

SAMPLE DESIGN OF THE 2004 CANADIAN NUTRITION SURVEY

Yves Béland¹, Johane Dufour, Larry MacNabb and Fritz Pierre

ABSTRACT

In order to address priority health data gaps, Statistics Canada has launched the Canadian Community Health Survey (CCHS) in 2000. As part of the CCHS biennial strategy, the provincial survey component of the second cycle of the CCHS will focus on nutrition related issues. The survey will cover persons of all ages living in private dwellings in the 10 provinces. This new survey will collect information using a 24-hour dietary recall, with repeated measures, approach in order to estimate the usual dietary intake of the Canadian population. This in turn will allow for the estimation of distribution patterns in regard to the consumption of key nutrients required for optimal health. As well, data on food insecurity and some anthropometric measurements for body weight measurement will also be collected. All this will be completed with the collection of a series of health status (chronic conditions, general health, etc.), health determinants (smoking, alcohol, physical activity, etc.) and socio-demographic characteristics. To meet the requirement of producing intake distributions for 15 key age-sex domains of interest a sample of 30,000 respondents will be selected from two sample frames. Data collection will begin in January 2004 and will extend over 12 months to reduce seasonal effects. This paper will describe several key aspects of the sample design of this new survey.

KEY WORDS: Cross-sectional, Health survey, Nutrition survey, 24-hour dietary recall, Repeated measures

RÉSUMÉ

Dans le but de remédier aux principales lacunes statistiques en ce qui a trait à la santé des canadiens, Statistique Canada a mis en place l'Enquête sur la santé dans les collectivités canadiennes (ESCC) durant l'année 2000. Dans le cadre de la stratégie biennale de l'ESCC, la composante provinciale du deuxième cycle de l'enquête portera sur la nutrition. Cette enquête vise les Canadiens de tout âge résidant dans des logements privés dans les 10 provinces. Cette nouvelle enquête recueillera de l'information via un journal alimentaire de 24 heures avec seconds rappels afin d'estimer l'apport alimentaire courant de la population canadienne. Il sera possible d'estimer les distributions de consommation pour différents éléments nutritifs requis pour une bonne santé. De plus, on collectera des données sur la précarité alimentaire de même que quelques mesures anthropométriques reliées aux mesures de poids corporel. Toutes ces données seront accompagnées d'une série de données portant sur l'état de santé (conditions chroniques, santé générale, etc.), de déterminants de la santé (tabac, alcool, activité physique, etc.) et de caractéristiques sociodémographiques. Afin de produire des distributions pour la consommation alimentaire de 15 groupes d'âge-sexe d'intérêt, un échantillon de 30 000 répondants sera sélectionné à partir de deux bases de sondage. La collecte des données débutera en janvier 2004 et s'échelonnnera sur une période de 12 mois pour réduire les effets saisonniers. Cet article décrit plusieurs aspects importants du plan d'échantillonnage pour cette nouvelle enquête.

MOTS CLÉS : Enquête nutrition; enquête santé; journal alimentaire de 24 heures; mesures répétées; transversal.

1. INTRODUCTION

The CCHS was developed as part of the *Health Information Roadmap* initiative (CIHI; 1999a, 1999b), a five-year plan to modernize and standardize health information across the country. Statistics Canada, the Canadian Institute for Health Information and Health Canada support the series of projects that make up the *Roadmap* initiative. In recognition of the important contributions this survey program has and will be making to health policy research and planning, Treasury Board recently approved ongoing funding for the CCHS, making it a core survey program at Statistics Canada.

The CCHS program includes two surveys: a health region-level survey every two years (2001-Cycle 1.1, 2003-Cycle 2.1, etc.) with a total sample of over 100,000 respondents and a provincial-level survey every two years (2002-Cycle 1.2,

¹ Yves Béland (yves.beland@statcan.ca), Johane Dufour, Larry MacNabb and Fritz Pierre, 16th floor, R.H. Coats, Tunney's Pasture, Ottawa, Ontario, K1A 0T6.

2004-Cycle 2.2, etc.) with a total sample of 20 to 30,000 respondents (Béland; 2002). The provincial-level survey is designed to focus in depth on a particular topic. During consultations for the development of the CCHS, nutrition was frequently mentioned as a major data gap, a high priority issue, especially in light of the overweight and obesity concerns for Canadians. The last national nutrition survey for Canada was conducted in 1972. Therefore it was decided that year two of Cycle 2 (referred to as Cycle 2.2) would focus on nutrition.

2. SURVEY OBJECTIVES

The major objective of Cycle 2.2 is to estimate the distribution of usual dietary intake in terms of foods, food groups, dietary supplements, nutrients and eating patterns among a representative sample of Canadians at national and provincial levels for the following 15 key age/sex domains of interest:

- < 1; 1-3 and 4-8
- 9-13; 14-18; 19-30; 31-50; 51-70 and 71+ crossed by gender.

This is the first time the CCHS target population has included children.

Other main objectives include:

- gathering selected anthropometric measurements for body weight assessment;
- measuring household food insecurity among various Canadian population groups;
- collecting data on selected health conditions and socio-economic and demographic characteristics of respondents.

3. QUESTIONNAIRE CONTENT

Expert consultation has played a major role in the development of the Cycle 2.2 survey content. This includes advice from the many sources including the CCHS Nutrition Expert Group composed of experts in the field of nutrition research and policy, the Population Health Surveys Advisory Committee of Statistics Canada, the United States Department of Agriculture (USDA), the Health Canada Food and Nutrition Surveillance System Working Group as well as others.

To meet the identified objectives the survey content will revolve around a 24-hour dietary recall. Cycle 2.2 will be using a computer assisted application developed by the USDA that is currently being used in the US National Health and Nutrition Examination Survey (<http://www.cdc.gov/nchs/nhanes.htm>). Through consultation with Health Canada nutrition experts the instrument has been modified to fit the Canadian market in both official languages.

Other survey topics to be covered:

- General health
- Physical activity (aged 12 and older)
- Sedentary activity (12 to 17)
- Children's physical activity (6 to 11)
- Self-reported height and weight (10% sample aged 18 and older)
- Vitamin and mineral supplements
- Measured height and weight (2 and older)
- Women's health
- Fruit and vegetable consumption (6 months and older)
- Chronic conditions (includes osteoporosis for persons aged 50 and older)
- Smoking
- Food security
- Socio-demographics (e.g. language, ethnic origin, education, etc.)
- Labour force
- Income.

4. PROPOSED SAMPLE DESIGN FOR THE NUTRITION COMPONENT

4.1 Target Population

This CCHS component will cover persons of all ages living in private occupied dwellings in the ten provinces. The survey excludes from its target population persons living in the three territories, on Indian Reserves and Crown lands, residents of institutions, full-time members of the Canadian Forces and residents of some remote areas.

4.2 Sample Size and Sample Allocation

To meet the survey objectives of estimating usual dietary intake distributions for specific domains of interest for each province, and given the budget allocated to the survey, a sample of 30,000 responding units is desired. A two-step strategy was used to allocate the sample to the provinces. First and in order to estimate intake distributions 80 sample units were allocated to each domain of interest (14 age/sex groups) in each province (note that intake distributions for the <1 age group are required at the national level only). Thus, 1,120 units were assigned to each province in the first step for a total of 11,200. The remaining 18,800 units were allocated to the provinces using a power-allocation scheme using a power $q=0.7$ (Bankier, 1988). There is one exception to this: the province of Prince Edward Island was not assigned sample units in the second step. The total sample size of any given province is found by adding the sizes obtained in the two steps. Table 1 gives the details of the provincial allocation for the CCHS Nutrition component.

Table 1 - Provincial sample sizes for the CCHS Nutrition component

Province	1st step 80/domain	2nd step power $q=0.7$	Total Sample
Newfoundland & Labrador	1,120	630	1,750
Prince Edward Island	1,120	0	1,120
Nova Scotia	1,120	940	2,060
New Brunswick	1,120	810	1,930
Quebec	1,120	4,000	5,120
Ontario	1,120	5,620	6,740
Manitoba	1,120	1,050	2,170
Saskatchewan	1,120	960	2,080
Alberta	1,120	2,160	3,280
British Columbia	1,120	2,630	3,750
Canada	11,200	18,800	30,000

Note: Sample sizes will be inflated before going in the field to account for non-response and vacant dwellings.

4.3 Sample Frames and Sampling of Households

The sample of 30,000 individuals will be selected from two frames: an area frame and a list frame. Two frames are necessary to ensure the minimum number of 80 individuals required in each domain of interest in each province. The area frame will be used to provide almost 75% of the sample (~22,000 units) while the remaining sample will be selected from the list frame.

The area frame, as designed for the Canadian Labour Force Survey, covers almost the entire population. A sample of dwellings will be selected under a multistage stratified cluster design (Statistics Canada, 1998). For those areas selected in the first stage of the design, a list of dwellings is prepared and maintained in the field. A sample of dwellings is then selected at the second stage from each list. The households in the selected dwellings become part of the sample. To get a base sample of 22,000 responding households, approximately 31,000 dwellings will be selected from the area frame to account for vacant dwellings and non-response.

In order to ensure the minimum sample in each age/sex group and especially for the younger age groups, a secondary frame will be used. The secondary frame is necessary because it would be difficult to find a sufficient number of households with young persons from the area frame as household-level information is not available prior to data collection. This secondary frame will be created using the household information of respondents from the regional component of the CCHS (Cycle 2.1) which started its 12-month data collection in January 2003. Households in which there were at least one individual aged 18 or less at the time of CCHS 2.1 data collection will become part of a list of dwelling addresses. That list of dwelling addresses will then be used to supplement the area frame sample of households to increase the probability of finding households with individuals aged 18 or less. The list will be stratified by province and urban/rural zone, and a sample of municipalities and/or cities will be selected at first stage. A sample of dwelling addresses will then be selected at second stage. The proposed sampling scheme from the list frame should minimize the travels of the field interviewers. To get a base sample of 8,000 responding households, approximately 13,000 dwellings will be selected from the list frame to account for vacant dwellings and non-response.

4.4 Sampling of persons

Interviewing more than one person in a same household allows for economies in the cost of collection, since a large part of these costs are attributable to the process required to reach the household. However, strong similarities observed among members of the same household can lead to an undesired cluster effect for certain important survey characteristics. Moreover and probably the most important disadvantage, considering the length of the interview, is that the response burden of the household is increased. For the Nutrition component it has been decided to select only one person per sampled household.

However and as the chances of being part of a sample are inversely proportional to the number of persons in the household, ensuring the minimum number of individuals per domain of interest per province, especially in the younger age groups, is an interesting challenge. In particular, selecting only one person per household with equal probability underrepresents persons coming from large households, typically parents and children, and overrepresents persons coming from small households, often single people and the elderly.

For the Nutrition survey, it was decided to select one person per household using varying probabilities of selection that vary by age and by sampling frame. Several scenarios using various parameters were simulated with the objective of identifying an optimal approach that would guarantee the minimum number of individuals in each domain of interest in each province without generating extreme sampling weights at the end. Table 2 gives by age and by sampling frame the weights used to determine the probabilities of selection of individuals in sampled households.

Table 2 – Relative Probabilities of Selection for Person-level Sampling Strategy by Age by Frame

Frame	Weight for selection						
	< 1	1-3	4-8	9-13	14-18	19-30	31+
Area	1	3	3	3	3	3	1
List	1	2	1	1	1	0	0

The person-level sampling strategy as described above combined with the household-level sampling strategy using two frames will provide the minimum number of 80 responding units per domain of interest for all provinces but Prince Edward Island. Table 3 shows the expected distribution of a simulated CCHS Nutrition sample when selecting one person per household with varying probabilities of selection from 22,000 households selected from the area frame and 8,000 households from the list frame.

4.5 Data Collection

Data collection, which is scheduled to begin in January 2004, will span a 12-month period in order to eliminate possible seasonal effects and to spread out the interviewer workload in the field. The interview length is estimated to be 60 minutes including the 24-hour recall module (30 minutes). All interviews will be face-to-face and conducted using the computer-assisted interviewing method. The height and weight measures of selected respondents will be collected at the end of the interview. In addition to the exact measures, self-perceived height and weight will also be asked of 10% of the sample

(and controlling for key age/sex groups) in order to allow for the calculation of correlation coefficients between the two sets of measures. An overall response rate of 80% is anticipated with this Nutrition survey. In order to ensure a good representation of every day of the week in the final sample, field work will be monitored on a monthly basis.

Table 3 – Expected CCHS Nutrition Sample Distribution by Age and Sex by Province

Province	Age / Sex domain of interest								
	<1 all	1-3 all	4-8 all	9-13 m/f	14-18 m/f	19-30 m/f	31-50 m/f	51-70 m/f	71+ m/f
Newfoundland/Labrador	21	90	144	181	231	218	310	398	160
Prince Edward Island	16	79	114	128	152	137	173	209	111
Nova Scotia	29	136	206	241	289	212	317	392	234
New Brunswick	26	120	189	217	253	238	319	376	186
Québec	89	340	522	614	688	629	838	975	422
Ontario	101	451	677	750	857	834	1,133	1,226	716
Manitoba	34	154	213	235	276	275	344	370	265
Saskatchewan	28	125	192	235	288	239	320	376	276
Alberta	54	219	325	384	427	481	596	524	268
British Columbia	66	240	390	433	536	450	607	646	385
Canada	463	1,954	2,974	3,420	3,996	3,713	4,956	5,492	3,023

Interviewers will receive 3.5 days of training before going out into the field. One of the advantages of using the 24-hour recall application as developed by the USDA is that a trained nutritionist is not required to conduct the interview. It should be noted that this is one of the most complex computer assisted applications ever implemented in the field by Statistics Canada. To deal with this much of the interviewer training will focus around using the application and practising many scenarios which may arise in the field. Cycle 2.2 of the CCHS also presents a couple of other challenges to interviewers. Specifically collecting physical measurements of height and weight and food recall details from children.

In order to ensure accuracy and consistency among the measured height and weight several procedures will be put in place. First high quality scales will be used which do not require calibration, are easy to use and incredibly accurate considering their portability (50 grams). The accuracy of the scales will be assessed at the beginning and end of the survey to ensure that their functioning does not degrade. Weight estimates will also be monitored throughout collection to ensure that further training is not required.

Measuring height is a slightly more complicated procedure. Due to the constraint that interviewer staff are not trained health professionals the procedure must be non technical and non invasive. The procedure was developed with experts in the field to meet this objective. A training video will be developed to ensure consistency among interviewers across the country. Further to measure inter-interviewer variability a test requiring interviewers to measure the same test subject will be conducted before and after collection begins. Problems will also be monitored throughout collection to assess the need for further training during collection.

Regarding the procedure for interviewing children the following methods will be implemented. All respondents aged 12 and older will provide their own information. For children aged 6 to 11 the interview will be conducted with assistance from the parent. This will ensure to the extent possible that details regarding foods eaten not in the parent's presence are collected. For children under the age of 6 only the parents will be providing the information. It is felt that for these age groups parents have much more control over what their children eat. In the instances where parents cannot provide the details, such as meals eaten at a daycare, parents will be asked to contact the persons responsible to fill in the details as much as possible.

All measurements procedures worked well during the field test which was carried out on a sample of 450 households in three urban areas in the country. The procedures for interviewing children using the 24-hour dietary recall are the current practices used by the USDA and they also performed well during the field test.

4.6 Assessing the Intra-Individual Variability

Within-person variation in an individual's nutrient intake data occurs naturally due to the day-to-day variety in people's food intake in industrialised countries. Unfortunately this within-person variation increases the variance of the distribution of observed intakes (Nusser et al., 1996) which will harm the estimation of the population "at risk" (end tails of the intake distributions). This is because it includes both the within-person (day-to-day) variation and the individual-to-individual variation, thus leading to estimates of inadequacy or excess that are likely to be higher than the true prevalence. As it is important for this survey to get accurate prevalence estimates of the populations at risk in terms of several nutrients for more in-depth studies, the distributions of observed intakes must be adjusted to more closely reflect only the individual-to-individual variability in intakes.

In partnership with Health Canada nutrition experts it is proposed to collect a second 24-hour dietary recall measure on a subsample of individuals in order to allow for adjusting the intake distributions. Based on the Health Canada provincial-federal Nutrition surveys conducted in the 90's and on provincial needs in regards to reporting requirements for several age/sex groups on food consumption patterns it is recommended to collect a second measure on a subsample of 50 individuals in each of the 14 domains of interest for each province; thus a total of 7,000 individuals would be re-contacted (Junkins and Vigneau; 2003). Those individuals would be re-contacted 3 to 10 days after the first interview for a 30-minute interview conducted over the phone (only the 24-hour recall module would be administered). Interviewers will set up appointment at the end of the first visit.

CONCLUDING REMARKS

The 2004 Canadian Nutrition survey will provide reliable, timely information about dietary intake, nutritional wellbeing, physical activity and key determinants that inform programs, policies and activities of federal and provincial governments. The development and the implementation of the sample design have presented several challenges. The data collection will be monitored closely to ensure quality. The results are anticipated in the fall of 2005.

ACKNOWLEDGEMENTS

The authors would like to thank Lorna Bailie and Jean-Louis Tambay for their useful comments.

REFERENCES

- Bankier M. (1988). Power Allocations: Determining Sample Sizes for Subnational Areas. *The American Statistician*, Vol. **42**, 174-177.
- Béland, Y. (2002). Canadian Community Health Survey – Methodological overview. *Health Reports*, **13(3)**, 2002, p. 9-14 (Statistics Canada, Catalogue no. 82-003).
- Canadian Institute for Health Information (1999a), *Health Information Roadmap: Beginning the Journey*(1-895581-32-X).
- Canadian Institute for Health Information (1999b), *Health Information Roadmap: Responding to Needs*. (1-895581-30-3).
- Junkins, B. and Vigneau, M. (2003). Number of repeat recalls for CCHS Nutrition Focus Survey. Health Canada, internal document.
- Nusser S.M., Carriquiry A.L., Dodd K.W. and Fuller W.A. (1996). A Semiparametric Transformation Approach to Estimating Usual Daily Intake Distributions. *Journal of American Statistical Association*, **91**: 1440-1449.
- Statistics Canada (1998), *Methodology of the Canadian Labour Force Survey*. Catalogue no. 71-526-XPB