

## ISSUES WITH HIGH COST AREAS IN THE CANADIAN LABOUR FORCE SURVEY SAMPLE REDESIGN

Edward J. Chen, John Lindeyer and Guy Laflamme<sup>1</sup>

### ABSTRACT

The Canadian Labour Force Survey (LFS) is a monthly household survey conducted by Statistics Canada to provide estimates of employment, unemployment and other key labour market indicators, such as the unemployment rates. The coverage of the survey includes all geographic areas in the ten provinces. Included in some provinces are high vacancy areas (cottage country areas), remote areas in the vast north, and some other areas where data collection costs are high due to inaccessibility such that flying in and overnight stays are required. In the post-census 2001 redesign, all aspects of the LFS design were reviewed and a new strategy was developed to deal with problems in the high collection cost areas.

This paper describes a study undertaken to improve the current LFS design in the expensive areas with an aim to reduce collection costs with minimal impact on the overall LFS data quality. Information from 2001 Census of Population, geography and the ongoing LFS data were used to evaluate and develop a new sample design in these areas.

KEY WORDS: High cost, high vacancy area, remote regions, sample design.

### RÉSUMÉ

L'Enquête sur la population active canadienne (EPA) est une enquête auprès des ménages élaborée par Statistique Canada qui fournit des estimations d'emploi, de chômage ainsi que d'autres indicateurs clés du marché du travail tels que le taux de chômage. La couverture de l'enquête comprend toutes les régions géographiques des dix provinces du Canada. Certaines provinces comportent des secteurs à faible densité et les régions éloignées du grand nord où les coûts de collecte de données sont élevés dû à la difficulté d'accès nécessitant des voyages par avion et des séjours de plus d'une journée. Dans le cadre du remaniement post-censitaire de 2001, tous les aspects du plan de sondage de l'EPA ont été revus et une nouvelle stratégie a été développée afin de régler les problèmes avec les régions à collecte coûteuse.

Cet article décrit une étude entreprise pour améliorer le plan de sondage actuel de L'EPA dans les secteurs coûteux dans le but de réduire les coûts de collecte avec un impact minimal sur la qualité des données de l'EPA. De l'information provenant du Recensement de la population de 2001 ainsi que de l'EPA en cours a été utilisée pour évaluer et développer un nouveau plan de sondage pour ces régions.

MOTS CLÉS : Coût élevé; plan de sondage; régions éloignées; secteur à haut taux de vacance.

### 1. INTRODUCTION

The Canadian Labour Force Survey (LFS) plays a central role in the national statistical system and is one of Statistics Canada's mission critical surveys. Every month, the LFS provides official estimates of labour market conditions which are among the most timely and important measures of the overall performance of the Canadian economy. The key estimates published include the unemployment rates and employment totals at national, provincial and sub-provincial levels. The LFS is the largest monthly household survey conducted by Statistics Canada (STC) and its survey frame, sample and processing systems support a wide and expanding range of other household surveys.

The LFS sample is redesigned every ten years following the latest Census of Population and extensive consultations with stakeholders – including the provincial focal points and major data users. The objectives of the LFS redesign are to update the household survey frame, stratify the population to meet changing user needs, and allocate the sample to reflect

---

<sup>1</sup> Edward J. Chen (Edward.Chen@statcan.ca), John Lindeyer and Guy Laflamme, Household Survey Methods Division, Statistics Canada, R.H. Coats Building, Ottawa, Canada, K1A 0T6.

population changes since the last redesign. In addition, the new design must reflect the Agency's fiscal reality to seek operational efficiency and to generate permanent cost savings. Streamlining the survey in high collection cost areas is one way to achieve this goal.

The LFS uses a complex two-stage area design where the first stage unit is called cluster and the second and ultimate stage unit is dwelling. This paper describes a study undertaken to examine the current high collection cost areas in Canada – the scope of study included high vacancy areas, remote areas in the vast north and other areas with high collection costs due to inaccessibility where overnight stays or flying in are often required in order to collect the household data.

## **2. HIGH COLLETION COST AREAS**

### **2.1 Review of the Current Design**

A review of the current design with respect to the high collection cost areas – namely the high vacancy areas, remote areas and other expensive areas – was conducted with input from the regional offices across Canada. In part, much of the review work commenced at the time when the Agency launched the Strategic Streaming Initiative (SSI) to seek operational efficiency. The Agency's SSI was a process for reviewing opportunities to enhance work flows and operations in different parts of the organization. The SSI objective was to try to do the work as efficiently as possible – without sacrificing quality. The focus of the SSI project in the LFS was on reviewing collection costs – identifying areas with high collection costs that posed relatively little impact if they were removed. The work can be summarized as follows:

### **2.2 High Vacancy Areas**

The high vacancy areas, in most cases, are the cottage country areas where many dwellings are classified as seasonal, secondary or unoccupied dwellings. Many of these vacant dwellings are also located near the shorelines of lakes or in isolated areas such as islands or difficult terrain where accessibility is also a major issue.

With input from the regional offices, a rigorous process was conducted to identify high cost areas and potentially exclude them from the collection process in the current design. Two years of vacancy rates from the LFS response files were used to first identify the high vacant clusters. We then consulted with the regional offices to determine the collection costs and the effort required for this preliminary list of clusters. At the same time, some high cost clusters were suggested by the regional offices for potential removal. A final determination to exclude these clusters from collection was based on several criteria: vacancy rates, collection costs, and their impacts on the sub-provincial estimates such as the Employment Insurance and Economic Regions (EIER) estimates. Altogether, a monthly sample of about 360 households was excluded from collection from April 2003 to April 2005 during the last two years of current design. A special estimation procedure was employed to compensate for the exclusion of these clusters.

Significant cost savings were achieved under the current design by the removal of these clusters. More importantly, these excluded areas, together with other areas that were identified but were not excluded, amass a major knowledge base of high cost areas in Canada. These lists of clusters or areas are important input to the 2004 LFS sample redesign.

### **2.3 Remote Areas**

The current design in the remote areas was reviewed with the aim to develop a more efficient design in these high cost areas. The current design uses census enumeration areas (EAs) or places as first stage sampling units, with any EA or place with fewer than 10 households or 25 persons omitted from the frame. Many of these northern areas are sparsely populated and difficult to access, with flying in or overnight stays often required. In addition, telephone coverage in remote areas was reviewed but found to be of poor quality. The remote areas comprised about 1% of the LFS sample, but with considerably higher collection costs a streamlined design is warranted.

### **2.4 Other High Cost Areas**

Other high collection cost areas often encountered in the regions are the isolated places or difficult to access areas such as mountainous regions. Low telephone coverage and lack of availability of interviewers in the region also contribute to higher cost. Some of these areas might have higher vacancy rates but the collection costs are really related to the availability of local interviewers.

## **3. NEW DESIGN AND STRATIFICATION RESULTS**

### 3.1 Identification of Exclusion Areas

In this redesign, a new set of clusters covering the entire country was created thanks to improvements in the geographical database available at Statistics Canada. Using the geographical information, these clusters were formed by grouping census blocks with the census and geographical information, such as vacancy rates and population density, computed for each cluster.

It's necessary to develop options and identify areas for exclusions in the household survey frame. An option to exclude the high vacancy areas and remote areas entirely from the household frame coverage was evaluated. It was felt that such an omission would introduce significant under-coverage, especially in the northern regions. Instead, the option of excluding some extreme problem areas and/or low population density areas was adopted. A complete and thorough process was conducted to define the high vacancy areas and remote areas for study. The new clusters located in the current remote areas were reviewed for their road networks, distance to major urban areas and accessibility. The scope of remote study areas included all of these difficult clusters, but it also included all clusters within the census sub-division where necessary. The scope of high vacancy areas included all rural clusters outside of Census Metropolitan Areas (CMAs) and outside of the defined remote areas.

Criteria for exclusions of areas from the coverage of household frame were developed. A different set of criteria was adopted for the high vacancy areas and remote areas so that the impacts of the exclusions could to the overall household survey frame could be minimized. Many excluded clusters were also reviewed on a case-by-case basis. The criteria of exclusions were:

**High Vacancy Areas:** The exclusions in these areas were largely based on extremely high vacancy rates (often higher than 75%), as well as some of the excluded clusters and their surrounding areas determined in the SSI for the current design. Many of these clusters are located in very isolated regions where there will not be any major residential developments or population growth in the next ten years. Some excluded clusters were also located on islands or known problem areas, and their removal has no impact because of low population.

**Remote Areas:** The exclusions in these areas were based largely on density and accessibility considerations because of the sparse population in the northern areas. Population per square kilometre and population per kilometre of street were computed for census blocks within remote clusters. Census blocks in remote clusters with surface area greater than fifty square kilometres, but with either no residents or less than 20 households per hundred kilometres of road, were excluded. The remaining portions of these clusters were then considered for accessibility and household counts before a decision was taken to either include them or exclude them. In addition to these exclusions, several remote clusters were also excluded from the frame because of its extremely high vacancy rates.

One small island in Nova Scotia and two water areas in Newfoundland and Labrador were excluded because of low population. Similarly, a large logging area located in the inland Gaspésie was excluded.

Table 1 summarizes the exclusions from the household survey frame coverage. Residents of Indian reserves remain excluded from the LFS target population. Included in the Appendix is a national map that displays the 2004 LFS design, with the large excluded areas.

<b>Exclusion Types</b>	<b>Clusters</b>	<b>Population</b>	<b>Dwellings</b>	<b>Households</b>	<b>Vacancy (%)</b>
<b>Indian Reserves</b>	<b>1,760</b>	<b>301,659</b>	<b>97,325</b>	<b>85,944</b>	<b>11.7</b>
<b>High Vacancy</b>	<b>41</b>	<b>16,309</b>	<b>28,353</b>	<b>6,902</b>	<b>75.7</b>
Newfoundland and Labrador	7	1,520	3,821	624	83.7
Nova Scotia	1	249	257	91	64.6
Ontario	25	11,382	15,263	4,892	67.9
Manitoba	8	3,158	9,012	1,295	85.6
<b>Remotes</b>	<b>357</b>	<b>13,384</b>	<b>22,090</b>	<b>4,918</b>	<b>77.7</b>
Newfoundland	5	67	419	27	93.6

and Labrador					
Quebec	122	898	1,804	442	75.5
Ontario	79	5,284	14,142	2,153	84.8
Manitoba	30	852	2,007	306	84.8
Saskatchewan	18	883	1,172	288	75.4
Alberta	55	3,197	1,317	872	33.8
British Columbia	48	2,203	1,229	830	32.5
<b>Small Island</b>	<b>1</b>	<b>15</b>	<b>2</b>	<b>1</b>	<b>50.0</b>
<b>Inland Gaspésie</b>	<b>42</b>	<b>0</b>	<b>120</b>	<b>0</b>	<b>100.0</b>
<b>Water Areas</b>	<b>2</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0.0</b>
<b>Total:</b>	<b>2,203</b>	<b>331,367</b>	<b>147,890</b>	<b>97,765</b>	<b>33.9</b>

### 3.2 New Design in High Vacancy Areas

The new design in the high vacancy areas consisted of first grouping of high vacant clusters into separate strata and ensuring there would be sufficient sample yields. Any cluster with a vacancy rate higher than 50% from outside of Census Metropolitan Areas (CMAs) and remote areas was considered as a potential candidate for inclusion in the high vacancy strata. The vacancy threshold of 50% was lowered to 40% in two special cases. 1) Efforts were taken to include the current excluded areas but with lower vacancy rates reported from census information because of changing cluster boundary. 2) Some difficult-to-access clusters reported from regional office consultations but with lower vacancy rates were also included. On the other hand, some clusters with vacancy rate higher than 50% were not included in the high vacancy strata since they were found to be located near major urban areas or because they were not in any known cottage country areas – especially in Ontario and Quebec.

Table 2 presents the stratification results for the high vacancy strata in the 2004 LFS redesign. High vacancy strata were created in all provinces except Prince Edward Island and New Brunswick where there was not sufficient population in these areas to form separate strata. In Ontario and Quebec, more than one stratum was created, respecting the Department of Human Resources and Skills Development's (HRSDC) Employment Insurance and Economic Regions (EIERS) as much as possible. In these two provinces, the high vacancy clusters are found in more concentrated cottage country areas.

**TABLE 2: Stratification Results for High Vacancy Strata in the 2004 LFS Redesign**

Province	Strata	Clusters	Population	Dwellings	Households	Vacancy (%)
<b>Newfoundland and Labrador</b>	1	37	13,825	16,192	5,184	68.0
<b>Nova Scotia</b>	1	22	10,111	10,761	4,196	61.0
<b>Quebec</b>	2	132	52,084	64,483	23,546	63.5
<b>Ontario</b>	3	234	103,444	143,172	43,991	69.3
<b>Manitoba</b>	1	28	12,097	14,516	5,119	64.7
<b>Saskatchewan</b>	1	42	17,472	23,395	7,077	69.7
<b>Alberta</b>	1	26	11,089	11,718	4,552	61.2
<b>British Columbia</b>	1	64	28,625	33,514	12,512	62.7
<b>CANADA</b>	11	585	248,747	317,751	106,177	66.6

Sample yields from the 2004 LFS design sample allocation were examined in each of the high vacancy strata. The stratum inverse sampling ratio (ISR) was then increased to as much as double its initial value with the constraint that the desired sample take under the new ISR in each stratum should be at least thirty households. The desired sample takes per cluster respect the general rule of about ten households in rural area design.

Unlike the sample selection method in other rural clusters where six clusters were often selected using randomized PPS systematic sampling, a sample of three or four clusters was selected using PPS sampling after the units were sorted by the vacancy rates in the stratum. Then a sample of dwellings was selected using systematic sampling. In several strata, one selected high vacant cluster was further dropped from collection during the first three years of the redesign. Again, this

cluster was dropped after consultation with the regional offices who confirmed it be costly or problematic. At the same time, high cost clusters in the rural areas were identified and some were dropped from collection. An estimation procedure is employed to compensate these omissions.

Under the new design, the sample yields in the high vacancy strata have been reduced by about fifty percent. The sample size reduction from the high vacancy strata was then re-distributed to the surrounding strata in the same EIER. In this way, the sample allocation to the EIER remains largely constant while shifting sample from higher cost areas to the less costly ones.

### 3.3 New Design in Remote Areas

The new design in the remote areas consists of first grouping remaining remote clusters into strata. Sample yields from the 2004 design sample allocation were examined in each of the remote strata. The stratum inverse sampling ratio (ISR) was then increased to as much as double its initial value with the constraint that the desired sample take under the new ISR in each stratum should be at least twenty households.

Table 3 presents the stratification results for the remote strata in the 2004 LFS redesign. Remote strata were created in all provinces except Prince Edward Island, New Brunswick and Nova Scotia where there was not sufficient population in these areas to form separate strata. In Ontario, Alberta and British Columbia, more than one stratum was created, respecting the EIER as much as possible.

**TABLE 3: Stratification Results for Remote Strata in the 2004 LFS Redesign**

Province	Strata	Clusters	Population	Dwellings	Households	Vacancy	Coverage
<b>Newfoundland and Labrador</b>	1	24	10,190	3,386	3,012	11.0%	99.3%
<b>Quebec</b>	1	75	32,716	13,393	11,373	15.1%	97.3%
<b>Ontario</b>	2	144	55,880	32,477	21,595	33.5%	91.4%
<b>Manitoba</b>	1	41	14,932	7,337	5,310	27.6%	94.6%
<b>Saskatchewan</b>	1	28	12,431	5,154	3,988	22.6%	93.4%
<b>Alberta</b>	2	98	49,066	18,278	15,673	14.3%	93.9%
<b>British Columbia</b>	2	88	35,875	15,838	13,390	15.5%	94.2%
<b>CANADA</b>	10	498	211,090	95,863	74,341	22.5%	94.0%

Two or more (but less than six as in other type of strata) clusters were selected in each stratum using PPS sampling after the units were sorted by the number of households. Then a sample of dwellings was selected using systematic sampling. No additional selected remote clusters were identified to be dropped from collection even though the design allows such flexibility.

With the same consideration as the high vacancy strata, the sample yields in the remote strata have been reduced by about fifty percent. Furthermore, the remote stratum in the Newfoundland and Labrador and one remote stratum in Alberta are excluded from collection for the next three years. The sample yields from the initial allocation are re-allocated to the same EIER regions but outside of the remote areas as much as possible. In this way, the sample allocation to the EIER remains largely constant while shifting sample from higher cost areas to less costly ones.

## 4. DISCUSSION

The new design successfully reduced sample yields in high vacancy and remote areas by about fifty percent, while maintaining coverage in these areas. The new design allows for maximum flexibility in the high cost areas so that the sample can be restored when there is sufficient growth in the area or the financial situation improves.

The more efficient design in the high cost areas is one of the major achievements in the 2004 LFS redesign. The design and sample reduction strategy in the high cost areas plays an important role in repaying the corporate investment in funding the redesign activities. In addition, less sample size reduction was needed in order to repay the investment while the impact to the overall data quality was reduced to minimum. The smaller sample reduction largely satisfies the needs of the LFS stakeholders while the new design also satisfies the Regional Offices' desire for efficiency in high cost areas.

Sample maintenance in the high cost areas can now be better monitored. We have better knowledge of the problem collection areas and are better able to react to collection problems encountered in the regions. Moreover, many other household surveys conducted by Statistics Canada can better allocate their samples and reduce their collection costs accordingly.

### ACKNOWLEDGMENTS

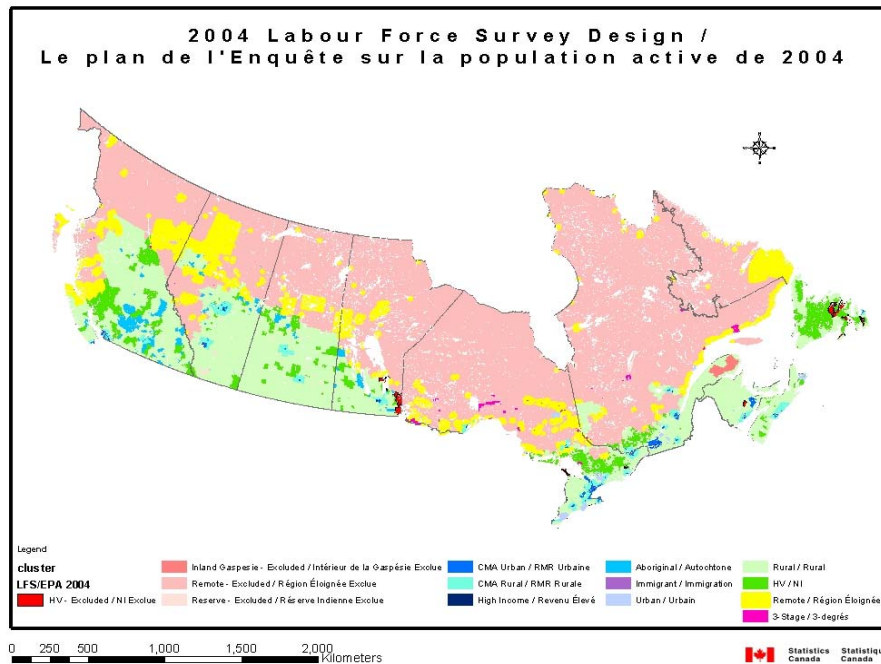
The authors would like to acknowledge many useful discussions and input from the 2004 LFS Redesign Committee. In particular, the direction and input from M.P. Singh, Jack Gambino, Harold Mantel and technical assistance from T.P. Liu are highly appreciated.

### REFERENCES

Statistics Canada, Methodology of the Canadian Labour Force Survey (1998), Catalogue no. 71-526-XPB.

G. Laflamme and C. Turmelle (2004), Redesign of the Canadian Labour Force Survey, Proceedings of the Joint Statistical Meeting, Toronto.

### APPENDIX: 2004 Labour Force Survey Design



Note: The remote cluster sizes in this map are enlarged to show the location of the remote clusters.