THE METHODOLOGY OF THE CANADIAN TRAVEL SURVEY, 1971

A. Ashraf
Household Surveys Development Division

The Canadian Travel Survey, 1971 was the largest survey on travel of Canadian residents. This paper describes some important aspects of the methodology. Particular emphasis is given to the development of definitions in relation to the methodology, the sampling technique and interview strategy.

1. INTRODUCTION

The Canadian Travel Survey, 1971 (CTS 71) was the largest national household survey designed to collect data on domestic travel by Canadians. It was primarily developed, designed and managed by Statistics Canada on behalf of the Canadian Government Office of Tourism, Department of Industry, Trade and Commerce.

The travel industry, at least historically, has not been treated as an industry, partially due to its fragmentation and distribution throughout other more homogeneous industries such as transportation, accommodation, recreation and service industries. The statistics about these industries have been collected and analysed for their own purposes without specific references to their tourism components.

The travel industry in the past has been fulfilling its own needs for data on domestic travel by specialized and localized market research type surveys. The government agencies, for their part, have conducted surveys if and when warranted for their marketing needs. This piece-meal approach to the data of vital importance was less than satisfactory to the travel industry planners and government policy makers.

The governments, both Federal and Provincial, as well as the industry have commissioned and conducted a number of market research studies, but those, until very recently, were largely uncoordinated and uncomparable due to varying emphasis in objectives and concepts. There was, however,

an almost complete lack of base data on domestic travel on a national scale.

The Federal-Provincial Conference on Travel and Tourism in 1967 resolved to sponsor a travel survey at a national scale to provide the industry and governments with the much needed data on domestic travel for a more realistic policy formation about the industry.

2. OBJECTIVES

The primary objective of the Canadian Travel Survey 1971 was to provide the base data covering the whole calendar year 1971 on the volume and value of travel by Canadians in Canada in such depth and geographic breakdown as is practicable. This broad statement of the objective was spelled out in terms of volume, to mean, an estimate of the total number of Canadians that travelled during the calendar year, value to mean, an estimate of the total expenditure in dollars that incurred as a result of the above travels, depth to mean, further details of volume and value in meaningful components, and finally geographic breakdown to mean, the estimates to be produced would be at national levels, down to the level of geography where they could be useful. These levels would be Provinces and sub-provincial areas designated as Origin and Destination Zones.

The secondary objective of the CTS 71 was to develop suitable methodology and concepts in this relatively new context and to lay the foundation for future systematic development of travel and tourism statistics.

The objective of this paper is to present a comprehensive description of the CTS 71. A brief discussion of alternatives in arriving at the concepts for the subject matter and the sample design is included to provide a general background to the actual decisions. The description of various sampling stages, use of panels, origin-destination zones, etc. is dealt with in separate sections, in order to emphasize their particular functions in fulfilling the objectives of the CTS 71. The details are excluded to facilitate a conceptual understanding of the

methodology. However, they are contained in the Procedures and Interviewers Manuals.

3. CONCEPTS, DEFINITIONS AND CHARACTERISTICS

Travel is often described as a discretionary mobility of individual(s) in time and space. In this process, the individual(s) may or may not incur expenditure, prior to this departure (for the purposes of his departure), in transit and at his destination, should there be one. In the broadest sense, then travel has four basic dimensions, namely, individual, time, distance and currency. Each ingredient in itself can be described in further components. It therefore follows that if all components were measured to the last detail we would indeed have a very comprehensive knowledge about travel. In practice, however, this is seldom practicable. The demands of the policy makers, having generated the need for data with some general statements about objectives are rather elaborate, and this is where conceptual and definitional problems arise. The definitions have to be operationally feasible and the characteristics have to be readily identifiable. This conflict of interest, or perhaps the conceptual differences among the policy makers, economists and statisticians, is by no means unique to the CTS 71. It is in fact a perennial conflict between the policy makers, who are primarily concerned with the immediacy for information, having generated the need for data, and the economists and statisticians who are primarily concerned with the lucidity of concepts, feasibility of definitions and the meaningfulness of the data (particularly in the case of a new series) in relation to its analytical plans, and of course, ultimate objectives.

The concepts, definitions and characteristics that were employed in the design of the CTS 71 methodology evolved as a result of a series of discussions with the knowledgeable persons in the industry, government and the academic world in Canada and elsewhere, and experiment of these various concepts in the field before they were selected for the CTS 71. The Travel Research Planning Committee of the Federal-Provincial Conference on Tourism came out with "Standard Definitions and Classifications of Travellers and Traveller Accommodations" in October, 1971.

Three concepts that presented the most controversy were trip, stop and destination. The geographic diversity of Canada made the task of giving any uniformity to these concepts even more difficult. It was clearly understood that local travel was not important for this study, however, certain cities being larger in population and area than some provinces, there seemed to be a pull in opposite directions. Similarly, it was agreed that with certain exceptions, "stops" have a definite economic impact. It was argued that if there exists a reason for a stop, then efforts may be made to prolong the stop and thus enhance the regional equity. A destination on the other hand can be associated with specific areas in a large number of cases, but in certain isntances there in fact is no specific destination. The question was of a realistic designation of a destination to all such trips in such a way that it will be both practically feasible and analytically meaningful.

The determination of characteristic of the trips and population was another area where very little development had indeed taken place. It was obvious that in order to measure the demand side of the travel industry, characteristics should be chosen that would provide a meaningful measure.

It is thus that we arrive at our first constraint, namely, the extent of demand. Pragmatically speaking, for an undefined and ill described industry we should be seeking knowledge about the extremeties of the "industry" as we visualize it to be before seeking its various elements. The planners on the other hand, having had nothing to go on for so long, were understandably impatient with a methodical and thorough investigation which, to them, will be another period of uncertainty or indecision. However, at the federal level, accommodation of such desires from both public and private sector means trade-offs between highly competitive alternatives. The definition of trip was to embody a compromise. It is very difficult to arrive at the definition of trip without excluding certain type of trips. For example, an overnight trip to a nearby summer cottage, say 150 miles away. However, adopting a distance criteria may exclude the former and include the latter. On the other hand, a trip may be defined by time, e.g., all overnight trips. In this case, however,

we will be able to get both of the aforementioned trips, if, and only if, a night was spent at the cottage, otherwise both will be excluded, along with all other same day excursions, sight-seeing and business trips. One may at this point quite reasonably ask why not use both elements of the trip concept in defining it, e.g., say 100 miles away from home and/or overnight. This definition looks extremely good on the surface, since it will only exclude the trip to the nearby cottage if no night was spent there. However, it may include all other overnight lodgings that may have little relevance to travel.

Of the above three approaches, the distance criteria offers a greater merit, in as much as shorter trips would not be reported together with all non-travel related trips. The next question then is the distance criteria that should be nationally applicable. The political geography of Canada is quite challenging in this respect. The boundaries of Prince Edward Island and Ontario being what they are, how does one determine a minimum equitable trip distance? A '100 mile away from home' trip definition in P.E.I. would include much of the travel outside the province; a 25 mile trip definition in Ontario, on the other hand, may be a shopping trip in a place like Toronto. In this case P.E.I. would have virtually no data on inter-provincial travel, while Ontario may have a lot of data on inter-city travel, neither of which would be very satisfactory.

The optimal strategy would appear to employ a combination of time and distance criteria, such that the inclusion of less relevant trips and the exclusion of relevant trips is minimized. This strategy is then applied to all characteristics. It would be rather lengthy to write the deliberations of all the important characteristics. The definitions that were developed are given in Appendix I for reference.

4. SAMPLE DESIGN

The basic design is a stratified multistage replicated sample of households. The selected households were interviewed on four occasions during the one year period. The sampling frame excludes institutions, military

establishments, and other special areas such as hospitals, nursing homes, etc. Each province and territory is considered independently in the design, so as to allow for provincial estimates.

The sample size was approximately 13,000 households from a population of some 5.5 million households in 1971, a sample of 0.24 percent households. The allocation scheme excluded Prince Edward Island, Northwest and Yukon Territories, wherein the sample size was increased to allow for an estimated minimum reliability for the number of trips. The sample size was determined by the budgetary constraints.

The sample sizes by provinces and territories were as follows:

Newfoundland	288
Prince Edward Island	144
Nova Scotia	468
New brunswick	360
Quebec	3546
Ontario	4608
Manitoba	648
Saskatchewan	576
Alberta	936
British Columbia	1224
Yukon Territory	175
Northwest Territories	175
Canada Total	13148

A more detailed description of the design is in the following sections.

5. STRATIFICATION

The Canadian population is relatively speaking, densely populated in few urban areas, while the rest is scattered in rural areas. A random sample of households under such conditions is not possible due to two reasons. One, there does not exist an "Address Register" of all

dwellings/households that can be used as a sampling frame, and two, the characteristics of population in all areas of a province are not necessarily homogeneous. It was therefore deemed desirable to consider a stratified, multistage design. The stratified form allows us to group together those areas of population that are homogeneous with respect to the characteristics that are related to the objectives of the study. The total area can then be divided into a given number of sub-sets that are initially exclusive and exhaustive, and an independent sampling scheme may be devised for each stratum.

In the CTS 71, therefore, stratification was one of the design features. The stratification variables were not all readily available. The stratification variables should be related to the travel habits of the population, and/or in part, the populations propensity to travel. Whatever the size of the stratum, a fairly large portion of the population in that region should have this characteristic. The travel as discussed in the preceding section may be for any reason associated with business or leisure. It may then be assumed that the population in the vicinity of business locations may be prone to more frequent business travel than away from it. This is not to say that population away from business areas would not travel for business purposes, but their travel would be less frequent.

On the basis of the above considerations, each province was stratified into two types of areas: Large Urban and Non-large Urban. This was the first stage of stratification, and therefore a given province now had one or morelarge Urban and Non-large Urban strata. The large areas comprising LUs and NLUs were further stratified into a number of other strata of a more manageable number and size for practical as well as design reasons. The practical reasons were, of course, the cost of various field operations, whereas the design reasons were the variances within and between strata and stages of sampling. Again, due to non-availability of information on travel related characteristics, the development of a cost and variance model was not possible. The factors that were considered instead were geography, population density, cost of listing, enumeration, etc.

A number of strata were constructed in each province comprised of contiguous Enumeration Areas. An EA is a Census Enumerator's Assignment and consists of approximately 150-250 households. The geographical boundaries of an EA respects all other physical and political boundaries. The number of EAs per stratum and the size of each stratum was decided on the basis of factors such as the overall sample allocation for NLU and LU areas and the expected number of households per EA.

The demographic data available at the time of the design was from the 1966 Census. The change in the growth and structure of the population by 1971 was inevitable. The exact magnitude of the change could not be measured at the time. The strata were therefore constructed of equal sizes, and the sample size for each stratum was approximately equal.

Large Urban Areas

The LUs were urban centres such as St. John's in Newfoundland, Regina in Saskatchewan, etc. A complete list of all LUs is given in Table 2. Within each LU there was sufficient heterogeneity in the socio-economic structure of the population to warrant further stratification. Guidance was derived from Census of Population designated areas such as Census Tract which are areas of socio-economic uniformity. Large Urban areas were further stratified to account for economic variations that may be pertinent to the area. Income and size of the family were two criteria that were used for this purpose. In certain LUs the second stage of stratification entailed designating areas of very high income. The contiguous Census Tracts with average household income of $8,000 or more were grouped to construct these strata. This stratification was carried out in those areas where such areas were large enough in proportion to the rest of the area to permit creation of such stratum, otherwise such desparities in population characteristics were ignored. The areas were instead stratified into strata of equal size of contiguous Census Tracts that registered differences. For example, a city such as Montreal had 21 strata in all, 2 of very high income households, 2 of Apartment Buildings and the 17 were all other households areas.

Non-Large Urban Areas

The NLU areas, were the areas other than the LU areas. The stratification of NLU areas was done on the basis of two factors, cost and similarity of characteristics in the contiguous EAs in the population. The basic unit was an EA which was also the first stage sampling unit. The number of strata for each province was arrived at by considering the sample size, expected number of EAs per stratum and the enumerator assignment.

6. ORIGIN-STOP-DESTINATION ZONES

The Origin-Stop-Destination Zones are sub-provincial areas or more specifically a grouping of strata for which tabulation was required. Therefore the delineation of the Origin-Stop-Destination Zones (O-D) zones was carried out in conjunction with the stratification. It may not appear to be a design problem, however, from the tabulation requirements that were made at the time for the various geographic levels, the O-D zones were indeed to conform with the strata boundaries. As noted in the discussion on concepts and definitions, a trip includes areas where stops are made and one of those stops may be the destination of the trip. It is where the interview took place. The stops, including the destination may be anywhere. Clearly some level of aggregation was warranted here. The Tourist Industry and Provincial governments of course had Tourist Regions in mind. These Tourist Regions were quite numerous and furthermore, their boundaries were not particularly related to any other boundaries. Considering the sample size and the assumption of average number of trips by a household, a maximum of 35 O-D zones was agreed upon. The 35 zones were arrived at by considering factors such as population homogeneity and density, geographic contiguity, tourist attraction areas as defined by provinces and the overall conformity of these areas to a number of strata. This allowed for the summation of characteristics over a given number of strata that correspond to an O-D zone.

The above factors were not necessarily applied to each province, but any one or more of them were used as a criteria. For example in British

Columbia, Vancouver (p. 892,286) and Victoria (p. 300,000) is collapsed to one O-D due to their highly urbanized nature in relation to the rest of the province. The northern parts form another zone due to the population engagement in ranching or forest industry. Aimilarly, that portion of British Columbia known as the Fraser Valley (excl. Victoria) were combined to form an O-D zone as the main industries in these parts are dairy farming, market farming and fishing.

For each province, distinction between LU and NLU has been maintained whenever possible. For example in Ontario, Windsor, London, Kitchener-Waterloo areas (p. 612,368) are combined to form an O-D zone while Toronto (p. 2,158,496), Ottawa (p. 384,397) and Sudbury (p. 117,075) each form O-D zones by themselves.

In the above design of O-D areas, each area by definition is an origin, stop or destination with the exception of Rocky Mountain areas which are Stop/Destination areas, only. The trips originating from this area are combined with the usual O-D zone. Similarly Hull, Quebec, is a special case so that statistics on the National Capital Region may be produced. The distribution of Origin-Stop and Destination areas by province was as follows:

Newfoundland	2	
Prince Edward Island	1	
Nova Scotia	2	
New Brunswick	2	
Quebec	7	
Ontario	9	
Manitoba	2	
Saskatchewan	2	
Alberta	3	(2 O-D plus 1 Destination only)
British Columbia	5	
Canada	35	

7. SAMPLE SELECTION METHODS

The basic design was the multistage stratified replicated sample. The multistage design, as the name implies, is the division of a population in groups that are in a descending order of density. A city, for example, may be considered in two groups, i.e. localities and/or blocks within localities. A sample may then be selected of localities and blocks, rather than of blocks that may well be all over the city or clustered in one area.

In the NLU areas, the first stage unit was an EA, the second stage was the household, and the third and final stage was a person within the household. In LU areas, the census tracts were the first stage, blocks the second, household the third, and the person the fourth. Within LU areas, wherever the apartment building necessitated a stratum, the apartment building was the first stage, household the second, and the person the third and final stage.

First Stage Units

The first stage units were selected with probability proportional to size in all strata. In the LU non-apartment strata, two census tracts were selected with PPS (Census 1966), whereas in apartment strata, the total number of apartments to be selected in each stratum were divided into two replicates of equal sizes. The measure of size in this case was the number of apartments in each apartment building. Each of the two replicates were then selected independently with PPS. In the NLU, two equal number of replicates of EAs were selected with PPS. The measure of size was the population as of Census 1966. There were some strata where the number of EAs to be selected were not divisible by 2. In those cases, two strata were used to balance the number of EAs to be selected. For example, if two strata required the selection of 18 EAs or 9 EAs each, the 10 EAs were selected from one stratum and 8 EAs from the other, providing us with two replicates of equal sizes each, 4 and 5 in this case.

Second Stage Units

In the LUs where the second stage units were the city blocks, the selection procedure was the same as described above for the selection of first stage units for EAs and apartment buildings. That is, two independent replicates of 8 blocks were selected with PPS from each stratum. At this stage, the sample of city blocks, EAs and apartment buildings were delineated on the maps for the listing of dwellings in each one of them. All such maps were then sent to the appropriate Regional Offices for that purpose. The listing operation provided the more recent count of dwellings in each selected area. The sampling ratio were then calculated based on the recent listing and the expected number of households to be selected for that area. In each case a systematic random sample of households was selected.

Third Stage Units

The third stage unit was the person within a household, except in non-apartment LU strata, where this would be the fourth stage unit. The procedure in each case was the same. One person, 14 years of age and over was selected randomly from each household. The list of selected households was provided to each enumerator, prior to the date for the first interview. The household information was designed to be recorded on a household card. The household composition was required to be listed in two sections. One section required the listing of all household members that were aged 14 years or more as of the date of interview, in the descending order of age. Each entry on this section, therefore, was numbered from 1 onwards. A table of random numbers was used to determine which number, i.e. person is then selected. The person thus selected would then be interviewed for the trips that he took with or without other members of the household. The other section of the household information card would list all persons under 14 years of age in the household.

The sample design can thus be summarized as follows:

Table 1: Sample Design

Stage of Sampling \ Stratum	Large Urban Areas			Non-Large Urban Areas
	Apartment Buildings	Residential Homes	Residential Homes (Very high income)	Congiguous EAs
First Stage	Apartment Building	Census Tract	Census Tract	EA
Second Stage	Apartments	City blocks	City blocks	Households
Third Stage	Persons	Households	Households	Persons
Fourth Stage	None	Persons	Persons	None

8. INTERVIEW METHODS

The CTS 71 was designed to collect data on travel for the calendar year 1971. In the pilot study prior to the main CTS in 1971 it was determined that depending on the method of aid to recall, the optimum recall period lies somewhere between 12 and 16 weeks. The interview scheme, therefore had three basic features, personal interview, staggered over a year so as to be approximately 12 to 16 weeks apart and an aid to recall left with the respondent. The aid to recall selected for the CTS 71 was a pocket diary, designed so that essential information about the travel may be noted. The essentials of a trip thus recorded, it would then be not only more convenient to convey such trip information to the enumerator at the time of the interview but also less susceptible to loss of information. In order to stagger the length of recall, the entire sample size was randomly divided into three "panels" of equal sizes. The time between the interviews of each panel was distributed so as to allows different lengths of

recall for each of the 4 rounds of interview. The interviews were scheduled for the spring, summer and fall of 1971 and early (January) 1972. The interview scheme was as follows:

Table 2: Interview Scheme

Round	Panel	Interview Week
First	1	March 8-12
	2	March 29-April 2
	3	April 26-April 30
Second	1	June 7-11
	2	July 5-9
	3	August 2-6
Third	1	September 6-10
	2	October 4-8
	3	November 1-5
Fourth	1	January 10-14
	2	January 17-21
	3	January 24-28

The concept of bounded recall was introduced by way of New Years Day. Plans called for leaving the diary with the respondent at the beginning of the year. The operational problems, however, did not allow this to happen until the second round of interviews. An additional week was allowed to follow up the cases wherever the respondent was temporarily absent, but could be reached a few days later.

9. ESTIMATION

The method of estimation that follows is a straightforward application of multistage stratified sampling. The estimates incorporate the adjustment of weights in two steps. Firstly, the weights are adjusted at the

stratum level to account for non-response. The assumption here is that the characteristics of the respondent are similar to those of non-respondents. Secondly, the provincial population is estimated from the sample and then compared with the known population for 1971. The weights are then adjusted by the ratio of the known population to the estimated. The weights thus adjusted are used to estimate characteristics and the estimate of the variance of the estimated characteristics as shown in the following development.

A - Notations

W = Weight
a = Census Tract
b = Block, Apartment Building in Enumeration Area
c = Household
d = Person
t = Trip
e = Province
f = Stratum

In each stratum two replicates of equal sizes were selected. Estimate of a characteristic for a stratum is obtained by taking a mean of the estimates from the replicates. In the estimation procedure described below, the notation for replicates and various steps leading to the estimates are omitted for the sake of brevity.

B - Weights

In accordance with the design, the weight for each household is defined as the product of weights of different sampling stages. In order to better understanding of the estimation procedure, reference should be made to section 7 where selection methods are described:

i) In the Large Urban areas, then the weight for a household in High and Moderate income sub strata is defined as:

$$W_{ef} = W_a \cdot W_b \cdot W_c \cdot W_t$$

where

W_t = $\frac{\text{Total number of persons in the household}}{\text{Total number of persons on the trip}}$

W_d = Total number of persons in the household over 14 years of age.

W_c = Inverse of the probability of selection of (the interviewed) households within a block.

W_b = Inverse of the probability of selection of blocks within a Census Tract.

W_a = Inverse of the probability of selection of Census Tracts within a stratum.

The weights for the Apartment Building sub-strata are defined in much the same way as above except there are only three stages we therefore have

$$W_{ef} = W_b \cdot W_c \cdot W_t$$

where W_b = Inverse of the probability of the selection of Apartment Buildings.

ii) In the Non-Large Urban areas the weights are defined as

$$W_{ef} = W_b \cdot W_c \cdot W_t$$

where W_b = Inverse of the probability of the selection of Enumeration Areas.

C - Adjustment of Weights

The weights are adjusted for two factors: first for non-response and second for the "known" population totals.

The adjustment for non-response is made by assuming that the interviewed households have characteristics similar to those that cannot be interviewed. The adjustment is made at the "block" level by the following method:

Let W'_c = Inverse of the probability of the selection of households

s = Number of households selected in a block

s' = Number of households interviewed in a block

s^* = Number of households vacant in a block.

We then define W_c, the adjusted household weight

$$W_c = W'_c \left(\frac{s - s^*}{s'} \right)$$

The adjustment for "known" population totals is made at the provincial level. The adjustment factor is the ratio of the Census population projection and the estimate of population from the CTS.

Let P_e = The Population Projection for 1971 from the Census for province e

$\hat{P}_e$ = The estimate of population from the CTS 71 for province e

γ_e = The Adjustment factor for "known" population for the province e

where

$$\gamma_e = \frac{P_e}{\hat{P}_e}$$

Then W^*_{ef}, the weight adjusted for "known" population total is given by

$$W^*_{ef} = W_{ef} \cdot \gamma_e$$

D - Estimates

The estimate $\hat{X}_e$ of a characteristic is therefore defined as

$$\hat{X} = \sum_f W^*_{ef} x_{ef}$$

$$= \frac{1}{2} \left(\sum_f {}^1W^*_{ef} {}^1x_{ef} + \sum_f {}^2W^*_{ef} {}^2x_{ef} \right)$$

where the superscripts 1 and 2 denote the replicate 1 and 2 respectively.

The estimate of variance, $\hat{V}$ of $\hat{X}$ is obtained as follows $\hat{V}(\hat{X}) = \frac{1}{4} ({}^1\hat{X}_{ef} - {}^2\hat{X}_{ef})^2$

10. ACKNOWLEDGEMENT

Acknowledgement is due to Mr. R. Platek and Mr. P.F. Timmons who directed the development and implementation of the methodology of the Canadian Travel Survey 1971, to Dr. M.P. Singh for his invaluable suggestions in improving the quality of this manuscript and finally to the referee for a very thorough and constructive criticism.

Appendix I - Definitions

Trip: 1) Travel to a plate 100 miles or more away from home excluding --
2) Travel to a place 25 miles or more away from community boundary

Destination: A place farthest away from the origin of the trip.

Stop: A place between origin and destination of a trip where either a night was spent or an expenditure was made of $1.50 or more, per person on the trip.

Time Period: Calendar year 1971. This included all the trips terminating in the year 1971.

Distance: Anywhere.

Expenditure: All expenditure in relation to the trip that would not have been incurred had the trip not taken place. This included expenditure at origin in preparation of trip, on the trip at stops and destination, on transportation, accommodation, sightseeing, entertainment, gifts, local transportation, etc.

Population: Canadian residents in the calendar year 1971 excluding --

Population Characteristics: Age, sex, marital status, income, occupation, education, type of employer, automobile ownership, etc.

RESUME

L'enquête sur les voyages des Canadiens pour 1971 a été la plus importante enquête effectuée dans ce domaine. Le présent exposé décrit quelques-uns des principaux aspects de la méthodologie. On insiste plus particulièrement sur l'élaboration de définition en fonction de la méthodologie, la méthode d'échantillonnage et les techniques d'interview.