THE REDESIGN OF THE CANADIAN LABOUR FORCE SURVEY

I.P. Fellegi, G.B. Gray, R. Platek
Dominion Bureau of Statistics, Ottawa

The Canadian Labour Force Survey was established in 1945 and designed, at that time, to provide quarterly estimates of labour force characteristics at the national level. Since 1952 it has been conducted monthly and estimates have been published also for five regions (a region being a province or a group of provinces). The method of enumeration is through interviewing during a specified week each month (the survey week) and collecting information pertaining to the previous week (reference week).

The sample design of the survey was originally based on the Census of 1941, and although it has been revised several times since, many of its features are still based on the Census of 1941. We are now in the process of redesigning the survey, introducing the new design province by province.

Table 1 of the handout provides a somewhat crude summary description of the main features of the old design. All cities with a population of more than 30,000 persons were selected with certainty and they constituted the so called self-representing areas of the survey. In each of these cities a two-stage sample of segments (mostly city blocks) and households was selected, each selected segment yielding an expected 5 households. The remaining parts of the country were divided into primary sampling units (p.s.u.'s) and within each province p.s.u.'s with similar overall characteristics were grouped into strata. The p.s.u.'s in a stratum were not necessarily contiguous and there was considerable variation in the number of p.s.u.'s per stratum as well as the total population of the stratum. The population of strata was about 100,000-150,000 persons. One p.s.u. was selected with p.p.s. in each stratum within which in successive stages a sample of segments, clusters and households was selected.

The following comparisons between the Labour Force Survey and the CPS should be emphasized:

1) Our p.s.u.'s were not natural units as the counties are in the CPS. Had we used counties as p.s.u.'s we would have had much too few p.s.u.'s due to the relatively small population of Canada.
2) Cities of 30,000 population with our 1% sampling ratio provide a sample of about two enumerator assignments. In the CPS, cities of more than 250,000 persons are made self-representing and this yields a sample of about one small enumerator assignment.
3) Our strata were much smaller than those in the CPS (about one-third as large) yet even so our strata yielded more than 4 enumerator assignments while the CPS strata usually yield about 1 enumerator assignment. In both surveys 1 p.s.u. is selected in each stratum.

With the creation of a sampling research group, a more systematic investigation of the old design has begun a few years ago. The desirable objectives of a new design have been developed partly through the observation of several serious shortcomings and imbalances in the old design:

1) An exceedingly important lack of balance was noticed between the self-representing areas and the non-self-representing areas with respect to cost and variance. Table 3 of the handout indicates that in Alberta the cost of the survey in the non-self-representing areas accounted for about one third of the cost but three quarters of the variance.
2) A study of the components of variance revealed that about 50% of the variance in the non-self-representing areas is accounted for by the first stage of sampling. This clearly indicated that we had too few p.s.u.'s in the sample and in comparison, too large a sample within. In fact, the number of p.s.u.'s in the sample from the non-self-representing areas, was 78 in total, employing 288 enumerators. By selecting 288 p.s.u.'s and a sample of one enumerator assignment in each, we could increase the number of p.s.u.'s in the sample while maintaining the same number of enumerators and the same overall sample size. In fact, the number of p.s.u.'s in the sample has but a small affect on the total cost, as long as the enumerators reside in the p.s.u. they enumerate, and the sample take in the p.s.u. provides at least one enumerator assignment. Consequently one of the early decisions was to increase the number of p.s.u.'s in the sample and to reduce the take within to one enumerator assignment. In line with this decision we have also decided to lower the population level above which the cities will be self-representing, again to the point where a self-representing city yields at least one enumerator assignment. This includes in the new design all cities of more than 15,000 population removing most of the industrialized urban population from the non-self-representing areas. This, of course, makes the task of stratification in the non-self-representing areas considerably easier.
3) The measures of size used for p.p.s. sampling at the various stages which are based on 1951 Census information are also outdated. This results in a considerable variance in the sample take which, of course, is reflected in the variance of the main estimates as well.
4) For some time, now, we have been concerned

with the bias of the so-called collapsed stratum variance estimates due to selecting only one p.s.u. in each stratum. Paradoxically, improving the stratification in the new design would have meant increasing this bias (which basically is a between strata variance). The selection of two p.s.u.'s per stratum, for a given set of sampling ratios, would necessitate the creation of twice as large strata. We knew that for the contemplated sampling ratios even under the two p.s.u. stratum scheme our strata would not have a population larger than 40,000-80,000 persons. The 1961 Census information was utilized in an empirical study investigating the relative efficiency of the two schemes. We have found that there is very little to be gained by creating strata with populations less than 40,000-80,000 persons and so we decided to have the luxury of unbiased variance estimates by selecting two p.s.u.'s from each stratum. It should be emphasized that this decision was taken in light of the fact that even so our strata would be rather small; 40,000-80,000 persons is about the size of a county. In the CPS strata have a population of more than 300,000 persons and only one p.s.u. is selected in each stratum. Again, the quantitative difference between the higher overall sampling ratio in Canada and the substantially smaller one, less than one-tenth as large in the U.S.A. resulted in a qualitative difference in the design.

5) Studies based on 1961 Census data indicated that our stratification (essentially unchanged since 1945) is hopelessly outdated. The socio-economic map of Canada has changed a great deal in the meantime. In the new design we wanted to form strata which had a good chance of staying internally homogeneous even in a few years' time. If a stratum with a large population is to be formed, the practise is usually to put into the same stratum areas which are sometimes separated by considerable distances, because one seldom finds enough neighbouring municipalities with similar characteristics to make up a stratum. Thus local developments which take place after the stratification affect different parts of the stratum differently thereby contributing to the gradual deterioration of the original stratification. We have found that due to the relatively small size of our strata and due to the fact that most of the industrialized urban population is removed from the non-self-representing areas and is sampled directly, it was possible to create strata which are made up of geographically contiguous areas. Therefore, we decided that wherever possible we will form geographically compact strata.

Ihe procedure of stratification was carried out within Economic Regions. These are subdivisions of provinces, defined as areas of "structural homogeneity" according to such factors as soil characteristics, production and marketing possibilities, commercial and industrial potential. It has been decided to carry out the stratification process in each Economic Region separately for two main reasons. Firstly, estimates may be required in the future, for areas smaller than a province, which may well be an Economic Region or a group of them. Secondly, the Economic Region as an area, is more conveniently manageable than the whole province at the redesign stage as well as in future revisions of the Labour Force Survey. The immediate problem in stratifying an Economic Region was to select the characteristics which should be used in the stratification. Ideally these characteristics should fulfill the following conditions:

a) The characteristic is relatively stable over time

b) The characteristic is related to some Labour Force characteristic

c) The number of persons having this characteristic varies from area to area so it discriminates between areas

d) The number of persons having this characteristic accounts for a sizable part of the population of the region.

Conditions a) and b) were taken into account by exercising our judgement in selecting characteristics (i.e., employment by industry, average wages and salaries, etc.). As for c) and d) the characteristics were evaluated in a particular Economic Region by computing the so-called "Importance Factor" defined as

$\frac{n\sigma^2}{N}$ where

n=total number of persons in the Labour Force in an Economic Region having the particular characteristics

N=total number of persons in the Labour Force in a particular Economic Region

σ^2=variance of the characteristics between areas (municipalities) in the Economic Region.

This measure was calculated for a number of characteristics in each region and the relative size of the Importance Factor determined the importance of a given characteristic. Then three or four of the most important characteristics were used for stratification purposes.

Considering the repetitive character and the volume of the operations involved in stratification, it was essential to have a uniform method, which could be followed by clerks without much supervision, to ensure both speed and efficiency. The building blocks used in setting up strata were enumeration areas which are the smallest units for which Census data are available. Each enumeration area was compared with the whole Economic Region with respect to the "Important" characteristics. This comparison might show, for example, that a

given E.A. is higher than average with respect to one characteristic, lower than average with respect to the second, etc. Plotting these comparisons a pattern is obtained for each E.A. The enumeration areas whose pattern was similar (on the basis of visual inspection by clerks) were grouped into one stratum with due regard to the desired objectives of geographic contiguity and approximately equal size of strata. Within each stratum p.s.u.'s were formed. These again were geographically compact and were, of course, representative of the entire stratum: a maximum deviation of 5% from the corresponding stratum average of the important characteristics was allowed. The proportion of urban and rural population in the primary sampling unit was even more strictly controlled. Due to the variation in the size and geographical distribution of urban centres it was often necessary to split urban centres between two or more primary sampling units. This splitting of urban centres was not made geographically but on a ratio basis. Throughout the whole operation a set of punched cards was maintained (one for each E.A.) containing the values of the important characteristics as well as room for strata and p.s.u. codes. This enabled us to utilize the computer and clerks to maximum efficiency and statisticians were only used to polish the draft strata-p.s.u. frame prepared by clerks. Table 4 contains a summary description of the frame of the new design.

In each province, institutions (such as Hospitals, Schools, Hotels, Military Establishments) and also Remote Areas were designated as special areas. Special sampling and enumeration procedures were used for these areas.

6) Concerning the self-representing areas preliminary cost and variance analysis indicated that the old design was not far from the optimum. Here the main problem seemed to be that in some parts of a city due to its local development (such as urban renewal, new apartment buildings, etc.) the original household counts become outdated quite fast and quite drastically. Any revision of the original household counts would alter the probabilities of selection and would therefore disrupt the whole systematic sample, spread throughout the entire city. For example, the selection of a new sample of segments in a city like Montreal where almost 500 of them are in the sample, would have been a considerable job. A feature of the new design to which we attach considerable importance, will be the division of the larger cities into strata, we call them sub-units, and the independent selection of a systematic sample of segments in each sub-unit. A regular programme of checking building permits will then indicate if a large development occurs in any of the sub-units. Whenever this occurs we will revise the measures of size but only in the affected sub-units. Other features of the design in the large cities include special treatments for institutions and large apartment houses. Time does not permit to go into details.
7) The old design of the Labour Force Survey was established to provide national estimates. Later on separate estimates were also published for five regions where some of the regions were the larger provinces. The estimates for the smaller provinces, though unpublished, were made available to the provincial governments with due warnings concerning their reliability. There was an increasing demand, however, for strengthening the provincial estimates. In the new design a compromise allocation of the sample into the ten provinces was worked out in such a way that the estimates of unemployed in each province would have a coefficient of variation of not more than 13% (with the exception of Prince Edward Island whose total population is only 100,000). Like all compromises this one is also difficult to explain rigorously. It was arrived at by considering various alternatives and choosing the one which did not deviate too far from the optimum design and yet brought the provincial estimates within sight of being publishable. In comparison the CPS publishes estimates only for four regions of the U.S.A. but then the Bureau of the Census is fortunate that there are 50 states in the U.S.A. rather than 10.

Having made the broad decisions outlined above the details of the new design emerged on the basis of guidance provided by the cost and variance study. I shall attempt to outline our general approach. As usual in studies of this kind a mathematical model was established for both the cost of the survey and the variance of the resulting estimates. Each of these was to be a function of the same variables with respect to which the design was to be optimized. The mathematical model for the cost was designed to reflect field costs only since Head Office costs were assumed to be given and fixed.

Since a preliminary study indicated that in the self-representing areas the old design was already close to the optimum, the cost function set up to represent the cost of enumeration in the self-representing areas was rather simple and shall be omitted here. The cost in the non-self-representing areas was split into two main parts:

1) enumeration costs - This is that part of the cost which may be thought to be proportional to the number of households in the sample. It includes, therefore, the cost of the actual interviews, the travel from household to household within clusters and also the cost of training the enumerators, since this latter is proportional to the number of enumerators which in turn is directly proportional to the sample size.
2) travel cost - The travel cost itself can be broken into various components corresponding

to travel between sampling units at the various stages of sampling with the exception that there is no travel between p.s.u.'s since one enumerator enumerates each p.s.u. and that there is an additional component of travel from home to area (that is the travel of the enumerator from his residence to the first household to be enumerated each day and back to his residence at the end of the day).

A detailed description of the cost function appears in the Appendix. It was assumed that the total field cost may be represented as the sum of the cost in the rural and urban areas.

The variance function pertaining to either the urban or the rural portion of a province appears in the second part of the Appendix. It is split up into four components, each corresponding to a stage of sampling. In each selected p.s.u. the urban and rural areas are subsampled independently. Therefore, the variance components are additive over the urban and rural parts except at the p.s.u. level where an additional covariance term occurs between the urban and rural subsamples.

The symbols appearing in the functions are either constants or variables belonging to one of the following five categories:

1) constants based on a detailed study of the enumerators' records of time and mileage. The information was supplemented by map studies. These constants are averages referring to a particular province (i.e. average time spent in enumerating a household).
2) constants derived from other sources, for example intercensal population estimates or information pertaining to variance components derived under the old design.
3) constants for which a provincial average was not satisfactory. Population density is an example: it varies within a province to such an extent that the average is hardly useful. Several different values were substituted for these constants and their affect on the optimum allocation was examined.
4) the basic variables in terms of which the optimization was carried out. In the present study these were the weights at the various stages of sampling.
5) variables which were functions of the basic ones.

The operation of optimization was an exercise in non-linear programming. Various boundary conditions had to be satisfied to take into account known restrictions. Some of these restrictions were imposed somewhat arbitrarily, because we felt that the empirical formulae were applicable only within certain ranges of the variables. Also the assumption that certain quantities are constants is valid only within a limited range of the variables. For example it was thought that the average time spent in enumerating a household is not sensitive to changes in the design and was therefore regarded as a constant. The details of the operation of optimization will be omitted here. It was carried out by a high speed computer (IBM 7074).

In closing, we would like to say a few words about the results. At this time, the survey is operating under the new design in three of the ten provinces. A full-scale field test had been completed in another two provinces. By the end of 1965 the survey will operate entirely under the new design. The results quoted in Table 3 refer to Alberta, because it was the first province to be redesigned and hence the only one with a few months of data under the new design. It is not an entirely typical province, in that the difference under the old design between the variance in the self-representing and non-self-representing areas is substantially greater in the other provinces (those outside the Prairies). Since the gain in the new design is made in the non-self-representing areas, the reduction of the variance in Alberta is smaller than what we expect in the other provinces. Even under these relatively unfavourable conditions Table 3 shows that under the new design the variance of the estimated employed under the new design has been reduced by 20% in the self-representing, 70% in the non-self-representing areas for an overall reduction of 55%. The variance of the estimated unemployed has been reduced by almost 25%. All of this variance reduction was effected while also reducing the field cost of the survey by 14%. The amount of information per unit cost ([1/variance]/cost) has increased by 155% for employed and 56% for unemployed.

TABLE 1
OLD DESIGN

Stage of Sampling	SELF-REPRESENTING AREAS			NON-SELF-REPRESENTING AREAS		
	Nature of Units	Size of Units (pop.)	Method of Selection	Nature of Units	Size of Units (pop.)	Method of Selection
Stratum	Metropolitan area or Special area	30,000+	Certainty	Group of similar p.s.u.'s (generally) not contiguous	80,000-200,000	Certainty
First Stage (P.S.U.)	None	None	None	Groups of Heterg. municipalities	10,000-22,000	One unit selected with p.p.s.
Second Stage (Segment)	City block(s)		Systematic with p.p.s. to household count	Census enumeration area(s)	Rural approximately 500 Urban 1,000 to 20,000	Systematic p.p.s. area sub stratification
Third Stage (Cluster)	None	None	None	Small area with recognizable boundaries	Multiple of 4-8 H.H.'s	Random (p.p.s. for multiple clusters)
Fourth Stage (Household)	Household	3-4	Random systematic to yield expected 5 H.H.'s as of design	Household	3-4	Random in multiple clusters

TABLE 2
NEW DESIGN
(in the Province of Alberta; some variation from province to province)

Stage of Sampling	SELF-REPRESENTING AREAS			NON-SELF-REPRESENTING AREAS		
	Nature of Units	Size of Units (pop.)	Method of Selection	Nature of Units	Size of Units (pop.)	Method of Selection
Stratum	Metropolitan area or Special area	15,000+	Certainty	Group of similar p.s.u.'s (geographically contiguous)*	35,000-55,000	Certainty
First Stage (P.S.U.)	Census Tracts	15,000	Certainty	Rural enumeration areas and nearby small urban*	3,200-5,500	Two units selected with p.p.s.
Second Stage (Segment)	City block(s)		p.p.s. systematic	Rural enumeration area and small urban or part of it	Rural 500 Urban approximately 800	Systematic p.p.s. within urban and rural
Third Stage (Cluster)	None	None	None	Small area with recognizable boundaries	Multiple of 3 or 4 H.H.'s	Random systematic (p.p.s. for multiple clusters)
Fourth Stage (Household)	Household	3-4	Random systematic	Household	3-4	Random systematic in multiple clusters

* Note: In the old design p.s.u.'s were formed first and then combined into strata; in the new design strata were formed first and then divided into p.s.u.'s.

TABLE 3

Cost, Variance and Information per Unit Cost of the Estimated Employed and Unemployed in Alberta by Type of Area and Design

		Cost in thousand dollars	EMPLOYED		UNEMPLOYED	
			Variance[4] in millions	Information per cost[5] in 10^{-10}	Variance[4] in millions	Information per cost[5] in 10^{-10}
Old design[1]	Self-representing	2.35	20.84	0.20	2.39	1.78
	Non-self-representing	1.33	45.46	0.17	6.94	1.09
	Total	3.68	66.30	0.04	9.33	0.29
New design[2]	Self-representing	1.55	16.59	0.39	3.24	2.00
	Non-self-representing	1.63	13.47	0.46	3.67	1.67
	Total	3.17	30.07	0.10	6.91	0.46
Ratio (New/Old)[3]	Self-representing	0.66	0.80	1.91	1.36	1.12
	Non-self-representing	1.23	0.30	2.75	0.53	1.54
	Total	0.86	0.45	2.55	0.74	1.56

(1) The old design in Alberta utilized a 1% overall sampling ratio in the self-representing areas and a 0.67% sampling ratio in the non-self-representing areas.

(2) The sampling ratio in the new design is 0.8% in both self-representing and non-self-representing areas.

(3) The ratios in the last three rows are in units.

(4) For the sake of making the variances comparable, they were adjusted to refer to the same level of employment and unemployment under both designs.

(5) The information per unit cost here refers to (1/Variance)/Cost.

Appendix to the "Redesign of the Canadian Labour Force Survey"

The Cost Function

The cost function for non-self-representing units is split into enumeration and travel costs while the travel cost is further split into three components. In each case, the time spent is given in terms of hours spent and this is followed by the cost in dollars. The selection of two p.s.u.'s per stratum is assumed here.

$$(1)\quad T_e = \left(\frac{P_U}{W_U}\cdot\frac{1}{P_{4U}} + \frac{P_R}{W_R}\cdot\frac{1}{P_{4R}}\right) t_H$$

= Enumeration Time (1)

$$C_e = T_e(r_h + r_m S_H) = \text{Enumeration Cost} \qquad (2)$$

and (2)(i) $T_{HA} = 2N\, d_{HA}\, \frac{1}{S_1}$ = Travel time between home and area. (3)

$$C_{HA} = T_{HA}\,(r_h + r_m S_1) = \text{Cost of travel between home and area.} \qquad (4)$$

$$(ii)\quad T_{SS} = (M_U + M_R - N)\, d_{SS}\, \frac{1}{S_2}$$

= Time required for segment to segment travel. (5)

$$C_{SS} = T_{SS}\,(r_h + r_m S_2)$$

= Cost of segment to segment travel. (6)

$$(iiiJ)\quad T_{CC:J} = (C_J - M_J)\, d_{CC:J}\, \frac{1}{S_{3J}}$$

= Time required for cluster to cluster travel (broken down by J=U (urban) and J=R (rural)). (7)

$$C_{CC:J} = T_{CC:J}\,(r_h + r_m S_{3J}) \qquad (8)$$

The Variance Function

The variance function for non-self-representing units is split up into two main parts, urban and rural and within each it is split up into four components as follows: (a) between p.s.u.'s, (b) between segments, (c) between clusters, and (d) between households. In addition to these four components of variance there exists a covariance between urban and rural p.s.u. totals given by (e). The four components of variance for either the urban or rural areas (denoted by J) and the covariance in (e) are given by:

$$(a)\quad V_{1J} = B_J\,(W_{1J}-1)\,\frac{1}{1-\frac{P_{1J}}{P_{0J}}}$$

$$[1 + (P_{1J}-1)\,\delta_{1J}] \qquad (9)$$

$$(b)\quad V_{2J} = B_J\,(W_{2J}-1)\,W_{1J}\,\frac{1}{1-\frac{P_{2J}}{P_{1J}}}$$

$$\{[1+(P_{2J}-1)\delta_{2J}] - \frac{P_{2J}}{P_{1J}}[1+(P_{1J}-1)\,\delta_{1J}]\} \qquad (10)$$

$$(c)\quad V_{3J} = B_J\,(W_{3J}-1)\,W_{1J}W_{2J}\,\frac{1}{1-\frac{P_{3J}}{P_{2J}}}$$

$$\{[1+(P_{3J}-1)\delta_{3J}] - \frac{P_{3J}}{P_{2J}}[1+(P_{2J}-1)\delta_{2J}]\} \qquad (11)$$

$$(d)\quad V_{4J} = B_J\,(W_{4J}-1)\,W_{1J}W_{2J}W_{3J}\,\frac{1}{1-\frac{P_{4J}}{P_{3J}}}$$

$$\{[1+(P_{4J}-1)\delta_{4J}] - \frac{P_{4J}}{P_{3J}}[1+(P_{3J}-1)\delta_{3J}]\} \qquad (12)$$

$$(e)\quad (CV)_{1:UR} = r_{UR}\sqrt{V_{1U}\cdot V_{1R}} \qquad (13)$$

The symbols in formulae (1)-(13) are either constants or variables. They refer to averages in a particular province and may be classified into one of the following five categories.

Category 1: Constants based on a detailed study of enumerators' records of time and mileage. The information was supplemented by map studies.

t_H = average time spent in enumerating a household, including travel between households within clusters (but not between clusters or between segments).

S_H = average speed of travel between successive households within a cluster.

t = average time per enumerator per round trip (time spent by an enumerator during a day starting from his place of residence, travelling to the first sample household, enumerating and travelling during his work and finally back to his place of residence).

$\beta_1, \beta_2, \beta_{3J}$ = constants of proportionality estimated from empirical results under the assumption that the distances travelled from home to area, from segment to segment and from cluster to cluster are proportional to the square root expressions of formulae (3), (5) and (7) respectively.

Category 2: Constants derived from sources other than enumerators' records or map studies.

P_J = population, 14 years of age and over (excluding non-enumerable persons such as inmates, armed forces personnel, etc.) within the urban or rural part of a province (estimated using the intercensal estimates of the population and the estimated proportion of the population living in urban or rural areas).

P_{4J} = average size of household in type of area J (urban or rural)

r_h = hourly rate of pay

r_m = rate of pay per mile

H = size of enumerator assignment (number of households)

r_{UR} = correlation coefficient between the estimates (generally unemployed in our studies) derived for the urban and rural parts of a p.s.u.

p_J = proportion of persons with a certain specific labour force characteristic (generally unemployed in our study) in type of area J.

$B_J = P_J p_J (1-p_J)$; (binomial variance).

Category 3: Constants for which one provincial average was not satisfactory. Several different values were substituted for these constants and their effect on the optimum allocation was examined.

ρ_J, ρ = population density (urban or rural or overall)

P_{2J} = average population in urban or rural segments

P_{3J} = average population in urban or rural clusters

Note: there are natural limitations concerning the values P_{2J} or P_{3J} may assume. Segments are to be formed by combining Census Enumeration Areas and so the average population of a segment must be a multiple of the average population of Enumeration Areas. Similarly, the availability or lack of natural boundaries impose restrictions on the size of clusters.

Category 4: The basic variables, in terms of which the optimization was carried out.

W_J = overall weight (inverse of overall sampling ratio) in type of area J (urban-rural)

W_1 = p.s.u. weight (inverse of sampling ratio at first stage)

W_{2J} = segment weight (inverse of sampling ratio at second stage in urban-rural areas)

W_{3J} = cluster weight (inverse of sampling ratio at third stage in urban-rural areas)

Note: in the final analysis both the cost and the variance functions were expressed in terms of constants and the seven basic variables listed above.

Category 5: Functions of the basic variables.

$K = \frac{1}{2H}\left[\frac{P_U}{W_U P_{4U}} + \frac{P_R}{W_R P_{4R}}\right]$ = number of strata in province (the selection of two p.s.u.'s per stratum is assumed to have been decided)

$P_{1J} = \frac{P_J}{2K\, W_1}$ = average population per p.s.u. living in urban or rural areas.

$d_{HA} = \beta_1 \sqrt{\frac{P_{1U}+P_{1R}}{\rho}}$ = average distance from home to area. (β_1 is a regression coefficient)[1]

$d_{SS} = \beta_2 \sqrt{\frac{P_{1U}+P_{1R}}{\rho} / \left[\frac{P_{1U}}{W_{2U}P_{2U}} + \frac{P_{1R}}{W_{2R}P_{2R}}\right]}$ = average distance between sampled segments within a p.s.u. (β_2 is a regression coefficient)[1]

$d_{CC:J} = \beta_{3J} \sqrt{\frac{P_{2J}}{\rho_J} / \frac{P_{2J}}{W_{3J}P_{3J}}}$ = average distance between sampled clusters within urban or rural segments. (β_{3J} is a regression coefficient)[1]

$S_1 = a_1 + b_1 d_{HA}$ = speed of travel between home and area (a_1, b_1 are regression coefficients)[2]

$S_2 = a_2 + b_2 d_{SS}$ = speed of travel between sampled segments (a_2, b_2 are regression coefficients)[2]

$S_{3J} = a_{3J} + b_{3J} d_{CC:J}$ = speed of travel between sampled clusters within urban or rural segments (a_{3J}, b_{3J} are regression coefficients)[2]

$n_{SJ} = \frac{P_J}{W_1 W_{2J} P_{2J}}$ = number of urban or rural segments in the sample in the province.

$n_{CJ} = \frac{P_J}{W_1 W_{2J} W_{3J} P_{3J}}$ = number of urban or rural clusters in the sample in the province.

$v_{SJ} = C_{1J} + d_{1J} \frac{P_{2J}}{W_{3J} W_{4J} P_{4J}}$ = average number of visits per urban or rural segment during survey week (C_{1J}, C_{2J} are regression coefficients)[3]

$v_{CJ} = C_{2J} + d_{2J} \frac{P_{3J}}{W_{4J} P_{4J}}$ = average number of visits per urban or rural cluster during survey week (C_{2J}, d_{2J} are regression coefficients)[3]

$M_J = n_{SJ} v_{SJ}$ = total number of visits to all urban or rural segments in the sample.

$C_J = n_{CJ} v_{CJ}$ = total number of visits to all urban or rural clusters in the sample.

$P_{OJ} = \frac{P_J}{K}$ = urban or rural population per stratum

$\delta_{rJ} = e_J (P_{rJ}^{-f_J} - P_{OJ}^{-f_J})$ = intraclass correlation between pairs of persons within r-th stage units within stratum (e_J, f_J are regression coefficients)[4]

N = number of "round trips" by all enumerators during a survey week. An enumerator may make one or more round trips during a day travelling from his home to sampled households, between sampled households and back home. N is estimated as the solution of the (linear) equation

$$Nt = T_e + T_{HA} + T_{SS} + T_{CC:U} + T_{CC:R}$$

after substituting for the quantities on the right hand side expressions (1), (3), (5) and (7).

Note 1: distances travelled between sampled r-th stage units are assumed to be directly proportional to the square root of the area in which they are located (i.e. the area of the (r-1)-st stage unit) and inversely proportional to the square root of the number of sampled r-th stage units in the area. The area of an (r-1)-st stage unit was estimated as the ratio of population over density of population. The home to area distance was assumed to be proportional to the square root of the area of the p.s.u. since an enumerator generally resides in his assigned p.s.u.

Note 2: the average speed of travel, within the range of distances involved, was assumed to be linearly dependent on the distance travelled.

Note 3: the average number of visits to be made to a sampled unit was assumed to be linearly dependent or the number of households to be enumerated in the unit. The effect of callbacks is hoped to be incorporated at this point.

Note 4: the intraclass correlation model applied here has been used by many authors (i.e. Hansen, Hurwitz, Madow: Sample Survey Methods and Theory, vol. 1, pp.307) but without the term involved P_{OJ}. This correction term was added to make the intraclass correlation between persons within stratum reduce to zero. It has little effect if P_{rJ}/P_{OJ} is small, however it has a noticeable effect at the p.s.u. level. This correction improved the fit between the curve and computed values of the intraclass correlation.