Continuous Sampling Method of Conducting Origin-Destination Surveys

WARREN B. LOVEJOY, The Port of New York Authority

The Port of New York Authority is using a continuous sampling method in conducting origin-and-destination surveys at its bridges and tunnels in the New York Metropolitan area. This paper is a discussion of the results of the first year's experience in the actual operation of this new type of survey.

The paper briefly summarizes the statistical rationale underlying the continuous sampling method, describes the various steps necessary in planning a survey of this type, the operating procedures which have been developed for the field work, and the new methods employed in processing the data. The various advantages which this type of survey provides over previous "one-shot" surveys conducted by the Authority are discussed, and an appraisal is provided of the results of the first year of the survey operation in the light of these expected advantages.

●REASONABLY ACCURATE and up-to-date origin-and-destination information is one of the primary tools used in the planning of vehicular facilities. Since 1935 the Port of New York Authority, in connection with its planning for future vehicular facilities across the Hudson River in the Metropolitan New Jersey-New York area, has conducted periodic roadside surveys of origins and destinations of vehicles using its tunnels and bridges. In the past, because of the expense of conducting these surveys, reliance has had to be placed on so-called "one-shot" surveys, taken approximately every three years, in which one to three days of traffic have been surveyed as representative of the entire year's traffic.

Although this method of conducting surveys has given over the years a substantial body of origin-destination information, the method has some serious drawbacks. The surveys have generally been conducted in October because observation has seemed to indicate that October traffic patterns are fairly representative of a yearly average traffic pattern. Actually, although volumewise this is true, it may well be that because of seasonal variations October is not typical of an annual traffic pattern at all. Furthermore, much of the work concerns peak-hour analysis and inasmuch as the peaks are usually summer peaks, perhaps from this standpoint it would have been better to catalog the traffic flow in the summertime rather than an average annual flow, which it was thought was being obtained in October.

Another difficulty was that the day, or two, or three days, on which the survey was conducted might for some reason or another have been quite unrepresentative of the rest of the days of the year. For instance, on one Sunday there was an unusual amount of traffic going between New Jersey points and the Yankee Stadium, and a later check showed that there was a football game at the Yankee Stadium that day which was of particular interest to New Jersey residents. Certainly the information obtained on that day was not typical of all of the Sundays of the year. Similarly, there have been severe weather conditions on some of the survey days, and there is little reason to doubt that traffic flows are different in rainy weather than they are in good weather. Therefore, the pattern obtained in the survey taken on a rainy day may have been quite unrepresentative of the average day in the year.

Another difficulty in this so-called "one-shot" survey was the job of hiring and training the 200 to 300 interviewers required to do the job. No matter how carefully this problem was approached, there were numbers of men involved who had very little aptitude, training, or inclination for this type of work. As a result, the accuracy of the results secured were open to considerable question. This is true not only of the interview information obtained in the field, but also of the accuracy of the coding of the results, as selected numbers of these men also had to be used in coding.

Continuous Sampling Technique Provides Numerous Advantages

Recently, a new sampling technique called "continuous" sampling has been developed by statisticians and market researchers in other fields to a point where its use in the traffic field, and more specifically in O-D surveys, has become practical. Briefly, this new technique is based on a carefully designed and controlled probability sample which builds up the interviews obtained to the required number for any desired degree of reliability by sampling over a considerable length of time rather than by sampling heavily in a short period of several days. Thus, in the continuous survey, by obtaining a few hundred interviews each day a substantial number of interviews has been accumulated over the course of a year and these have been obtained under all of the varying conditions which exist in the field throughout the year.

There are a number of distinct advantages to be derived from use of this new technique. Among the more important of these, as they apply to the Authority origin-destination survey problem, are the following:

1. By spreading the sampling over a long period, this system avoids the great danger inherent in the previous "one-shot" surveys that seasonal variations or unpredictable variations of one kind or another might make the day or the few days surveyed quite unrepresentative of an average day or days. In continuous sampling, enough days and seasons of the year are covered so that it can be said with certainty that the survey results are representative of normal activity. By the same token, it is possible, by this new technique to measure seasonal variations in traffic patterns, about which there presently is no real information.
2. This new technique will provide up-to-date information on origin-destination patterns at all times. By summarizing the results of the survey quarterly, semi-annually, or annually, information on changes in the O-D patterns caused by the construction of new vehicular facilities in the metropolitan area, or by developments in other modes of transportation, will be provided almost as soon as the changes occur. This will enable isolation of the effect on the Authority's own facilities of each of these changes, thus greatly improving knowledge of traffic movements and ability to forecast the size and timing of future diversions. Current data will be available at all times for incorporation into studies and forecasts that are required almost continually in the course of the working year in connection with various planning studies that are undertaken by the Authority.
3. Building the size of the sample over a long period, rather than by intensive sampling during one day or several days, eliminates the necessity for hiring large numbers of unskilled, uninterested, temporary employees as interviewers and coders. Rather, use is now being made of a group of four interviewers who continuously sample throughout the year. These men are Port Authority career employees who are well trained in interviewing methods, have developed a thorough knowledge of the geography of the metropolitan area, and have become skilled coders of the interview data, thus producing more reliable and meaningful survey results than have been obtained in the past.
4. By spreading the sampling period over the course of a year, the office work load of coding, verifying, tabulating, etc., which used to be tremendous and require the hiring of numerous temporary personnel under the old "one-shot" system, is now spread evenly throughout the year, thus eliminating the uneven periods of activity of the past.
5. Because continuous sampling avoids intensive sampling at any facility at any one time, there is less chance of causing congestion at the facilities while obtaining the O-D information. Furthermore, skilled interviewers greatly reduce the time required for each interview. These features of the new type survey have appealed greatly to the operating personnel at the bridges and tunnels and have enabled getting more complete information on each interview than used to be possible under the old method, when it was often necessary to concentrate simply on the answers to one or two questions from each motorist in order to keep the traffic flowing.

DEVELOPMENT OF STATISTICAL SAMPLE DESIGN

Having decided that the advantages possible with this new survey technique were well

worth investigating, the aid of S. T. Hitchcock, of the U.S. Bureau of Public Roads, and Leslie Kish, of the Survey Research Center, University of Michigan, was enlisted in developing a statistical sample design. Before any work could be done in the development of this design, certain preliminary information had to be provided. One of the advantages of a probability sample is that by adhering to the mathematical rules of probability, it is possible to estimate with considerable accuracy the precision that the survey results will attain. Before the sample size required and the detail of the construction of the sample could be determined, therefore, it was necessary for the Port Authority to decide what level of precision was required for its planning purposes. It was determined that it would be satisfactory to have a level of precision which would enable saying that 95 times out of 100 if the survey found that 10 percent of the vehicles had a certain origin-and-destination pattern, the actual percentage of vehicles with this origin-and-destination pattern would lie somewhere between 9 and 11 percent.

The second type of information to be provided before the sample could be designed concerned the characteristics of the population being measured. Fortunately, because all of the facilities involved are toll facilities, there exist very accurate measurements, volumewise, of the characteristics of the traffic over each of the facilities, thus, it was possible to provide the Bureau of Public Roads and Dr. Kish with data which showed seasonal, daily, and hourly variations in traffic volumes on each of the tunnels and bridges operated. In fact, for the three trans-Hudson facilities this same type of information was available for each toll lane. Information also had been developed which showed, for each of the facilities, the expected number of lanes that would be open at any hour of the day in which the interviewers would be there. Finally, there were forecasts of future annual, monthly, daily and hourly traffic volumes which could be expected at each of the facilities.

Inasmuch as statistical formulas are generally of interest only to statisticians, no attempt is made to specify the various statistical formulas which can be used to describe the sample design used. The design, however, can be described in more general terms. First, it is a self-weighting probability sample. That is, the sample is so designed that the number of interviews obtained is proportionate to the volume of vehicles moving through the facility at the time the interviews are being taken. Thus, there is no need to expand or weight the numbers of interviews in order to make them proportionate to traffic volumes. Although this feature of the design makes for somewhat more difficult and rigorous operation, it was decided to use it because of the great savings in time which could be achieved by eliminating the necessity for weighting of the interviews by hour, by lane, by direction, etc.

The sample can be described as a stratified multi-stage probability sample. This means that the entire amount of traffic is divided into several strata, which are felt to be as homogeneous as they can be made. Thus, the traffic has been divided into weekday, Saturday, and Sunday traffic, and within each of these day types has been split into 8-hr shifts running from 11 P. M. to 7 A. M., 7 A. M. to 3 P. M., and 3 P. M. to 11 P. M. These hourly groupings were selected after studying the hourly volume records that were available and analyzing the traffic in order to identify similar types of travel.

Perhaps the meaning of the word "multi-stage" can be explained by briefly citing the various stages through which the final selection of a vehicle to interview is carried. The first stage is selection of a shift at a facility in which, for instance, it is selected at random that on a certain day an interviewer will be at the George Washington Bridge on the 7 A. M. to 3 P. M. shift.

The next stage is selection of a location at the facility where the man will start his interviewing. This is necessary because there are several toll plaza locations at this facility and the man can not rotate from one location to another readily. Suppose, picking at random, it is determined that the interviewer will start at the westbound Main Plaza. This then is the second stage of selection.

There are a number of toll lanes at this location so that for the third stage of selection another random selection must be made as to the lane at which the interviewer will start to operate. The interviewer rotates from lane to lane each hour on a definite prearranged pattern, but selection of the lane in which he will start is based on a random

procedure so that the laws of probability can be applied.

The fourth and final stage is selection of the specific car in the lane in which the interviewer is located. This is done by specifying for that lane, for that hour, for that location, and for that facility, the interviewing intensity the interviewer must attain. By interviewing intensity is meant that he will take every fourth car, or every sixth car, or whatever is indicated by the procedure.

At each of the four stages a random procedure is introduced so that the probability feature of the sample is maintained through each one of the stages. Furthermore, a known probability of selection is calculated for each stage of selection, and it is determined that the product of the probabilities of each one of the four stages will give the over-all probability it has been determined is necessary to obtain a sample of the required size by the end of the year.

PLANNING FOR OPERATION OF THE SURVEY

Once the sample design had been developed the next phase in planning the survey was to determine the various operating procedures and methods to be followed. A great deal of field testing and checking was done in order to arrive at the necessary methodology; as a matter of fact, the procedures are still being modified from time to time according to lessons learned from the day-to-day conduct of the survey. First it was necessary to determine how much interviewing time could be expected over long periods from each interviewer without loss of efficiency due to fatigue or lack of interest. After a good deal of experimentation it was decided to have the interviewers interviewing 44 min out of each hour during the 8-hr period of the shift. In each hour, 16 min were to be taken as a relief period and also as the time to rotate from one lane to another. In addition, the men are given sufficient time for their lunch period. It was found that a schedule of this type could be maintained by interviewers throughout the entire 8-hr period day after day without undue fatigue.

The second step was determination of the number of interviewers who would be needed for the conduct of the survey. This, or course, is proportional to the number of interviews required and the interviewing speed possible. Again, after a good deal of field work, it was determined that an average rate of 40 interviews an hour was possible. This does not mean that in very busy periods more than 40 interviews can not be secured. (As a matter of fact, close to 100 interviews have been obtained from one interviewer in an hour.) It does mean, however, that it was felt that throughout the conduct of the survey, on the average, a rate of about 40 interviews per hour was achievable. It was also felt that because of the repetitious nature of the interviewing job and the often difficult physical conditions surrounding it, much more efficiency would be obtained from the men if they also were used in the office for coding the material obtained from the interviews in the field. It was determined that in the annual distribution of the work approximately 60 percent of an interviewer's working days would be in the field and 40 percent would be in the office coding the interview material. Thus, four interviewers could obtain the requisite number of interviews and, at the same time, very nearly carry the entire load of coding the interviews.

Based on the sample design (which, incidentally, is developed in 13-week groups and therefore totals four seasonal groups during the year), working schedules had to be developed for both the interviewing time and office time of the interviewers. Interviews had to be obtained at all six of the Port Authority facilities at all hours and on all types of days. Therefore, it is necessary for the men to work around the clock, to work weekends as well as weekdays, and to rotate from one facility to another from day to day. As far as possible in working up these schedules an attempt was made to insure that the men would work five days and be off two regularly. In some cases these exact results could not be achieved, but over the 13-week basic period each man had the same number of working and off days. An endeavor also was made to have the men work the same hours for an entire week. Controlling on time this way, it was necessary to send the men to a different facility almost every day; therefore, use of a vehicle by each man was an absolute requisite for the job and the men receive extra compensation for the use of their private cars.

Field experience showed that uniformed interviewers achieve a better response from the public than non-uniformed men, so uniforms were provided for each of the interviewers. These uniforms consist of both summer-weight and winter-weight trousers, shirts, tunics, overcoats, and foul-weather gear. A complete set of the uniforms is provided each man to start, and he is given a specified annual allowance for maintenance and replacement of uniforms.

Because most of the equipment required for the successful conduct of this survey also is required in "one-shot" surveys as well, it is not covered here. It is sufficient to say that the interviewing sheets used are specially designed for ease in recording the information and also for ease in coding and then key-punching the information into IBM cards for processing later. Survey forms are shown in the Appendix. A good deal of safety equipment (cones, lights, etc.) has been assembled to safeguard the men when they are working on the plazas.

Finally, in hiring the interviewers success was achieved in obtaining men by promotion directly within the Port Authority staff. Thus, the interviewers are career employees who are interested in doing a good job in furthering their Port Authority careers. The salary classification of these men was set high enough to attract the kind of personnel needed, in spite of the known physical difficulties encountered by the men in periods of bad weather and the inconvenience of the shift work which the job entails. The men have to be responsible, as they are working on their own with only the over-all supervision that can be given from the office and by occasional spot checks. They also have to be in good shape physically and be able to stand a good deal of monotony without losing sight of the aim of collecting and coding accurate data. The performance of these men to date has been a revelation.

SURVEY EXPERIENCE TO DATE

How has the survey actually worked out? Operations were started on December 30, 1957, and so there has been a period of testing under all weather conditions and traffic conditions expected to be encountered throughout the survey period.

From both an operational and a statistical point of view the survey is turning out to be much as anticipated in the planning. The number of interviews secured during the first eleven months of 1958 was somewhat less than had been anticipated, primarily because the traffic itself was less than had been forecast. Even with the relatively slow start for the first six months of the year, however, it is expected to total within 6 or 7 percent of the 100,000 interviews it was planned to obtain during the first year of operation of the survey. A response rate of approximately 99 percent has been achieved. Of the 1 percent of refusals encountered, very few have engendered unpleasant situations, most of the refusals being because of lack of time.

In aiming at an average interviewing rate of 40 interviews per hour, different interviewing intensities had to be assigned for each hour in the day depending on anticipated hourly traffic volumes. The interviewers have been able to maintain the assigned rates virtually 100 percent of the time without causing traffic congestion. Furthermore, despite an extremely severe winter and very rainy conditions for most of the spring and summer, the survey has shown that although bad weather conditions certainly do not make life easy for the interviewers, they do not make it impossible to obtain the correct number of interviews.

The various precautions taken for the safety of the interviewers have proved to be effective. Working with the police forces at the bridges and tunnels in all cases, it has been found that strategic placing of traffic cones and the good lighting provided all of the plazas have been sufficient to protect the men under all conditions, even when the weather is bad at night. It is gratifying to say that there have been no injuries or "near misses" during the time that the survey has been in progress.

To determine whether the sample is really representative of the entire traffic flow, some careful auditing of the survey results has been done. Preliminary tabulations have been made for each of the first two 13-week periods in such a way that it has been possible to check survey coverage hourly by direction, by toll lane, by facility, by day of the week, and by license plates of the vehicles interviewed. As a result of this

auditing, it was discovered that in the first quarter there had been oversampling of one of the locations at the George Washington Bridge, with a resultant undersampling of the other locations. This situation has now been adjusted. Except for this one error, the audit has shown that the correct proportionate distribution of traffic hour-by-hour at each facility has been obtained in the sample. Check has also shown that the procedure for rotation between the lanes has yielded the correct number of interviews from each lane. It also has been found that the proper directional distribution of the traffic has been achieved, not only on a daily basis but also in peak hours versus non-peak hours. All in all, the results of this auditing have confirmed the representativeness of the sample.

Origin-destination patterns as revealed by these preliminary tabulations have been compared with the results of the "one-shot" survey of October 1956. The results of the first 13-week period (roughly January through March) are quite similar to the 1956 results. However, the second 13-week period (April through June) has evidenced seasonal variations of sufficient magnitude to warrant a thorough investigation of the extent to which origin-destination patterns vary from one season of the year to the next. Analysis of the third and fourth 13-week periods is now proceeding with this purpose in mind.

As far as coding of the interviews is concerned, originally an average of 200 or less interviews was coded per day by the interviewers when they were working in the office. That has now been increased to an average of 450 interviews a day. By a process of verification it has been found that the accuracy of the coding has been extremely high. The method of coding verification used is modeled along the lines of statistical quality control procedures and it has been determined that original expectations of accuracy have been more than met. Less than one error is found in each 125 interviews coded; most of the errors that have been found have been extremely minor in nature.

The cost of conducting the continuous survey for a year is only two-thirds that of a "one-shot" survey covering a weekday, a Saturday, and a Sunday. The survey is currently being carried on with four interviewers and the equivalent of one man-year of statistician time and one man-year of clerical time. In addition, of course, there is the cost of tabulating the material. Here, too, it is felt that savings will be achieved by the orderly scheduling of the processing work load throughout the year, which is possible with the continuous survey.

FUTURE PLANS

Because this is such a new type of traffic survey and because it has been in operation for a relatively short time, little actually is known about the method and what it can produce. Undoubtedly many improvements in efficiency, accuracy, and other aspects can and will be made based on the experience being gathered now. For instance, it already has been determined that by slightly changing the intensity of interviewing on certain shifts, the same amount of information as is now being obtained can be gathered with approximately 50 fewer field shifts each year. This will allow use of more of the interviewers' time in coding and in keeping the various records that must be maintained in order to produce the accuracy desired.

There are many other ways in which improvements may be possible in the future. Detailed analysis of the variances obtained in the sampling for different hours of the day, for different types of days, or at different facilities, may well lead to changing the sample design in such a way as to obtain more efficient use of the interviewers or perhaps to get along with fewer interviews and still achieve the same level of accuracy as sought from the start. It may be found unnecessary to interview as often during non-peak hours as has been done so far. Once there has been accumulated over the course of a year or two a sufficient body of information to give reliable annual and seasonal relationships, the information thereafter may possibly be kept current by a much lighter sampling than was required to build up the information in the first place. It may be found that seasonal variations are at a minimum at certain times of the year and that, having a more or less homogeneous population during these periods, they need be sampled only sparsely. Also, peak-hour information possibly should be

emphasized and there may, therefore, be need to change the survey from time to time in order to concentrate on peak periods rather than giving total over-all coverage such as is presently being obtained. In other words, the procedures and operating methods now being used really should be regarded as first steps in a new venture—steps that will eventually lead to a much more efficient way of accomplishing the desired end. It is intended to continue modifying the methods until it is felt that the maximum efficiency possible has been achieved.

POSSIBLE APPLICATION TO NON-TOLL FACILITIES

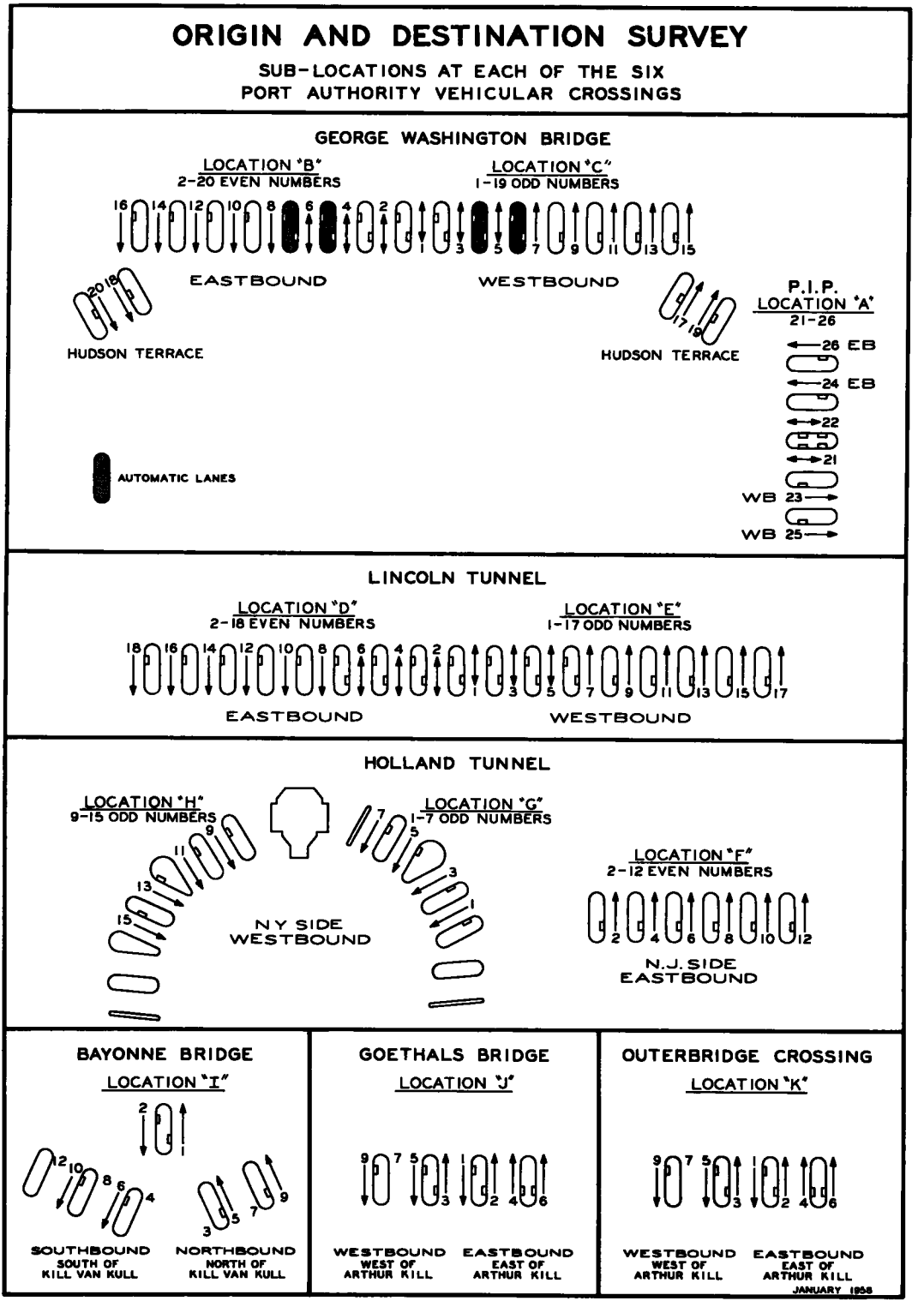
Because all of the Port Authority facilities are toll facilities, they are made to order for the conduct of a continuous sampling survey. The vehicles all must slow down and funnel into separate lines as they approach the toll booths. Because they must stop to pay their tolls anyway, the drivers do not object to stopping for the few additional seconds required to answer the survey questions.

This does not mean, however, that this technique cannot be expanded to apply to freeways and other types of vehicular facilities where the situation is not so ideally suited to this type of survey work. For instance, a continuous sample design could be worked out for origin-destination survey work on other types of highways, whether of the controlled access type or not. Just as the reported survey is now rotating from one facility to another, and from one location to another at a facility, and from one lane to another at each location, so, on a freeway, the interviewers could rotate from one entrance or exit to another in a planned and controlled manner so as to obtain all the necessary origin-and-destination information covering the total traffic on the freeway without the necessity for blocking traffic on the freeway itself.

This same method could be applied to origin-and-destination surveys taken along cordon lines or rings around metropolitan areas, where a small team of interviewers would rotate from one point of access to another in a prearranged and carefully controlled method, interviewing vehicles either on stop signals or in some other way. In this way, a good picture of the total origin-and-destination pattern of all of the vehicles crossing a cordon line could be obtained over a period without seriously disrupting traffic, as so often must be done in the more conventional type of roadside O-D survey.

As a matter of fact, this same continuous sampling method can be applied to other types of information gathering than origin-destination information surveys. A classified count of traffic entering and leaving a city or crossing a cordon can be achieved in the same way by rotating a small force of enumerators from one point to another on a controlled basis, thus building up over a period of time the information required as to the average movement of vehicles over the given line. Just as it is felt that the Port Authority survey work has just begun to scratch the surface of the usefulness of this tool, so it is felt that for many other applications this continuous sampling method can be used to great advantage, and that its realm of usefulness and its different possible applications will continually expand as users come to understand more and more of the implications of the method.

Appendix--Information and Field Forms



P ...192
10-58

ORIGIN AND DESTINATION SURVEY

FACILITY AND SHIFT ASSIGNMENT

WEEK NO. _____

WEEK BEGINNING _____

DAY TYPE	DATE	INTERVIEWER NO.	FACILITY	SHIFT	SURVEY START	SURVEY FINISH	SELEC. RATE 1:	RANDOM START	FACILITY LOCATION	LANE ROTATION BEGIN	LANE ROTATION SEQUENCE

PA 2190-A
12-57

ORIGIN AND DESTINATION SURVEY

GEORGE WASHINGTON BRIDGE

Page ☐

Fac.	Year	Month	Week	Day	Day Type	Shift	Hour	Dir.	Lane
0 1	5 8								

Type	Pass.	Lic.	Purpose		East of Hudson River Street Address Borough or Town	West of Hudson River City or Town
1. Auto 2. Truck 3. Tr-Trlr.	No. of people in vehicle	1. N.Y. 2. N.J. Others Abbrev.	1. Work 2. Shop 3. Recr. 4. Other			
					0	5
					0	5

LANES OPEN

No.	Time

Remarks:

Weather _____

Other _____

_____

_____
Interviewer

PA 2202
10-58

ACCURACY CONTROL RECORD FOR O/D SURVEY

Facility _____ 195____

SEAS. (1)	DAY TYPE (2)	WK. (3)	DATE (4)	SHIFT (5)	HR. (6)	LANE (7)	NAME OF INTER. (8)	TIME FACTOR (9)	NO. OF INTERVIEWS SEC. (10)	REF. (11)	NON-RES. (12)	TOTAL (13)	LANE FACTOR (14) Actual	Assum	(A)	LANE ADJ. (B) (15)	ADJ. NO. OF INTER. (16)= (13)− (15)	REMARKS (17)

PA 2201
10-58

SUMMARY OF SHIFTS COVERED - O/D SURVEY

Month_____ 19____

DATE	FACILITY	SHIFT	INTERVIEWER	NUMBER OF INTERVIEWS SECURED	REFUSALS	NON-RESPONSE	TOTAL	DATE CODED	CODER NO.	DATE VERIFIED	VERIFIER NO.

10-58

CODING VERIFICATION CONTROL RECORD FOR O/D SURVEY

VERIFIER	ORIG. CODER	DATE	FACILITY	SHIFT	POPULATION N	LOT n	TYPE OF ERROR EHR (AK), SKVK	REJECT (R) or ACCEPT (A)	WHR (AK), NKVK	REJECT (R) or ACCEPT (A)	LIC. PLATE