

STATISTICS CANADA'S NEW HOUSEHOLD SURVEY STRATEGY

Jack Gambino, Jean-Louis Tambay and Guy Laflamme¹

ABSTRACT

In this paper we begin with a review of Statistics Canada's current approach to conducting household surveys. Throughout its history, Statistics Canada has conducted the vast majority of its surveys using an area frame approach with the monthly Labour Force Survey sample design as the key component in a complex system. With the increased use of telephone interviewing, advances in technology and the creation of an Address Register (AR) for the census, alternatives to the current approach have become more feasible and attractive. As a result, there have been investigations of such alternatives in the past few years. In the second part of the paper, a new approach to household surveys, featuring increased integration, common core content, a master sample and possible use of an AR-based list frame are described along with some of the technical issues that the new approach raises.

KEY WORDS: Household survey; Master sample, Survey redesign

RÉSUMÉ

L'article commence par une brève revue de l'approche courante de Statistique Canada pour les enquêtes auprès des ménages. La plupart des enquêtes auprès des ménages menées par Statistique Canada utilisent une base aréolaire pour laquelle le plan de sondage de l'Enquête sur la population active est une composante clé. Grâce à une utilisation accrue des entrevues téléphoniques, aux avancées technologiques, et à la création d'un registre des adresses (RA), des alternatives à l'approche courante deviennent plus réalisables. Conséquemment, de telles alternatives ont récemment été étudiées. Dans la deuxième partie de l'article, nous décrivons une nouvelle approche pour les enquêtes auprès des ménages, utilisant une intégration accrue, un contenu central commun, un échantillon maître et l'usage d'une base de sondage de type liste basée sur le RA. Des questions méthodologiques soulevées par cette nouvelle approche sont présentées.

MOTS CLÉS : Échantillon maître; enquête auprès des ménages; remaniement d'enquête.

1. INTRODUCTION

The history of household sample surveys at Statistics Canada begins with the introduction of the Labour Force Survey (LFS) in 1945. The LFS began as a quarterly survey and became a monthly survey in 1952. From the outset, the LFS was used as a vehicle for other surveys. In fact, the first LFS supplementary survey (i.e., a survey whose questions are asked right after the LFS questions), on overcrowding, was conducted in February 1946, which was only the second time the LFS was run. The LFS has been used in this way ever since. In addition, some surveys are conducted by interviewing LFS respondents after they have completed their six months in the LFS sample. These respondents are sometimes referred to as "rotate-outs" since they are households that were in the LFS and have been replaced by a new set, or rotation group, of households. Finally, some surveys use the LFS design (specifically its sampling mechanism) to select households that have not been interviewed by the LFS.

The use of the LFS as a general-purpose vehicle for household surveys increased over time. It became even more attractive to use the LFS in this way following major LFS sample size increases in the mid-1970s and in 1989. As a result, the burden on respondents continued to grow. As the demand for new surveys continued to increase, it became clear that the infrastructure would eventually not be able to meet the demand. In this paper we describe the strategy that is being

¹ Household Survey Methods Division, Statistics Canada, Ottawa, Canada, K1A 0T6, gambino@statcan.ca, jean-louis.tambay@statcan.ca, guy.laflamme@statcan.ca

developed at Statistics Canada to change the household survey program to enable it to satisfy the needs of the current suite of surveys as well as future ones. In section 2, we describe the current situation, and we present alternatives in section 3. The remaining sections describe the features and benefits of the proposed new strategy.

2. THE CURRENT SITUATION

In the introduction, we presented an LFS-centric view of the household survey program at Statistics Canada, and this reflects the bulk of the surveys conducted by the agency. The exceptions include surveys that use random digit dialing (RDD), the General Social Survey being the major example of this, and post-censal surveys, such as the Aboriginal Peoples Survey (APS), which are conducted after each population census and use the census as a frame. In this section, we will focus on the LFS design and how it is used by the LFS and other household surveys.

Since its inception, the LFS has used an area frame and a multistage design to select its sample. The LFS is redesigned every ten years following census years that end in a 1. One notable trend in the design of the LFS has been the gradual reduction in the number of stages of sampling. In the current design, there are only two stages of sampling in most areas. The first stage consists of a sample of clusters, where each cluster is a small geographical area (e.g., adjacent city blocks) containing about 200 dwellings. Clusters are grouped into strata. In each stratum, a sample of six clusters is selected using probability-proportional-to-size (PPS) sampling. Each sample cluster is associated with a rotation group. Rotation group 1 corresponds to units rotated in the sample in January and July, rotation group 2 to units in February and August, and so on.

A current list of the dwellings in each selected cluster is obtained and a sample of these dwellings is selected systematically. The people in the selected dwellings are in the survey for six consecutive months. The dwellings are then “rotated out” and replaced by another set of dwellings from the same cluster. In each stratum, this rotation of dwellings occurs in one cluster per month according to the rotation group numbers assigned to the clusters. Thus one-sixth of the LFS sample is replaced each month. When a cluster is exhausted, it is replaced by another cluster in the same stratum.

Because of its relatively large sample size (currently 54,000 households per month, one-sixth of which are newly rotated-in), the LFS is an attractive vehicle for conducting other household surveys. For them, the cost of using the LFS or its infrastructure is far smaller than starting from scratch. As a result, the LFS has been used extensively by other household surveys throughout its history. As we noted in the introduction, there are three ways of using it. First, there are supplementary surveys which use current LFS respondents. An example of this is the Travel Survey of Residents to Canada, which asks its questions right after the interviewer finishes asking the LFS questions. Second, some surveys contact LFS respondents one or more months after they have completed their six months in the LFS. An example of this is the Youth in Transition Survey which used rotate-outs for part of its sample. Third, some surveys select their own sample from the LFS frame, avoiding dwellings that have already been, or are scheduled to be, in the LFS. The Survey of Household Spending is an example of this type of survey. The first approach is acceptable for small surveys whose questionnaire imposes a relatively light respondent burden. The second is preferred by surveys that need to target specific sub-populations and/or that would benefit from the contacts already established with the household. The third approach is used more by surveys that impose a larger respondent burden and that cover the general population.

Although this approach to conducting household surveys has been very successful, it can no longer meet the demand for household surveys in an effective way. There is a limit to how many supplements LFS respondents are willing to participate in. Increasing their respondent burden eventually leads to lower response rates. The LFS infrastructure cannot handle, in a timely and cost-effective way, all the surveys that Statistics Canada’s clients would like it to conduct. Furthermore, although the LFS infrastructure forces a certain amount of coordination across the surveys that use it, there is scope to improve how we manage sample, content and processing systems. Therefore, it becomes natural to think of alternatives that would provide more sample to meet the demand, while coordinating all aspects of collection and processing across surveys. The ensuing redesign of the household survey program also provides an opportunity to bring greater coherence to survey content by standardizing the way questions that are asked in several surveys, such as income, are worded.

Before discussing the major alternatives that have been considered, we briefly mention the use of telephone numbers as a frame source. In Canada, the number of households that have only cell phones is increasing rapidly. Since most cell phone payment plans bill users by the minute, there has been a reluctance to conduct interviews by cell phone. As a result,

surveys that use Random Digit Dialing do not cover households that have only cell phones. In addition, surveys that use telephone billing files to construct a list frame of telephone numbers also miss these households since the billing files cover only landline telephones. Thus the days of telephone-only surveys appear to be numbered unless some of these things change. For example, if we were able to obtain billing files for both cell phone and landline numbers, we would then be able to construct a more complete list frame.

For completeness, we mention the Address Register (AR) since it has started to play an important role in household surveys, and this role is expected to grow. The AR was initially created for the 1991 Census of Population, and its coverage and quality have improved with every census since then. The AR is a list of residential addresses throughout the country. Overall, for the 2006 census, it covered 89 percent of all residential addresses, with an even higher percentage of addresses in urban areas. In the current LFS design, the AR is used to provide initial lists of addresses for clusters in areas where it is felt that the AR coverage is adequate for this purpose. The overall coverage of the AR is expected to go well beyond 90 percent by the next census in 2011, which suggests the possibility of using the AR as a list frame in some places. Thus in areas where the AR coverage is close to complete, it becomes possible to replace the current stratified two-stage design by a stratified single-stage design. This would complete the long-term trend in the reduction of stages of sampling alluded to earlier.

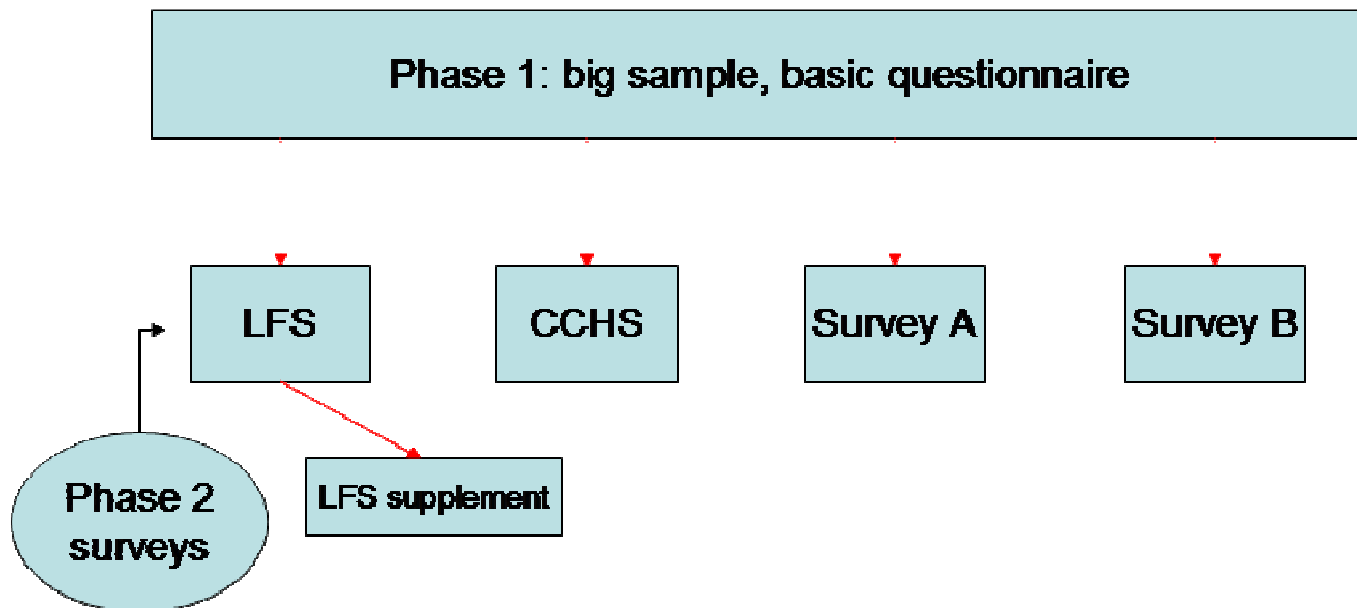
3. ALTERNATIVES

From 2001 to the present, inspired by Singh (2001), several alternatives to the current approach to conducting household surveys have been studied (Statistics Canada, 2007). Here, we present only the two major ones, since most of the others were variations on these. A common theme among the alternatives is the creation of a master sample. One can view the current LFS as a master sample from which other surveys (supplements, rotate-outs) select their own samples. In the alternative options, the master sample is a broadening of the current set-up. The meaning of this becomes clearer when we describe the two alternatives.

3.1 Alternative 1: Pure two-phase design

In the first alternative, each year, a very large master sample of dwellings, say 300,000 of them, receive a basic, short questionnaire. The questionnaire covers only core content such as demographics, immigration, education and income. This basic information is then used to select subsamples. Thus the master sample constitutes a first-phase sample, which is followed by subsequent phases of sampling. This is illustrated in Figure 1 below.

Figure 1: Pure two-phase design



In the figure we see that the LFS and several other surveys become second-phase surveys. The survey labelled CCHS is the Canadian Community Health Survey, which currently has a sample of 65,000 households per year. The LFS, as well

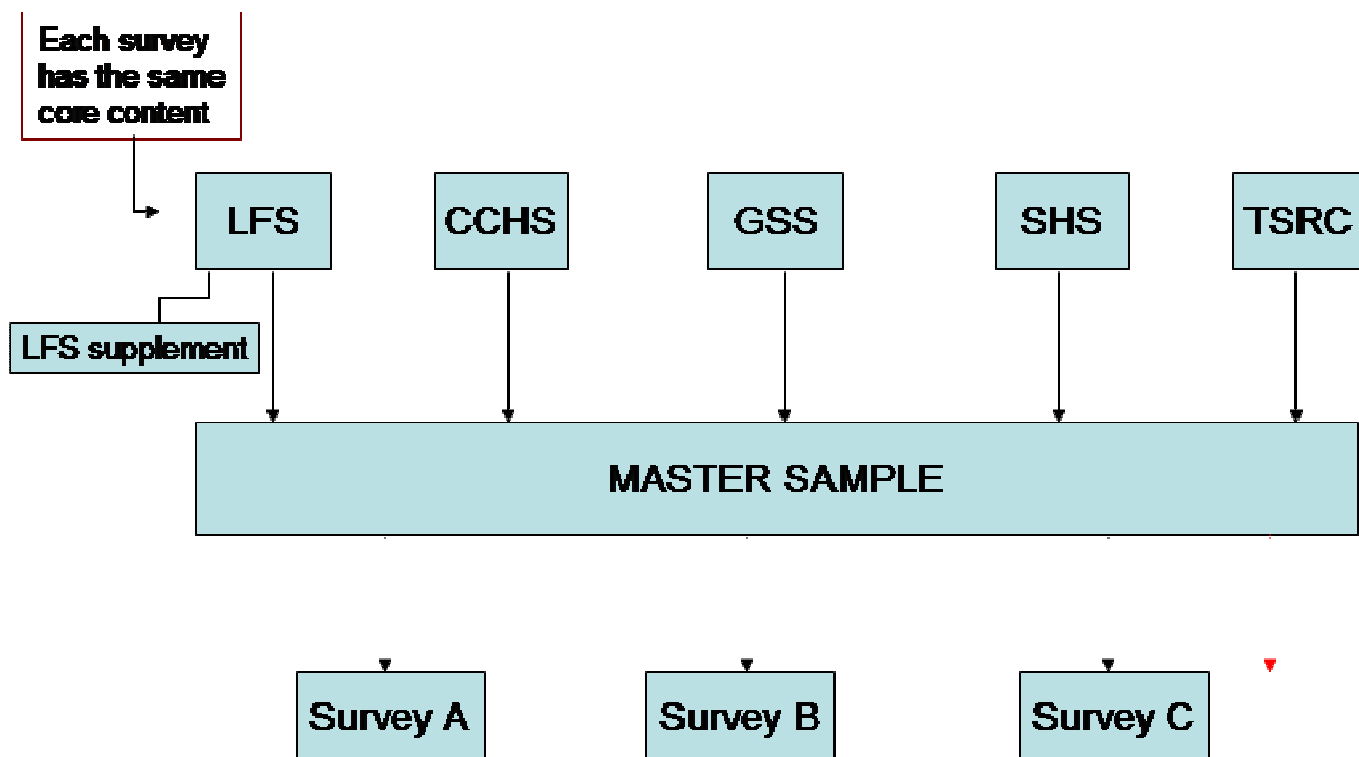
as other second-phase surveys, can continue to have their own supplements. The advantage of this alternative is that information from the first phase can be used to select sample from targeted populations or to bring efficiencies to the sample design. For example, if Survey A is a survey of immigrants, it can select its sample with little wasted, since immigrant status is one of the core questions asked in the first phase.

While this alternative is very attractive, it was ultimately rejected because of cost. Even if the first phase is conducted by mail, the need for telephone and personal follow-up of nonrespondents increases the total cost of the program too much to make this a feasible option. Hence, we have decided to focus our efforts on a second alternative which we describe next.

3.2 Alternative 2: The master sample as a union of surveys

The second alternative, which is illustrated in Figure 2 below, can be viewed as turning the approach in the first alternative on its head. Major surveys such as the LFS and CCHS, and possibly others, are conducted on their own as they are now, and then their samples are combined to form a master sample. To be able to use the master sample as a source for selecting future sample, we change the content of the first-phase surveys (i.e., the LFS, CCHS, etc.) so that they all have a common core set of questions. This common core would be the same as the one in Alternative 1. It would not require major changes for these surveys since they collect much of this information already. Of course, aside from the common core content, the respective surveys would continue to ask their usual questions, so the effect and the effort required will be relatively minor. Compared to Alternative 1, the extra cost, other than start-up costs, to the overall household survey program is relatively modest since the creation of the master sample becomes essentially a “head office exercise”. The start-up costs include developing and implementing the common core content and developing methodologies and systems to integrate the samples into a master sample and then use the master sample to select subsamples.

Figure 2: Master sample as a union of surveys



In Figure 2, we assume that, in addition to the LFS and CCHS, three other surveys contribute to the master sample. They are the General Social Survey (GSS), the Survey of Household Spending (SHS) and the Travel Survey of Residents to

Canada (TSRC). The latter is currently an LFS supplement, and we are considering options for making it a stand-alone survey in order to free up the LFS sample for labour-related supplements and to increase the size of the master sample. A major consideration is the impact this would have on costs. It is felt that, for the creation of a master sample to be a worthwhile effort, at a minimum, it should comprise the samples of the LFS, CCHS and TSRC. Because the GSS is currently an RDD survey, it may not be feasible to use it as a first-phase survey due to cost. The SHS is already a first-phase survey, but it is fairly burdensome for respondents, hence it may only be suitable as a source of samples for surveys whose content fits in well with SHS content.

4. FEATURES OF THE PROPOSED ALTERNATIVE

In this section, we review some of the features and benefits of the proposed alternative.

We saw in the previous section that a requirement of the new approach is replacing some of the content of existing surveys such as the LFS and CCHS with a common set of questions. As a result, Statistics Canada initiated a review of concepts and definitions related to a variety of variables such as income and family. This is a worthwhile activity regardless of whether the new household survey strategy ever gets implemented since it allows us to harmonize some key concepts and definitions used by many household surveys. The goals of this effort go beyond harmonization. One of the outcomes will be the development of off-the-shelf modules for variables such as income and education. Future surveys will then have access to fully developed modules, for both content and editing, that they can easily incorporate into their questionnaires.

The existence of a master sample that is used by most household surveys both encourages and, to some extent, imposes better management of data collection over time. In addition, since the master sample is the parent of these surveys, it makes sense to have a common mechanism, and therefore processing systems, for the survey steps beginning with the selection of units from the master sample. How far to extend the commonality, i.e., just sample selection versus including nonresponse adjustment, weighting, and so on, has not yet been determined.

As part of the new household survey strategy, we are investigating how to better use multiple modes of data collection since both the first-phase and subsequent surveys will use a variety of modes. In fact, the LFS and CCHS already use a mix of telephone and personal interviewing. A major inspiration has been the recent experience of the 2006 Census of Population, which successfully used mail, telephone, personal and internet modes for data collection. The census managed to efficiently move cases between collection modes, interviewers and regional offices.

A major challenge under a master sample scheme is sample coordination. Rules for who gets what sample and when, and the associated allocation of costs, need to be worked out. The rules are not obvious. For example, the CCHS may want to reserve a part of the master sample for which it is the parent for itself because it expects to conduct a follow-on survey in the future. However this may prevent other surveys from using the same sample because of concerns about response burden. Another type of problem arises when a second-phase survey needs a large sample from a targeted subpopulation. For example, a survey may want a large sample of households with children, effectively using up all such households in the master sample. If other surveys then want to draw from the master sample, they have to choose between two less desirable options. They can avoid households that were in the children's survey and accept a sample with very few children (and that is potentially biased), or they can ignore the fact that the households were in the children's survey and risk low response rates for that part of their sample. This shows that a master sample is not a panacea. Nevertheless, it is an improvement to the current situation, even in the situations described in this paragraph.

5. FURTHER STUDIES

The development of a new household survey strategy has included several studies, which we briefly describe here.

Since one of the sample design options is to do away with clustering and introduce a list frame in some areas, we are investigating the impact of this change on the sampling variance. Preliminary results show that there is a small but significant benefit for labour-related variables such as employment and unemployment. We expect bigger gains for continuous variables such as income. We are investigating a compromise design which would do away with clusters but, instead of selecting individual dwellings, we would sample "mini-clusters" consisting of a handful of dwellings. This has

the advantage of keeping travels costs close to those under a clustered design while having some of the benefits of a list frame.

Sample coordination strategies for second-phase surveys are being investigated. This is being done at both the theoretical and practical level. For the latter, a pilot test will be conducted in 2008. In the test, the combined LFS-CCHS sample will be used as a master sample prototype. A second-phase sample will be drawn from it and the sampled households will receive the GSS questionnaire.

Finally, despite the increasing coverage problems with telephone-based frames mentioned earlier, we still expect to make some use of such frames, usually in a dual-frame framework, in some areas. Consequently, we are conducting a study of the impact of this coverage bias on the bias of estimates. We are also studying sample allocation issues under a dual-frame approach in the context of multiple surveys.

ACKNOWLEDGEMENTS

The authors would like to thank Tracey Leesti and Julie Trépanier of Statistics Canada for comments and suggestions that improved the paper.

REFERENCES

Singh, M.P. (2001). The Canadian Annual Population Survey. Internal Statistics Canada document.

Statistics Canada (2007). Statistics Canada's Household Survey Strategy. Internal Statistics Canada document.