Catalogue no. 11-522-XIE

Statistics Canada International Symposium Series - Proceedings

Symposium 2006 : Methodological Issues in Measuring Population Health

2006

Statistics Canada Statistique Canada

Canada

Proceedings of Statistics Canada's Symposium 2006
Methodological Issues in Measuring Population Health

Sample Design of the Canadian Health Measures Survey

Suzelle Giroux[1]

Abstract

This paper describes the sample design used to satisfy the objectives and logistics of the Canadian Health Measures Survey. Among the challenges in developing the design were the need to select respondents close to clinics, the difficulty of achieving the desired sample size for young people, and subsampling for measures associated with exposure to environmental contaminants. The sample design contains solutions to those challenges: the establishment of collection sites, the use of more than one sample frame, and a respondent selection strategy.

KEYWORDS: health survey, direct measures, area frame, collection site, multi-stage sample

1. Introduction

The recently developed Canadian Health Measures Survey (CHMS) is intended to estimate the prevalence and distribution of health indicators such as excess weight, obesity, physical activity, nutrition and exposure to environmental contaminants. With interviews and direct measures, the survey will collect data to help estimate the number of Canadians who have certain chronic conditions and risk factors. The most recent survey of this type, the Canada Health Survey, was conducted in 1978 and 1979.

2. Characteristics of the Survey

2.1 Collection

Collection was scheduled to begin in March 2007 and continue over a two-year period. The data will be collected in two stages. First, interviewers will visit respondents in their homes and ask them questions about various aspects of their health and other socio-demographic variables. Second, respondents will be asked to make an appointment with the CHMS clinic, where health professionals will take physical measurements such as height, weight and blood pressure, assess their physical fitness and take blood and urine samples for laboratory analysis. After the visit to the clinic, respondents will be given a preliminary health report and asked to wear an accelerometer for one week and return it in a postage-paid envelope. Respondents will be reimbursed for the cost of travelling to the clinic.

2.2 Logistics

The CHMS's mobile clinic consists of two large trailers connected by a walkway. The clinic will be moved from site to site, remaining operational at each site for six to eight weeks. The clinic's capacity for each six- to eight-week period is expected to be about 350 respondents.

Collection sites were established to ensure that travel time and distance to the clinic would be reasonable. The method used to select the sites is described in the next section.

[1]Suzelle Giroux, Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa, Ontario, Canada, K1A 0T6 (sgiroux@statcan.ca)

3. Sampling Plan

3.1 Target population and sample size

The target population consists of persons aged 6 to 79 at the time of the survey, in private households. Excluded from the target population are persons living on Indian reserves and Crown lands, regular members of the Armed Forces, institutional residents and residents of remote areas.

The survey's analytic objectives are to obtain reliable national estimates for 10 age-sex groups, i.e., five age groups (6-11, 12-19, 20-39, 40-59 and 60-79) for each gender (male and female). The desired number of respondents visiting the clinic was set at 5,000, divided evenly among the 10 age-sex groups. With that number of respondents, it will be possible to estimate a prevalence of 10% or more for each group, with a coefficient of variation of 16.5%, based on a design effect of 1.5.

3.2 Sample frame

Collection sites were established using the Labour Force Survey's (LFS) area frame. LFS clusters, small geographic units that usually contain some 150 to 200 dwellings, were grouped together to form collection sites on the basis of the following criteria:

1. the site must have a population of at least 10,000;
2. the site's boundaries must not cross provincial or census metropolitan area (CMA) boundaries;
3. the maximum distance from the centre of the site to its boundaries must be:
    a. less than 50 kilometres if the site is within a CMA;
    b. less than 100 kilometres if the site is outside a CMA.

Using these criteria, 257 collection sites were established. About 3.7% of the Canadian population is excluded.

3.3 Sample design

A multi-stage sample design is used to reach the target population. First, a sample of sites will be selected. Second, at each site, dwellings and then persons will be selected. The method used to select the sites is described in the next section.

3.3.1 Collection site selection

With the desired number of respondents set at 5,000 and clinic capacity estimated at 350 per site, the number of sites to be selected is 5,000 / 350 = 15.

Since this is a small number of primary sampling units and we need to ensure effective representation across Canada, the 257 collection sites were stratified on the basis of the five major geographic regions in Canada. The 15 collection sites were allocated in proportion to the size of the target population within each region. The regional breakdown of the sites is provided in Table 1.

Table 1: Allocation of the sample of collection sites

Region	Population aged 6 to 79, 2001 Census	Number of sites in the target population	Number of sites in the sample
Atlantic	2,061,425	36	1
Quebec	6,560,375	50	4
Ontario	10,248,545	61	6
Prairies	4,538,970	77	2
British Columbia	3,540,000	33	2
Total	**26,949,315**	**257**	**15**

Within each region, the sites were ranked on the basis of whether they were in a census metropolitan area and on the basis of population size. The sites were selected using systematic sampling proportional to size within each region.

3.3.2 Selection of dwellings and persons

One of the survey's challenges is to obtain an equal number of respondents for each age-sex group, since household composition is not always known at the time of selection. The most difficult respondents to reach are those in the 6-11 and 12-19 age groups.

Two options are being considered for the selection of dwellings and persons for the CHMS.

3.3.2.1 Option 1

In this option, two sources are used to sample dwellings and then persons for each site.

Dwelling selection

The first source is cycle 3.1 of the Canadian Community Health Survey (CCHS 3.1). Dwellings that are located in the 15 collection sites selected and whose households include people in the 6-11 or 12-19 age group are selected.

Since this source does not provide enough respondents in those two age groups, another sample of dwellings from the LFS frame is used as well. The LFS infrastructure is used to select those dwellings: first, clusters are selected, and then dwellings are chosen systematically within the selected clusters. The number of LFS dwellings taken is based on the number of CCHS 3.1 dwellings and the number of dwellings required to have equal numbers of respondents in each age-sex group. Since the composition of the households in the LFS dwellings is not known in advance, it is necessary to sample more dwellings than necessary and reject households that do not have any members in the hard-to-reach age groups.

Person selection

In the interview, the respondent is asked to provide information about members of the household (including age and sex). Selection criteria with probabilities that vary by age group are used to choose individuals for the survey.

For dwellings selected from the CCHS 3.1, only one person, in either the 6-11 age group or the 12-19 group, will be chosen from each household. If there is no one between the ages of 6 and 19, no one will be selected and the household will be rejected.

For dwellings selected from LFS clusters, two persons will be chosen from each household if a child aged 6 to 11 is present; that is, a child between 6 and 11 and a person from one of the other age groups (12-79) will be selected. If

there are no children aged 6 to 11, only one person will be chosen from those who are between 12 and 79. In addition, a random rejection scheme is used for households with members between 20 and 59, because those age groups are much larger than the others and some households have to be rejected to ensure that the sample size is the same for all age-sex groups.

Advantages

This option takes advantage of the LFS infrastructure, such as the enumeration of dwellings in certain clusters, which reduces costs. The recent enumeration of clusters also makes it possible to include new construction (new dwellings) and thus achieve better coverage. In addition, the use of LFS infrastructure prevents survey overlap and reduces the response burden.

Disadvantages

Using two sources for the sample (CCHS 3.1 and LFS) and adding a sampling stage (clusters) prior to the selection of LFS dwellings reduces the sample's efficiency. There is an additional burden for households and persons selected from the CCHS 3.1. Moreover, the small number of CCHS 3.1 households with persons between 6 and 19 means a larger LFS sample, which results in a heavier burden and higher costs for households that are contacted and subsequently rejected.

3.3.2.2 Option 2

In this option, the Census is used as the sample frame. For each of the 15 sites, the list of addresses of dwellings identified in the 2006 Census is obtained, and the information about the age and sex of the occupants at the time of the Census can be used.

Dwelling selection

For each site, the dwellings are stratified by the age of household members. Age is determined using the birth dates of household members and the collection start date for the site in question. The stratification scheme, which takes the hardest-to-reach age groups into account, is defined as follows:

Stratum 1: at least one person aged 6 to 11 is present in the dwelling; or else
Stratum 2: at least one person aged 12 to 19 is present in the dwelling; or else
Stratum 3: at least one person aged 60 to 79 is present in the dwelling; or else
Stratum 4: at least one person aged 20 to 39 is present in the dwelling; or else
Stratum 5: at least one person aged 40 to 59 is present in the dwelling; or else
Stratum 6: dwellings that are not in the target population based on household composition, or dwellings that are vacant at the time of the Census.

In each stratum, a random sample of dwellings will be selected.

Person selection

As in option 1, the respondent is asked in the interview to provide information about the current members of the household (including age and sex). Persons who were present at the time of the Census may no longer be there at the time of collection.

Two persons will be chosen from each household if a child aged 6 to 11 is present; that is, a child between 6 and 11 and a person from one of the other age groups (12-79) will be selected. If there are no children aged 6 to 11, only one person will be chosen from those who are between 12 and 79. Selection criteria with probabilities that vary by age group are used; the criteria also differ by stratum.

Advantages

Household composition in dwellings at the time of the Census is known, which helps to stratify the frame and identify the desired age groups. It will not be necessary to reject households, at least early on, since collection is starting less than a year after the Census. The sampling plan is more efficient since only one source is used to sample individuals and one of the sampling stages used in option 1 (clusters) is eliminated.

Disadvantages

The strata's capacity to provide the desired age groups will decline as the length of time since the date of the Census increases. Another source will have to be found so that new construction and dwellings missed in the Census can be added.

Recent developments

At the time this article was written, option 2 had been chosen as the method of selecting dwellings and persons. Census data were obtained and analyzed so that the sample design could be developed and implemented in time for the start of collection.

3.4 Other considerations

3.4.1 Environmental contaminants

Since it would be too costly to analyze certain environmental contaminants for all 5,000 respondents, subsamples will be selected from the original sample. Additional tests will be performed on the blood and urine samples of the selected persons. The subsamples are independent, vary by age group and, wherever possible, include only one member of the household, when a child aged 6 to 11 is present. To subsample those persons, selection intervals were defined for each age group and contaminant. For each contaminant, a random number is generated for each household. If the random number falls within the selection interval set for the contaminant and the household respondent's age group, the respondent is subsampled, and an analysis will be performed to measure the levels of the contaminant.

In cases where two persons per household are selected for the survey, a collocation method is used in the subsampling process to ensure that the two persons are not subsampled for the same measure. Members of the 6-11 age group will be chosen from one end of the selection interval (0 to p), and members of the other age groups will be chosen from the other end (1-p to 1).

3.4.2 Morning appointments or afternoon/evening appointments

Persons selected for the survey will be randomly assigned either a morning appointment or an afternoon/evening appointment. Some tests require the respondent to fast for 12 hours. Only respondents with morning appointments will be asked to do so, since it would be difficult for those with afternoon or evening appointments (burden). The random assignment of appointments will ensure that there is no bias associated with respondents choosing appointments.

3.4.3 Repetition of physical measurements

As a check on the quality of some of the physical measurements taken at the clinic, some respondents will be selected to have one of the three groups of measurements repeated at the end of the appointment. The measurements included in this quality check are the anthropometric measurements (height, sitting height, weight, waist circumference, skin folds), hand grip strength, and the sit-and-reach test. The repeated measurements will be compared with the originals to assess measurement error.

3.4.4 Collection order

Data will be collected one site at a time, in sequence. This means that collection at the 15 sites will take place over a period of two years. The order in which collection takes place is based on both chronological distribution – i.e., the sites are divided evenly between the first and second years for each region (except the Atlantic region, which has only one site) – and seasonality wherever possible, subject to operational constraints.

4. Next Steps

In the next few weeks – i.e., in the winter of 2007 – there will be a technical test of the survey's components, followed by a dress rehearsal with volunteers. Collection is scheduled to begin at the first site in March 2007.

Acknowledgements

The author would like to thank Sarah Maude Dion, Nicolas Lavigne and Rebecca Morrison for their assistance in developing the CHMS's sampling plan.

References

Giroux S. (2004), "Study on the number of sites for the Canadian Health Measures Survey", unpublished report, Statistics Canada

Giroux S., Lavigne N. (2005), "Sampling strategy for the Canadian Health Measures Survey, Part 1: selection of collection site", unpublished report, Statistics Canada

Laflamme G., Dochitoiu (2005), "Creating sites for the Canadian Health Measures Survey", unpublished report, Statistics Canada

Lavigne N. (2006), "Stratégie de sélection des personnes", unpublished report, Statistics Canada

Statistics Canada (1998), "Methodology of the Labour Force Survey", Catalogue 71-256-XPB, Ottawa : Statistics Canada

Tambay J.L., Mohl C. (1995), "Improving sample representativity through the use of a rejective method", 1995 ASA proceedings of the Survey Methods Section