Survey Methodology, June 1987
Vol. 13, No. 1, pp. 93-108
Statistics Canada

Sample Design for the Health and Activity Limitation Survey

D. DOLSON, K. McCLEAN, J.-P. MORIN, and A. THÉBERGE[1]

ABSTRACT

The Health and Activity Limitation Survey is part of the program to establish a data base on the disabled population in Canada. The sample design used for the part of the survey covering the population not living in institutions is described. In addition, the methods used to determine the sizes of the samples and to select the samples are presented.

KEY WORDS: Disability; Stratified sampling; Two-stage sampling; Optimum allocation; Sampling without replacement.

1. INTRODUCTION

As part of the program to obtain more information about Canada's disabled population, the Health and Activity Limitation Survey (HALS) was conducted in the fall of 1986. It is designed to obtain information concerning the nature of the problems experienced by that population and, in general, their daily activities (at home, at work, at school, during travel, and so on). The survey is divided into two parts: one covers the population living in institutions and the other, which is the subject of this article, covers the non-institutional population.

Canada has been divided into 238 subprovincial areas (SPAs). All Quebec and Ontario municipalities with more than 125,000 residents and all municipalities in the other provinces with more than 75,000 residents are included as SPA's. The other areas are made up of groups of census subdivisions respecting geographical contiguity and the provincial boundaries. The number of these areas in each province is proportional to the square root of the population, minus the previously defined municipalities. One of the main objectives of the survey is to generate statistics on the disabled population at the SPA level so that the population's various needs can be analysed in detail. In addition, estimates will be produced for three age groups - namely, children (under 15 years of age), adults (15 to 64 years of age) and seniors (65 years of age and older).

The data was collected in two stages. The first stage involved a multipart question (question 20) included on form 2B of the 1986 Canadian Census of Population. This question asked about the respondents' limitations in various types of activities and their own assessments of their conditions. A copy of question 20 is given in the Appendix. The second stage was implemented some time after the census. It involves a screening questionnaire and follow-up to collect information on the problems and activities of disabled respondents.

The main purpose of the first stage is to separate respondents into two groups: those who answered "yes" to at least one part of question 20 and those who answered "no" to all parts. The aim is to identify beforehand a large part of the potential disabled population, in order to focus survey resources on the target group. However, previous surveys have shown that this question will not identify the entire target population. (See Dolson *et al.* 1984 and Dolson *et al.* 1986.)

---

[1] D. Dolson, K. McClean, J.-P. Morin, and A. Théberge, Social Survey Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6.

The second stage is HALS. Personal interviews are conducted for the "yes" stratum and telephone interviews are conducted for the "no" stratum. From an operational point of view, the interviews are in two parts - the screening questionnaire and the follow-up.

The screening questionnaire is designed to identify respondents for whom the follow-up questionnaire is relevant. The questionnaire for adults covers the seventeen activities of daily living (ADLs) used in the Canadian Health and Disability Survey in 1983 and 1984, repeats Part (a) of question 20 from the Census, and includes a few questions on mental illness and handicaps (see the Appendix). If an affirmative answer is given to at least one of these questions, the interviewer proceeds with the follow-up; if not, the interview is terminated. Part (a) of the Census question is asked again because there may have been a change in status, either because the response in the Census was given by a proxy, or because the respondent has reassessed his or her own condition.

The screening section in the questionnaire for children includes questions on special aids, activity limitations, attendance at a special school and health conditions or problems. A "yes" answer to at least one of these questions prompts a follow-up interview. The Census question is not repeated because all interviews regarding children require a proxy and the question on activity limitations is equivalent to Part (a) of Census question 20.

The second section of this article describes how the population of Canada has been divided into various subpopulations for estimation purposes. The third section covers the HALS sample design. The fourth section deals with the file of geographic information and projected demographic data for 1986 that was used to create the survey frame. The fifth section explains how the sampling was done.

2. POPULATIONS COVERED

Permanent residents of general and psychiatric hospitals, special care centres or institutions for the elderly or chronically ill, institutions for the physically handicapped and orphanages or children's homes are the subject of a distinct part of the survey - namely, HALS (Institutions). This article will look at the part of the survey covering that portion of the Canadian population not covered by HALS (Institutions) and not residing in jails, military camps, young offender facilities, naval vessels, penal or correctional institutions and collective dwellings in the "others" category (for example, circuses and non-religious communes).

Each enumeration area (EA) whose population is not totally excluded from the survey is classified in one of the following five survey frames:

1. Indian reserves where the 1981 Census was conducted using canvassers;
2. Other Indian reserves;
3. Canvasser EAs;
4. EAs in the Whitehorse, Yellowknife, Pine Point, Hay River and Fort Smith SPAs;
5. All other EAs.

The order of priority for belonging to a frame is 1-2-4-3-5. This means that an EA that is an Indian reserve and situated in the Whitehorse SPA is classified as an Indian reserve.

Each EA is divided in two, with the "yes" EA made up of those persons who would answer "yes" to the Census question, and the "no" EA made up of those who would answer "no" to it. A different sample design is used for each of the five survey frames: all of the "yes" EAs and none of the "no" EAs are selected in the first frame; all of the "yes" EAs and a sample of the "no" EAs are selected in the second frame; none of the "no" EAs and a sample of the "yes" EAs are selected in the third frame; all of the EAs are selected in the fourth frame; and a sample of the "yes" EAs and a sample of the "no" EAs are selected in the fifth frame.

3. SURVEY DESIGN

The sampling method presented in this section was used for survey frames three and five. Because our space is limited, the sample design used for the second survey frame will not be described in this article. (For more information on the HALS methodology, see Dolson *et al.* 1986.)

3.1 Sample Design

Each province is divided into subprovincial areas (SPAs), which are themselves divided into enumeration areas (EAs).

Each EA is divided into a "yes" EA and a "no" EA, the first containing those persons who would answer "yes" to Census question 20, the second containing those persons who would answer "no" to that question. In each SPA, the "yes" EAs are stratified into large and small EAs on the basis of the criterion explained in the fourth section of this paper. Persons belonging to a "yes" EA are associated with a stratum and an SPA in addition to their EA, while persons belonging to a "no" EA are associated only with their EA. In each province, the population is subdivided into three age groups: children (under 15 years of age), adults (15 to 64 years of age) and seniors (65 years of age and older).

The sampling method involves using a two-stage stratified sample design for the "yes" EAs in each SPA and a two-stage sample design for the "no" EAs in the province. The primary units are the EAs and the secondary units are the respondents.

All persons who completed Census form 2B in a "yes" EA selected for the sample are interviewed, along with a third of those in the "no" EAs selected.

3.2 Sample Allocation

This sample design must allow us to minimize sampling costs for a given maximum coefficient of variation of the estimates and a given variance for the estimator $\hat{B}$ of the relative bias B. We define B as the ratio of the number of "no" persons with a characteristic of interest in the province, T_0, to the number of "yes" persons with a characteristic of interest in the province, T_1. By "no" person, we mean an individual who would answer "no" to all parts of Census question 20, and by "yes" person, an individual who would answer "yes" to at least one part of the question.

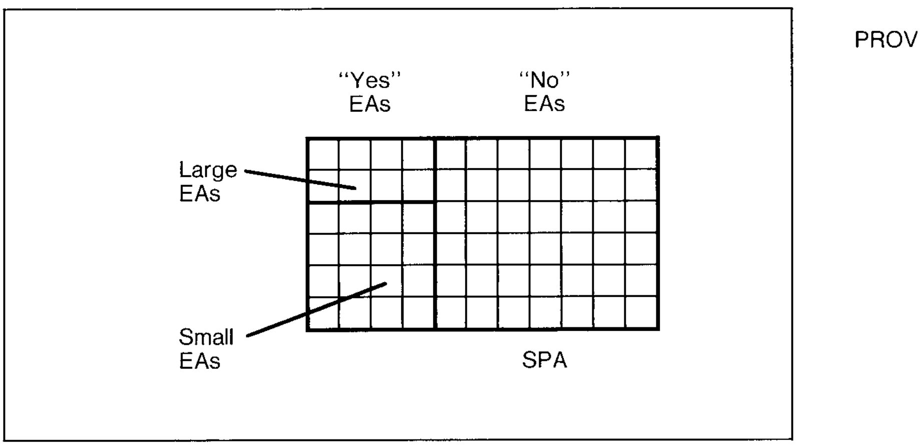

Figure 1. Illustration of Sample Design.

Let N_0 be the number of "no" EAs in the province; N_{jk}, the number of "yes" EAs in stratum j and SPA k in the province; n_0 and n_{jk}, the corresponding sample sizes; and c_0 and c_{jk}, the corresponding unit sampling costs. If we have an N_p SPAs in the province, we therefore want to minimize

$$\sum_{k=1}^{N_p} (c_{1k} n_{1k} + c_{2k} n_{2k}) + c_0 n_0$$

given

$$CV^2(y_k) \leq CV_*^2; \ \mathrm{Var}(\hat{B}) = \mathrm{Var}_*(\hat{B});$$

$$n_{jk} \leq N_{jk}; \ n_{2k} = \lambda_k n_{1k}; \ n_0 \leq N_0$$

$$(j = 1, 2; \ k = 1, \ldots, N_p)$$

where λ_k is the ratio of the expected number of disabled persons in the small EAs to the expected number of disabled persons in the large EAs of SPA k, y_k is the estimated number of "yes" persons who have a characteristic of interest in SPA k, and values marked with an asterisk are constants.

If the sampling fraction in the "yes" EAs is f_1, M_{ijk} is the number of "yes" persons in EA i of stratum j of SPA k in the province and p_{ijk} is the probability of a characteristic of interest for a "yes" person in EA i of stratum j in SPA k, then

$$E(y_k) = Y_k = \sum_{j=1}^{2} \sum_{i=1}^{N_{jk}} M_{ijk} p_{ijk},$$

$$\mathrm{Var}\ (y_k) = \sum_{j=1}^{2} \left\{ \frac{N_{jk}^2}{n_{jk}} \left(1 - \frac{n_{jk}}{N_{jk}}\right) S_{jk}^2 + \frac{N_{jk}}{n_{jk}} \left(\frac{1 - f_1}{f_1}\right) \sum_{i=1}^{N_{jk}} M_{ijk} S_{ijk}^2 \right\},$$

where

$$y_k = \sum_{j=1}^{2} \sum_{i=1}^{N_{jk}} \frac{N_{jk}}{n_{jk}} M_{ijk} p_{ijk},$$

$$S_{jk}^2 = \frac{1}{N_{jk} - 1} \sum_{i=1}^{N_{jk}} \left(M_{ijk} p_{ijk} - \left(\sum_{i=1}^{N_{jk}} \frac{M_{ijk} p_{ijk}}{N_{jk}} \right) \right)^2,$$

$$S_{ijk}^2 = \frac{M_{ijk}}{M_{ijk} - 1} p_{ijk} \left(1 - p_{ijk}\right).$$

After a few algebraic manipulations, we obtain

$$\mathrm{Var}\ (y_k) = \frac{1}{n_{1k}} \left\{ N_{1k}^2 S_{1k}^2 + \left(\frac{1 - f_1}{f_1}\right) N_{1k} \sum_{i=1}^{N_{1k}} M_{i1k} S_{i1k}^2 + \frac{N_{2k}^2 S_{2k}^2}{\lambda_k} \right.$$

$$\left. + \left(\frac{1 - f_1}{f_1}\right) \frac{N_{2k}}{\lambda_k} \sum_{i=1}^{N_{2k}} M_{i2k} S_{i2k}^2 \right\} - \sum_{j=1}^{2} N_{jk} S_{jk}^2 .$$

We can therefore write $CV^2(y_k)$ as

$$CV^2(y_k) = \frac{\text{Var}(y_k)}{Y_k^2} = \frac{A_k}{n_{1k}} - B_k. \tag{3.1}$$

Furthermore, B (the relative bias) and $\hat{B}$ (its estimator) are given by

$$B = \frac{T_0}{T_1} = \frac{\sum_{i=1}^{N_0} M_{i0} p_{i0}}{\sum_{k=1}^{N_p} Y_k},$$

$$\hat{B} = \frac{t_0}{t_1} = \frac{\sum_{i=1}^{n_0} \frac{N_0}{n_0} M_{i0} p_{i0}}{\sum_{k=1}^{N_p} y_k},$$

where M_{i0} is the number of "no" persons in EA i in the province and p_{i0} is the probability of a characteristic of interest for a "no" person in EA i.

Assuming that t_0 and t_1 are independent, then

$$\text{Var}(\hat{B}) = B^2 \left(\frac{\text{Var}(t_0)}{T_0^2} + \frac{\text{Var}(t_1)}{T_1^2} \right). \tag{3.2}$$

After a few algebraic manipulations, if f_0 is the sampling fraction in the "no" EAs, we obtain

$$\text{Var}(t_0) = \frac{1}{n_0} \left\{ N_0^2 S^2 + N_0 \left(\frac{1 - f_0}{f_0} \right) \sum_{i=1}^{N_0} M_{i0} S_i^2 \right\} - N_0 S^2$$

where

$$S^2 = \frac{1}{N_0 - 1} \sum_{i=1}^{N_0} \left(M_{i0} p_{i0} - \left(\sum_{i=1}^{N_0} \frac{M_{i0} p_{i0}}{N_0} \right) \right)^2, \quad S_i^2 = \frac{M_{i0}}{M_{i0} - 1} p_{i0} (1 - p_{i0})$$

which can be written in the form

$$\text{Var}(t_0) = \frac{A_0}{n_0} - B_0. \tag{3.3}$$

Furthermore, assuming that the y_k's are independent, we have

$$\text{Var}(t_1) = \sum_{k=1}^{N_p} \text{Var}(y_k).$$

Using equation (3.1), this expression can be written as

$$\text{Var}(t_1) = \sum_{k=1}^{N_p} \left(\frac{A_k Y_k^2}{n_{1k}} - B_k Y_k^2 \right). \tag{3.4}$$

From (3.2), (3.3) and (3.4), we obtain

$$\text{Var}(\hat{B}) = B^2 \left\{ \left(\frac{A_0}{n_0 T_0^2} - \frac{B_0}{T_0^2} \right) + \sum_{k=1}^{N_p} \left(\frac{A_k Y_k^2}{n_{1k} T_1^2} - \frac{B_k Y_k^2}{T_1^2} \right) \right\}$$

$$= \frac{1}{n_0} \left(\frac{B^2 A_0}{T_0^2} \right) + \sum_{k=1}^{N_p} \frac{1}{n_{1k}} \left(\frac{B^2 A_k Y_k^2}{T_1^2} \right) - B^2 \left(\frac{B_0}{T_0^2} + \sum_{k=1}^{N_p} \frac{B_k Y_k^2}{T_1^2} \right)$$

$$= \frac{X}{n_0} + \sum_{k=1}^{N_p} \frac{W_k}{n_{1k}} - Z.$$

The optimization problem can be re-expressed as the problem of minimizing

$$\sum_{k=0}^{N_p} c_k n_k$$

subject to

$$0 < a_k \leq n_k \leq b_k \quad (k = 0, 1, 2, \ldots, N_p) \tag{3.5}$$

and

$$\sum_{k=0}^{N_p} d_k / n_k = e \tag{3.6}$$

where, for $k = 1, 2, \ldots, N_p$,

$$n_k = n_{1k}, \; c_k = c_{1k} + c_{2k}\lambda_k, \; a_k = \frac{A_k}{CV_*^2 + B_k}, \; b_k = \min(N_{1k}, N_{2k}/\lambda_k).$$

In practice, rather than using $b_k = \min(N_{1k}, N_{2k}/\lambda_k)$ we define

$$b_k = \frac{N_{1k} N_{2k} (1 + \lambda_k)}{\lambda_k^2 N_{1k} + N_{2k}},$$

then, if $n_k > N_{1k}$, sample sizes are given by $n_{1k} = N_{1k}$ and

$$n_{2k} = \lambda_k n_k + \frac{(n_k - N_{1k}) N_{2k}}{N_{1k} \lambda_k},$$

while, if $\lambda_k n_k > N_{2k}$ sample sizes are given by $n_{2k} = N_{2k}$ and

$$n_{1k} = n_k + \frac{(\lambda_k n_k - N_{2k}) N_{1k} \lambda_k}{N_{2k}}.$$

Thus, we consider $N_{2k} / (N_{1k} \lambda_k)$ small EAs to be equivalent to one large EA. On average, there are as many disabled persons in one large EA as in $N_{2k} / (N_{1k} \lambda_k)$ small EAs.

Proceeding in this way, it is not always true that $n_{2k} = \lambda_k n_{1k}$. However, we avoid CVs higher than target values, when, for example, small EAs remain to be observed (even if all the large EAs have been selected).

For some values of k, it is possible that $a_k \geq b_k$. If this is the case, we set $n_k = b_k$. Let

$$E_1 = \{k = 0, 1, 2, \ldots, N_p \mid n_k = a_k\},$$

$$E_2 = \{k = 0, 1, 2, \ldots, N_p \mid n_k = b_k > a_k\},$$

$$E_3 = \{k = 0, 1, 2, \ldots, N_p \mid a_k < n_k < b_k\},$$

$$E_4 = \{k = 0, 1, 2, \ldots, N_p \mid n_k = b_k \leq a_k\}.$$

The solution exists if

$$\sum_{k=0}^{N_p} d_k / b_k \leq e,$$

and it takes the form

$$n_k = \begin{cases} a_k & (k \in E_1) \\ b_k & (k \in E_2 \cup E_4) \\ K (d_k / c_k)^{1/2} & (k \in E_3) \end{cases} \tag{3.7}$$

where

$$K = \frac{\sum_{k \in E_3} (d_k / c_k)^{1/2}}{e - \sum_{k \in E_1} d_k / a_k - \sum_{k \in E_2 \cup E_4} d_k / b_k}, \tag{3.8}$$

since the n_k ($k \in E_3$) minimize $\sum_{k \in E_3} c_k n_k$, subject to the constraint

$$\sum_{k \in E_3} d_k / n_k = e - \sum_{k \in E_1} d_k / a_k - \sum_{k \in E_2 \cup E_4} d_k / b_k.$$

What are the sets E_1, E_2, E_3 and E_4 corresponding to the solution? Set E_4 is easy to determine. We must have

$$a_k < (d_k / c_k)^{1/2} K < b_k \quad (k \epsilon E_3), \qquad (d_k / c_k)^{1/2} K \geq b_k \quad (k \epsilon E_2),$$

$$(d_k / c_k)^{1/2} K \leq a_k \quad (k \epsilon E_1). \tag{3.9}$$

Determining the sets involves trying each of the possibilities for E_1, E_2 and E_3 until a value for k which satisfies (3.9) is obtained. To reduce the number of possibilities to be examined, note that, if for $k' \geq k$,

$$b_k' \, (c_k' / d_k')^{1/2} \geq b_k \, (c_k / d_k)^{1/2} \quad (k, k' \epsilon \{0, 1, \ldots, N_p\}), \tag{3.10}$$

then there is a k^* such that $E_2 = \{0, 1, 2, \ldots, k^*\}$, or $E_2 = \{\ \}$, while, if for $k' \geq k$,

$$a_k' \, (c_k' / d_k')^{1/2} \geq a_k \, (c_k / d_k)^{1/2} \quad (k, k' \epsilon \{0, 1, \ldots, N_p\}), \tag{3.11}$$

then there is a k^{**} such that $E_1 = \{k^{**}, k^{**} + 1, \ldots N_p\}$ or $E_1 = \{\ \}$.

3.3 Parameter Estimation

To calculate the optimum sample allocation, the following quantities must be determined:

P_1 = proportion of HALS screened-in individuals who replied "yes" to Census question 20,

P_2 = proportion of HALS screened-out individuals who replied "yes" to Census question 20, and

P_3 = proportion of HALS screened-in individuals who replied "no" to Census question 20.

Since these parameters cannot be computed directly using data from the Canadian Health and Disability Survey, a test called the "calibration study" was carried out in September and October 1985.

Census question 20 was included, without abbreviation, as a supplementary question in the September Labour Force Survey (LFS). It was asked to a sample of approximately 36,000 individuals. The questions on the 17 ADLs and a question on mental handicaps were added as a supplement to the October LFS and were asked of the same individuals.

For each five-year age group, the weighted values from the calibration study were used to estimate the probability of an affirmative response, $P(yes)$, to Census question 20. The HALS screening questionnaire differs from that used in the calibration study. In HALS, there are more questions on mental and psychological problems and part (a) of Census question 20 is asked again. Therefore, we did not depend on the calibration study alone to calculate the parameters.

4. 1986 GEOGRAPHIC AND DEMOGRAPHIC FILE

4.1 Description of Available Information

When the sample allocation was done in the spring of 1986, the following information was available for use in calculation of population projections by age group and EA:

1. population projections by age group and province in 1986;
2. estimated population by age group and CD in 1984;
3. population by age group and EA in 1981;
4. conversion file to establish the correspondence between the 1981 and 1986 EAs;
5. estimated numbers of dwellings by EA in 1986.

The conversion file is structured according to the concept of equivalent sets. Each equivalent set is the smallest region consisting of EAs that has not had its boundaries altered. For example, if three 1981 EAs were reorganized as two 1986 EAs, the group of three 1981 EAs (or the group of two 1986 EAs) is an equivalent set.

The four methods described in the next subsection are designed to produce population projections by age group and by equivalent set in 1986. If an equivalent set is made up of several 1986 EAs, the projected population for the equivalent set can be divided proportionally among the EAs using the estimated numbers of dwellings by EA in 1986.

4.2 Estimation Methods

For province p, let

$ES_{l,k}$ = the l-th equivalent set of the k-th CD ($l = 1, 2, \ldots, N_k$; $k = 1, 2, \ldots, N_p$),

$ES_{l,k;81}(j)$ = population of $ES_{l,k}$ in the j-th age group in 1981 ($j = 1, 2, \ldots, 16$),

$CD_{k;84}(j)$ = estimated population of the k-th CD in the j-th age group in 1984,

$\hat{P}_{86}(j)$ = projected population in the j-th age group in the province in 1986.

For the three methods that follow, the first step is to calculate $\hat{CD}_{k;86}(j)$, the projected population of the j-th age group in the k-th CD in 1986. We assume there exists K_j' ($j = 1, 2, \ldots, 16$) such that

$$\hat{CD}_{k;86}(j) = K_j'(\hat{CD}_{k;84}(j)) \quad (k = 1, 2, \ldots, N_p; j = 1, 2, \ldots, 16),$$

$$\sum_{k=1}^{N_p} \hat{CD}_{k;86}(j) = \hat{P}_{86}(j) \quad (j = 1, 2, \ldots, 16).$$

This implies that

$$\hat{CD}_{k;86} = \frac{\hat{P}_{86}(j)\, CD_{k;84}(j)}{\sum_{k=1}^{N_p} CD_{k;84}(j)}.$$

The first method of estimating $ES_{l,k;86}(j)$ involves assuming the existence of K_j ($j = 1, \ldots, 16$) such that

$$\hat{ES}_{l,k;86}(j) = K_j ES_{l,k;81}(j) \quad (l = 1, \ldots, N_k; j = 1, \ldots, 16),$$

$$\sum_{l=1}^{N_k} \hat{ES}_{l,k;86}(j) = \hat{CD}_{k;86}(j) \quad (j = 1, 2, \ldots, 16).$$

We will say that this method uses the simple model. We obtain

$$\hat{ES}_{l,k;86}(j) = \frac{\hat{CD}_{k;86}(j)\, ES_{l,k;81}(j)}{\sum_{l=1}^{N_k} ES_{l,k;81}(j)} \quad (l = 1, \ldots, N_k; j = 1, \ldots, 16).$$

With this simple model, the estimated total population of $ES_{l,k}$ in 1986 is

$$\sum_{j=1}^{16} \frac{\hat{CD}_{k;86}(j)\, ES_{l,k;81}(j)}{\sum_{l=1}^{N_k} ES_{l,k;81}(j)}.$$

If one thinks that a better estimate, $\hat{ES}_{l,k;86}(tot)$ of this quantity can be produced by independent means (for example, using the estimated number of dwellings in $ES_{l,k}$ in 1986), then more elaborate models can be used to estimate $ES_{l,k;86}(j)$. The multiplicative model is specified by the following equations:

$$\hat{ES}_{l,k;86}(j) = K_j (ES_{l,k;81}(j)) + e_l' \quad (l = 1, \ldots, N_k; j = 1, \ldots, 16),$$

$$\sum_{l=1}^{N_k} \hat{ES}_{l,k;86}(j) = K(\hat{CD}_{k;86}(j)) \quad (j = 1, \ldots, 16),$$

$$\sum_{l=1}^{N_k} e_l' = 0,$$

$$\sum_{j=1}^{16} \hat{ES}_{l,k;86}(j) = \hat{ES}_{l,k;86}(tot) \quad (l = 1, \ldots, N_k).$$

One can interpret e_l as the net intra-CD migration for the l-th equivalent set.

The third model, called the additive model, is given by the following equations:

$$\hat{ES}_{l,k;86}(j) = ES_{l,k;81}(j) + e_l + f_j \quad (l = 1, \ldots, N_k; j = 1, \ldots, 16),$$

$$\sum_{l=1}^{N_k} \hat{ES}_{l,k;86}(j) = \hat{CD}_{k;86}(j) + D \quad (j = 1, \ldots, 16),$$

$$\sum_{l=1}^{N_k} e_l = D,$$

$$\sum_{j=1}^{16} \hat{ES}_{l,k;86}(j) = \hat{ES}_{l,k;86}(tot) \quad (l = 1, \ldots, N_k).$$

This model involves the assumption that the population increases (or decreases) for each age group in each of the equivalent sets in a CD can be decomposed into two terms - one which depends only on the equivalent set and not on age (e_l), and one which depends only on age and not on the equivalent set (f_j).

A final trivial model involves simply formulating

$$\hat{ES}_{l,k;86}(j) = ES_{l,k;81}(j) \quad (l = 1, 2, \ldots, N_k; j = 1, \ldots, 16).$$

4.3 Evaluation of Estimation Methods

The four methods were evaluated using data for the period 1976-1981. We used the 1976 projection of the population by age group and province in 1981, ($\hat{P}_{81}(j)$), the population by age group and EA in 1976, a 1976-1981 conversion file and the pre-Census estimate of the number of dwellings per EA in 1981. Since there are no estimates for population by age group and CD in 1979 (the equivalent of $CD_{k;84}(j)$), we set

$$\hat{CD}_{k;81} = \frac{\hat{P}_{81}(j)\,\hat{CD}_{k;84}(j)}{\sum_{k=1}^{N_p} \hat{CD}_{k;84}(j)}.$$

For $\hat{ES}_{l,k;81}(tot)$, which is needed for the multiplicative and additive models, we used

$$\hat{ES}_{l,k;81}(tot) = \frac{\sum_{j=1}^{16} ES_{l,k;76}(j) \quad \sum_{j=1}^{16} \hat{CD}_{k;81}(j)}{\sum_{l=1}^{N_k} \sum_{j=1}^{16} ES_{l,k;76}(j)}.$$

Table 1
Comparison of the Four Methods

Prov.	EFF_S	EFF_M	EFF_A
Nfld.	0.890	0.891	0.887
P.E.I.	0.903	0.914	0.919
N.S.	0.960	0.972	0.912
N.B.	0.870	0.868	0.884
Que.	0.778	0.764	0.818
Ont.	0.932	0.930	0.916
Man.	0.892	0.904	0.912
Sask.	0.732	0.749	0.801
Alta.	0.818	0.827	0.860
B.C.	0.713	0.716	0.775
Yukon	0.770	0.768	0.840
N.W.T.	1.252	1.246	1.157

For each province p, an efficiency measure was calculated for the simple, multiplicative and additive models relative to the trivial model:

$$EFF_m = \frac{\sum_{k=1}^{N_p} \sum_{l=1}^{N_k} \sum_{j=1}^{16} \left((\widehat{ES}_{l,k;81}^{(m)}(j) - ES_{l,k;81}(j) \right)^2}{\sum_{k=1}^{N_p} \sum_{l=1}^{N_k} \sum_{j=1}^{16} \left((\widehat{ES}_{l,k;81}^{(T)}(j) - ES_{l,k;81}(j) \right)^2} \quad (m = S, M, A),$$

where $\widehat{ES}_{l,k;81}^{(m)}(j)$ with $m = S, M, A$ and T are the projections obtained by means of the simple, multiplicative, additive and trivial models respectively. Some values obtained are given in Table 1.

The simple model gives the worst results for one province and one territory, the multiplicative model for two provinces and the additive model for seven provinces and one territory.

The simple model is the best for five provinces, while the multiplicative model is best for two provinces and one territory and the additive model is best for three provinces and one territory.

Since the simple model also has, as its name implies, the advantage of simplicity, it is the one that was chosen.

4.4 Method of Stratification by Enumeration Area Size

If simple random sampling were used to select EAs within each subprovincial area (SPA), disabled persons belonging to an EA with many disabled residents would have less chance of being selected than those in a small EA - that is, an EA with few disabled persons. To avoid excessive differences in selection probabilities, the population of EAs in each SPA is stratified according to the number of disabled persons in the EAs, and then proportional allocation is used. With proportional allocation, the number of EAs selected is proportional to the number of disabled persons for each stratum.

Using the results of earlier surveys, a link was established between the age distribution of the population of an EA and the number of disabled persons expected in the EA. Since the number of disabled persons is unknown, the variable used for stratification and sample allocation is the expected number of disabled persons.

In the case under consideration here, there are only two strata - one for large EAs and one for small EAs. Since proportional allocation is being used, we employed a criterion found in Raj (1968) to determine the optimum dividing line between large and small EAs. This criterion gives the optimum dividing line as the average of the average size of the small EAs and the average size of the large EAs.

5. SAMPLE SELECTION

It was necessary to draw samples for the three populations (children, adults and seniors) among the large and small "yes" EAs of each SPA, both for frame three and for frame five, and among the "no" EAs of each province for frame five. When an SPA contained fewer than two large EAs or fewer than two small EAs, we selected all of the EAs in that SPA for the three populations. The "yes" and "no" samples were created independently, using the one-pass algorithm described by Bebbington (1975). The samples from the three populations for the "yes" and "no" components were nested to minimize the total number of EAs selected.

The following table shows the sizes obtained for the samples by province for each age group.

Table 2

Sample Sizes by Province and Age Group

Province	Children		Adults		Seniors	
	Number of "yes" EAs selected	Number of "no" EAs selected	Number of "yes" EAs selected	Number of "no" EAs selected	Number of "yes" EAs selected	Number of "no" EAs selected
Nfld.	880	136	405	154	476	173
P.E.I.	242	242	111	217	82	166
N.S.	1257	157	434	130	438	115
N.B.	1142	162	459	146	453	138
Que.	4749	153	1070	114	1488	133
Ont.	6085	158	1304	116	1542	120
Man.	1082	203	457	169	367	144
Sask.	2291	265	942	241	921	193
Alta.	2762	190	909	176	1389	222
B.C.	3117	170	752	125	948	119

6. DISCUSSION

The postcensal survey is a relatively new survey method that will no doubt undergo extensive development in the next few years. This type of survey allows for a great deal of flexibility in data collection and use of large samples scattered throughout the country, with reasonable costs and timeliness. The Health and Activity Limitation Survey is the first postcensal survey of its size in Canada.

The sample design presented in this article is an attempt to maximize use of the opportunities offered by the postcensal approach, with optimum use of the available resources. One of the major problems inherent in the proposed method is control of sample size. Sample allocation is determined before the census is taken; this means that all calculations must be done using projections based on the previous census. In this context, the actual size of a sample made up of a group of small areas selected on the basis of the projection results may vary considerably from its expected size.

Therefore, on the one hand, one may obtain a sample that is inadequate with respect to the quality requirements for the estimates. On the other hand, the resources allocated to data collection may be exceeded. In order to prevent these problems, we implemented the following strategy. A target number of interviews for each population was calculated for the "yes" sample. This number was based on the sample size required to produce estimates that would satisfy our quality criteria. However, for the reasons mentioned above, we selected more EAs than were necessary to obtain the target number of interviews. For reasons of cost, if the real number of interviews to be conducted, as calculated in the field, was higher than the target number, a subsample of EAs were excluded from the survey. Only for the Halifax Regional Office (covering Prince Edward Island, Nova Scotia and New Brunswick) was the number of interviews in the "yes" sample substantially higher than the target number. The decision was therefore made to exclude certain EAs from this part of the sample. In order to know which EAs would be excluded, it was necessary to know the target number and the real number of interviews for each EA. For 40 per cent of the EAs, the real number of interviews had to be imputed since this information was not available in time.

For this imputation, the total real number of interviews was known for each census commissioner district. The portion of this total not already allocated to EAs with known numbers of interviews was distributed among the EAs requiring imputation, in proportion to the target number of interviews.

We then calculated, for each population, the difference between the real number and the target number of interviews for each of the two strata of each SPA. A positive difference (real-target) indicated a population for which some EAs could be excluded from the survey. In each stratum, the EAs were divided into three groups (1, 2 and 3), in accordance with whether they had been selected for three, two or only one of the populations respectively. The EA file was then sorted by stratum and by group in ascending order, with the order of the EAs within each group being random. Each EA was considered successively and was suppressed for the three populations if:

1) a positive difference remained non-negative after suppression of the EA;
2) a negative difference was not further reduced.

In this way, each positive difference was reduced to a number as close as possible to zero, considering the random order of the EAs.

ACKNOWLEDGEMENTS

The authors would like to thank D.A. Binder for his contribution with regard to the sample design and P. Reed for the computer work involved in the sample allocation. The authors would also like to thank the referees for their useful comments.

APPENDIX

Question 20 of Census Form 2B

20. a) Are you limited in the kind or amount of activity that you can do because of a long-term physical condition, mental condition or health problem: (See Guide)

At home?
☐ No, I am not limited
☐ Yes, I am limited

At school or at work?
☐ No, I am not limited
☐ Yes, I am limited
☐ Not applicable

In other activities, e.g., transportation to or from work, leisure time activities?
☐ No, I am not limited
☐ Yes, I am limited

b) Do you have any long-term disabilities or handicaps?
☐ No
☐ Yes

Screening Questions for HALS (Questionnaire for Adults)

1. Do you have any trouble hearing what is said in a normal conversation with one other person?
2. Do you have any trouble hearing what is said in a group conversation with at least three other people?
4. Do you have any trouble reading ordinary newsprint, with glasses if normally worn?
5. Do you have any trouble seeing clearly the face of someone from 12 feet/4 metres (example: across a room), with glasses if normally worn?
7. Do you have any trouble speaking and being understood?
8. Do you have any trouble walking 400 yards/400 metres without resting (about three city blocks)?
9. Do you have any trouble walking up and down a flight of stairs (about 12 steps)?
10. Do you have any trouble carrying an object of 10 pounds for 30 feet/5 kg for 10 metres (example: carrying a bag of groceries)?
11. Do you have any trouble moving from one room to another?
12. Do you have any trouble standing for long periods of time, that is, more than 20 minutes? Remember, I am asking about problems expected to last 6 months or more.
13. When standing do you have any trouble bending down and picking up an object from the floor (example: a shoe)?
14. Do you have any trouble dressing and undressing yourself?
15. Do you have any trouble getting in and out of bed?
16. Do you have any trouble cutting your own toenails?

17. Do you have any trouble using your fingers to grasp or handle?
18. Do you have any trouble reaching in any direction (example: above your head)?
19. Do you have any trouble cutting your own food?
20. Because of a long-term physical condition or health problem, that is, one that is expected to last 6 months or more, are you limited in the kind or amount of activity you can do . . .

    (i) at home? (ii) at school or at work? (iii) in other activities such as travel, sports, or leisure?
21. Has a school or health professional ever told you that you have a learning disability?
22. From time to time, everyone has trouble remembering the name of a familiar person, or learning something new, or they experience moments of confusion. However, do you have any ongoing problems with your ability to remember or learn?
23. Because of a long-term emotional, psychological, nervous, or mental health condition or problem, are you limited in the kind or amount of activity you can do?

    (i) at home? (ii) at school or at work? (iii) in other activities such as travel, sports, or leisure?

REFERENCES

BEBBINGTON, A.C. (1975). A simple method of drawing a sample without replacement. *Applied Statistics*, 24, 136.

CARTER, R.G., GILES, P.D., and SHERIDAN, M.J. (1982). Description and rationale for the screen tests for the January 1983 Disability Survey. Disability Data Development Project, Health Division, Statistics Canada.

HOUSE OF COMMONS (1981). Obstacles, Report of the special committee on the disabled and handicapped. Ottawa.

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

DOLSON, D., GILES, P., and MORIN, J.-P. (1984). A Methodology for surveying disabled persons using a supplement to the Labour Force Survey. *Survey Methodology*, 10, 187-197.

DOLSON, D., McCLEAN, K., MORIN, J.-P., and THÉBERGE, A. (1986). Methodology report of HALS. Working Paper, Statistics Canada.

GRABOWIECKI, F. (1982). Discussion of the target population for the Disability Survey. Disability Data Development Project, Health Division, Statistics Canada.

GRABOWIECKI, F. (1983). Content of Statistics Canada's Disability Survey. Technical Report, Health Division, Statistics Canada.

LAZARUS, G., and NESICH, R. (1985). A report on the methodology of the Canadian Health and Disability Survey. Working Paper, Statistics Canada.

McDOWELL, I. (1981). An examination of the OECD survey questions in a Canadian Study. *Revue d'épidémiologie et de santé publique*, 29, 412-429.

MORIN, J.-P., and DOWLER, L. (1986). Proposition d'une méthodologie pour l'ESLA-institutions. Working Paper, Statistics Canada.

MORIN, J.-P. (1986). Comparaison initiale de l'ESIC et de l'ESG. Working Paper, Statistics Canada.

WORLD HEALTH ORGANIZATION (1980). International classification of impairments, disabilities and handicaps. Geneva, Switzerland.

RAJ, D. (1968). *Sampling Theory*. New York: McGraw-Hill.

WILSON, R.W., and McNEIL, J.M. (1981). Preliminary analysis of OECD disability on the pretest of the Post-censal Disability Survey. *Revue d'épidémiologie et de santé publique,* 29, 469-475.