Catalogue no. 11-522-XIE

Statistics Canada International Symposium Series - Proceedings

Symposium 2004: Innovative Methods for Surveying Difficult-to-reach Populations

2004

Statistics Canada Statistique Canada

Canada

Proceedings of Statistics Canada Symposium 2004
Innovative Methods for Surveying Difficult-to-reach Populations

METHODOLOGICAL CHALLENGES IN A SURVEY ON THE ETHNIC AND CULTURAL DIVERSITY OF THE CANADIAN POPULATION

Valérie Bizier, Jennifer Kaddatz and Danielle Laroche[1]

ABSTRACT

There are many challenges in conducting a survey of a culturally diverse population, especially when the population is as multicultural as Canada's. This paper describes the methodological problems that the Ethnic Diversity Survey team encountered in choosing the sampling plan, developing the questionnaire, collecting the data, weighting the data and estimating the variance, and the various methods used to address the challenges inherent in a survey of this type.

KEYWORDS: Ethnic Origins; Postcensal Survey; Questionnaire Design; Rare Populations.

1. BACKGROUND

1.1 Reasons for developing the Ethnic Diversity Survey

The ethnic and cultural composition of Canada's population has changed substantially over the last century, as immigration is playing a key role in population growth. Up to 1971, a majority of newcomers to Canada were from European countries. Since then, the number of immigrants from Asia, Africa, the Caribbean, Central America and South America has risen sharply, increasing the cultural and racial diversity of Canadian society.

Because of this trend, the country's leaders need objective information about ethnicity to support strategic decisions aimed at ensuring social cohesion and the inclusion of all Canadians in Canadian society. Up to now, the census question on the ethnic or cultural origin of the respondent's ancestors has been the main source of data on ethnic and cultural diversity in Canada. The trouble with that question is that it measures only one aspect of ethnicity. A person's ethnicity is influenced by a number of other factors, such as ethnic identity, race, nationality, language profile and religion.

Moreover, there has been a change in the way people report their origins in recent years. The number of people who report having ancestors of Canadian origin has increased from census to census, which suggests that the question is often misunderstood.

To measure ethnicity more effectively and fill some statistical gaps related to ethnicity, Statistics Canada and Canadian Heritage joined forces to develop the Ethnic Diversity Survey (EDS).

The Ethnic Diversity Survey (EDS) had four goals.

- explore various ways of measuring ethnicity to assist in future data collection;
- better understand how Canadians of different ethnic backgrounds interpret and report their ethnicity;
- provide information about ethnic diversity in Canada;
- examine how people's backgrounds affect their participation in the social, economic and cultural life of Canada.

[1]Valérie Bizier, Statistics Canada, R.H. Coats Building, 15 Q, 120 Parkdale Ave., Ottawa ON, Canada, K1A 0T6;
Jennifer Kaddatz, Statistics Canada, 600-300 W. Georgia Street, Vancouver BC, Canada, V6B 6C7;
Danielle Laroche, Statistics Canada, R.H. Coats Building ,15 Q, 120 Parkdale Ave., Ottawa ON, Canada, K1A 0T6.

1.2 Target population

The EDS's target population consisted of persons aged 15 or over, living in private dwellings in the 10 provinces, which included Canadian citizens, landed immigrants and non-permanent residents.

The territories, remote areas and collective dwellings had to be excluded from the target population because of collection constraints and costs. Native peoples and residents of Indian reserves were also excluded because a survey of Aboriginal peoples was being conducted at almost the same time, and Statistics Canada wanted to minimize the response burden for those who might have been selected for both surveys. In 2001, the survey's target population was estimated at 23,092,645.

2. SAMPLING PLAN

2.1 Choosing the sample frame

To achieve the survey goals, it was essential to survey both the majority ethnic groups and the ethnic minorities in Canada. The usual frames, such as area frames or dwelling frames, were inappropriate for the survey because ethnic minorities are generally rare, scattered populations.

The postcensal approach seemed to be the best option for creating the frame. It involves using the census database to locate the target population. It has several benefits. It is an economical method of collecting data from small, scattered subgroups of the population since the Census makes it easier to locate respondents for collection. And since it is easier to locate small subpopulations, the postcensal approach makes it possible to generate estimates for small domains. The Census also contains a great deal of related information that is useful in preparing the sample for collection and improving estimation methods.

In Canada, the population census is conducted every five years; the most recent one was on May 15, 2001. It is a modified "de jure" census, since in most cases people are enumerated at their usual place of residence. In general, two questionnaires are used in the Census. A short questionnaire containing six questions, most of them demographic, is distributed to four dwellings out of five, and a long questionnaire with about 60 questions is distributed to one dwelling in every five. The questions included in the long form cover subjects such as demography, ethnic origin, use of languages, race, religion, education, activity, income and housing.

Since the variables of interest for identifying the target population and selecting the sample for the EDS are collected by the long questionnaire, the database for that questionnaire was used to select the EDS sample.

2.2 Sample design

A stratified two-phase sample design was used for the EDS. Phase 1 was the systematic distribution of the long census questionnaire to one in every five households in every enumeration area (EA) in Canada. In Phase 2, the Phase 1 respondents were assigned to various strata, and a systematic sample of individuals was selected from each stratum.

As noted earlier, to achieve the survey's goals, it was essential to cover both majority and minority ethnic groups in the target population, i.e., to ensure that both groups were well represented in the sample. Another prerequisite was to have proper representation of individuals declaring Canadian origin and of ethnic minorities of both European and non-European origin, since the latter groups have different characteristics. On the basis of responses to the census question on ancestors' ethnic and cultural origins, the seven strata listed below were created. (Note that multiple responses containing both European and non-European origins were assigned to the non-European stratum because of the predominance of visible minorities.)

<u>Majority ethnic groups (three strata)</u>

1. Individuals of Canadian origin only
2. Individuals of Canadian origin and British and/or French origin

3. Individuals of British and/or French origin only

Ethnic minorities (four strata)

4. Individuals of Canadian origin and European origin
5. Individuals of Canadian origin and non-European origin
6. Individuals of European origin
7. Individuals of non-European origin

Another important point to consider when studying ethnicity is the time when the respondent or his/her ancestors first arrived in Canada. As a result, the answers to the questions on the birthplace of the respondent and his/her parents were also used in stratification. Respondents were assigned to one of the following strata: first generation, second generation, or third-plus generation. Here, "first generation" refers to respondents born outside Canada, "second generation" to respondents born in Canada with at least one parent born outside Canada, and "third-plus generation" refers to respondents born in Canada to parents born in Canada.

Intersecting the ethnic-origin groups with the birthplace groups produced 21 strata. However, to ensure that there were enough individuals in each stratum, some generational strata were combined. For example, the strata consisting of first-generation and second-generation individuals reporting only Canadian origins were consolidated. Then, for coverage reasons, an unclassified stratum was added. It contained individuals who could not be assigned to any other stratum because they had not answered one of the questions required for stratification. In the end, 15 strata were used to select the Phase 2 sample. They are listed in Table 1 in section 2.3.

After the strata were determined, the sample frame was sorted by province, federal electoral district, enumeration area and household number. Selecting the sample systematically from each stratum ensured that the sample would be fairly well distributed geographically and the risk of selecting more than one person in a household would be low.

2.3 Sample allocation

The target for the survey was to have at least 40,000 respondents. Since the expected response rate was 70%, a sample of about 57,200 was needed. The coefficient of variation (CV) was used as a measure of reliability. In each stratum, the initial sample size was determined on the basis of a targeted minimum proportion of 4%, a maximum CV of 12.5%, a design effect of 1.2 and a response rate estimated from the results of the September 2001 pilot test.

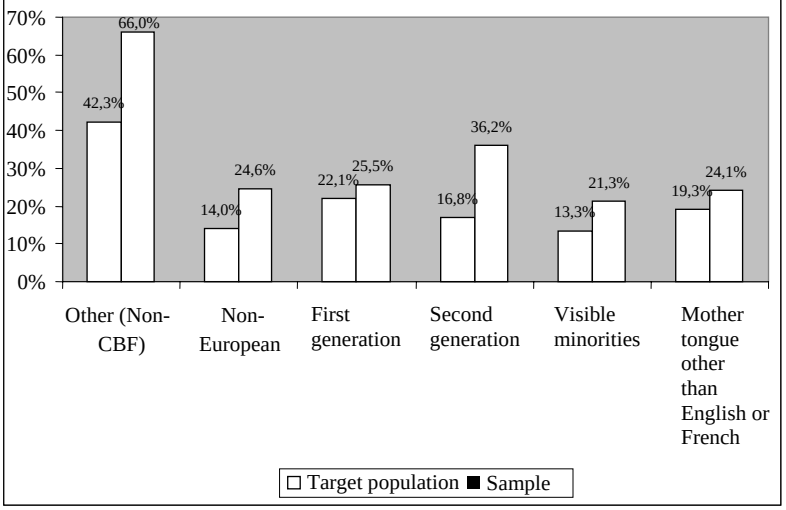
Table 1 shows the number of individuals selected from each stratum in Phase 2 and the sampling fraction in each stratum.

Table 1: Distribution of the Phase 2 sample by stratum and sampling fraction
Ethnic Diversity Survey (2002)

Stratum	Sample size (n_h)	Population size (N_h)	Sampling fraction (n_h/N_h)
Unclassified (non-response in the Census)	2,630	1,835,218	1/697
Canadian origin only – Generations 1 and 2	3,196	268,199	1/83
Canadian origin only – Generation 3+	3,218	4,406,439	1/1369
Canadian and British and/or French origin – Generations 1 and 2	2,970	311,703	1/104
Canadian and British and/or French origin – Generation 3+	2,986	1,727,043	1/578
British and/or French origin – Generations 1 and 2	2,986	1,587,629	1/531
British and/or French origin – Generation 3+	2,988	3,713,414	1/1242
Canadian and European origin – Generations 1 and 2	4,365	255,984	1/58
Canadian and European origin – Generation 3+	4,248	583,874	1/137
Canadian and non-European origin – All generations	4,314	103,539	1/24
European origin – Generation 1	5,032	1,750,284	1/347
European origin – Generation 2	4,258	1,636,840	1/384
European origin – Generation 3+	4,258	1,971,425	1/462
Non-European origin – Generation 1	5,324	2,493,333	1/468
Non-European origin – Generation 2+	4,379	447,721	1/102
Total	57,152	23,092,645	-

Given that the sampling fractions vary considerably from stratum to stratum, the EDS sample has a number of special characteristics. As an illustration of the impact that sample stratification and allocation had on the composition of the sample, the chart below (Figure 1) provides a comparison of the proportions of characteristics in the sample and in the target population.

Figure 1: Characteristics of the sample and the target population



The chart shows that the proportion of persons with at least one origin other than Canadian, British or French is higher in the EDS sample than in the target population (66.0% in the sample and 42.3% in the population).

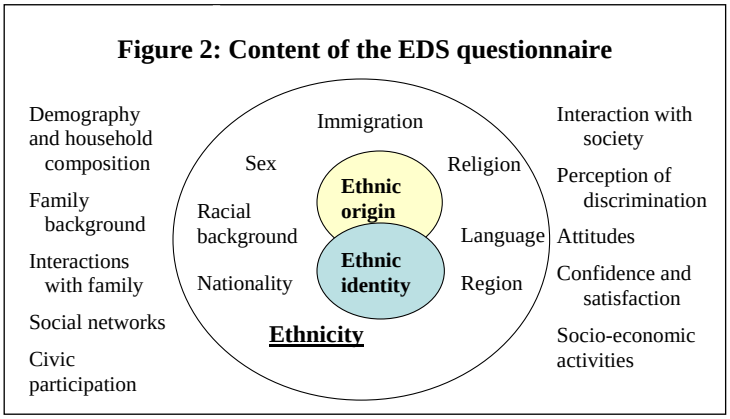
It also shows that the sample contains proportionally more first-generation individuals, second-generation individuals, persons who belong to a visible minority and persons whose mother tongue is neither English nor French.

3. QUESTIONNAIRE DESIGN

In designing a questionnaire, one has to consider a number of important factors associated with the survey. It is important to take the specific characteristics of the target population into account in order to ensure that the questions are relevant, clear and easy to understand for all individuals in the population. As noted in the previous section, the EDS's target population was very diverse. In particular, it was necessary to take account of ethnic and cultural differences of individuals, the fact that many of the respondents were immigrants, and concepts associated with ethnicity, such as ancestry and ethnic identity, which can be difficult to understand for some people. It was also important to consider the fact that some of the respondents whose families had been in Canada for several generations felt that the ethnicity questions were of little or no concern to them.

Another key issue in designing a questionnaire is the sensitivity of the information being collected. In the case of the EDS, some questions that were important for the survey, such as religion, perception of discrimination, and hate crimes, might be sensitive for some people. It was necessary to ensure that the wording of such questions did not create an excessive burden for respondents. We also had to create an atmosphere of trust with respondents if we were to have any hope at all of getting an adequate response from them.

Figure 2: Content of the EDS questionnaire



To assist in developing the questionnaire's content, expert groups were formed and consulted on a number of occasions. In particular, they helped to determine the conceptual framework of ethnicity, to identify other issues that were likely to influence or to be influenced by ethnicity, and to word the questions so that they would be clear and relevant to respondents. They also recommended that each sensitive question in the survey be preceded by an introduction to explain why the question was being asked.

Figure 2 shows the various subjects covered by the EDS questionnaire. The subjects that are part of the conceptual framework of ethnicity are inside the circle, and the subjects that are likely to influence or to be influenced by ethnicity are outside the circle. To test the questionnaire, qualitative studies involving focus groups and face-to-face interviews with feedback were conducted. The studies were carried out five times in English and French across Canada. Each series of studies helped us better understand the way people responded to the questions, identify certain problems and improve the questionnaire. In all, some 20 focus groups and about 150 personal interviews were held with people of various ethnic backgrounds.

A national pilot test was also carried out in September 2001. Nearly 1,500 interviews in English and French were conducted during the test. The test helped identify problems that had not been noticed before, refine the questionnaire and test the collection procedures. Following the test, it was decided that a letter of introduction would be sent to the selected individuals before collection so that they would have a better understanding of the survey's purpose and feel that it was of more concern to them.

As the reader will have noticed in Figure 2, some of the subjects covered in the questionnaire were very specific, and as a result, the questions had to be asked of the respondents directly. This represented an extra challenge since a number of people selected for the survey spoke neither English nor French. In addition, it was noticed during the pilot test that the amount of non-responses due to a language barrier was higher than predicted by the Census. It was therefore decided that the EDS questionnaire would be translated into seven non-official languages: Mandarin, Cantonese, Italian, Punjabi, Portuguese, Vietnamese and Spanish. The choice of these languages was based on the responses to the census questions on knowledge of official languages and languages spoken at home. The seven languages were the ones most often spoken at home by sampled individuals who reported that they knew neither English nor French.

Translating a questionnaire into non-official languages involves a number of challenges. Not all languages have words to describe the same concepts. The fact that none of the EDS team members spoke the non-official languages made testing the questionnaires difficult. Fortunately, other Statistics Canada employees who spoke the languages agreed to take part in testing the questionnaires. In addition, some of the languages are extremely difficult to read, even for people who speak them fluently. Hiring interviewers was somewhat difficult because many people who spoke one or more non-official languages very well were unable to read them. Eventually, however, we did manage to recruit the necessary staff to carry out data collection.

4. DATA COLLECTION

The data were collected by means of computer-assisted telephone interviewing (CATI) from Statistics Canada's regional offices in Halifax, Sherbrooke, Sturgeon Falls, Toronto, Winnipeg and Vancouver. Collection took place from April to August 2002; this was nearly a year after the Census, since we had to wait for the census data on ethnic origin to be coded. Some CATI components that had seen little or no use in previous Statistics Canada surveys were implemented for the EDS. First, the EDS's CATI application included a call planner to assist in managing appointments and official languages cases so that each selected person would have the best possible chance of being interviewed. Second, because of the time that had elapsed between the Census and data collection for the EDS, a tracing component was added to the application to help track down respondents who had moved since the Census.

Another feature of the EDS application was the questions that had a large number of response choices, such as the questions concerning ethnic or cultural group, language and country. Since the coding of write-in responses takes a great deal of time and effort, the EDS team wanted to minimize the number of questions that did not involve pre-set response choices. For questions with few possible choices, the choices were displayed on the same screen as the question so that the interviewer could easily locate the right response. However, for questions with more than 30 choices, this approach was impossible. For those questions, the EDS application programmers had to store the response choices separately from the screen containing the question and provide a search tool to help the interviewer find the right response from the long list of possible responses. The search tool was designed so that the user could locate the response by typing its first three letters (alphabetic mode) or a group of at least three letters within it (trigraph mode).

The EDS application was also designed so that certain responses would be inserted into subsequent questions. This made it easier to probe or follow up on responses and narrow down the number of possible response choices for some questions, such as the language question. For example, to obtain more information about how a person's ethnicity affects his or her social life, the respondent's most important ethnic origins were inserted in follow-up questions such as "As far as you know, how many of your friends have … ancestry?".

Because of the resource it would have required, however, the application was not developed in non-official languages. Interviewers assigned to those languages had to conduct their interviews with a paper copy of the translated questionnaire while at the same time entering the information in the CATI application in English or French.

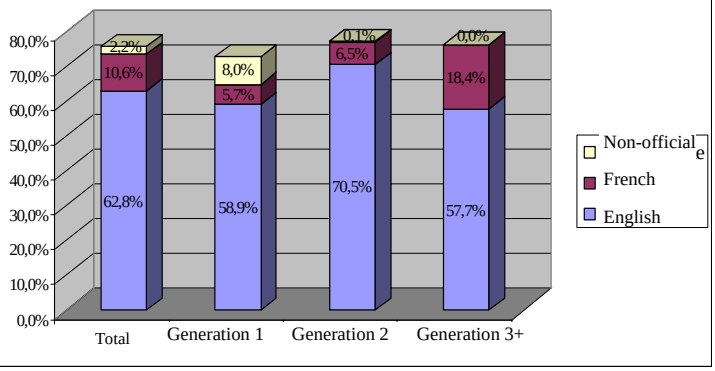
Some non-official language cases were identified prior to collection using census information. They were sent directly to the Toronto or Vancouver regional offices, which handled most of the non-official language interviews. As was noted during the pilot test, however, other non-official language cases were impossible to identify with census information. As a result, a different strategy had to be developed to handle those cases, since they were not sent to the regional offices with sufficient linguistic resources and since the Blaise system does not allow cases to be transferred from one regional office to another. For those cases, the interviews were conducted via NetMeeting. This procedure required more coordination between regional offices since it involved setting up a connection, through Statistics Canada's internal network, between the computer from which the interview was being conducted and a computer located in the regional office where the case could be accessed.

In addition, it had been impossible to develop a feature in the call planner that would automatically redirect non-official language cases to an interviewer with the required language skills. This meant that the interviewers assigned to non-official language interviews had to access their cases manually and manage their appointments themselves. In summary, while the interviews in non-official languages took more time and organization to administer (50 minutes compared with 35-45 minutes for English or French), it was worthwhile translating the questionnaire into several languages, as indicated by the response rates.

The overall response rate was 75.6%, which is excellent since the survey took place a year after the Census and since the sample was composed of rather mobile people. The high response rate was due to the tracing carried out prior to collection, the letter of introduction sent to the selected persons, the training provided to interviewers and the translation of the questionnaire into various languages.

Figure 3 provides an idea of the impact that translating the questionnaires had on response rates. It shows the proportion of the response rate that is attributable to interviews in English, French and a non-official language, overall and by generation.

Figure 3 : Proportion of response rate by interview language



According to the chart, interviews in non-official languages boosted the response rate by 2.2%, which is substantial. For first-generation respondents, the translation of the questionnaire had a much greater impact, as it raised the response rate by 8%.

It is worth noting that of the 13,000 people who did not respond to the EDS, just over 900, most of them first-generation immigrants, were unable to do so because of a language barrier (about 7% of non-response). The amount of non-response for this group could have been reduced by 40% if the collection period had been extended; at the end of collection, 365 cases were on the list for a non-official language interview.

5. WEIGHTING

The final weight assigned to each survey respondent consisted of the sampling weight multiplied by a non-response adjustment factor and a post-stratification adjustment factor. The sampling weight, also referred to as the initial weight, is the inverse of the probability of being selected for the survey. For the EDS, that probability is the product of the probability of having received a long questionnaire in the last census (about one in five) and the probability of being selected in Phase 2, which is the number of persons selected in the stratum over the total number of persons in the stratum.

Since the Census provided a great deal of information about non-respondents, the propensity-to-respond method was used to adjust the weights for non-response. The technique involves using a logistic regression model to predict each person's probability of responding to the survey. Then individuals are assigned to classes based on response probabilities, and the weights of the individuals are adjusted by the inverse of their class's response rate.

As noted above, 13,000 persons did not respond to the EDS. They had various reasons for not responding. We found that they could be divided into two groups: people who were not contacted, and other non-respondents. Since the two groups had distinct characteristics, it was decided to make two adjustments for non-response, one for each group.

In post-stratification, the survey weights were adjusted to the census totals for geography, selected age groups, sex and selection strata. The post-stratification adjustment factor is the ratio of the known total for the population to total weight after non-response adjustment in each post-stratum formed by intersecting geography, age groups, sex and selection strata.

6. VARIANCE ESTIMATION

The bootstrap method was used to estimate the variance. This re-sampling method involves:

1) selecting a sufficient number M of simple random subsamples of size $n_h - 1$ (with replacement) from the main sample, independently for each stratum h;
2) assigning a weight to the individuals selected into the subsamples based on the initial weight, the subsampling with replacement, and the various adjustments to the weights in the main survey;
3) estimating, within each subsample, the characteristic for which a variance approximation is desired;
4) calculating the empirical variance of the M estimates produced.

The bootstrap estimate of the variance for the characteristic being studied is the empirical variance of its M estimates. Hence we have:

$$\hat{V}_B(\hat{\theta}) = \frac{1}{M}\sum_{i=1}^{M}(\hat{\theta}_{Bi} - \hat{\theta})^2$$

where $\hat{\theta}$ and $\hat{\theta}_{Bi}$ are, respectively, the estimate of the characteristic obtained using the survey weights and the estimate obtained from the bootstrap weights of subsample i.

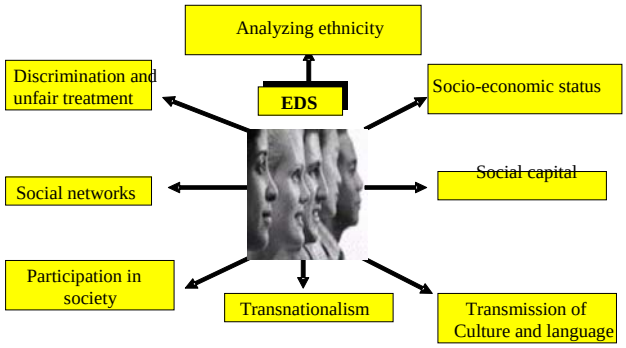
The specific choice of the subsample size, $n_h - 1$, in each stratum simplifies the formula for the initial bootstrap weights (Rao, Wu and Yue (1992) and Rao and Wu (1988)). With regard to the non-response adjustment, the results reported by Langlet, Faucher and Lesage (2003) were considered, and the only adjustment made was in the non-response classes obtained during weighting.

7. THE NEXT CHALLENGE

The next challenge for the Ethnic Diversity Survey is data analysis. An overview of some potential analysis topics is provided in Figure 4.

To this end, an analytical file has been made available to researchers in Statistics Canada's research data centres. Release of a public use microdata file is also planned for May 2005.

Figure 4: Analysis topics based on EDS data



8. CONCLUSION

Conducting a survey of a diverse population presented a number of methodological challenges. We had to choose a sampling plan that would allow us to reach certain minorities in the population, take respondents' cultural, ethnic and linguistic differences into account in developing the content, and secure the resources to collect the desired information. The following were some of the factors that made it possible to carry out the Ethnic Diversity Survey successfully: using the Census as the sample frame, translating the questionnaires into non-official languages, and using a CATI application that is adapted to improve the collection of complex data. As Canada's ethnic, cultural and linguistic diversity is certain to increase, these practices may prove helpful in the development of future surveys.

REFERENCES

Langlet, E. R., Faucher, D. and Lesage, E. (2003), "An Application of the Bootstrap Variance Estimation Method to the Canadian Participation and Activity Limitation Survey", *2003 Proceedings of the American Statistical Association*, Section on Survey Research Methods, Alexandria, VA: American Statistical Association, pp. 2299-2306.

Rao, J.N.K., Wu, C.F.J. and Yue, K. (1992). "Some Recent Work on Resampling Methods for Complex Survey", *Survey Methodology*, 18, no. 2, pp. 209-217.

Rao, J.N.K., and Wu, C.F.J. (1988), "Resampling inferences with complex survey data", *Journal of the American Statistical Association*, 83, pp. 231-241.