (4.89%), ii) *Grade 11 - 13* for the LFS and *Graduate from High School* for the Census (3.84%) and iii) *Grade 9 and 10* for the LFS and *Graduate from High School* for the Census (3.64%). The first two types of discrepancies can be explained by the difference in the wording of the LFS and Census' questions.

The LFS and the Census' weighted percentage distributions of matched individuals for the second education variable (highest degree / certificate / diploma obtained above elementary / secondary school) are given in Table 5.13. Again, the results are based on 9,743 persons. The LFS and Census percentages are close, except for the individuals that did not receive any other education above elementary / secondary school or that received other education that could not be counted towards a degree, certificate or diploma from an educational institution (category "None"). The LFS percentage for this category is 53.48% while it is about 5 percentage points lower for the Census (48.45%). For all other categories, the difference between the LFS and the Census percentages ranges from 0.21 to 2.04 percentage points. The percentage of equal values is still low (68.5%) and can again be explained by the way the recategorization of the Census variable was done.

**Table 5.12: Weighted Percentage Distribution of Matched Individuals by Highest Grade of Elementary / Secondary School (%)**

	Grade 8 or Lower	Grade 9 and 10	Grade 11 - 13 <sup>1</sup>	Graduate from High School
<b>LFS</b>	11.71	14.77	9.35	64.18
<b>CENSUS</b>	11.34	11.14	10.93	66.59

<sup>1</sup> Individuals in this category did not graduate from high school.

**Table 5.13: Weighted Percentage Distribution of Matched Individuals by Highest Degree / Certificate / Diploma Obtained Above Elementary / Secondary School (%)**

	LFS	CENSUS
None <sup>1</sup>	53.48	48.45
No Post Secondary Degree / Certificate / Diploma <sup>2</sup>	9.28	11.32
Trades Certificate / Diploma from Vocational School or Apprenticeship Training	9.32	9.81
Non-University Certificate / Diploma from Community College, CEGEP, School of Nursing etc.	11.18	12.83
University Degree Below Bachelor's Level	1.83	2.13
Bachelor's Degree	10.72	10.51
University Degree / Certificate Above Bachelor	4.20	4.96

1 This category includes i) individuals that did not receive any other education above elementary / secondary school and ii) individuals that received other education that could not be counted towards a degree, certificate or diploma from an educational institution.

2 This category includes individuals that received any other education that could be counted towards a degree, certificate or diploma from an educational institution, but did not obtain a secondary degree / certificate / diploma.

### 5.3. Summary of Results

The analysis of responses at the dwelling level (first part of the study) showed that the LFS dwellings' characteristics (tenure and type of dwelling) are similar to those of the Census. The main findings of the analysis of response at the individual level (second part of the study) are:

1. LFS reflects fewer employed persons and more unemployed persons compared to the Census giving a higher unemployment rate for the LFS;
2. Responses regarding the age/sex groups are similar for both surveys (this was expected because these variables were used for matching individuals);
3. LFS records more married persons and fewer single, separated or divorced than the Census;
4. The Census registers more individuals in the highest grades of schooling than does the LFS.

## 6. CONCLUSION AND FUTURE PLANS

The project that matched LFS data with the 1991 Census of population data was undertaken to conduct a number of studies on the quality of the LFS data. The three studies that were carried out showed interesting results, and the LFS, as well as other surveys, can gain from them.

The study on the characteristics of LFS nonrespondents (first study) showed that nonrespondents are different from respondents. The main results were that: 1) nonresponse is higher in smaller households, 2) in one-person households, nonresponse is higher for adults aged between 20 and 44, and inactive persons are more likely to respond while employed persons are more prone to nonresponse, and 3) in households of two or more persons, nonrespondents have a slightly lower unemployment rate and a slightly lower household income. All of the differences pointed out here between nonrespondents and respondents can potentially introduce bias in the LFS estimates. These results can then be used to help determining more appropriate post-stratification classes for the weighting system; for example, the household size could be used as an additional benchmark variable. Also, the results can be used to evaluate edit and imputation systems, for example, helping to determine weighting classes for nonresponse adjustment.

The study on coverage (second study) concluded that the portion of the target population which is not covered by the LFS is widely different from the covered portion of the population. It was observed that: 1) LFS dwellings erroneously identified as vacant are mostly apartments/flats and they are associated with smaller household, and 2) the LFS persons missed are more likely to be unemployed, young adults, male and unmarried. The latter implied a larger unemployment rate for the uncovered portion of the population. As coverage problems can introduce substantial bias into the LFS estimates, unoccupied dwellings should be examined thoroughly. The results of the study suggest that a program such as the vacancy check should be restored.

The last study, which compared the LFS and the Census responses, indicated that responses of both surveys are not similar when the two questionnaires do not have the same wording of questions. For instance, the results showed that the LFS records more married persons and fewer single, separated or divorced than does the Census. In view of reducing the differences observed between the LFS and the Census responses, and in view of the harmonisation of questions among Statistic Canada surveys, some questions of the LFS questionnaires could be reviewed.

The results of the three studies can also be used for other purposes such as helping the interviewers in different ways (some of the results could be helpful when developing the interviewer manual). The profile of LFS nonrespondents could be used to improve LFS data collection procedures, and follow-up procedures could be attuned to the characteristics of nonrespondents. Geography Division can also benefit from our project and ameliorate the quality of the Address Register. For example, the residential addresses and the "PROV/FED/EA/HHLD" Census keys could be subject to a 100% validation check (a more rigid quality control process) to reduce spelling errors in addresses and data capture errors in the Census keys. Furthermore, every person that wishes to use the AR as a survey frame can certainly gain from our matching procedures and findings.

Accordingly, the Census Division, as well as LFS supplementary surveys, could gain from this project and should get the databases to evaluate the quality of their data. As the LFS-Census 1991 project has many positive aspects, repeating this project should be considered again in the future.

---





## **ACKNOWLEDGMENTS**

The authors would like to thank Ritu Kaushal, Mei Cheung, Marek Wysocki, Michelle Simard, Gopa Ray, Jean-François Bastien, Angèle Vaillancourt and Juliette Gauthier who contributed to the realisation of this project. They would also like to thank Jack Gambino, Jocelyn Tourigny and Rick Ciok for their valuable comments that helped improve the quality of the paper.



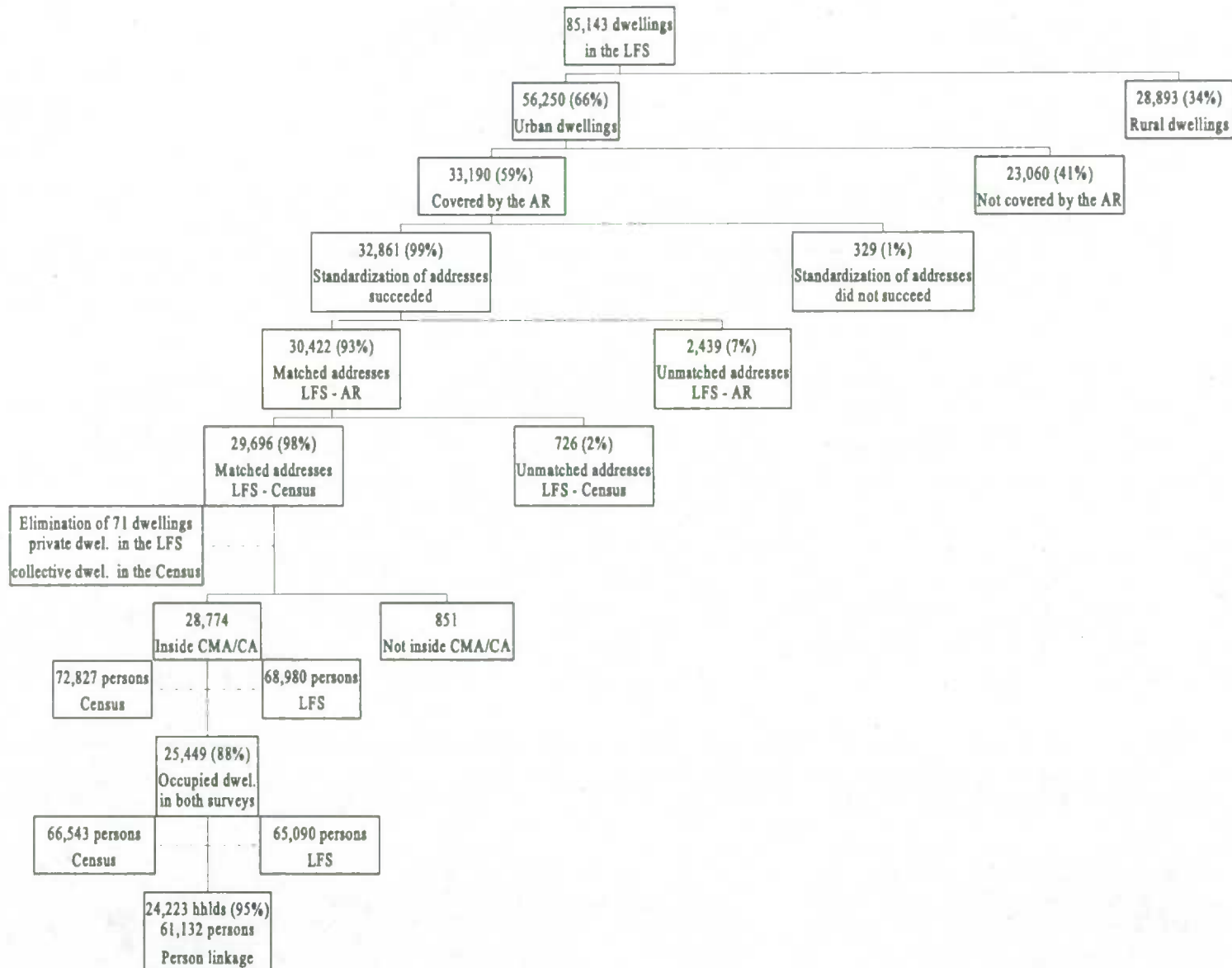
## REFERENCES

- Allard, B. and Dufour, J. (1994). *Characteristics of Respondents and Nonrespondents in the Labour Force Survey*, internal report, Statistics Canada.
- Brown, C. (1992). *Census Evaluation of the Address Register Person Coverage and Characteristics of Persons and Dwellings Added*, internal report, Statistics Canada.
- Gambino, J. (1996). *The 1995 Labour Force Survey Sample Design*, internal report, Statistics Canada.
- Kaushal, R. and Dufour, J. (1995). *1991 Census-LFS linkage: Quality and Coverage Studies*, internal report, Statistics Canada.
- Lafrance, B., Charland, K. and Brown, C. (1994). *Evaluation of Statistics Canada's Residential Address Register*, Proceedings of the USBC Annual Research Conference.
- Singh, M.P., Drew, J.D., Gambino, J. G. and Mayda, F. (1990). *Methodology of the Canadian Labour Force Survey 1984-1990*. Statistics Canada Publication, Catalogue number 71-526.
- Swain, L., Drew, J.D., Lafrance, B. and Lance, K. (1991). *The Creation of a Residential Address Register at Statistics Canada*, Proceedings of Statistics Canada Symposium 1991, Spatial Issues in Statistics, November 1991.
-

---



# **APPENDIX A** **FLOW CHART OF THE LINKAGE PROCESSES**



**APPENDIX B**  
**MATCH RATES AT THE CMA/CA LEVEL**

CMA/CA	Province	Match rate (%)
St-John's	Newfoundland	57.89
Halifax	Nova Scotia	64.40
Moncton	New Brunswick	86.29
Saint John	New Brunswick	54.31
Fredericton	New Brunswick	85.67
Québec	Québec	72.20
Hull	Québec	80.27
Montréal	Québec	86.12
Trois-Rivières	Québec	72.96
Chicoutimi	Québec	77.76
Sherbrooke	Québec	58.75
St. Catharines	Ontario	89.51
Sudbury	Ontario	57.98
Thunder Bay	Ontario	83.30
Sault Ste-Marie	Ontario	86.13
Windsor	Ontario	74.86
Sarnia	Ontario	89.09
Toronto	Ontario	89.19
Hamilton	Ontario	91.11
Kingston	Ontario	77.35
London	Ontario	88.98
North Bay	Ontario	81.62
Oshawa	Ontario	86.28
Ottawa	Ontario	90.79
Peterborough	Ontario	70.65

<b>CMA/CA</b>	<b>Province</b>	<b>Match rate (%)</b>
Kitchener	Ontario	83.87
Guelph	Ontario	92.24
Brantford	Ontario	91.24
Winnipeg	Manitoba	89.24
Saskatoon	Saskatchewan	74.25
Regina	Saskatchewan	84.36
Edmonton	Alberta	76.40
Red Deer	Alberta	86.99
Calgary	Alberta	89.34
Lethbridge	Alberta	89.74
Matsqui	British Columbia	92.24
Vancouver	British Columbia	94.36
Victoria	British Columbia	94.45
Kelowna	British Columbia	77.57
Chilliwack	British Columbia	92.16

## APPENDIX C PERSON LINKAGE PROCESS

### C.1. Description of the Process

To perform the person linkage process, variables from both the LFS and the Census were compared. As mentioned in this paper, most of the Census variables have both imputed and unimputed values while the LFS variables have a mix of both types of values. To account for this, new variables combining the imputed and unimputed values were created (as explained in section 2.2.). The age and the education level were treated differently. For age, only the imputed value was available for the Census; so, an unimputed value was derived using the unimputed month, decade and year information. When the imputed and unimputed ages were unequal, the person-to-person matches were done first using the unimputed age; and if the match was unsuccessful, the imputed age was used instead of the unimputed one. For the education variables, two variables had to be created (using the information known) for the Census since the LFS had two of them.

In this Appendix, labels are generally employed instead of a full description to identify variables. The list below provides the label for each of the LFS and Census variables used for matching.

#### LFS variables:

- RO (Regional Office)
- docket (docket number)
- age
- sex
- status (labour force activity)
- dwelling (type of dwelling)
- tenure (owned/rented flag)
- marst (marital status)
- education (H381 and H382)

#### Census variables:

- RO-docket number derived from the Census household number
- imputed and unimputed ages
- mthb2u (month of birth unimputed)
- decdbu (decade of birth unimputed)
- yearbu (last digit of year of birth unimputed)
- imputed and unimputed sexes
- imputed lftag (labour force activity)
- imputed and unimputed types of dwelling
- imputed and unimputed tenures
- imputed and unimputed marst
- hlos (highest level of schooling derived)

Note: H381 is the highest grade of elementary/secondary school completed and H382 relates to any other education (highest degree/certificate/diploma obtained).

In order to verify if individuals within a household were the same for the LFS and the Census, a distance measure, called *score*, was calculated between individuals. A weight was added to the score each time the LFS and the Census recorded equal values for the same variable. Naturally, the most important variables such as age and sex should have higher weights than the other variables. Also, an unimputed value was given more influence on the score than an imputed one. The score between two individuals was calculated following six steps:



- 1- let score = 0
  - 2- if sex from the LFS and sex from the Census are equal and not missing then
    - if sex from the Census is unimputed then score + 2
    - if sex from the Census is imputed then score + 1
    - else if (sex from the LFS and sex from the Census are unequal and not missing and sex from the Census is imputed) then score + 0.5
  - 3- if dwelling type from the LFS and dwelling type from the Census are equal and not missing then score + 0.5
  - 4- if marital status (marst) from the LFS and marital status from the Census are equal and not missing then
    - if marst from the Census is unimputed then score + 1
    - if marst from the Census is imputed then score + 0.5
  - 5- if tenure from the LFS and tenure from the Census are equal and not missing then score + 0.5
  - 6- if (imputed and unimputed Census ages are equal and not missing) or (unimputed Census age is missing, but unimputed Census year of birth is available (from decdbu and yearbu), and imputed Census age is not missing) then
    - if (imputed Census age = LFS age + 1) and mthb2u=2 (year of birth <= June 3)  
then score + 5
    - else if imputed age from the Census = age from LFS - 1 and mthb2u=3 (year of birth >= June 4) then score + 3
    - else if imputed Census age = LFS age then score + 5
    - else if (imputed Census age - 3 <= LFS age <= imputed Census age + 3) then score + 3

(Exception: in the one case where the unimputed age was missing, and the derived age (from the year of birth) was different from the imputed age, the unimputed age was used in the above decision rule.)

    - else if imputed Census age is not missing but unimputed age is missing then
      - if imputed Census age = LFS age then score + 2.5
      - else if (imputed Census age - 3 <= LFS age <= imputed Census age + 3) then score + 1.5
    - else if imputed and unimputed Census ages are unequal but not missing then
      - if imputed Census age and LFS age equal then **score2** = score + 2.5
      - else if imputed Census age - 3 <= LFS age <= imputed Census age + 3  
then **score2** = score + 1.5
      - if unimputed Census age and LFS age are equal then score + 5
      - else if (unimputed Census age - 3 <= LFS age <= unimputed Census + 3) then score + 3
-

**Note:** In the last else, when imputed and unimputed ages from the Census are not equal and non-missing, two different scores were calculated: the original one (score) with the unimputed age and another with the imputed age (score2). If the original score (score) calculated between two specific individuals was lower or equal to 4, then score2 (calculated with the imputed age) was considered. If the original score was greater than 4, then score2 was not used.

In all other situations not mentioned above, nothing was added to the score.

A program (in SAS language) was created to calculate the scores between individuals and to identify matched individuals. In simple terms, the program compares each individual from a specific household in a data set with all individuals from that same household belonging to the other data set and it chooses the best match. Then, the program adds the match to the matching file (a new file created for the purpose of the linkage process) if it was considered to be a good one. The following steps describe in more details the person linkage procedure:

1. Selection of a dwelling from a file containing the 25,449 dwelling identifiers (file created for the use of linkage).
2. Identification of all the individuals from the LFS and from the Census in the dwelling chosen in Step 1 and calculation of the household size for both surveys.
3. Identification of the household (LFS or Census) with the smaller size. If the household sizes are equal, then the LFS household is chosen. For each individual within that household, a score between the individual and all the other persons in the household from the other survey is calculated. The match with the highest score is considered to be the best match for the individual and a value of 1 is assigned to a variable named BONMATCH in the program for that match. If there is more than one maximum, then the first one is chosen.

In this step, a file is created and includes all of the possible combinations for the matches. So for each household, (LFS household size \* Census household size) records are written to the file.

4. If two individuals (coming from the LFS or the Census) were matched with the same person (coming from the other survey), then the matches have to be rebuilt since no duplicate is allowed. So this step checks if there are duplicates, as regards to the individuals, for the good matches (the ones with BONMATCH=1). If three persons or more (coming from one of the surveys) were matched with the same individual (coming from the other survey), then the program goes immediately to step 7 and the person-to-person matches have to be resolved manually. Otherwise, if two persons (coming from one of the surveys) were matched with the same individual (coming from the other survey), the matches were rearranged automatically so that there was no duplicate. The following explains how this was done.

Suppose we had:

LFS:	3 persons => Person A, Person B, Person C
CENSUS:	4 persons => Person W, Person X, Person Y, Person Z

Suppose that for each LFS person, the person-to-person matches with the Census individuals were:

---

LFS	CENSUS	SCORE
A	X	9
B	X	9
C	Z	6

Both LFS persons A and B were paired with the same Census person, X, and persons W and Y were not chosen. To rearrange the matches, the program followed the steps:

- look at the scores calculated between the persons implicated in the duplicates (in this example, A and B) and the persons not chosen from the other survey (in this case, persons W and Y).

LFS	CENSUS	SCORE
A	W	8
A	Y	4
B	W	6
B	Y	7

- identify the highest score for each person that was involved in the duplicates (A and B). This gave us two matches (in this case, A with W for a score of 8; and B with Y for a score of 7).

- calculate the sum of scores obtained by changing one of the duplicate matches and calculate the sum of scores obtained by changing the other:

LFS	CENSUS	SCORE	SUM
A	X	9	
B	Y	7	16

LFS	CENSUS	SCORE	SUM
A	W	8	
B	X	9	17

- use the set of pairings that maximized the sum of scores (here, A with W and B with X). (Situations where the two sums of scores are equal are described further below.)

So the final matching for this household was:

LFS	CENSUS	SCORE
A	W	8
B	X	9
C	Z	6

The example above is a simple one; the program could solve more complex situations but could not resolve situations where three individuals or more were paired with the same person. Those situations happened rarely and were resolved manually.

This step ended when all matching were distincts (except for situations mentionned in the previous paragraph).

### Situations when the sums of scores were equal

Suppose we had the same situation as before but equal sums of scores were found when calculating them by changing one of the duplicate matches and by changing the other:

LFS	CENSUS	SCORE	SUM
A	X	9	
B	Y	7	16

LFS	CENSUS	SCORE	SUM
A	W	7	
B	X	9	16

Some of those situations could be handled by calculating a score for the education variables (named *distedu*). Since these variables were only available from the long questionnaire (2B), only the Census persons who received the longer questionnaire had them; that was 19.5% of the Census persons. When the education variables were available, the calculation of the new score was as follows:

- the new score DISTEDU was initialized to 0,
- if the LFS and the Census values were equal as regards to each of the education variables, then DISTEDU+1,
- calculation of the sum of scores for the two possible combinations of distinct matches.
- the match having the maximum sum of scores was considered to be the best match (in the example below, A with X and B with Y). If the sum of scores were equal, the first set of matches was chosen.

LFS	CENSUS	SCORE	SUM	DISTEDU	SUM
A	X	9		2	
B	Y	7	16	1	3

LFS	CENSUS	SCORE	SUM	DISTEDU	SUM
A	W	7		0	
B	X	9	16	1	1

In this case, the final set of matches would be:

LFS	CENSUS	SCORE
A	X	9
B	Y	7
C	Z	6

5. All person matches were now distinct. The next step was to verify if a household in LFS and the corresponding one in the Census (step 2) represented the same household, and if members of the households were the same. The corresponding households from the two surveys represented the same



household if there was at least one good match between persons from both surveys. A good person match was observed when the two following conditions were met:

If household sizes for LFS and for the Census are equal or differ by 1, then the person match is a good one if score > 4 or score2 >= 4 (if a score2 was calculated).

If household sizes for LFS and for the Census differ by more than 1, then the person match is a good one if score > 7 or score2 >= 7 (if a score2 was calculated).

Since persons of 70 years of age or over are not interviewed by the LFS while they are included in the Census, non-paired individuals in the Census being 70 years old and over were not counted in the household size. A person was 70 years old or over if his/her imputed age or his/her unimputed age was greater or equal to 70. There were therefore 9 unmatched persons with only one of the two ages (unimputed or imputed) with a value greater or equal to 70. These persons were considered to be 70 years old and over.

6. Matches that could not be handled by the computer were verified manually and added to the matching file if they followed the same conditions as the ones described above.

7. Steps 1 through 6 were repeated for the next dwelling in the list until all dwellings were processed.

## C.2. Performance of the Person Linkage Process

As mentioned in section 2.2., the person linkage was performed on 25,449 occupied dwellings common to both the LFS and the 1991 Census and 24,223 (95.2%) of them were matched. Among these matched households, 61,132 individuals were paired between the two surveys. Table 1 shows more information on the performance of the process.

**Table 1: Performance of the Person Linkage Process**

CENSUS DATA					
	Attempted	Matched	%	Not matched	%
Households	25,449	24,223	95.2	1,226	4.8
Persons	66,543	61,132	91.9	5,411	8.1
LFS DATA					
	Attempted	Matched	%	Not matched	%
Households	25,449	24,223	95.2	1,226	4.8
Persons	65,090	61,132	93.9	3,958	6.1

As described in the first part on this Appendix, a LFS household was matched to a Census household if there was at least one acceptable match between individuals from both datasets. So all individuals within a matched household were not necessarily found in the LFS and in the Census. Table 2 presents the distribution of matched households by the number of non-matched individuals for LFS, for the Census and for both. The number of non-matched persons in matched households for both surveys combined applies to the smallest of the two household sizes, i.e.: (minimum between the household size in LFS and the one in the Census) minus (number of matched persons).

**Table 2: Distribution of Households  
by Number of Non-Matched Individuals**

Individuals not matched	CENSUS AND LFS	
	Number	%
0	23,494	97.00
1	716	3.00
2	11	0.05
3	1	0.004
4	0	0.00
5	1	0.004

Individuals not matched	LFS	
	Number	%
0	22,981	94.90
1	1,112	4.60
2	92	0.40
3	21	0.10
4	12	0.05
5	5	0.02

Individuals not matched	CENSUS	
	Number	%
0	22,143	91.40
1	1,694	7.00
2	244	1.00
3	75	0.30
4	39	0.20
5	15	0.10
6	7	0.03
7	5	0.02
8	1	0.004

It can be seen from Table 2 that for most of the matched households, all their members were found in both the LFS and the Census. About 5% of the matched households coming from the LFS have one or two individuals that were not paired and a few (not even 0.2%) have more than two individuals that were not matched. These individuals represent LFS overcoverage since they were not found in the Census database. Regarding matched Census households, a little more than 8% of them have one, two or three individuals that were not found in the LFS and less than 0.4% have four or more individuals not paired. As the number of original persons in the Census was higher (66,543 for the Census compared to 65,090 for the LFS), more persons were missed in the LFS. Individuals in the Census that have not been paired with an individual from the LFS are part of the undercoverage measure.

Another interesting result with regard to the matched households is the difference between LFS household sizes and Census household sizes. As the households coming from the Census have sometimes more members than the LFS households, it was decided to calculate the difference between household sizes as follows: Census household size minus LFS household size. The results are presented in Table 3.

**Table 3: Distribution of Households  
by Difference between Household Sizes**

Census HHld size minus LFS HHld size	Matched Households	%
-5	1	0.00
-4	8	0.03
-3	13	0.05
-2	72	0.30
-1	466	1.92
0	22,204	91.67
1	1,128	4.66
2	225	0.93
3	57	0.24
4	31	0.13
5	7	0.03
6	8	0.03
7	3	0.01

The difference between household sizes ranges from 0 to 7 people, but most of the households have more or less the same number of members in the LFS and in the Census. In fact, 91.67% of the 24,223 matched households have the same number of members in each dataset. The other 8.33% are either households with missing individuals in the LFS or households with missing individuals in the Census. Fortunately, most of them have no more than two individuals that were not found in the other dataset, that is 7.81% of the matched households.

---





## APPENDIX D

### NEW CATEGORIZATION OF THE CENSUS EDUCATION VARIABLE

As mentioned in the text, the LFS and the Census do not have the same education variables: the LFS has two of them and the Census has only one variable derived. To be able to compare education levels between the LFS and the Census, the Census categories of the education variable were classified as the LFS' are. Table D.1 gives a list of the categories for each LFS education variables and Table D.2 shows the correspondence between LFS and Census variables.

**Table D.1: List of Categories for Each LFS Education Variables**

**Highest Grade of Elementary / Secondary School Completed**

Grade 8 or Lower

Grade 9 and 10

Grade 11 - 13

Graduate from High School

**Highest Degree / Certificate / Diploma Obtained Above Elementary / Secondary School**

None <sup>1</sup>

No Post Secondary Degree / Certificate / Diploma <sup>2</sup>

Trades Certificate / Diploma from Vocational School or Apprenticeship Training

Non-University Certificate / Diploma from Community College, CEGEP, School of Nursing etc.

University Degree Below Bachelor's Level

Bachelor's Degree

University Degree / Certificate Above Bachelor

**1** This category includes i) individuals that did not receive any other education above elementary / secondary school and ii) individuals that received other education that could not be counted towards a degree, certificate or diploma from an educational institution.

**2** This category includes individuals that received any other education that could be counted towards a degree, certificate or diploma from an educational institution, but did not obtained a secondary degree / certificate / diploma.



**Table D.2: Correspondence Between the LFS and Census Education Variables**

CENSUS	LFS 1 <sup>1</sup>	LFS 2 <sup>2</sup>
Grade 11 - 13	Grade 11 - 13	None
Grade 5 - 8	Grade 8 or Lower	None
Grade 9 - 10	Grade 9 and 10	None
Grade 1 - 4	Grade 8 or Lower	None
None or Kindergarten	Grade 8 or Lower	None
Other Non-Univ. Cert. or Diploma	Graduate from High School	Non-University Certificate / Diploma from Community College, CEGEP, School of Nursing etc.
Other Non-Univ. with Trades	Graduate from High School	Trades Certificate / Diploma from Vocational School or Apprenticeship Training
Without Other Non-Univ. Trades Certificate or Diploma	Graduate from High School	No Post Secondary Degree / Certificate / Diploma
Secondary School Graduation	Graduate from High School	None
Trades Certificate or Diploma	Graduate from High School	Trades Certificate / Diploma from Vocational School or Apprenticeship Training
Bach. of First Professional Degree	Graduate from High School	Bachelor's Degree
Doctorate	Graduate from High School	University Degree / Certificate Above Bachelor
Master's Degree	Graduate from High School	University Degree / Certificate Above Bachelor
Univ. Cert. Above Bach. Level	Graduate from High School	University Degree / Certificate Above Bachelor
Univ. with Non-Univ. with Non-University Certificate	Graduate from High School	Non-University Certificate / Diploma from Community College, CEGEP, School of Nursing etc.
University with Trades	Graduate from High School	Trades Certificate / Diploma from Vocational School or Apprenticeship Training
Univ. with Cert. Below Bach.	Graduate from High School	University Degree Below Bachelor's Level
Univ. with Non-Univ. without Cert., Diploma or Degree	Graduate from High School	No Post Secondary Degree / Certificate / Diploma
Univ. without Non-Univ. without Cert., Diploma or Degree	Graduate from High School	No Post Secondary Degree / Certificate / Diploma
Univ. without Univ. with Trades	Graduate from High School	Trades Certificate / Diploma from Vocational School or Apprenticeship Training
Univ. without Non-Univ. with Non-Univ. With Univ. Cert. Below Bach	Graduate from High School	University Degree Below Bachelor's Level
Not Applicable	-----	-----

<sup>1</sup> Highest Grade of Elementary / Secondary School Completed

<sup>2</sup> Highest Degree / Certificate / Diploma Obtained Above Elementary / Secondary School