SSC Annual Meeting, June 2007
Proceedings of the Survey Methods Section

ONTARIO SURVEY ON THE PREVALENCE OF HYPERTENSION (OSPH): PRACTICAL ASPECTS IN SAMPLE ALLOCATION AND COLLECTION

L. Stratychuk[1] and Jean Dumais[2]

ABSTRACT

In 2005, the Heart and Stroke Foundation of Ontario (HSFO) launched a study of the prevalence of hypertension in Ontario. The HSFO desired results publishable for four ethnic groups. Additional constraints included the HSFO's budget and desired sample size, the mandatory inclusion of certain geographic areas, as well as the anticipated non-response rates. This paper concentrates on the sample allocation and data collection aspects of the survey.

KEYWORDS: hypertension, applied survey methodology, sample allocation

RESUMÉ

En 2005, la Fondation des maladies du coeur d'Ontario (FMCO) a lancé une étude sur la prévalence de l'hypertension en Ontario. La FMCO voulait des résultats publiables pour quatre groupes ethniques. Des contraintes additionnelles incluaient le budget de la FMCO et la taille d'échantillon finale désirée, l'inclusion obligatoire de certaines régions géographiques, et les taux de non-réponse anticipés. Cet article se concentre sur les aspects de la répartition de l'échantillon et de la collecte de données.

MOTS CLÉS: hypertension, méthodes d'enquêtes appliquées, répartition de l'échantillon

1. BACKGROUND

In the fall of 2004, the Ottawa Heart Institute (OHI) engaged the Statistical Consultation Group (SGC) at Statistics Canada to assist in the creation and implementation of a survey of hypertension and associated co-morbidities among Ontarians, which was funded by the Heart and Stroke Foundation of Ontario (HSFO). Previous survey data for blood pressure in Ontario was quite dated, where there was no information about the link between ethnicity and blood pressure.

The primary responsibility of the SCG entailed the creation of a sampling plan for the survey, and ultimately associated weights for the survey data. The OHI was the source of subject matter expertise for this endeavour, and responsible for the actual implementation and associated field work for this survey.

2. GOALS

The OHI and the HSFO had two key objectives in the creation of this survey. First, a final sample of at least 2500 Ontarians was desired. Secondly, it was hoped that the results of the survey could be broken down by the ethnic groups of interest, namely: "Blacks", "South Asians", "East Asians", and the residual "Rest" category of Caucasians and all other groups. For ethnic groups, the "Black" group refers to individuals of African or Caribbean descent, the "South Asian" group refers to those originally from the Southern Asian countries of India, Pakistan, Sri Lanka and Bangladesh, the "East Asian" group refers to those originally from the East Asian countries of China, Japan, Korea, Vietnam. The "Rest" category includes the Caucasian, as well as the other groups that wouldn't fit into

[1] Lori Stratychuk, Statistics Canada, Social Survey Methods Division, R.H. Coats Bldg. 15th Floor, 100 Tunney's Pasture Driveway, Ottawa, Ontario K1A 0T6, Lori.Stratychuk@statcan.ca
[2] Jean Dumais, Statistics Canada, Social Survey Methods Division, R.H. Coats Bldg. 15th Floor, 100 Tunney's Pasture Driveway, Ottawa, Ontario K1A 0T6, Jean.Dumais@statcan.ca

the other three categories. Though strictly inadequate, in what follows the terms "ethnic group" or "ethnicity" will be used to refer to those four target populations. Links between country and origin and prevalence of hypertension were reported in recent papers such as [1] Wolf-Meyer *et al* (2003) and it was hoped to investigate this claim.

3. UNUSUAL CHALLENGES

There were numerous challenges in this project that are rather foreign to Statistics Canada. First and foremost, no frame was readily available. The Statistical Consultation Group was working as if it were external to Statistics Canada, as the data were to be collected by the Ottawa Heart Institute. The SCG had access to the expertise available at Statistics Canada, but to none of the internal data resources. Therefore a method needed to be devised to create a list of the individuals to be sampled.

The lack of survey-trained staff was another challenge. Staff of varied backgrounds had to be hired and trained for very specific and different tasks (canvassing and listing, interviewing, physical and medical data collection). This necessitated having individuals qualified to collect this particular kind of information—in this case nurses. Finally, the survey was to be descriptive of all of Ontario, and therefore the issue doing a survey over the various geographies of the province was yet another hurdle to be overcome. In the case of the last challenge, it was decided to exclude Northern Ontario (roughly the areas north and west of Sudbury). While this area represents 89% of the land mass in Ontario, it only comprises 6% of the population; the cost of collecting information in such a sparely populated area was found to be prohibitive. Any future references to "Ontario" in the rest of the text refer to the southern part of Ontario which was eligible for selection.

4. SAMPLING PLAN

One of the key objectives of the survey was to have estimates publishable for each of the four ethnic groups of interest. Ideally this would involve the sample being split equally among the 4 ethnic groups; however the distribution of individuals in these groups were quite skewed. According to the 2001 Census results, Blacks represented 3% of the Ontario population, both the South Asian and East Asian group each comprised 5%, where the remaining 87% of the Ontario population was comprised of the Caucasians/Others group (with this being split into 84% Caucasians and 3% Others). The idea of two-phase sampling in order to drastically enrich the ethnic representation of the sample was quickly dismissed due to cost concerns. In the end, it was decided that the ethnic groups of interest would be targeted through the publicly information available from the 2001 Census.

Relevant Census Geography Definitions
Census Sub-Division (**CSD**) is the general term for municipalities (as determined by provincial legislation) or areas treated as municipal equivalents for statistical purposes.
Dissemination Area (**DA**) is a small, relatively stable geographic unit composed of one of more blocks. It is the smallest standard geographic area for which all census data are disseminated. DAs completely partition all of Canada. DAs range in size from a few hundred, up to over a thousand individuals.

The Census collects a wealth of information regarding individuals and households every 5 years, where portions of this information are released for varying levels of geography and levels of detail of information. Among the publicly available information, there are the population counts by ethnic background for the Dissemination Areas (DAs). It was decided to take advantage of this information in order to enrich the ethnic representation of the sample.

Based on the requirements of the survey, as well as the available resources, it was decided that a multi-stage sampling plan would match the requirements best. Specifically, first the broad geographic areas of Census Subdivisions (CSDs) were selected, which were then stratified by ethnicity where possible. Next dissemination Areas (DAs) were selected, the selected DAs were listed, and a systematic sample of households was then chosen. From these households, one individual was randomly chosen from the household.

Although the sample selection procedure at each of the stages appears very linear, the process of establishing the sample selection procedure and the final sample sizes was somewhat of an iterative process. As is often the case, the choices made at one stage were altered by the requirements of sampling in the subsequent stages.

4.1 Selection of Census Subdivisions

Since the blood pressure and physical measurements were to be taken at a clinic and not the respondent's home (in exceptional cases, blood pressure and physical measurements were taken at the respondent's home), it was advantageous to first select CSDs which would serve as an anchor point for where to locate the clinics. Because of its size, the CSD of Toronto was selected with certainty. Additionally, the CSD of Ottawa was selected with certainty with the original intent of being a pilot site to work out any kinks in the process. Another ten CSDs were selected from Ontario with Probability Proportional to Size (PPS) with total population as the size measure; using PPS selection had the benefit of making population dense areas more likely to be selected, which helped to keep collection costs lower.

At the outset, it was hoped that the survey would yield an absolute minimum sample of 2500 respondents, with a more hopeful objective of 3000. Based on very pessimist response rate projections—much lower than is normally projected for survey with face-to-face interviewing, an initial sample size of 7000 was chosen. In order to yield a starting point, the sample sizes for each CSD were calculated assuming a simple proportional allocation between CSDs (Table 1, column (A)). The desired sample sizes for each CSD were then rounded to an even hundred, based on the anticipated capacity of the clinics (column (B)). Subsequent examination of the CSDs showed that certain ones were "rich" in ethnic diversity (where this will described in greater detail in the subsequent section), and the decision was made to steal sample from the CSDs "poor" in ethnic diversity and give it to the CSDs "rich" in ethnic diversity (column (C)). Later, but prior to any field work had began, the addition of the CSD of Sudbury was requested in order to give some northern representation to the sample. This necessitated another round of "stealing" sample, this time to give to Sudbury (column (D)).

Table 1: Sample Sizes for each selected Census Subdivision

Census Subdivision (CSD)	Sample Size…			
	(A) With a Simple Proportional Allocation	(B) Revised subject to Operational Constraints	(C) Revised to augment sample in areas rich in desired ethnicities	(D) Revised to include Sudbury
Barrie	314	300	300	300
Brampton	1003	1000	800	800
Guelph	323	300	300	300
Lakeshore	88	100	100	100
Markham	644	600	600	600
Mississauga	1887	1900	**1200**	**1100**
Niagara on the Lake	34	100	100	100
Oshawa	427	400	**400**	**300**
Ottawa	510	500	**500**	**400**
South Stormont	37	100	100	100
Stratford	90	100	100	100
Toronto	1641	1600	**2500**	**2300**
Sudbury	*0*	*0*	*0*	*500*

4.2 Selection of Dissemination Areas

In order to create a balance between the high monetary cost of a sampling fraction that is too low, and the negative impact of intra-class correlation on the variance of estimates resulting from a sampling fraction that is too high, the number of Dissemination Areas (DAs) selected from each CSD was based on having an average sampling fraction of roughly 5%. For example, according to the 2001 Census, in the CSD of Markham there are 280 DAs with an average population of 743. With 37.2 (=5% of 743) individuals to be selected for each sampled DA, and a total sample size for Markham of 600 already established, it follows that 16 DAs had to be selected for the CSD of Markham in order to follow the rule of an average sampling fraction of 5%. In practice, the size of DAs varies from a few hundred to over a thousand and therefore the sampling fraction within the DA varies from the desired 5% target.

Next each DA within each of the selected CSDs was designated as Black, South Asian, East Asian or Rest (Caucasians/Others) based on which ethnic group was which had the largest percentage share of the population for that DA during the 2001 Census. Sample selection took place in 2005; the 2006 Census had not yet taken place and therefore the results from the 2001 Census were the most recent available. Although there may be some movement in households over time, the evolution of the ethnic composition of a neighbourhood is a slow process, where it is very unlikely that the ethnic representation found in the 2001 differs drastically from what was found in 2005. It was determined that the CSDs of Toronto, Markham, Mississauga and Brampton had enough DAs of ethnic interest (e.g. Black, South Asian or East Asian) in order to warrant stratification of the DAs by ethnicity in order to target those DAs known to contain more individuals of the desired ethnic background.

For the CSDs where stratification by ethnicity was not possible (Barrie, Guelph, Lakeshore, Niagara on the Lake, Oshawa, Ottawa, South Stormont, Stratford and Sudbury), the selection of DAs was simply done using PPS with total population as the size measure. For the CSDs which were stratified by ethnicity (Toronto, Markham, Mississauga and Brampton), the additional step of allocating the number of DAs to the 4 ethnic strata was necessary. The division of the DAs between the strata was a compromise between what a proportional allocation would dictate, and what an equal allocation would dictate. Having decided the number of DAs to be selected in each ethnic stratum, DAs were then selected using PPS with total population as the size measure.

4.3 Selection of Households and Individuals

Once the CSDs and their DAs have been selected, field staff went out and enumerated the dwellings in the selected DAs. The listing of the dwellings occurred just prior to the scheduled collection of data so as to have the most current and therefore best list of dwellings possible. Once the listing of the households was completed, the forms were reviewed and submitted to quality control. From the list of dwellings, a systematic sample of dwellings was then selected in order to achieve the desired sample size for that DA.

Once a dwelling had been selected in the sample, an interviewer went to this dwelling to make face-to-face contact. One member of the dwelling was selected to represent the household eligible adults using the random method of choosing which person most recently had a birthday.

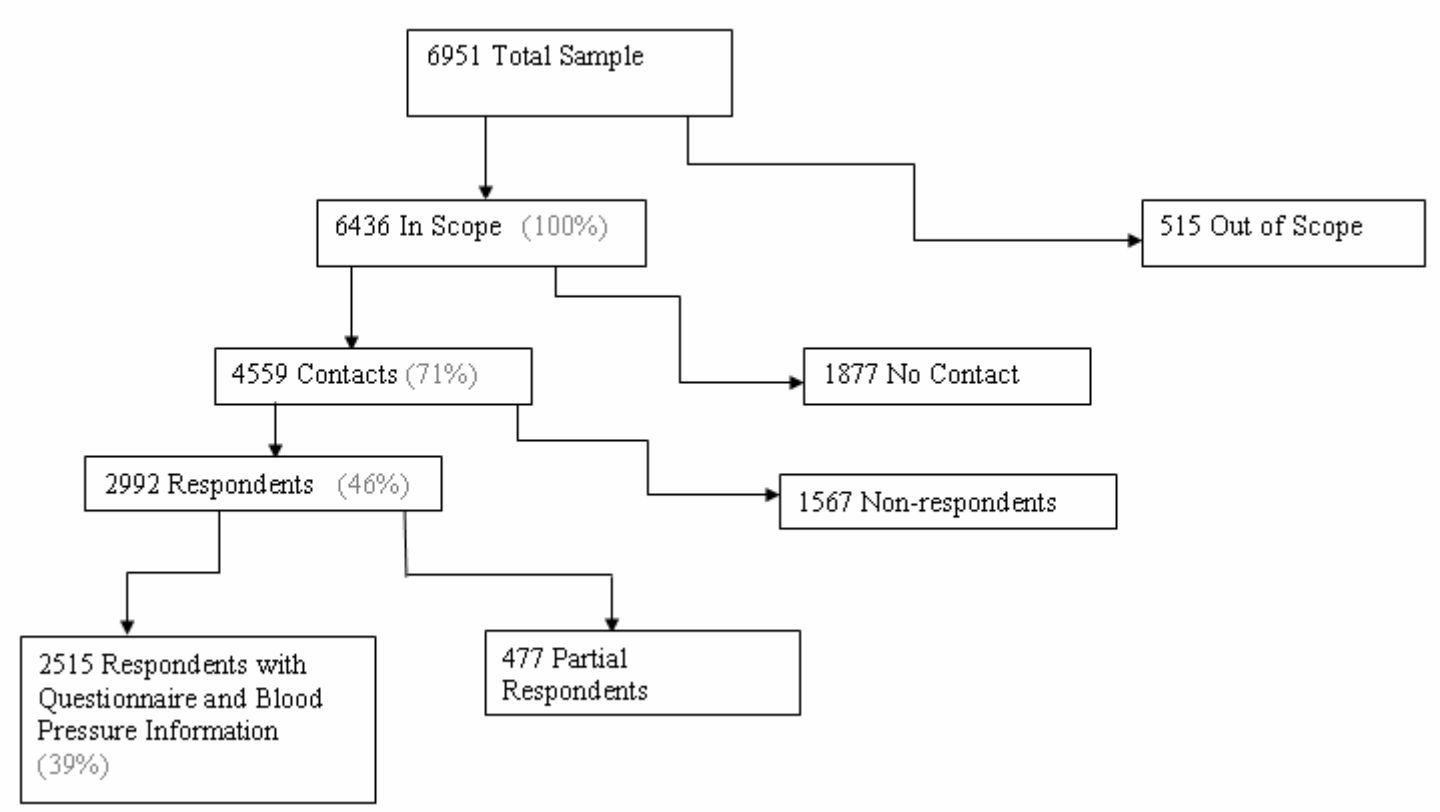
5. COLLECTION OF DATA AND RESPONSE RATES

Initial contact was made by interviewers visiting the selected dwellings. During a face-to-face interview at the respondents' home, the interviewers collected demographic information and some elements of the respondents' medical history. The interviewers then made appointments for the individuals to visit the survey clinics in order to get blood pressure and other physical measurements, such as height and weight. Roughly, 85% of the people who made an appointment followed through and attended their clinic appointment. The propensity of individuals to attend their scheduled visit at the clinic was possibly aided by individuals being reimbursed for their travel expenses to the clinic.

In the end, the response rates for the survey were lower than what had been anticipated (Figure 1). Out of the 6951 dwellings that were initially sampled (where 49 from the original figure of 7000 "disappeared" due to technical issues such as rounding during sampling size calculations), 515 of the contacted individuals were identified as out of scope (e.g.,

outside of age range, pregnant and over 20 weeks gestation, etc.). However contact was made with only 73% of all selected dwellings. From the 4559 individuals who were contacted, 2992 responded. Among those, 2515 answered to background questionnaire and visited the clinic for the BP and physical measurements, thereby giving a response rate of "complete" information of 39%, which is fairly low for a face-to-face interview.

Figure 1: Flow chart of response rates



There are numerous reasons to explain the difficulties in collection, many of these are related to the lack of survey experience by the field staff. As evidenced by the contact rates, the initial contact with the households was quite difficult. Getting the initial participation proved to be difficult, however once converted to a respondent, continued response was assisted with the addition of monetary incentives.

More importantly, significant gains were made in the ethnic representation of the sample. If no ethnic stratification had been performed then the ethnic distribution of the unweighted survey data would have the same results as from the 2001 Census (Table 2). However the extent to which the "minority" ethnic groups are over-represented in the unweighted file is a measure of how successful the survey was in increasing the ethnic representation of the sample. For example, Blacks represent 3% of the Ontario population (according to the 2001 Census) but account for 9% of the survey respondents. Note that although Black accounts for 9% of the unweighted survey respondents, in the final weighted file, they still account for 3% of the population since results were post-stratified to agree with Census totals. This result holds for all other ethnic groups with their respective population shares.

Table 2: Percentage Distribution of Ontarians by Ethnicity

Ethnic Group	Census 2001 (%)	Unweighted OSPH (%)
Rest (Caucasian/Others)	87.4	66.4
Black	3.2	8.7
Chinese	4.6	9.9
South Asian	4.9	15.0

6. CONCLUSIONS

The OSPH reached the two major objectives named at the outset. First, the desired sample size of at least 2500 Ontarians was reached. Response rates for household surveys conducted in person have always been the highest when compared to other modes of data collection. Response rates reported by Primatesta *et al.*, Razak *et al.* or Wolf-Maier *et al.* (see references below) varied from 60% to 80%. NHANES enjoyed, during the 1980-1999 period, low- to mid- ninety percent response rates. Survey takers – public and private- agree that such high response rates are a thing of the past as all have experienced declining household and individual response over the recent years. Further analysis of the sampling and collection processes should shed light on the most important hindrances to participation.

Second, the goal of enriching the ethnic representation of the sample was also accomplished. Results from the OSPH presented at the annual Canadian Cardiovascular Congress in October 2007 were very well received and confirmed progress made in identification, treatment and awareness of hypertension. The question of whether the sample was sufficiently enriched by ethnicity remains. Initial analysis of the data indicates that there are some significant differences in blood pressure based on ethnicity. If even greater ethnic representation had been achieved with the sample, perhaps even more detailed analysis would have been possible. Further analysis of the data will reveal if enough was done to enrich the ethnic representation of the sample, not enough, or perhaps even too much.

REFERENCES

Agyemang, C., I. van Valkengoed, et al. (2006) "Factors associated with hypertension awareness, treatment and control among ethnic groups in Amsterdam, The Netherlands: The SUNSET study", Journal of Human Hypertension, 20:874-81.

Cochran, W.G. (1977). *Sampling Techniques* (3rd. ed.). New York: John Wiley.

Kish, L. (1965). *Survey Sampling*, New York: John Wiley

Lohr, S. L. (1999). *Sampling: Design and Analysis*, Thomson Publishing.

Ong, K.L et. Al, *Prevalence, Awareness, Treatment, and Control of Hypertension Among United States Adults 1999-2004, American Heart Association*, DOI: 10.1161/01.HYP.0000252676.46043.18

Primatesta, P., L. Bost, N.R. Poulter (2000) "Blood pressure levels and hypertension status among ethnic groups in England" Journal of Human Hypertension, 14:143-8.

Razak, F., S.S. Anand, et al. (2007) "Defining obesity cut points in a multiethnic population" Circulation, 115:2111-8.

Statistics Canada (2003). *Survey Methods and Practices*, Catalogue No. 12-587-XPE.

Statistics Canada (2002), *2001 Census Dictionary*, Catalogue No. 92-378-XPE

Wolf-Maier, K., R.S. Cooper, et al. (2003) , "Hypertension prevalence and blood pressure levels in 6 European countries, Canada, and the United States", JAMA. 289(18):2420-2.