TRANSPORTATION RESEARCH RECORD 886

New Concepts in Data Analysis

TRANSPORTATION RESEARCH BOARD

NATIONAL RESEARCH COUNCIL

NATIONAL ACADEMY OF SCIENCES
WASHINGTON, D.C. 1982

Transportation Research Record 886
Price $7.40
Edited for TRB by Naomi Kassabian

modes
1 highway transportation
2 public transit

subject areas
12 planning
13 forecasting

Library of Congress Cataloging in Publication Data
National Research Council. Transportation Research Board.
New concepts in data analysis.

(Transportation research record; 886)
1. Traffic surveys–Data processing–Addresses, essays, lectures. 2. Traffic surveys–United States–Data processing–Addresses, essays, lectures. I. National Research Council (U.S.). Transportation Research Board. II. Series.
TE7.H5 no. 886 380.5s 83-6290 [HE369] [388.3'142'02854]
ISBN 0-309-03473-6 ISSN 0361-1981

Sponsorship of the Papers in This Transportation Research Record

GROUP 1–TRANSPORTATION SYSTEMS PLANNING AND ADMINISTRATION
Kenneth W. Heathington, University of Tennessee, chairman

Transportation Forecasting Section
George V. Wickstrom, Metropolitan Washington Council of Governments, chairman

Committee on Transportation Information Systems and Data Requirements
Joseph M. Manning, Wellesley, Massachusetts, chairman
George A. Bonina, H. W. Bruck, Donald M. Chapman, Kenneth J. Dueker, David J. Dunlap, Robert T. Dunphy, Robert A. Ferlis, James Finane, Alexander French, Philip N. Fulton, Larry R. Goode, William A. Lesansky, Ginger Levin, James J. McDonnell, Morton A. Meyer, Philip D. Patterson, Alan E. Pisarski, Rolf R. Schmitt, John L. Skinner, Jr., Arthur B. Sosslau, Robert D. Studer, Robert N. Tap, Jack Williams

James A. Scott, Transportation Research Board staff

The organizational units, officers, and members are as of December 31, 1981.

Contents

Authors of the Papers in This Record

Brown, Cleveland, Marketing and Planning Department, Southeastern Michigan Transportation Authority, 660 Woodward Avenue, Detroit, MI 48226

Crabtree, Larry D., Motor Statistical Division, R. L. Polk and Company, Polk Building, 431 Howard Street, Detroit, MI 48231

Krause, Gary, Southeastern Michigan Transportation Authority, 660 Woodward Avenue, Detroit, MI 48226

Lambert, Timothy, Marketing and Planning Department, Southeastern Michigan Transportation Authority, 660 Woodward Avenue, Detroit, MI 48226

Parvataneni, Rai, Marketing and Planning Department, Southeastern Michigan Transportation Authority, 660 Woodward Avenue, Detroit, MI 48226

Sheskin, Ira M., Department of Geography, University of Miami, Coral Gables, FL 33124

Stopher, Peter R., Schimpeler-Corradino Associates, 300 Palermo Avenue, Coral Gables, FL 33134

Origin-Destination Travel Survey for Southeast Michigan

RAI PARVATANENI, PETER STOPHER, AND CLEVELAND BROWN

A small-sample origin-destination survey of randomly selected households was conducted for southeast Michigan to update the existing regional travel data base. The data obtained include records of all trips made by tripmakers 5 years old and older for a 24-h weekday period, demographic information about the sampled household, and attitudinal information on several transportation-related issues from a randomly selected adult in the sampled household. The sample was drawn as a three-stage, stratified random sample of about 2500 households for a region containing approximately 1.6 million households. Despite the small size of the sample (0.16 percent), the trip rates were estimated to ±5 percent accuracy with 90 percent confidence. The rationale for the survey, the method of establishing the sample size, and the procedures for drawing the sample and executing the survey are described; a summary of some of the results is given. Of particular note, the survey measured an overall increase of 17 percent in trip rates over those reported in 1965, although the trip-rate changes varied significantly by both purpose and area type. In addition, compared with 1965, the survey measured a significant increase in car ownership but a decrease in household size. Some of the results of the attitudinal questions are provided, particularly those relating to fuel conservation, price increases, and supply limitations and to attitudes relating to financing of transit improvements. The attitudes measured in the survey in September through November 1980 are in contradiction to changes in federal policy.

During the past decade, large-scale surveys conducted in the 1960s have served as the source of household travel data used in local and regional transportation planning. However, the geographic and demographic characteristics of most urban regions have undergone substantial change, which has resulted in altered travel behavior. A technical council committee of the Institute of Transportation Engineers (1), through an analysis of trip rates of eight U.S. cities along with five Canadian and European cities, has shown a considerable increase in average household trip rates from the 1960s to the 1970s. Greater automobile availability and disposable income and resultant land development shifts have been identified as some of the factors that have caused altered travel behavior.

In recognition of changes in regional travel, supplemental surveys have been initiated to collect detailed current travel information for southeastern Michigan. The collection of this information has been approached in a manner that will enhance the utility of 1980 census data. The supplemental surveys include an on-board transit user survey, a transit screenline count survey, and a major regional travel survey, which is the subject of this paper. These efforts will result in an expanded and updated regional travel data base and provide a data source for transportation planning and implementation activities in the 1980s. The objectives of the regional travel survey are as follows:

1. To gather information on socioeconomic, demographic, and travel characteristics of members of selected households for enhancing the predictability of regional travel-demand models;
2. To evaluate the impact of the changing energy situation on individual travel habits;
3. To obtain such attitudinal data from automobile users about potential ridesharing and transit use as may be useful in the regional travel-demand models;
4. To gather information on the effectiveness of current transit and ridesharing promotional activities; and
5. To gather limited attitudinal data on issues relating to regional transportation policies.

This paper provides a discussion on limitations of previously existing regional travel data, new data requirements for alleviating some of these limitations, survey methodologies for gathering needed new data, and subsequent analyses of the newly acquired data.

LIMITATIONS OF EXISTING TRAVEL DATA

Prudent transportation planning relies on current descriptions of segmented households or individuals and their behavior at a selected level of aggregation that primarily includes socioeconomic and travel data. Further, an understanding of individual attitudes and perceptions towards various transportation-related issues would enhance the planning process in its effort to more precisely simulate individual travel needs. With such data as input, travel-demand models are used to forecast future travel volumes on specified transportation systems. Because development and operation of transportation systems involve large expenditures of funds, reliable travel-demand data should be employed carefully so that decisions on expenditures of capital funds are accomplished effectively and efficiently. The following discussion provides a detailed evaluation of the presurvey status of the travel-demand and related data needs, sources, and applications in southeast Michigan.

1965 Regional Origin-Destination (O-D) Survey

A comprehensive inventory of regional travel patterns was developed in southeast Michigan by the Detroit Regional Transportation and Land Use Study (TALUS) in 1965 as a special project of the Detroit Metropolitan Area Regional Planning Commission (RPC). The TALUS survey gathered O-D data from more than 40 000 households, which resulted in information on more than 340 000 trips. The survey area included Wayne, Oakland, Macomb, and parts of Washtenaw, Monroe, St. Clair, and Livingston counties from which a 4 percent sample of all households in the study area was obtained. Since 1965, this information has provided the basis for all regional land use and transportation planning, which includes the adopted Regional Transportation Plan. The data have served as primary input for current forecasts of regional population and employment. Since 1965, southeast Michigan has been subject to changes in demographic and socioeconomic characteristics. Changes include an increase in automobile availability, increased household disposable income, changes in composition of the work force, and deterioration of off-peak transit service. Such changes have had substantial effects on the travel volumes and patterns in the region.

A comparison of 1970 forecast work-trip attractions to 1970 census journey-to-work trip attractions has shown that estimates based on the TALUS data do not adequately predict travel in the region's outlying counties. The value of TALUS data alone for conducting on-going planning activities appears questionable.

1980 Decennial Census Survey

The 1980 Decennial Census Survey provides detailed socioeconomic descriptions of the region's households. Most of these data are obtained from a 100 percent sample. The data will be extensive, current, and reliable, and thus their use in conducting planning studies is warranted. In addition to socioeconomic data, the census survey collects limited journey-to-work information. These data are obtained from a 20 percent sample for metropolitan areas and will include information on travel modes for work trips, locations of primary work places, the total portal-to-portal travel times, and some information on ridesharing. Although this information provides for sound input to the development of regional transportation plans (particularly during the peak period because the data pertain to work trips), there are still many limitations to the census data. ITE Committee 6A-12 has reviewed both potential applications and limitations of the census data, the findings of which were presented in a paper entitled Preparation for the 1980 Census in the ITE Journal in March 1979 (2). The expected limitations of the 1980 census travel data are discussed briefly as follows.

Desired Aggregation of Travel Data

Although April 1, 1980, was census day, much of the travel-related tabulations will not be released by the Census Bureau until 1982 at the earliest. Further, because of the confidentiality protection given to respondents, data disaggregated to the household level cannot be released. Rather, the data are made available only at an aggregate level (census block, tract, etc.). Because disaggregate models are used predominantly in regional planning, this limitation severely compromises the maximum utility of census data.

Atypical Data Gathered

The Census Bureau does not obtain typical travel-to-work data. The census-reported data provide an overestimation of actual travel on a typical day, because on a typical day some 10-20 percent of workers may not commute to work from home for some reason or other. Adjustments have to be made to factor down the work travel reported by the census. Further, the census does not obtain work-schedule information, which can be very helpful in developing ridesharing promotional efforts. Thus, special efforts must be extended to gather additional information through supplemental surveys and monitoring of employment data.

Secondary Work Travel Data

The census survey does not obtain data on non-home-based work trips such as those from work to other places to execute work-related activities. Similarly, persons holding more than one job do not furnish information on secondary job-related trips.

Travel Data on Submodes

The census journey-to-work information does not adequately identify access and egress travel modes. In transit system planning, submodal information is essential. For example, in the design of park-and-ride services, it is necessary to estimate the volume of drivers who use park-and-ride lots to store their vehicles as contrasted with those who use the lots in a walk-and-ride or kiss-and-ride mode of travel.

Nonwork Travel Data

Discretionary nonwork travel accounts for more than 60 percent of the trips made in southeast Michigan. The census does not collect information on nonwork travel. Such information must be obtained through supplemental means.

Attitudinal Data

Last, even though issues such as energy concerns do play an important role in developing regional transportation policies and plans, the census does not gather information on individual or household attitudes and perceptions toward various transportation-related issues.

OTHER PERTINENT REGIONAL SURVEYS

Other transportation-related surveys have been conducted within the region for special purposes and generalized use, but these surveys have not collected information of sufficient detail and sample size for use in comprehensive transportation studies. The given data limitations suggested a need for conducting supplemental information-gathering activities in 1980 if having current and detailed information on regional travel behavior for all modes and trip purposes was desired and valuable.

SUPPLEMENTAL DATA NEEDS AND APPLICATIONS

In order to address limitations of the current data sources, a need exists to collect additional travel information. The supplemental data needs can be classified broadly into three categories and subclassified as detailed below:

1. Household characteristics
   a. Household composition
   b. Information on household members 16 years or older
   c. Gross household income
   d. Household vehicle availability
2. Person/trip data
   a. Tripmaker identification
   b. Trip O-D locations and starting and ending times
   c. Trip purpose
   d. Mode of travel
3. Attitudes and perceptions
   a. Related to transit use
   b. Transportation strategies
   c. Energy considerations

DETERMINING SAMPLE SIZE

In recognition of funding restraints, extra care was given to the design of a small sample that provides statistically accurate results. The critical variable for sample size determination was the household tripmaking rate. The existing trip-generation forecasting procedure consists of four linear regression equations with the independent variables of family life cycle, income, household size, and automobile availability. The four equations are for four area types defined as follows:

Area type 1: 10 or more employees per acre of usable land,
Area type 2: less than 10 employees and more than 5 dwelling units per acre of usable land,
Area type 3: less than 10 employees and from 0.5 to 5.0 dwelling units per acre of usable land, and
Area type 4: less than 10 employees and less than 0.5 dwelling unit per acre of usable land.

A procedure based on sampling-error computation was recently developed by M.E. Smith (3) for calculating the sample sizes from prior data on trip-generation rates. (Smith also showed that the samples calculated for trip rates will be more than adequate in general for trip distribution and mode-split modeling.) The procedure takes into account the contributions of different subgroups of the data to the total sampling error and produces an estimate of the minimum sample size needed to attain the required accuracy. The procedure requires that a sample size be computed on the basis of the required accuracy at the specified confidence level and that these calculations be done by estimating a pooled coefficient of variation over the identified subgroups (cells). Subsequently, the sample size may be readjusted on the basis of the subsample size in the "critical cell," which is defined as the cell that has the largest coefficient of variation. Application of the sampling procedure generates a sample size for each cell based on its contribution to the overall coefficient of variation. However, by using the distribution of households by cell from the base data, the expected sample size in each cell can be estimated. This will usually be different from the sample size based on the cell's contribution to the coefficient of variation and hence follows the need for readjustment. By applying Smith's procedure within the four area types treated as independent entities, the following sample sizes were computed. At the outset, a uniform accuracy level in each area type was assumed by specifying that the trip rates be estimated to within ±5 percent with 90 percent confidence for each area type. The sample sizes are given below:

Area Type	No. of Households
1	610
2	450
3	343
4	404
Total	1807

In no case does the expected sample provide a sufficient subsample in the critical cell of an area type. After a correction factor has been applied, again following Smith's procedure to give the optimum sample size for each critical cell, the adjusted sample sizes became as follows:

Area Type	No. of Households
1	1157
2	660
3	481
4	524
Total	2822

After the initial sample size had been derived, consideration was given to two other factors, namely, the magnitude of tripmaking in each area type and the political jurisdictional balance within the region, primarily between the counties. The number of households based on the regional forecasts was 84 484 for area type 1, 191 886 for area type 2, 1 034 090 for area type 3, and 344 023 for area type 4. A relatively large number of households in area type 3, coupled with the fact that the average trip rate for this area type is larger than the others, revealed that the trip-rate accuracy levels should be higher for area type 3 in order to improve the overall accuracy levels. Second, the sample should be somewhat spread more uniformly between political jurisdictions to be able to draw meaningful conclusions from the attitudinal data. Based on these factors, adjustments were made to the above sample sizes. The final sample consisted of the following:

Area Type	No. of Households
1	681
2	675
3	621
4	625
Total	2604

MULTISTAGE SAMPLING TECHNIQUE

To achieve a true random sample, a complete sampling frame consisting of a list of all households in the study region stratified by zonal area type should be used. However, no current listing exists of households for all seven counties of southeast Michigan. To overcome this problem without undertaking a complete, in-field enumeration of all households in the region, a three-stage random sampling process was used, in which the stage designs permitted use of extant lists of aggregations of households until the final stage, when enumeration would be a greatly reduced activity.

In the first stage, a stratified random sample of zones was selected with varying sampling fractions for four strata comprising the four area types. The population for this sample consisted of 1446 analysis zones, each of which was classified by area type. The second-stage sample was a sample of blocks from those zones selected in the first stage. This and the third stage used property description maps from the tax assessment and equalization departments of the counties. Although these maps varied from county to county in style, content, scale, and referencing system, all had a common system of delineating developed and partly developed land into blocks of land area that were completely surrounded by streets and had no streets passing through them. Also, the maps provided a numeric code for every subdivided parcel of land either by lot or by current property boundaries. All maps of this nature are kept reasonably current; most are current to within a matter of months. Traffic analysis zone boundaries were drawn on these maps and blocks within the zone enumerated. A random sample of blocks was drawn for each zone by using different sampling rates for each area type.

The third-stage sample consisted of parcels from the selected blocks. Each selected parcel was then located in the current tax records of the local taxing authority (city or county), from which its use could be established. If the use was found to be residential, the address was recorded from the tax records, and the parcel became part of the sample.

At each stage, the sampling was continued beyond the designated sample size to provide backup against in-field failures to obtain an interview from an original sample. In some instances a zone or a block contained no or too few residential units; this necessitated use of additional zones or blocks to complete the sample. In order to avoid potential bias, the random sampling procedure within each stage was extended to allow for such eventualities.

SURVEY

Selection of Survey Mechanism

A number of alternative mechanisms or techniques were considered for the survey, including a self-administered mail survey, a telephone interview, a combined telephone interview and mail survey, and an in-home personal interview. On balance, the in-home personal interview was deemed to be the preferred mechanism, in view of the purposes of the survey, the nature and length of questions to be asked, and the probable response to the survey. However, the objective of obtaining travel information for a 24-h weekday period from each household member 5 years old and older necessitated further consideration of procedures.

The traditional historical-record method of collecting travel data (i.e., requesting data on the previous 24 h for each eligible family member) was not considered satisfactory for several reasons. Principally, past experience with the method suggests that a number of trips (particularly short trips and non-home-based trips) are seriously under-reported and that 1980 lifestyles seemed likely to make it difficult for the interviewer to find a majority of eligible household members at home at the desired time of the interview. As a result, a travel diary was adopted and designed to be used on an appointed day by an eligible household member to record his or her travel on that day.

By using the travel diary, the survey mechanism was designed as a two-step process. First, the interviewer completed an interview with a randomly designated household member (by using a selection grid where the designated individual was defined on the basis of the day of the week and the numbers of adults and adult males then at home) that gathered attitudinal and demographic data. After rapport had been established with the respondent, travel diaries were distributed for each eligible household member and an explanation was given of how to complete them, the day for completion was set, and an appointment was made for the interviewer to return to collect the completed travel diaries. The second step was the return visit to collect the completed travel diaries; that visit was an opportunity to check the travel diaries for completeness, probe for missing trips, and provide a promised incentive for completing the travel diaries.

Conduct of Survey

Travel surveys always pose problems with respect to timing during the year, particularly in northern cities of the United States. Travel is known to be atypical during major school breaks, in the period from Thanksgiving through New Year, and in the period of winter from January through March, when snowstorms and other specific weather occurrences may cause major disruptions in travel. This limits travel surveys primarily to the period between Labor Day and Thanksgiving and from the beginning of April through mid-June. Because of the desire to collect the data as close as possible to the 1980 census (for purposes of compatability), the survey was scheduled for fall 1980. Interviewing commenced on September 6, 1980, and concluded on November 23, 1980. Retrieval of travel diaries continued into December and some mail and telephone followups for missing critical data continued into February 1981.

The execution of a complete interview took a significant amount of time as a result of several factors. First, the interview and distribution of travel diaries generally took 30-50 min to complete. Second, the return call to pick up the travel diaries required generally some 10-20 min at the household and frequently necessitated one or more calls back to obtain a complete set of travel diaries. Third, interviewers were required to make three calls at a household initially (at least one call on a weekday and one on a weekend day) before the household could be deemed a "no answer" and replaced from the backup sample. Fourth, although the multistage sampling produced a somewhat clustered sample, significant travel distances were involved, particularly in the outer counties of the region. As a result, interviewer productivity was severely constrained.

In the 11-week period of the survey, 2706 interviews were completed, for which 2502 complete sets of travel diaries were obtained. (The other 204 interviews had one or more travel diaries missing at the conclusion of all data collection.) To obtain the 2706 interviews, a total of 5309 sample addresses was generated. Table 1 shows the disposition of this total sample. The 2502 complete interviews represent 77.7 percent of the successful contacts and 92.7 percent of those contacts that resulted in completion of the attitude and demographic interview. A brief explanation of a few of the dispositions is useful to clarify the survey results. "No such address" was recorded when both of the neighboring addresses were found and it was clear that no intervening property existed. These represent outright errors in the tax rolls and the consequences of recent redevelopment. "Cannot find" was recorded when the address could have existed but neither the interviewer nor the supervisor was able to locate it. Most of these occurred in the outlying rural areas. "Noneligible respondent" was recorded for two primary situations. The first was when all household members were unable to speak or understand English; the second was when no adults in the household could be interviewed. The latter included households where the only adults present appeared to the interviewer unable to provide a coherent response or not rational, i.e., potentially under the influence of drugs or alcohol. The 121 "contact recorded, no interview found" occurred where an interviewer indicated that an interview had been completed, but no interview forms were found. Some of these were forms that were not retrieved from interviewers who for one reason or another were removed from conducting the survey. Finally, short demographic interviews were conducted on refusing households, originally intended as a check for nonresponse bias. The number of successful short interviews, at less than 10 percent of the refusals, proved too few for a nonresponse analysis, however.

During the conduct of the survey, telephone verifications were carried out on 15 percent of each interviewer's work to make sure that a valid interview had taken place, that incentives were provided, and that the interviewer had been courteous and polite. In addition, any missed data were requested at this time and the respondent was asked how long the interview took and if he or she had any comments to offer.

Certain elements of the survey were defined as

Table 1. Disposition of household sample addresses.

Disposition	Number	Percent
No answer	867	16.3
Uncompleted request for call back	83	1.6
No such address	336	6.3
Cannot find	53	1.0
Noneligible respondent	133	2.5
Under construction	5	0.1
Vacant	340	6.4
Business	113	2.1
Duplicate address	37	0.7
Contact recorded, no interview found	121	2.3
Subtotal	2088	39.3
Refusal	462	8.7
Termination	12	0.2
Short interview	41	0.8
Interview without complete travel diaries	204	3.8
Subtotal	719	13.5
Complete interview	2502	47.1
Total	5309	100.0

critical, and additional effort was made to obtain those elements. The critical elements were defined as (a) a completed travel diary from each eligible household member (i.e., if any travel diaries were not returned by a household, the interview was considered incomplete) and (b) completion of data on automobile availability, income, household size, and the variables used to define life cycle (ages of children and age of head of household). These elements were sought in follow-up activities. For the most part, collection of missing travel diaries continued through mid-December only. After that, the chances of recovering missing travel diaries were considered too low for the cost and effort required, and the probability of obtaining travel diaries containing information for some day outside the survey period would be too high and could lead to invalid results. Missing demographic data were sought by both mail and telephone follow-up. These procedures succeeded in completing an additional 48 surveys. An additional 95 interviews were missing income data only, and a multiple-classification analysis procedure was developed to estimate income for these interviews on the basis of area type, number of workers, and number of available vehicles.

The 2502 completed interviews, therefore, consisted of 2359 that were satisfactorily completed from the original interviews, 48 that were completed by additional solicitation for critical demographic data, and 95 that were complete except for income but for which income could be estimated from other data. In subsequent analysis, the computer was unable to match the interviews and travel diaries for 56 households, so the subsequent trip-rate analysis is based on 2446 of these complete interviews.

SURVEY DATA ANALYSIS

Although the trip-rate analysis is based on data from only 2446 households, information from 2706 households was used for the analysis of the attitudinal data. The data were expanded by political jurisdiction and area type based on the total number of households within each stratification.

Trip-Rate Data Results

By applying Smith's procedure to the trip data gathered from this survey, it was found that the accuracy levels achieved from the survey data are very much in concert with the assumptions made earlier during the design phases of the study. The results of home-based total vehicle trips, which were the key factor in the design of the survey, are shown in Table 2.

As seen from Table 2, there have been significant increases in average trip rates between 1965 and 1980 except for area type 2, which exhibited a decrease in trip rates. When the percent differences are weighted by the number of households in each area type, the home-based vehicular trips in 1980 are higher by about 17 percent than in 1965.

Demographic profiles relevant to trip-making behavior (automobile availability and household size) were generated and comparisons were made between 1965 and 1980. These comparisons indicated that the average household size has decreased from 1965 to 1980, i.e., from 3.5 persons per housing unit to 2.75. On the other hand, the distribution of automobile ownership as presented below indicates that the trends are toward owning two or more automobiles.

	Percent Who Own Automobiles			
	No. of Automobiles Owned			
Year	0	1	2	3+
1965	15.2	47.5	31.5	5.8
1980	11.6	37.8	38.2	12.4

Further analysis of the home-based vehicular trips broken down by home-based work and home-based other trips reveals that home-based work trip rates have decreased from those of 1965. It is doubtful that the drop in 1980 home-based work trip rates is due to fewer workers in the household. National trends show that the recent increase in workers per household is due to the increased numbers of working women. Lower home-based work trip rates cannot be attributed to high unemployment; in the surveyed sample, only 3.5 percent of the respondents were laid off from their regular jobs at the time of the interview. A plausible explanation for lower 1980 home-based work trip rates is the trip-chaining being done to offset the higher cost of gasoline. In other words, the non-home-based trip rates are higher in 1980 than they were in 1965. Whereas workers may have proceeded directly from work to home in the past, they are now more likely to stop for shopping or a visit with a friend rather than going home first. In fact, in the attitudinal part of the data, respondents indicated that they have been chaining trips due to the energy situation. A detailed analysis of the trip data has yet to be undertaken.

Attitude Survey Results

The survey included a variety of attitudinal questions. The following analysis presents some important issues related to energy and public transportation.

Conserving Energy (Behavior and Behavioral Intent)

The interviewer stated that he or she was going to read a list of things that people might do because of higher gasoline/diesel fuel prices or gasoline/diesel fuel shortages. After each one, please indicate whether

1. You started to do this regularly more than a year ago,
2. You started doing this regularly within the past year,
3. You would do this if gasoline/diesel fuel prices were to double next week,
4. You would do it if you could buy only 10 gallons (35 liters) of gasoline/diesel fuel a week for each registered vehicle starting next week, or
5. You would do it either if prices doubled or if gasoline/diesel fuel were rationed.

Summarized results are presented in Table 3. Responses to the first two items are combined to give the percentage of respondents who say they are already undertaking the stated action on a regular basis. The next three items are combined to provide the percentage of those who would consider taking the action regularly if the price or supply constraints became reality. It is useful to look at three sets of actions: Those that the majority (more than 50 percent) claim they are doing regularly now, those that the majority would expect to do if gasoline prices double or supply is restricted, and those that the majority would not expect to do under any of the stated conditions.

In the first category, as indicated by the "Am Doing" column (the sum of "more than a year" and "past year"), respondents claim that they are

Table 2. Comparison of trip rates: 1980 versus 1965.

Area Type	Total Sample Size	Accuracy Unadjusted (%)	Avg Home-Based Total Vehicle Trip Rate 1980	1965	Difference (%)
1	617	85	3.535	1.870	+89
2	574	90	3.428	3.915	-12
3	685	97	5.915	5.212	+13
4	570	96	6.605	5.188	+27

Table 3. Behavior and behavioral intent to undertake conservation actions.

Question	Percent Responding None	Am Doing	Would Do
Observe the 55-mph speed limit	4	89	7
Take a vacation closer to home	36	35	29
Shop less frequently	19	55	26
Carpool or vanpool to work or school	47	21	32
Cancel a vacation trip	44	15	41
Combine car journeys you used to make separately	13	62	25
Buy a car that gets better mileage	31	30	40
Take the bus or train to work or school	50	11	39
Have car tuned up regularly	4	90	7
Move closer to work or school	75	11	13
Walk or bicycle to work or school	69	15	16
Shop closer to home	14	69	16
Look for a job closer to home	68	17	14
Shop on the way to or from work or school	22	61	17
Cut down use of snowmobiles, power boats, or other recreational vehicles	57	22	22
Sell a car and not buy one in its place	77	5	19
Use a train, bus, or airplane for vacation trips	28	34	39
Take a bus or train more often for nonwork travel	46	11	42
Move to a place with better bus service	88	3	10

Notes: Due to rounding error, totals will not always be 100. N = 2180.

1. Having their cars tuned regularly (90 percent),
2. Observing the speed limit of 55 mph (89 percent),
3. Shopping closer to home (69 percent),
4. Shopping on the way to and from work or school (61 percent),
5. Combining car trips that used to be separate (62 percent), and
6. Shopping less frequently (55 percent).

Among these, not more than 22 percent of the respondents indicated that they had begun the action in the past year (which would encompass the sharp gasoline price increases in the fall of 1979), although actions 4, 5, and 6 were each reported as having been initiated in the past year by 17 percent or more of the respondents.

In the second category, as indicated by a combination of the "Am Doing" column and the "Would Do" column (the sum of "double price," "ration," and "either"), respondents indicated what they would expect to be doing if the price of gasoline doubled or if gasoline were rationed:

1. Using train, bus, or plane for vacation trips (72 percent);
2. Buying a car that gets better mileage (70 percent);
3. Taking a vacation closer to home (66 percent);
4. Canceling a vacation trip (56 percent);
5. Taking a bus or train more often for nonwork travel (54 percent);
6. Carpooling or vanpooling to work (53 percent); and
7. Taking a bus or train to work or school (50 percent).

Rationing seems to have much less effect on people's perceptions than price increase; at most, 10 percent of the respondents said that only rationing to 10 gallons would cause them to cancel a vacation.

The remaining six actions would not be considered by the majority of respondents. As indicated by the "None" column, respondents indicated they would not, under any of the stated circumstances, do the following:

1. Move to a place with better bus service (88 percent),
2. Reduce the number of cars they owned (77 percent),
3. Move closer to work or school (75 percent),
4. Walk or bicycle to work or school (69 percent),
5. Look for a job closer to home (68 percent), and
6. Cut down the use of recreational vehicles (57 percent).

These results tend to indicate, first, that trip-chaining and reductions in discretionary travel are the primary adjustments that people have been willing to make so far. This trend continues in those actions that people indicate a willingness to undertake next; three actions involve changes in vacation trips and one is nonwork travel. None of the energy scenarios is perceived as being harsh enough to produce a change in home location or job location, to reduce the number of cars owned, or to lead to dependence on nonmotorized travel for work trips. Even a shift to transit for the work trip is envisaged by only 39 percent of the respondents and 32 percent might carpool.

Perceived Effectiveness and Favorableness of Imposed Conservation Strategies

The interviewer stated: "I am going to read some suggested ways to reduce gasoline/diesel fuel consumption. For each one, please tell me how much you think each one would reduce your household's gasoline/diesel fuel consumption (column A) and whether or not you think it is a good idea as a way to reduce fuel consumption (column B)...." Results of this question are given in Table 4. In terms of reducing gasoline consumption, improved bus service (items F and G in Table 4) and gasoline rationing (C) are seen to be the most effective, while taxing gas-guzzling cars (B) and asking people to drive less (D) are likely to be the least effective. On the other hand, the strategies respondents favor most are discretionary ones and those that would return money to them, such as items D (yes, 73 percent), E (yes, 67 percent), F (yes, 81 percent), and G (yes, 85 percent). Likewise, disapproval is high for mandatory and economic disincentives, as shown by the responses to items A (no, 84 percent), B (no, 68 percent), and C (no, 77 percent).

General Awareness of Regional Transit Services

In reply to the question "How much would you say you know about public transit services in the southeast Michigan region?" the following responses were found (N = 2706):

Response	Percent
Very much	5.5
Some	26.9
Very little	43.9
Nothing	23.7
	100.0

Table 4. Perceived effectiveness and favorableness of imposed conservation strategies.

	Percent Responding					
	Column A				Column B	
	Very Much			Not at All		
Item	1	2	3	4	Yes	No
A. Add a 50¢/gal nationwide gasoline tax	27	17	16	40	16	84
B. Add a $100/year tax on gas-guzzling cars	23	12	12	53	32	68
C. Introduce nationwide gasoline/diesel fuel rationing of 10 gal/registered vehicle/week	33	19	14	34	23	77
D. Ask people to drive one-fifth less than now but not force them to do so	20	22	22	36	73	27
E. Give a $100/year tax rebate on cars that get more than 30 mpg	27	19	11	44	67	33
F. Twice as frequent bus service	32	19	10	39	81	19
G. No more than a 5-min walk to a bus stop	38	18	10	35	85	15

Notes: Due to rounding error, totals will not always be 100. N = 2180.

Table 5. Regional attitudes toward financing public transportation.

	Percentage							
	Strongly Agree				Strongly Disagree			
Item	1	2	3	4	5	6	Agree Total	Disagree Total
A. All costs for running and improving public transportation should come from fares paid by passengers	29	13	18	12	13	15	60	40
B. Federal government should subsidize running and improving local public transportation	31	24	15	8	6	16	70	30
C. In addition to fares, running and improving public transportation should be paid for by increase in gasoline taxes	5	7	10	9	14	55	22	78
D. In addition to fares, running and improving public transportation should be paid for by increase in sales tax	4	5	9	9	15	58	18	82
E. In addition to fares, running and improving public transportation should be paid for by increase in property tax	1	1	2	5	11	80	4	96
F. In addition to fares, running and improving public transportation should be paid for by increase in other vehicle taxes	5	9	10	9	13	52	24	74
G. In addition to fares, running and improving public transportation should be paid for by increase in income tax	2	4	5	6	11	72	11	89
H. Public transportation should be made free of fares for all riders	9	4	6	7	15	60	19	82

Notes: The first six columns represent all answers and will sum to approximately 100 percent, except for rounding errors. The last two columns are summaries of the first six and also add up to 100 percent. N = 2706.

The majority of the sample (67.6 percent) knows very little or nothing at all about public transit. (It was also noted that 67.4 percent have not used transit in southeast Michigan for at least one year, if ever.)

Importance of Major Improvements in Public Transit

In reply to "How important do you think it is to make major improvements in public transportation in southeast Michigan?" the following was found (N = 2706):

Response	Percent
Very important	63.5
Somewhat important	26.3
Not important	4.5
No opinion	5.7
	100.0

A majority of the respondents (89.8 percent) thinks it is very or somewhat important to make major improvements in public transportation. Interestingly, respondents admit to not knowing much about the existing transit system, but a majority thinks it is quite important to improve it.

Regional Attitudes Toward Financing of Public Transportation

"I am going to read some statements about paying for public transportation here in southeast Michigan. Please tell me how strongly you agree or disagree with each statement." Results of this question are given in Table 5. Respondents generally disagree with all of the suggested financing mechanisms. The majority agrees in only two cases: fares paid by the passengers (item A, 60 percent) and local public transit system subsidies by the federal government (item B, 70 percent). The majority of respondents disagrees with the other six financing options. An increase in the property tax is the most unpopular financing mechanism (item E, 96 percent disagree). These data indicate that most people in the region believe that someone else should pay for public transportation.

Summary of Observations

A brief summary of observations from the regional attitude survey data follows:

1. Trip-chaining and elimination of discretionary travel are behaviors that people assume to control the amount of energy they consume and save money rather than change their place of residence, the number of cars they own, their mode of transportation from automobile to walking or bicycling, or their use of recreational vehicles.

2. As a means of energy conservation, incentives to users of energy-saving modes (e.g., bus or carpool) are preferred by respondents to economic disincentives to people who do not use such modes.

3. Respondents agree that carpooling and bus travel save money and energy, but the majority of them view these modes as impractical for themselves.

4. Travel time and convenience to the traveler are favored over saving money and energy in the choice of a mode of transportation.

5. The majority of respondents thinks it is important to make major improvements to public transportation in southeastern Michigan, but when they were questioned about financing mechanisms, fares paid by passengers and federal government subsidy were the only two financing options favored by a majority of the sample. Thus, the respondents recognize a need for public transportation but feel that someone else should pay for it.

CONCLUSIONS

A small-sample, supplemental O-D survey was conducted successfully by using a personal home interview to collect attitudinal and demographic data and a travel diary to collect a 24-h travel record for all household members 5 years old and older. The sample size and distribution were based on the trip-rate variances estimated from 1965 data, with some modifications; a sample of about 2500 households was generated that achieved the desired accuracy of ±5 percent error with 90 percent confidence.

The trip rates exhibited from this survey show a 17 percent increase over the rates measured in 1965, which seems to be consistent with other recent surveys measuring trip rates. Within this 17 percent overall increase, a decrease was found in home-based work trips and increases in all other trips, particularly non-home-based trips. It is not clear, however, to what extent these measured increases are the result of real increases in trip-making or are the result of a different survey mechanism (the travel diary), which could be expected to provide a more accurate picture of trip-making.

The results of the attitude survey are, for the most part, unsurprising but serve to confirm a number of prevailing professional expectations and assessments, particularly in relation to transportation energy and the use of carpools and transit. Two points that deserve particular emphasis are, first, that 68 percent of the sample know very little or nothing about transit in the southeast Michigan region (this percentage does not change when the data area is expanded to the entire region), whereas less than 6 percent consider that they are very familiar with regional transit services and that federal subsidies are seen as the preferred mechanism to fund transportation improvements. This second finding is particularly relevant given current changes in policy occurring at the federal level with respect to transportation funding. It is also noteworthy that lack of knowledge of regional transit services seems to have little impact on the perception that transit improvements are needed; these are favored by almost 90 percent of respondents.

ACKNOWLEDGMENT

The opinions, findings, and conclusions expressed in this paper are ours and not necessarily those of the U.S. Department of Transportation, Federal Highway Administration, Urban Mass Transportation Administration, or the State of Michigan.

REFERENCES

1. Consistency of Origin-Destination (O-D) Characteristics through Time. ITE Journal, Oct. 1979, pp. 32-39.
2. Preparation for the 1980 Census. ITE Journal, March 1979, pp. 41-47.
3. M.E. Smith. Design of Small-Sample Home-Interview Travel Surveys. TRB, Transportation Research Record 701, 1979, pp. 29-35.

Pilot Testing of Alternative Administrative Procedures and Survey Instruments

IRA M. SHESKIN AND PETER R. STOPHER

Traditionally, pilot surveys have involved pretests of the survey instrument and administrative procedures to be employed in the main survey. Such pilot surveys usually have attempted to pretest a single version of the survey instrument and the administrative procedures and to seek appropriate refinements. By using examples from the Dade County On-Board Transit Survey and a Midwest regional travel survey, it is argued that an important and underused part of a pilot study is comparisons between various alternative administrative procedures or survey-instrument components, in which each alternative is foreseen to have both advantages and disadvantages. The pilot study is likely to provide considerable information on the relative merits of the alternatives tested and will lead to improved design of the final instrument or procedure. Such testing may lead frequently to decisions that can have extensive impacts on response rate, response quality, or survey cost.

Survey research is in many ways as much an art as it is a science. While it is possible to transfer general procedures from one spatial and temporal setting to another, each survey effort is to a large extent unique. Thus, every survey should be preceded by a pilot study (1, p. 205). Often, pilot studies have consisted only of a pretest of the questionnaire, perhaps even administered to a sample not representative of the population to be sampled in the main survey. In a university setting, this usually translates to the testing of the questionnaire on a captive classroom audience; in other settings often only an in-house test is performed. Four reasons may be seen for the employment of cursory pilot studies in most cases. First, it is possible that some researchers have not recognized the importance of a full-scale pilot study. Second, budgetary constraints often have obviated any large-scale pilot-study effort, frequently because the importance of budgeting for it was not recognized. Third, time considerations may make it infeasible to carry out a pilot study. Fourth, if the survey effort falls under the rules and requirements of the U.S. Office of Management and Budget (OMB), a pilot test on more than 10 people requires OMB approval. This approval is likely to involve sufficient lead time and delays to make a pilot test infeasible for all but extremely large censuses and surveys, which is surely in contradiction to the intent that lies behind the OMB role in survey approvals.

The purpose of this paper is to discuss the need to pretest alternative survey forms and the probable benefits that accrue. The major contention is that if two or more proposed procedures or proposed methods for asking a question are foreseen to have both advantages and disadvantages, both procedures should be tested in a pilot study. The need to test alternative procedures is highlighted by Dillman (2).

DADE COUNTY ON-BOARD TRANSIT SURVEY AND MIDWEST REGIONAL TRAVEL SURVEY

The discussion in this paper employs examples from pilot studies designed by us for two transportation surveys: the Dade County On-Board Transit Survey and a Midwest regional travel survey. A brief description of the purposes of each survey and the survey mechanisms follows.

The Dade County On-Board Transit Survey was designed to collect data from a random sample of bus passengers (3). The principal purposes of the survey were to provide the following:

1. A major test of a proposed monitoring and surveillance activity for the Metro Transit Agency (MTA) as called for by the Transportation Development Program (TDP) (4);
2. A partial supplement to the travel data collected by the 1980 census on trips to work and part of a data base for using the census data to update trip-rate estimates for nonwork trips;
3. Needed data on bus ridership in the central business district (CBD) (the current data base is seriously deficient in this part of the matrix);
4. Improved data to MTA for use in adjusting its revenue-based, patronage-estimating formula, particularly as needed after recent changes in transfer policies;
5. Data on the use of media by bus passengers, particularly as it relates to providing riders and potential riders with information on the bus system and the services available;
6. Part of the data needs for a recalibration of the Dade County modal-split model; and
7. Data on the perceptions of riders about the MTA system and specific elements of it and a basis for comparing bus-rider judgments (attitudes) with those of the general population of Dade County; the latter were collected in a separate survey in 1980 by MTA (5).

As is common in most U.S. urban areas, bus riders constitute less than 10 percent of the population of Dade County. Hence, any survey aimed specifically at bus riders would be highly inefficient if the sample were drawn from households, employees, or any other non-travel-specific grouping of the population. Thus, the survey mechanism was designed as an intercept survey of bus passengers. A dual survey mechanism was employed that included a brief form to be completed on the bus and a longer, take-home, mail-back survey (6).

The Midwest regional travel survey was designed to collect data from a stratified random sample of the population in seven counties. The principal purposes of the survey were to provide the following data:

1. The means of update trip-generation rates and modal-split models,
2. Attitudes of the population toward transportation and energy (7),
3. Attitudes toward possible changes in the transit system, and
4. Preferred methods of obtaining information on carpooling.

The trip-generation and modal-split models to be updated use certain demographic characteristics and income as input variables, so these characteristics must be measured to permit updating to be accomplished. Also, the survey coincided with a period of high unemployment in the southeast Michigan region (mainly connected with a low cycle in the automotive industry). Because of the potential effects of this on tripmaking, detailed information was required on employment status.

The selected survey mechanism was the home-interview survey. Two instruments were used. The first was an attitudinal demographic survey asked of a randomly selected adult household member. The second was a travel log distributed to each household member more than five years old and designed to obtain trip information for a 24-h weekday period.

TESTING ALTERNATIVE SURVEY FORMS

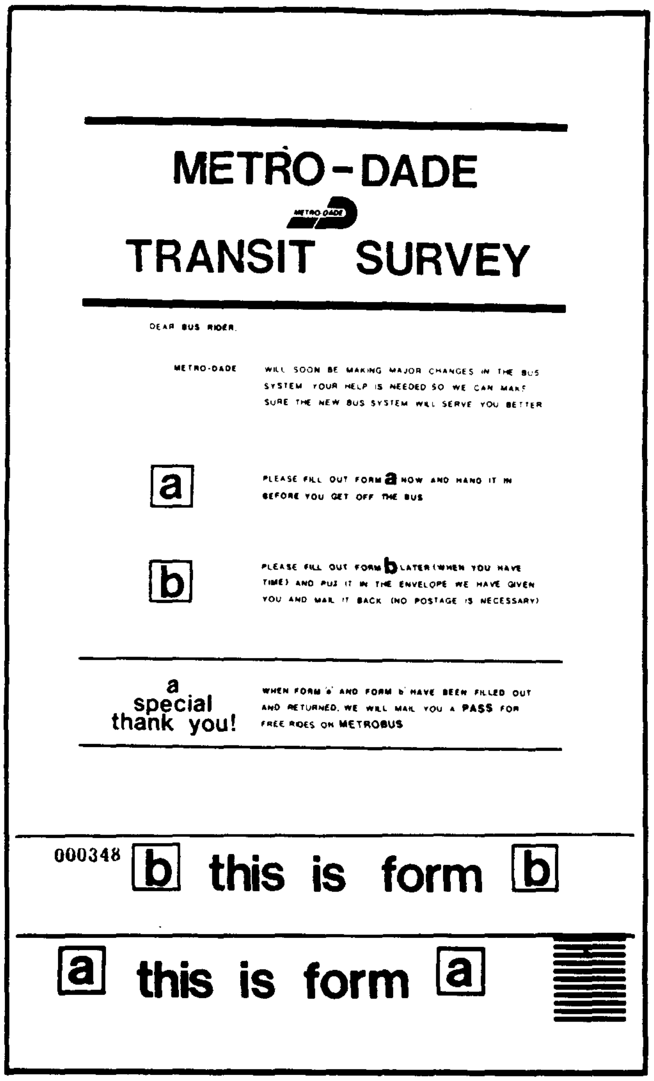
Dade County On-Board Transit Survey

Frequently, in the design of a survey instrument, two or more ways appear to be potentially useful to ask a given question or set of questions; or there may be several possible ways to request answers, e.g., by using 5-point, 6-point, or 7-point scales on judgmental questions. Similarly, many survey instruments may contain questions that are particularly crucial to the purposes of the survey but that are difficult to ask. For such situations, two or more alternative formats often will be developed for such questions, but choice among them may not be obvious. In either case, the most definitive test of the alternative formats is to use each one as part of the pilot study. To do this, a carefully structured scientific test of each alternative must be developed. An example is described in this section based on a set of problem questions in the Dade County On-Board Transit Survey.

In that survey, questions concerning respondents' perceptions of the times and costs of bus transportation versus alternative modes needed to be asked for purposes of recalibrating the Dade County modal-split model, but all suggested formats for asking these questions were viewed as difficult. Compounding this problem is the fact that bus riders are not a random sample of the population but rather are more likely to be members of specific sociodemographic groups in which problems of comprehension or concentration are likely to be more pronounced. This section describes in more detail the forms (including the alternative formats) used on the pilot study of the Dade County survey and the results obtained.

The survey instrument was designed as a two-part entity. An on-board form (form a) was printed on card stock (to make it easier to fill out while riding a bus) and was designed to be short enough to

Figure 1. Instruction sheet.



METRO-DADE

TRANSIT SURVEY

DEAR BUS RIDER,

METRO-DADE WILL SOON BE MAKING MAJOR CHANGES IN THE BUS SYSTEM. YOUR HELP IS NEEDED SO WE CAN MAKE SURE THE NEW BUS SYSTEM WILL SERVE YOU BETTER

a PLEASE FILL OUT FORM a NOW AND HAND IT IN BEFORE YOU GET OFF THE BUS

b PLEASE FILL OUT FORM b LATER (WHEN YOU HAVE TIME) AND PUT IT IN THE ENVELOPE WE HAVE GIVEN YOU AND MAIL IT BACK (NO POSTAGE IS NECESSARY)

a special thank you! WHEN FORM 'a' AND FORM 'b' HAVE BEEN FILLED OUT AND RETURNED, WE WILL MAIL YOU A PASS FOR FREE RIDES ON METROBUS

000348 b this is form b

a this is form a

fill out on a bus. Although a reply-paid panel was printed on this card so that mail return was possible, the form was designed to be placed in a receptacle at the exit door of the bus or handed back to the survey person. The second part was a longer, take-home form (form b) designed to be completed at home and mailed in in the reply-paid envelope provided. The whole package was stapled and included an instruction page and a letter from the county transportation coordinator (Figure 1). The instruction page explained briefly the purpose of the survey and instructed respondents that form a was to be completed on the bus but that form b was to be done at home and returned by mail. In addition, a free bus-pass incentive was offered to gain cooperation. The back of this page contained helpful county telephone numbers. The letter from the transportation coordinator stated the reasons for the survey and the importance of each person's contribution, reviewed the instructions for filling out the forms, and provided a telephone number for help, comments, or verification that this was a bona fide survey. The entire survey instrument was combined so that, when one looked at the instruction page, a 1-in tab from each of form a and form b showed below the top page. This simplified the problem for the respondent of finding each form. The major reason for the two-part form was to permit the evaluation of nonresponse bias (6).

In the pilot study, 2158 forms were distributed: 632 (29 percent) of the on-board forms and 380 (18 percent) of the take-home forms were returned, although due to time constraints only 301 of the take-home forms were computerized.

Two versions of the on-board form and three versions of the take-home form were devised. Because a possible "shadow effect" of one questionnaire on another existed, each on-board form (called the on-board short and the on-board long for reasons explained below) was combined in equal numbers with each take-home form (called the take-home short, take-home long, and take-home table). This produced the following six versions of the questionnaire:

1. On-board short/take-home short,
2. On-board short/take-home table,
3. On-board short/take-home long,
4. On-board long/take-home short,
5. On-board long/take-home table, and
6. On-board long/take-home long.

These versions were distributed in a systematic mix to consecutive bus riders as they boarded to assure that, as far as possible, the full range of six survey instruments was distributed at each bus stop.

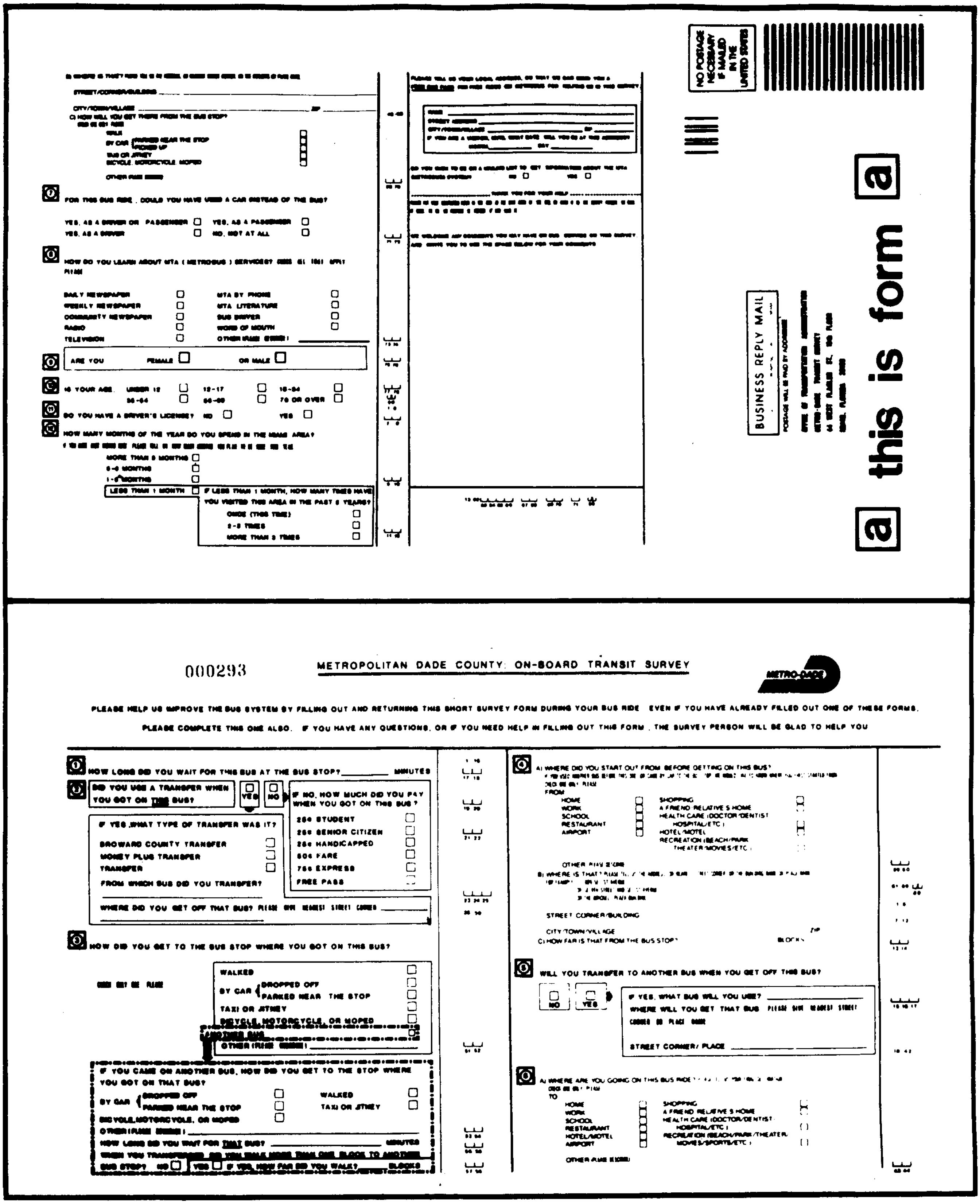
Alternative On-Board Forms

The major purpose of the on-board form was to elicit some response from persons who would not be bothered to take a form home, spend 45 min completing it, and remember to mail it (see Figure 2). Also, reading and writing on a moving bus is very difficult and many persons in Dade County, particularly the elderly, ride the bus for only a few blocks at a time. All these considerations seemed to dictate the use of an on-board form that was as brief as possible.

A competing force, however, was the importance of collecting origin-destination information by trip purpose from as many passengers as possible. Because the response rate would be higher on the on-board form than on the take-home form, the possibility of asking for origin-destination information on the on-board form presented itself. Obtaining such information is not simple because it means asking people for the addresses of their origin and destination as open-ended questions. This can have a number of negative impacts on the survey. First, the length of the document increases significantly (questions 4 and 6 in Figure 2). Second, these questions require writing words while one is on a moving bus rather than simply checking a box or writing one or two numbers on a line. Third, such questions very well may frustrate respondents who do not know the address of their origin or destination and they may simply stop filling out the form. Fourth, any self-administered survey is biased against the illiterate, but a semiliterate person may be able to handle a form on which he or she can read slowly and check boxes. Such a person would experience difficulty with the origin-destination questions.

Thus, it was decided to create two versions of the on-board form: the on-board short and the on-board long. The only difference between the two forms is that the on-board long contains the origin-destination questions. Figure 2 shows the on-board long form. Note that questions 4 and 6 (including the part of question 6 continued on the back of the form) occupy an entire column of the form and increase its length by about 33 percent.

Figure 2. On-board long form.



NO POSTAGE NECESSARY IF MAILED IN THE UNITED STATES

BUSINESS REPLY MAIL

a this is form a

000293

METROPOLITAN DADE COUNTY: ON-BOARD TRANSIT SURVEY

PLEASE HELP US IMPROVE THE BUS SYSTEM BY FILLING OUT AND RETURNING THIS SHORT SURVEY FORM DURING YOUR BUS RIDE. EVEN IF YOU HAVE ALREADY FILLED OUT ONE OF THESE FORMS, PLEASE COMPLETE THIS ONE ALSO. IF YOU HAVE ANY QUESTIONS, OR IF YOU NEED HELP IN FILLING OUT THIS FORM, THE SURVEY PERSON WILL BE GLAD TO HELP YOU

1 HOW LONG DID YOU WAIT FOR THIS BUS AT THE BUS STOP? ______ MINUTES

2 DID YOU USE A TRANSFER WHEN YOU GOT ON THIS BUS? YES NO

IF NO, HOW MUCH DID YOU PAY WHEN YOU GOT ON THIS BUS? 25¢ STUDENT / 25¢ SENIOR CITIZEN / 25¢ HANDICAPPED / 60¢ FARE / 75¢ EXPRESS / FREE PASS

IF YES, WHAT TYPE OF TRANSFER WAS IT? BROWARD COUNTY TRANSFER / MONEY PLUS TRANSFER / TRANSFER

FROM WHICH BUS DID YOU TRANSFER?

WHERE DID YOU GET OFF THAT BUS?

3 HOW DID YOU GET TO THE BUS STOP WHERE YOU GOT ON THIS BUS?

WALKED / BY CAR (DROPPED OFF / PARKED NEAR THE STOP) / TAXI OR JITNEY / BICYCLE, MOTORCYCLE, OR MOPED / ANOTHER BUS / OTHER

IF YOU CAME ON ANOTHER BUS, HOW DID YOU GET TO THE STOP WHERE YOU GOT ON THAT BUS?

BY CAR (DROPPED OFF / PARKED NEAR THE STOP) / WALKED / TAXI OR JITNEY / BICYCLE, MOTORCYCLE, OR MOPED / OTHER

HOW LONG DID YOU WAIT FOR THAT BUS? ______ MINUTES

4 A) WHERE DID YOU START OUT FROM BEFORE GETTING ON THIS BUS?

FROM: HOME / WORK / SCHOOL / RESTAURANT / AIRPORT / SHOPPING / A FRIEND OR RELATIVE'S HOME / HEALTH CARE (DOCTOR/DENTIST/HOSPITAL/ETC.) / HOTEL/MOTEL / RECREATION (BEACH/PARK/THEATER/MOVIES/ETC.) / OTHER

STREET CORNER/BUILDING

CITY/TOWN/VILLAGE ZIP

C) HOW FAR IS THAT FROM THE BUS STOP? BLOCKS

5 WILL YOU TRANSFER TO ANOTHER BUS WHEN YOU GET OFF THIS BUS? NO YES

IF YES, WHAT BUS WILL YOU USE?

WHERE WILL YOU GET THAT BUS?

STREET CORNER/ PLACE

6 A) WHERE ARE YOU GOING ON THIS BUS RIDE?

TO: HOME / WORK / SCHOOL / RESTAURANT / HOTEL/MOTEL / AIRPORT / SHOPPING / A FRIEND OR RELATIVE'S HOME / HEALTH CARE (DOCTOR/DENTIST/HOSPITAL/ETC.) / RECREATION (BEACH/PARK/THEATER/MOVIES/SPORTS/ETC.) / OTHER

FOR THIS BUS RIDE, COULD YOU HAVE USED A CAR INSTEAD OF THE BUS?

ARE YOU FEMALE OR MALE

IS YOUR AGE: UNDER 12 / 12-17 / 18-24 / 25-64 / 65-69 / 70 OR OVER

DO YOU HAVE A DRIVER'S LICENSE? NO YES

HOW MANY MONTHS OF THE YEAR DO YOU SPEND IN THE MIAMI AREA?

MORE THAN 9 MONTHS / 4-9 MONTHS / 1-4 MONTHS / LESS THAN 1 MONTH

IF LESS THAN 1 MONTH, HOW MANY TIMES HAVE YOU VISITED THIS AREA IN THE PAST 5 YEARS? ONCE (THIS TIME) / 2-5 TIMES / MORE THAN 5 TIMES

Table 1. Patterns of missing data for on-board form.

Question	Missing Answers: On-Board Long Form[a] No.	Percent	On-Board Short Form[b] No.	Percent
1. Waiting time	47	13.9	26	8.8
2. Transfer?	35	10.4	16	5.4
2. Type fare/transfer	37	10.9	29	9.9
3. Access mode	36	10.7	16	5.4
Avg. 1-3	39	11.5	22	7.4
4a. Origin purpose	37	10.9	NA	
4c. Distance to bus stop	167	49.4	NA	
6a. Destination purpose	47	13.4	NA	
6c. Egress mode	114	33.7	NA	
Avg. 4, 6	91	27.0	NA	
7. Captivity	45	13.3	35	11.9
8. Learn about bus	45	13.3	41	13.9
9. Sex	32	9.5	29	9.9
10. Age	35	10.4	26	8.8
11. Driver's license	55	16.3	32	10.9
12. Residence	74	21.9	49	16.7
Avg, 7-12	48	14.1	35	12.0
Avg, 1-3, 7-12	44	13.1	30	10.2
Avg, 1-12	58	17.1	NA	
Mailing list?	63	18.6	42	14.3
Comments not present	234	69.2	159	54.1

Notes: NA = not available. These questions were not asked on the on-board short form.
[a]No. distributed, 1079; no. of responses, 338; response rate, 31.3 percent.
[b]No. distributed, 1079; no. of responses, 294; response rate, 27.2 percent.

Note also the difficulty of these questions; the respondent must be able to find the antecedents of the demonstrative pronouns in questions 4b, 4c, and 6b.

It is important to note that the alternative of asking the origin-destination questions on the take-home form was used in all cases. Irrespective of the presence of these questions on the on-board form, the origin and destination of the trip were needed on the take-home form as an aid to recall the subject trip and a context-setting device for judgmental questions and questions on alternative modes.

Table 1 shows the results of the pilot study of the two versions. The number of on-board forms distributed was 2158, 1079 of each version. A 31.3 percent response rate (338 returns) was achieved for the on-board long; a 27.2 percent response rate (294 returns) was achieved for the on-board short. Because a fairly large sample (numerically) was obtained, it is possible to make statistical comparisons on some aspects of the responses. Although this is useful to distinguish between chance and systematic occurrences, it is not essential to the use of a well-designed pilot test, where reliance should be placed on qualitative assessments. These two response rates are significantly different at the 5 percent level but not at the 1 percent level (Z = 2.09), so the null hypothesis--that the addition of these two questions, although lengthening the form, would not discourage response--cannot be rejected at the 5 percent level.

In addition to the possible implications of the presence of the origin and destination questions on response rate, it is also possible, for reasons stated above, that there may be some effects on the quality of information received on the form. Many aspects of quality are difficult to assess. Thus, the surrogate variable used for judging quality is the percentage of missing answers to each question. It is recognized that this variable does not measure the quality or the accuracy of the information provided. Thus, Table 1 shows the percentage of respondents omitting answers to each question on each alternative form. In some cases, questions do not appear on Table 1 because the data were not punched in a manner that facilitated distinguishing between missing data and instances where no answer should appear because of a contingency question.

The table shows that although more on-board long forms were returned, the percentage of missing information was clearly greater on the long form. A t-test for examining for a significant difference between the average percentage of missing information for the 10 questions in common between the two forms (questions 1-3 and 7-12) shows (t = 1.79, alpha = 0.05) that the average percentage of missing information on the on-board long is significantly greater than on the on-board short. The percentage of missing information is greater both for the questions (1-3) that appear prior to the difficult origin-destination questions (4,6) and for the questions (7-12) that appear subsequently. In addition, the lengthening effect of these questions appears to have significantly reduced the percentage of respondents writing in comments (Z = 7.22, alpha = 0.05). One of three explanations is possible. First, because the origin-destination questions lengthened the form by 33 percent, respondents ran out of time and had to get off the bus. Second, respondents tired of filling out the form because it was longer. Third, after struggling to write words while they were on a moving bus for the address questions, respondents were reluctant to try to write words again in the Comments section.

Another problem with the on-board long form was that the origin-destination questions (4 and 6) were not completed well. On the on-board form, 69 percent of the responses included a usable address for the origin of the bus trip (question 4a). On the take-home form, 88 percent provided a usable origin address. This percentage might have been even higher, but no doubt some respondents completing the take-home form probably figured they had already answered the question on the on-board form and decided to skip it on the take-home form. Evidently, respondents who took the time to complete and mail back the rather complicated take-home form were not deterred by the address questions. Thus, even given the lower response rate on the take-home form, a satisfactory number of origin-destination addresses would be received on the final survey if these questions were omitted from the on-board form.

Two interesting sidelights may be noted. The first is the large percentage of missing information on questions 4c and 6c. This pointed to a design flaw in which too little space was left between 4b and 4c and between 6b and 6c, so that respondents read right over these questions. The second is that the contingency aspects of questions 2 and 3 proved too difficult for most respondents. These two questions were simplified and combined on the form for the main survey.

In addition to the tests and comparisons described, the survey designers spent a considerable amount of time reviewing individual questionnaires. They reviewed the consistency of answers among questions and the trips on which forms were given out and tried to obtain a subjective impression of the way in which forms had been completed. These reviews were also used in decisions to change or modify layouts, question-and-answer wordings, and formats.

In sum, the decision was made to produce a revised version of the on-board short form for the main survey. Although a significantly higher response rate (at the 5 percent but not at the 1 percent level) was achieved for the on-board long, the form yielded a significantly higher rate of missing information and significantly lower rate of comments. Also, tests of the take-home form seemed to yield sufficient origin-destination information for analysis purposes. That the lengthening of the form by 33 percent did not affect the response rate adversely is similar to the results shown below for the testing of alternative take-home forms. Although this result, to some extent, is at odds with conventional wisdom that states that longer forms should achieve lower response rates, it could be that both the long and the short versions of the on-board form were sufficiently short to lie within the tolerance range of the same population groups (2).

Alternative Take-Home Forms

One of the major purposes of the take-home form was to collect data to recalibrate the Dade County modal-split model. (See Figures 3-6.) Optimally, disaggregate behavioral modal-split models require individual perceptions of time and cost parameters for a selected mode and one or more alternative modes (8, Chap. 15). Because at least 13 modes can be identified in Dade County, it would obviously be beyond the patience of the vast majority of respondents to provide data on all alternative modes. Thus, an initial decision was made to query perceived time and cost parameters for the bus ride on which the respondents received the form and for three alternative modes. If a respondent provided data on at least one alternative, the response was usable for the modeling. The importance of this information as well as the obvious difficulties of asking questions about alternative modes prompted considerable attention to the modal-split questions.

Thus, three versions of the take-home form were designed. Figures 3-6 show the take-home long form. The take-home short form contains a subset of the questions on the long form (excluding the four sets of 18 mode-specific perceptual questions). The take-home table form asks in a matrix format the mode-specific time and cost questions that are asked as separate questions on the long and short forms. Each of the three versions may be separated into four sections:

Section I was devised as a warm-up section beginning with a set of perceptual questions designed to create interest (questions 1A-1M). Also included is a series of questions for devising marketing strategies (questions 2-5).

Section IV asks for information on education, income, automobile ownership, family structure (relationship, age, sex, driver's license), residential status, employment, and race. Such information is needed both for the Dade County modeling sequence and for federal reporting requirements.

Section II (questions 1-9) asks for detailed information on the bus trip the respondent was making when the form was distributed. This includes information on the land use and the address at the origin and destination, access and egress modes to the bus, and time and cost of the trip. In addition, on the long form, 18 perceptual questions are asked (question 10) about the bus ride on which the respondent received the form.

Section III asks the respondent to select three alternative modes and answer a series of questions, imagining that they had used the alternative modes instead of the bus for the trip on which they received the form. The manner in which these questions are asked varies by the version of the form. For the long form, the respondent is asked to look at a list of 13 modes and cross out the means of travel used on the day he or she received the form. Three alternative modes are then selected by the respondent to become Travel Means A, Travel Means B, and Travel Means C. The respondent must then be capable of translating his or her choices for A, B, and C to a separate page for each, where detailed questions about times, costs, and frequencies and perceptual questions are asked. (Note that, to conserve space, only the page for Travel Means A is shown in Figure 5. The pages for Travel Means B and C contain the same questions.)

The short form is designed in exactly the same manner as the long form except that the three sets of 18 perceptual questions about the alternative modes are omitted.

The table form (Figure 7) requests the same information as the short form, except that respondents are asked to fill in times, costs, and frequencies for the bus ride and three alternative modes in the cells of a matrix where the 13 modes form the rows and the modal characteristics form the columns.

All three of these formats display potential problems because either following the rather difficult procedure of translating the abstract notion of Travel Means A from one page to another or filling in the cells of a matrix is a difficult task for the bus-riding public, who may not be accustomed to filling out forms. Another difficulty is introduced because it is necessary to request people not accustomed to doing so to think hypothetically about a situation (modal choice) that they may not have thought about a great deal. This applies particularly to transit captives, who, because they lack an automobile, probably have never thought about the time and cost parameters of other modes.

Before the execution of the pilot study, the belief was that each form displayed some significant benefits. If respondents would persevere with the long form, the most information would be obtained. On the other hand, the long form was 10 pages long in comparison with 8 pages for the short form and 6 pages for the table form. If respondents could be shown to complete the matrix satisfactorily, a much shorter and simpler-looking form could be used. If the table proved unsatisfactory and the long form proved long enough to discourage response, the short form might represent the best alternative.

One other advantage of the table form was that it was possible to shade some of the cells in the matrix to indicate that no response should be placed there. On the long and short forms, all the time and cost questions had to be asked for Travel Means A, B, and C. Thus, if the respondent selected, say, walk for Travel Means A, he or she would be asked how much time was spent traveling in vehicles and finding parking. This would certainly serve to confuse some respondents. On the table form, the cells for these questions could be shaded out.

The overall response rate for the take-home form was 16.7 percent; 380 forms were returned of the 2158 distributed. Only 301 forms are included in the analysis because the others arrived too late for processing. Table 2 indicates that 97 of the 719 long forms (13.5 percent) were returned, 84 (11.7 percent) of the short forms, and 120 (16.7 percent) of the table forms. The proportion of table forms returned is significantly greater (alpha = 0.05) than both the proportion of long forms ($Z = 1.69$) and short forms ($Z = 2.72$). This is the expected result given that the table form was two pages shorter than the short form and four pages shorter than the long form. On the other hand, there exist no significant differences between the response rates of the long and short forms ($Z = 1.03$), although it is noteworthy that a greater response rate was achieved for the long form. In sum, if we

Figure 3. Take-home long form: Section I.

METRO-DADE

METROPOLITAN DADE COUNTY: TRANSIT SURVEY

b this is form b

000348

SECTION I

WE WOULD LIKE TO KNOW SOMETHING ABOUT YOUR FEELINGS ABOUT MTA (METROBUS) SERVICE HERE IN THE MIAMI AREA. PLEASE THINK ABOUT THE SERVICE YOU HAVE HAD ON MTA (METROBUS) BUSES WHILE YOU ANSWER THE QUESTIONS IN THIS SECTION

1. WE HAVE LISTED BELOW A NUMBER OF STATEMENTS THAT MIGHT BE USED TO DESCRIBE A PERSON'S FEELINGS ABOUT MTA (METROBUS) BUS SERVICE. WE WOULD LIKE YOU TO TELL US HOW STRONGLY YOU AGREE OR DISAGREE WITH EACH OF THE FOLLOWING STATEMENTS

[illegible]

	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
I LIKE THE COLOR OF THE MTA (METROBUS) BUSES	1	(2)	3	4	5	6

[illegible] (2) [illegible]

		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
A	THINKING ABOUT METROBUS IN GENERAL I AM VERY SATISFIED WITH THE SERVICE	1	2	3	4	5	6

NOW WE HAVE A NUMBER OF STATEMENTS ABOUT BUS SERVICE WHICH WE WOULD LIKE YOU TO ANSWER IN A SIMILAR WAY

		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
B	MOST BUS DRIVERS ARE POLITE TO THE PASSENGERS	1	2	3	4	5	6
C	HAVING TO WAIT TOO LONG FOR A BUS IS A MAJOR PROBLEM	1	2	3	4	5	6
D	BUS SCHEDULES ARE HARD TO FIGURE OUT	1	2	3	4	5	6
E	YOU CAN RELAX MORE IN A BUS THAN IN A CAR	1	2	3	4	5	6
F	YOU CAN USUALLY COUNT ON BUSES TO RUN ON TIME	1	2	3	4	5	6
G	WAITING FOR A BUS IN MIAMI WEATHER IS A BIG PROBLEM	1	2	3	4	5	6
H	BUS ROUTES GENERALLY SEEM TO GO WHERE I WANT TO GO	1	2	3	4	5	6
I	CRIME ON THE BUS IS A SERIOUS PROBLEM	1	2	3	4	5	6
J	BUS ROUTE MAPS ARE HARD TO FIGURE OUT	1	2	3	4	5	6
K	NEWSPAPER STORIES ABOUT THE BUS SYSTEM ARE FAIR	1	2	3	4	5	6
L	THE BUS SYSTEM IS GETTING BETTER	1	2	3	4	5	6
M	THE AGENCY THAT RUNS THE BUS SYSTEM SHOULD DEFINITELY NOT BE ALLOWED TO RUN THE NEW TRAIN SYSTEM	1	2	3	4	5	6

2. WE HAVE LISTED BELOW SOME IDEAS THAT MIGHT BE USED TO GIVE BUS RIDERS INFORMATION ABOUT THE BUS SYSTEM. WHICH 3 OF THESE IDEAS DO YOU FEEL ARE THE MOST IMPORTANT?

[illegible] (3) [illegible]

1 BUS SYSTEM INFORMATION CENTERS IN NEIGHBORHOODS
2 EXPANDED TELEPHONE INFORMATION
3 EXPANDED BILINGUAL TELEPHONE SERVICE
4 BETTER DESTINATION SIGNS ON THE OUTSIDE OF THE BUSES
5 MORE MAPS, SCHEDULES, AND NEWSLETTERS ON BUSES
6 MORE BUS SYSTEM INFORMATION IN NEWSPAPERS
7 MORE BUS SYSTEM INFORMATION ADVERTISING ON TV
8 A BUS RIDER'S GUIDE THAT TELLS YOU ABOUT BUS FARES, TRANSFER RULES ETC
9 A POSTED ROUTE MAP AT ALL BUS STOPS
10 A POSTED SCHEDULE AT ALL BUS STOPS
11 A HOT-LINE MTA INFORMATION PHONE AT MAJOR BUS STOPS
12 PERMANENT INFORMATION DISPLAYS AT SHOPPING CENTERS
13 DRIVERS WHO WILL GIVE PEOPLE MORE INFORMATION

3. DO YOU READ A MIAMI NEWSPAPER REGULARLY?

YES NO → IF NO, PLEASE GO TO QUESTION 4

IF YES, WHICH NEWSPAPER DO YOU GET OR READ REGULARLY?

- [ ] THE MIAMI HERALD
- [ ] MIAMI NEWS
- [ ] EL MIAMI HERALD
- [ ] DIARIO LAS AMERICAS
- [ ] MIAMI TIMES
- [ ] NEIGHBORHOOD NEWSPAPER
- [ ] COMMUNITY NEWSPAPER
- [ ] STUDENT NEWSPAPER

OTHER [illegible] __________

WHEN YOU READ A NEWSPAPER, WHAT SECTIONS/TOPICS DO YOU READ?

- [ ] WORLD AND NATIONAL
- [ ] LIVING OR FEATURES SECTION
- [ ] SPORTS
- [ ] FINANCIAL PAGES
- [ ] ARTS AND LEISURE
- [ ] COMICS
- [ ] DAILY TV PROGRAM SCHEDULE
- [ ] LOCAL NEWS
- [ ] FOOD TODAY
- [ ] EDITORIALS
- [ ] NEIGHBORS SECTION
- [ ] TROPIC MAGAZINE
- [ ] THE WEEKLY TV MAGAZINE
- [ ] SPECIAL SUNDAY SUPPLEMENTS

OTHER [illegible] __________

4. DO YOU LISTEN TO THE RADIO REGULARLY IN MIAMI?

YES NO → IF NO, PLEASE GO TO QUESTION 5

IF YES, WHICH RADIO STATIONS DO YOU LISTEN TO REGULARLY?

AM			FM		
[ ] WCMQ	1220		[ ] WAIA	97.3	[illegible]
[ ] WGBS	710	[illegible]	[ ] WAXY	105.9	[illegible]
[ ] WINZ	940	[illegible]	[ ] WCKO	102.7	[illegible]
[ ] WIOD	610	[illegible]	[ ] WCMQ	92.1	
[ ] WMBM	1490	[illegible]	[ ] WEDR	99.1	[illegible]
[ ] WNWS	790	[illegible]	[ ] WHYI	100.7	[illegible]
[ ] WOCN	1450	[illegible]	[ ] WINZ	94.9	[illegible]
[ ] WQAM	560	[illegible]	[ ] WLYF	101.5	[illegible]
[ ] WQBA	1140	[illegible]	[ ] WMJX	96.3	[illegible]
[ ] WRBD	1470	[illegible]	[ ] WQBA	107.5	[illegible]
[ ] WRHC	1550	[illegible]	[ ] WWWL	93.9	[illegible]
[ ] WVCG	1080	[illegible]	[ ] WYOR	105.1	[illegible]
[ ] WWOK	1260	[illegible]			

5. DO YOU WATCH TELEVISION REGULARLY IN MIAMI?

YES NO → IF NO, PLEASE GO TO QUESTION 6

IF YES, WHICH STATIONS DO YOU WATCH REGULARLY?

- [ ] CHANNEL 2 PBS-WPBT
- [ ] CHANNEL 4 CBS-WTVJ
- [ ] CHANNEL 7 NBC-WCKT
- [ ] CHANNEL 10 ABC-WPLG
- [ ] CHANNEL 6/33 IND-WCIX
- [ ] CHANNEL 23 IND-WLTV

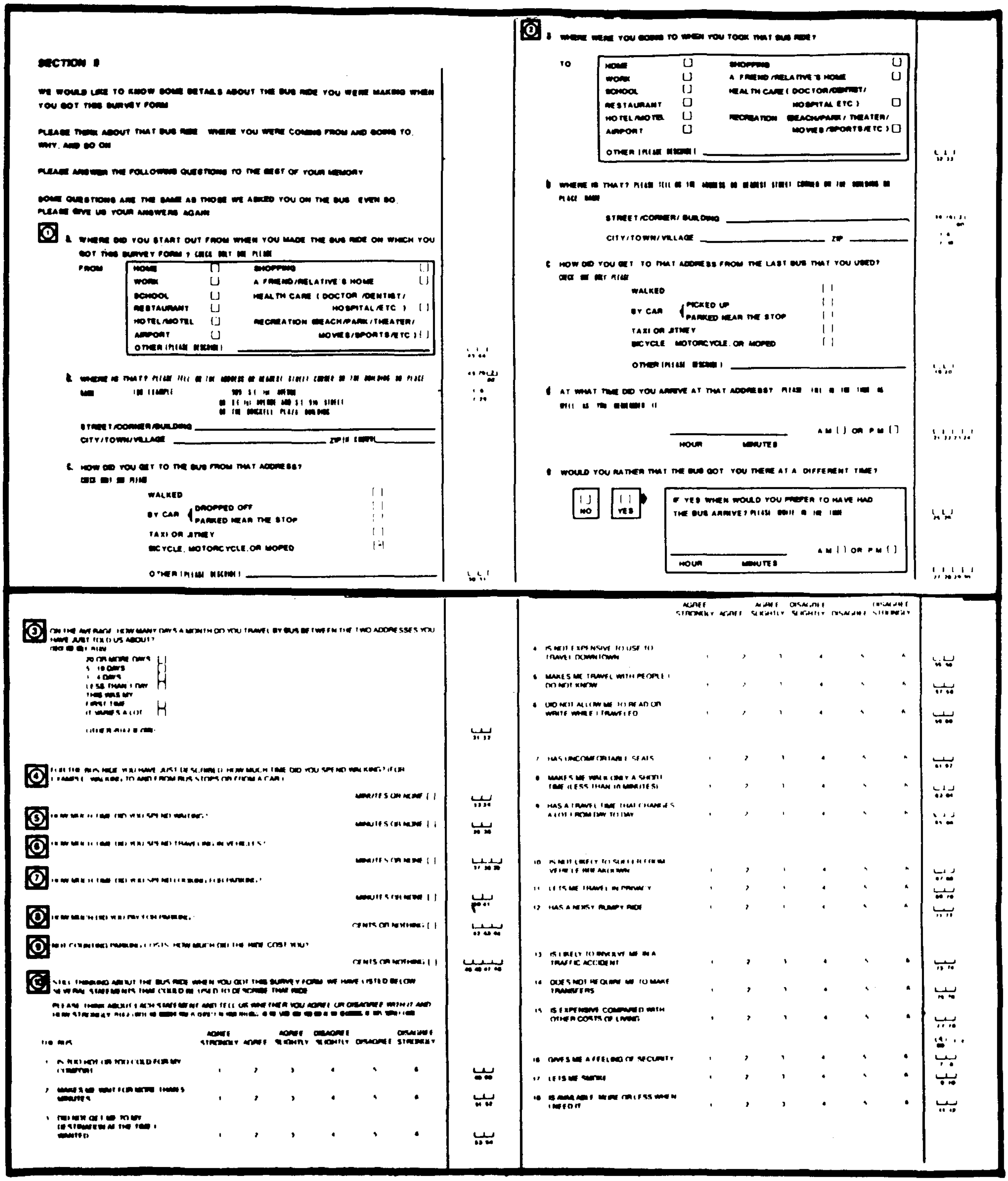
OTHER [illegible] __________

6. GENERALLY, HOW MANY DAYS A MONTH DO YOU RIDE MTA (METRO BUS) BUSES, WHEN YOU ARE IN MIAMI?

- [ ] 20 OR MORE DAYS
- [ ] 5 - 19 DAYS
- [ ] 1 - 4 DAYS
- [ ] LESS THAN 1 DAY
- [ ] THIS WAS MY FIRST TIME
- [ ] IT VARIES A LOT

OTHER [illegible] __________

Figure 4. Take-home long form: Section II.



SECTION II

WE WOULD LIKE TO KNOW SOME DETAILS ABOUT THE BUS RIDE YOU WERE MAKING WHEN YOU GOT THIS SURVEY FORM

PLEASE THINK ABOUT THAT BUS RIDE. WHERE YOU WERE COMING FROM AND GOING TO, WHY, AND SO ON

PLEASE ANSWER THE FOLLOWING QUESTIONS TO THE BEST OF YOUR MEMORY

SOME QUESTIONS ARE THE SAME AS THOSE WE ASKED YOU ON THE BUS. EVEN SO, PLEASE GIVE US YOUR ANSWERS AGAIN

1 a. WHERE DID YOU START OUT FROM WHEN YOU MADE THE BUS RIDE ON WHICH YOU GOT THIS SURVEY FORM? CHECK ONLY ONE PLACE

FROM: HOME [ ] WORK [ ] SCHOOL [ ] RESTAURANT [ ] HOTEL/MOTEL [ ] AIRPORT [ ] SHOPPING [ ] A FRIEND/RELATIVE'S HOME [ ] HEALTH CARE (DOCTOR/DENTIST/HOSPITAL/ETC.) [ ] RECREATION (BEACH/PARK/THEATER/MOVIES/SPORTS/ETC.) [ ] OTHER (PLEASE DESCRIBE) ______

b. WHERE IS THAT? PLEASE TELL US THE ADDRESS OR NEAREST STREET CORNER OR THE BUILDING OR PLACE NAME

STREET/CORNER/BUILDING ______
CITY/TOWN/VILLAGE ______ ZIP (IF KNOWN) ______

c. HOW DID YOU GET TO THE BUS FROM THAT ADDRESS? CHECK ONLY ONE PLEASE

WALKED [ ]
BY CAR: DROPPED OFF [ ] PARKED NEAR THE STOP [ ]
TAXI OR JITNEY [ ]
BICYCLE, MOTORCYCLE OR MOPED [ ]
OTHER (PLEASE DESCRIBE) ______

2 a. WHERE WERE YOU GOING TO WHEN YOU TOOK THAT BUS RIDE?

TO: HOME [ ] WORK [ ] SCHOOL [ ] RESTAURANT [ ] HOTEL/MOTEL [ ] AIRPORT [ ] SHOPPING [ ] A FRIEND/RELATIVE'S HOME [ ] HEALTH CARE (DOCTOR/DENTIST/HOSPITAL ETC.) [ ] RECREATION (BEACH/PARK/THEATER/MOVIES/SPORTS/ETC.) [ ] OTHER (PLEASE DESCRIBE) ______

b. WHERE IS THAT? PLEASE TELL US THE ADDRESS OR NEAREST STREET CORNER OR THE BUILDING OR PLACE NAME

STREET/CORNER/BUILDING ______
CITY/TOWN/VILLAGE ______ ZIP ______

c. HOW DID YOU GET TO THAT ADDRESS FROM THE LAST BUS THAT YOU USED? CHECK ONLY ONE PLEASE

WALKED [ ]
BY CAR: PICKED UP [ ] PARKED NEAR THE STOP [ ]
TAXI OR JITNEY [ ]
BICYCLE, MOTORCYCLE OR MOPED [ ]
OTHER (PLEASE DESCRIBE) ______

d. AT WHAT TIME DID YOU ARRIVE AT THAT ADDRESS?

______ HOUR ______ MINUTES A.M. [ ] OR P.M. [ ]

e. WOULD YOU RATHER THAT THE BUS GOT YOU THERE AT A DIFFERENT TIME?

[ ] NO [ ] YES → IF YES WHEN WOULD YOU PREFER TO HAVE HAD THE BUS ARRIVE? ______ HOUR ______ MINUTES A.M. [ ] OR P.M. [ ]

3 ON THE AVERAGE, HOW MANY DAYS A MONTH DO YOU TRAVEL BY BUS BETWEEN THE TWO ADDRESSES YOU HAVE JUST TOLD US ABOUT?

20 OR MORE DAYS [ ]
5–19 DAYS [ ]
1–4 DAYS [ ]
LESS THAN 1 DAY [ ]
THIS WAS MY FIRST TIME [ ]
IT VARIES A LOT [ ]
OTHER ______

4 FOR THE BUS RIDE YOU HAVE JUST DESCRIBED, HOW MUCH TIME DID YOU SPEND WALKING? (FOR EXAMPLE, WALKING TO AND FROM BUS STOPS OR FROM A CAR) MINUTES OR NONE [ ]

5 HOW MUCH TIME DID YOU SPEND WAITING? MINUTES OR NONE [ ]

6 HOW MUCH TIME DID YOU SPEND TRAVELING IN VEHICLES? MINUTES OR NONE [ ]

7 HOW MUCH TIME DID YOU SPEND ... ? MINUTES OR NONE [ ]

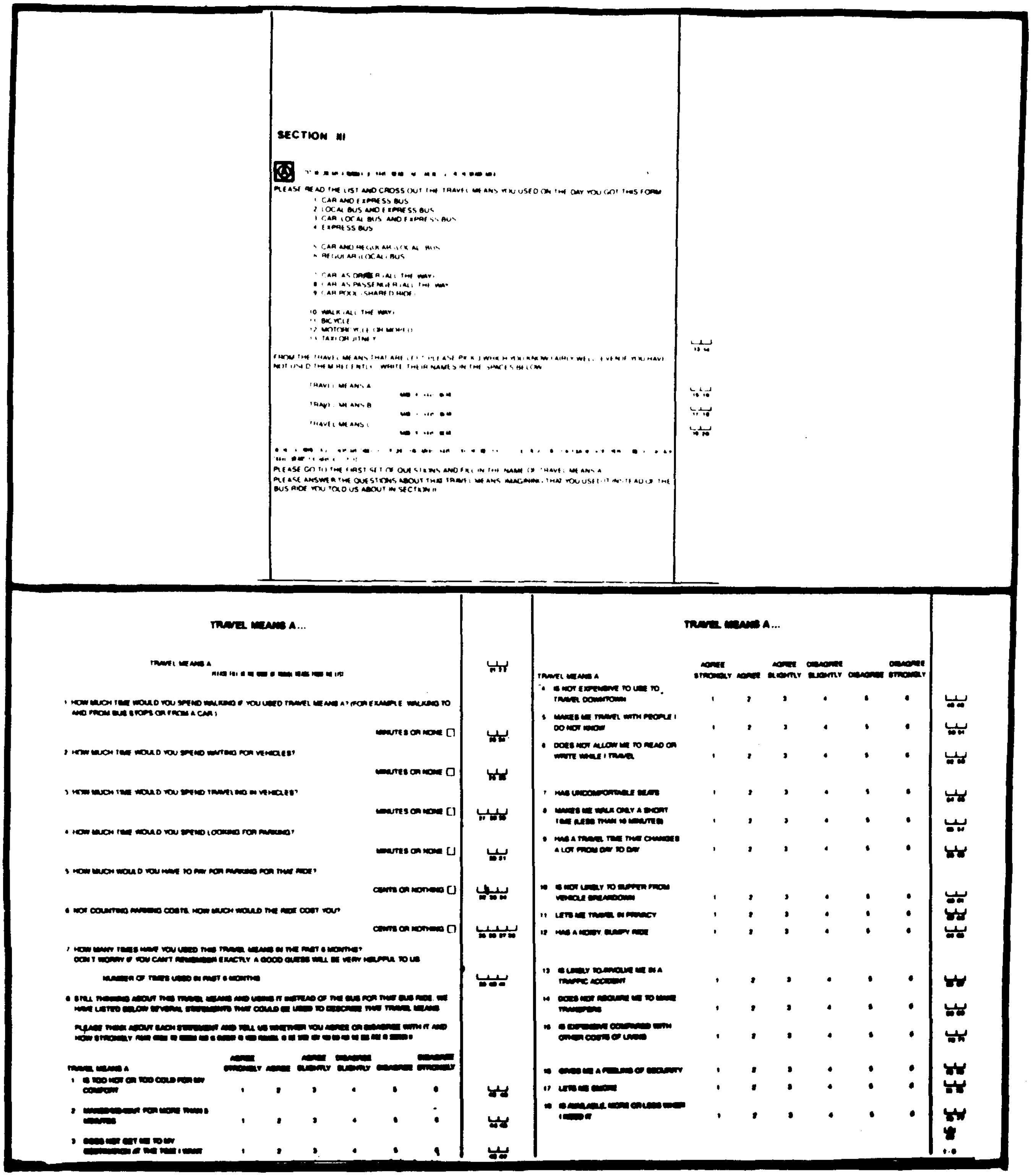
8 HOW MUCH DID YOU PAY FOR PARKING? CENTS OR NOTHING [ ]

9 NOT COUNTING PARKING COSTS, HOW MUCH DID THE RIDE COST YOU? CENTS OR NOTHING [ ]

10 STILL THINKING ABOUT THE BUS RIDE WHEN YOU GOT THIS SURVEY FORM, WE HAVE LISTED BELOW SEVERAL STATEMENTS THAT COULD BE USED TO DESCRIBE THAT RIDE. PLEASE THINK ABOUT EACH STATEMENT AND TELL US WHETHER YOU AGREE OR DISAGREE WITH IT AND HOW STRONGLY.

THE BUS	AGREE STRONGLY	AGREE	AGREE SLIGHTLY	DISAGREE SLIGHTLY	DISAGREE	DISAGREE STRONGLY
1. IS TOO HOT OR TOO COLD FOR MY COMFORT	1	2	3	4	5	6
2. MAKES ME WAIT FOR MORE THAN 5 MINUTES	1	2	3	4	5	6
3. DID NOT GET ME TO MY DESTINATION AT THE TIME I WANTED	1	2	3	4	5	6
4. IS NOT EXPENSIVE TO USE TO TRAVEL DOWNTOWN	1	2	3	4	5	6
5. MAKES ME TRAVEL WITH PEOPLE I DO NOT KNOW	1	2	3	4	5	6
6. DID NOT ALLOW ME TO READ OR WRITE WHILE I TRAVELED	1	2	3	4	5	6
7. HAS UNCOMFORTABLE SEATS	1	2	3	4	5	6
8. MAKES ME WALK ONLY A SHORT TIME (LESS THAN 10 MINUTES)	1	2	3	4	5	6
9. HAS A TRAVEL TIME THAT CHANGES A LOT FROM DAY TO DAY	1	2	3	4	5	6
10. IS NOT LIKELY TO SUFFER FROM VEHICLE BREAKDOWN	1	2	3	4	5	6
11. LETS ME TRAVEL IN PRIVACY	1	2	3	4	5	6
12. HAS A NOISY, BUMPY RIDE	1	2	3	4	5	6
13. IS LIKELY TO INVOLVE ME IN A TRAFFIC ACCIDENT	1	2	3	4	5	6
14. DOES NOT REQUIRE ME TO MAKE TRANSFERS	1	2	3	4	5	6
15. IS EXPENSIVE COMPARED WITH OTHER COSTS OF LIVING	1	2	3	4	5	6
16. GIVES ME A FEELING OF SECURITY	1	2	3	4	5	6
17. LETS ME SMOKE	1	2	3	4	5	6
18. IS AVAILABLE MORE OR LESS WHEN I NEED IT	1	2	3	4	5	6

Figure 5. Take-home long form: Section III.



SECTION III

PLEASE READ THE LIST AND CROSS OUT THE TRAVEL MEANS YOU USED ON THE DAY YOU GOT THIS FORM

1 CAR AND EXPRESS BUS
2 LOCAL BUS AND EXPRESS BUS
3 CAR, LOCAL BUS, AND EXPRESS BUS
4 EXPRESS BUS

5 CAR AND REGULAR (LOCAL) BUS
6 REGULAR (LOCAL) BUS

7 CAR AS DRIVER (ALL THE WAY)
8 CAR AS PASSENGER (ALL THE WAY)
9 CAR POOL (SHARED RIDE)

10 WALK (ALL THE WAY)
11 BICYCLE
12 MOTORCYCLE OR MOPED
13 TAXI OR JITNEY

FROM THE TRAVEL MEANS THAT ARE LEFT PLEASE PICK 3 WHICH YOU KNOW FAIRLY WELL, EVEN IF YOU HAVE NOT USED THEM RECENTLY. WRITE THEIR NAMES IN THE SPACES BELOW.

TRAVEL MEANS A

TRAVEL MEANS B

TRAVEL MEANS C

PLEASE GO TO THE FIRST SET OF QUESTIONS AND FILL IN THE NAME OF TRAVEL MEANS A.
PLEASE ANSWER THE QUESTIONS ABOUT THAT TRAVEL MEANS, IMAGINING THAT YOU USED IT INSTEAD OF THE BUS RIDE YOU TOLD US ABOUT IN SECTION II.

TRAVEL MEANS A...

TRAVEL MEANS A

1 HOW MUCH TIME WOULD YOU SPEND WALKING IF YOU USED TRAVEL MEANS A? (FOR EXAMPLE, WALKING TO AND FROM BUS STOPS OR FROM A CAR.)
MINUTES OR NONE ☐

2 HOW MUCH TIME WOULD YOU SPEND WAITING FOR VEHICLES?
MINUTES OR NONE ☐

3 HOW MUCH TIME WOULD YOU SPEND TRAVELING IN VEHICLES?
MINUTES OR NONE ☐

4 HOW MUCH TIME WOULD YOU SPEND LOOKING FOR PARKING?
MINUTES OR NONE ☐

5 HOW MUCH WOULD YOU HAVE TO PAY FOR PARKING FOR THAT RIDE?
CENTS OR NOTHING ☐

6 NOT COUNTING PARKING COSTS, HOW MUCH WOULD THE RIDE COST YOU?
CENTS OR NOTHING ☐

7 HOW MANY TIMES HAVE YOU USED THIS TRAVEL MEANS IN THE PAST 6 MONTHS?
DON'T WORRY IF YOU CAN'T REMEMBER EXACTLY. A GOOD GUESS WILL BE VERY HELPFUL TO US.
NUMBER OF TIMES USED IN PAST 6 MONTHS

8 STILL THINKING ABOUT THIS TRAVEL MEANS AND USING IT INSTEAD OF THE BUS FOR THAT BUS RIDE, WE HAVE LISTED BELOW SEVERAL STATEMENTS THAT COULD BE USED TO DESCRIBE THAT TRAVEL MEANS.
PLEASE THINK ABOUT EACH STATEMENT AND TELL US WHETHER YOU AGREE OR DISAGREE WITH IT AND HOW STRONGLY.

TRAVEL MEANS A	AGREE STRONGLY	AGREE	AGREE SLIGHTLY	DISAGREE SLIGHTLY	DISAGREE	DISAGREE STRONGLY
1 IS TOO HOT OR TOO COLD FOR MY COMFORT	1	2	3	4	5	6
2 ... FOR MORE THAN 5 MINUTES	1	2	3	4	5	6
3 DOES NOT GET ME TO MY ... AT THE TIME I WANT	1	2	3	4	5	6
4 IS NOT EXPENSIVE TO USE TO TRAVEL DOWNTOWN	1	2	3	4	5	6
5 MAKES ME TRAVEL WITH PEOPLE I DO NOT KNOW	1	2	3	4	5	6
6 DOES NOT ALLOW ME TO READ OR WRITE WHILE I TRAVEL	1	2	3	4	5	6
7 HAS UNCOMFORTABLE SEATS	1	2	3	4	5	6
8 MAKES ME WALK ONLY A SHORT TIME (LESS THAN 10 MINUTES)	1	2	3	4	5	6
9 HAS A TRAVEL TIME THAT CHANGES A LOT FROM DAY TO DAY	1	2	3	4	5	6
10 IS NOT LIKELY TO SUFFER FROM VEHICLE BREAKDOWN	1	2	3	4	5	6
11 LETS ME TRAVEL IN PRIVACY	1	2	3	4	5	6
12 HAS A NOISY, BUMPY RIDE	1	2	3	4	5	6
13 IS LIKELY TO ... ME IN A TRAFFIC ACCIDENT	1	2	3	4	5	6
14 DOES NOT REQUIRE ME TO MAKE TRANSFERS	1	2	3	4	5	6
15 IS EXPENSIVE COMPARED WITH OTHER COSTS OF LIVING	1	2	3	4	5	6
16 GIVES ME A FEELING OF SECURITY	1	2	3	4	5	6
17 LETS ME ...	1	2	3	4	5	6
18 IS AVAILABLE, MORE OR LESS WHEN I NEED IT	1	2	3	4	5	6

Figure 6. Take-home long form: Section IV.

(O) SUPPOSE THAT ALL OF THE TRAVEL MEANS LISTED BELOW ARE AVAILABLE TO YOU. AGAIN, PLEASE THINK ABOUT THE BUS RIDE ON WHICH YOU GOT THIS SURVEY FORM.
PLEASE READ THE LIST OF TRAVEL MEANS, AGAIN, THINKING ABOUT WHICH ONES YOU WOULD LIKE TO USE FOR THAT RIDE

1. CAR AND EXPRESS BUS
2. LOCAL BUS AND EXPRESS BUS
3. CAR, LOCAL BUS AND EXPRESS BUS
4. EXPRESS BUS
5. CAR AND REGULAR (LOCAL) BUS
6. REGULAR (LOCAL) BUS
7. CAR, AS DRIVER (ALL THE WAY)
8. CAR, AS PASSENGER (ALL THE WAY)
9. CARPOOL (SHARED RIDE)
10. WALK (ALL THE WAY)
11. BICYCLE
12. MOTORCYCLE OR MOPED
13. TAXI OR JITNEY

WHICH OF THE TRAVEL MEANS IN THE LIST ABOVE IS THE ONE YOU WOULD LIKE THE BEST FOR THAT RIDE? [illegible]

THE ONE I LIKE BEST IS: ____________

WHICH ONE OF THE TRAVEL MEANS WOULD YOU LIKE THE SECOND BEST FOR THAT RIDE?
THE ONE I LIKE SECOND BEST IS: ____________

AND WHICH ONE WOULD YOU LIKE THIRD BEST FOR THAT RIDE?
THE ONE I LIKE THE THIRD BEST IS: ____________

(O) STILL THINKING ABOUT THE BUS RIDE ON WHICH YOU GOT THIS FORM SUPPOSE THAT NO BUSES WERE RUNNING IN THE MIAMI AREA ON THAT DAY (THAT MEANS WE ASK YOU TO IMAGINE THAT TRAVEL MEANS 1 THROUGH 6 IN THE LIST ABOVE ARE NO LONGER AVAILABLE.)
ARE THERE ANY OTHER TRAVEL MEANS IN THAT LIST (FROM THOSE NUMBERED 7 THROUGH 13) THAT YOU THINK YOU MIGHT HAVE USED INSTEAD OF THE BUS?

☐ YES ☐ NO → IF NO, WOULD YOU SAY THAT USING THE BUS WAS THE ONLY WAY YOU COULD HAVE TRAVELED? YES ☐ NO ☐

(O) PLEASE TELL US A FEW THINGS ABOUT THE HOUSEHOLD YOU LIVE IN. WE WOULD LIKE TO KNOW HOW EACH PERSON IS RELATED TO YOU (IF AT ALL.)

[illegible]

PLEASE FILL IN EACH PERSON'S RELATIONSHIP TO YOU, THEIR AGE (OR YOUR BEST GUESS, IF YOU ARE NOT SURE), SEX, AND WHETHER OR NOT THEY HAVE A DRIVER'S LICENSE

PLEASE DO NOT INCLUDE ANYONE WHO SPENDS LESS THAN 6 MONTHS OF THE YEAR IN YOUR HOUSEHOLD
PLEASE START WITH YOURSELF AND FILL IN A LINE FOR EACH MEMBER OF YOUR HOUSEHOLD

	RELATIONSHIP	AGE	SEX	DRIVER'S LICENSE
1	MYSELF	☐	F ☐ M ☐	YES ☐ NO ☐
2		☐	F ☐ M ☐	YES ☐ NO ☐
3		☐	F ☐ M ☐	YES ☐ NO ☐
4		☐	F ☐ M ☐	YES ☐ NO ☐
5		☐	F ☐ M ☐	YES ☐ NO ☐
6		☐	F ☐ M ☐	YES ☐ NO ☐
7		☐	F ☐ M ☐	YES ☐ NO ☐
8		☐	F ☐ M ☐	YES ☐ NO ☐
9		☐	F ☐ M ☐	YES ☐ NO ☐
10		☐	F ☐ M ☐	YES ☐ NO ☐
11		☐	F ☐ M ☐	YES ☐ NO ☐
12		☐	F ☐ M ☐	YES ☐ NO ☐
13		☐	F ☐ M ☐	YES ☐ NO ☐
14		☐	F ☐ M ☐	YES ☐ NO ☐
15		☐	F ☐ M ☐	YES ☐ NO ☐

SECTION IV

WE WOULD LIKE TO KNOW A FEW DETAILS ABOUT THE PEOPLE WHO RIDE MTA (METROBUS) BUSES.
YOUR ANSWERS ARE COMPLETELY CONFIDENTIAL AND WILL NOT BE USED FOR ANYTHING OUTSIDE THIS STUDY OF THE BUS SYSTEM.

(1) PLEASE TELL US THE HIGHEST LEVEL OF EDUCATION YOU HAVE REACHED

- UP TO 6 YEARS OF SCHOOL ☐
- 7-9 YEARS OF SCHOOL ☐
- 10-12 YEARS OF SCHOOL ☐
- SOME COLLEGE ☐
- 2-YEAR DEGREE ☐
- BACHELOR'S DEGREE ☐
- SOME GRADUATE STUDY ☐
- GRADUATE DEGREE(S) ☐
- OTHER (PLEASE DESCRIBE): ____________

(2) HOW MANY MONTHS OF THE YEAR DO YOU SPEND IN THE MIAMI AREA?
[illegible]

- LESS THAN ONE MONTH ☐
- 1-6 MONTHS ☐
- 6-9 MONTHS ☐
- MORE THAN 9 MONTHS ☐ → HOW LONG HAVE YOU LIVED IN THIS AREA? [illegible]
  - LESS THAN 1 YEAR ☐
  - 1-3 YEARS ☐
  - 4-6 YEARS ☐
  - 7-10 YEARS ☐
  - OVER 10 YEARS ☐

(3) ARE YOU EMPLOYED NOW? ☐ NO ☐ YES
IF YES, IS YOUR JOB: FULL-TIME ☐ PART-TIME ☐ BOTH, FULL AND PART TIME ☐
IF NO, ARE YOU RETIRED? YES ☐ NO ☐

(4) DO YOU CONSIDER YOURSELF TO BE

- HISPANIC ☐
- LATIN AMERICAN ☐
- CARIBBEAN ISLANDER ☐
- AMERICAN INDIAN OR ALASKAN NATIVE ☐
- ASIAN ☐
- PACIFIC ISLANDER ☐
- BLACK ☐
- WHITE ☐
- OTHER (PLEASE DESCRIBE): ____________

(5) HOW MANY PRIVATE MOTOR VEHICLES (CARS, TRUCKS, MOTORCYCLE, ETC.) ARE AVAILABLE TO YOUR HOUSEHOLD IN THE MIAMI AREA?

- NONE ☐
- ONE ☐
- TWO ☐
- THREE ☐
- FOUR ☐
- FIVE OR MORE ☐

(7) PLEASE TELL US YOUR APPROXIMATE PERSONAL INCOME BEFORE TAXES, INCLUDING PENSIONS, INVESTMENTS EARNINGS, ETC. FOR THE YEAR

[illegible]

- $ 0 - $ 9,999 PER YEAR ☐
- $ 10,000 - $ 19,999 PER YEAR ☐
- $ 20,000 - $ 34,999 PER YEAR ☐
- $ 35,000 - $ 49,999 PER YEAR ☐
- $ 50,000 - AND OVER PER YEAR ☐

(8) ALSO, PLEASE TELL US (IF YOU KNOW IT) THE APPROXIMATE AMOUNT OF YOUR HOUSEHOLD'S INCOME BEFORE TAXES, INCLUDING PENSIONS, INVESTMENTS EARNINGS, ETC. FOR THE YEAR

[illegible]

- $ 0 - $ 9,999 PER YEAR ☐
- $ 10,000 - $ 19,999 PER YEAR ☐
- $ 20,000 - $ 34,999 PER YEAR ☐
- $ 35,000 - $ 49,999 PER YEAR ☐
- $ 50,000 - AND OVER PER YEAR ☐

THANK YOU FOR YOUR HELP

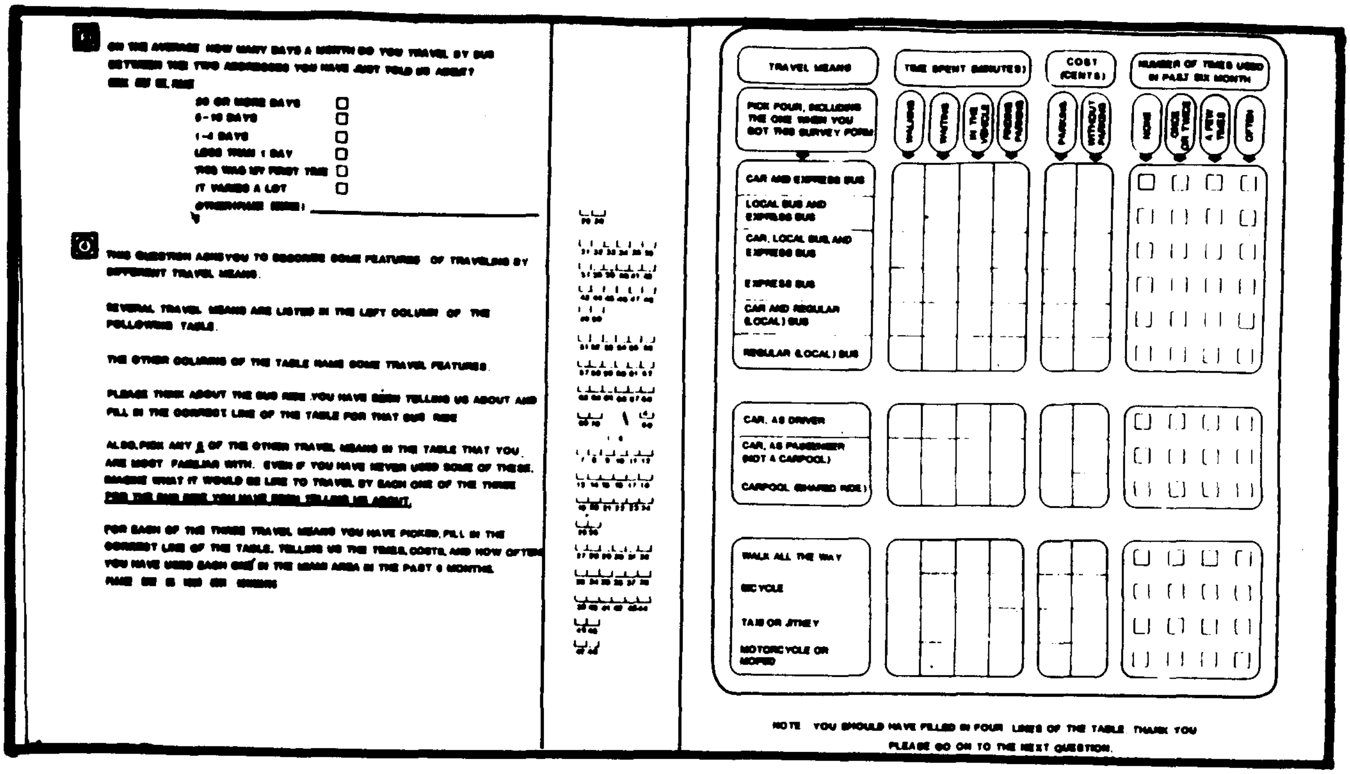
IF YOU HAVE ANY COMMENTS ON TRANSPORTATION IN DADE COUNTY, OR ON THIS SURVEY, PLEASE USE THE NEXT PAGE TO WRITE YOUR COMMENTS. WE APPRECIATE THE OPPORTUNITY TO HEAR FROM YOU.

PLEASE PLACE YOUR COMPLETE SURVEY FORM IN THE BUSINESS-REPLY ENVELOPE THAT WE PROVIDED FOR YOU AND DROP IT IN ANY MAIL BOX. NO POSTAGE IS NECESSARY.

IF YOU HAVE LOST THE ENVELOPE, PLEASE SEND THIS FORM TO:

OFFICE OF TRANSPORTATION ADMINISTRATION
METRO-DADE TRANSIT SURVEY
44 WEST FLAGLER ST., [illegible] FLOOR
MIAMI, FLORIDA [illegible]

Figure 7. Matrix page.



consider only the response rates, the table form appears best.

The quality of the information on each form, however, as measured by the percentage of missing answers for each question, leads to a different conclusion (Table 2). On the table form, an average of 39 percent of the data is missing compared with 31.9 percent for the short form and 30 percent for the long form. Although no significant differences exist (alpha = 0.05) between the average percentage missing on the short and long forms (t = 0.481) or table and short forms (t = 1.452), there is a significantly higher average percent missing on the table form than on the long form (t = 1.92). Thus, it would appear that, while the brevity of the table form induced a significantly greater percentage of persons to fill out the form, respondents obviously experienced difficulties with some of the questions.

Examining the percentage of missing information on various portions of the questionnaire reveals some insights into various aspects of questionnaire design and suggests some needed changes in the take-home form.

The attitude and marketing questions (Table 2) on pages 1 and 2 of the questionnaire were filled out relatively well on all three forms; 7.8 percent of respondents omitted answers to the attitude questions and 10.3 percent, to the marketing questions. In both cases, the long form has the least missing information, the short form the most, and the table an intermediate rate, although the differences in the rates are not great. An interesting sidelight is the unusually large number of respondents (27.9 percent) who did not answer question 1K about their perception of the fairness of newspaper stories on transit. Evidently many persons felt unqualified to answer, perhaps because they had not read any newspaper stories on the bus system.

Beginning with the bus trip parameters (Table 2) and continuing through the Means C times and costs, the superiority of the long form and the overwhelming problems of the table form become clear. For each group of questions, the average percentage of missing information on the table form is between 43 percent and 55 percent higher than on the long form. Also, in each case, the percentage of missing information on the short form is strikingly higher than on the long form. Two explanations for the lack of response to the questions in the matrix on the table form are possible. First, it is probable that many respondents were simply incapable of following instructions for the matrix and filling it in. Second, the instructions for the matrix occupy almost an entire column of the form and the matrix itself takes up one column (Figure 7). The table form contained 12 columns of questions. Respondents might have felt that it was not worth trying to figure out the matrix when it was only one question on the form, and anyway they had done their duty by answering the other questions.

There is an obvious explanation for the somewhat better results from the long form than the short form in spite of its greater length: The presence of the perceptual questions sparked respondents' interest in the form.

Again, in addition to the numerical and statistical analysis, individual forms were scrutinized carefully to look for a variety of possible indicators for change and for instrument selection. A common problem with subjective scaling questions is either receiving the same scale position selected for every statement or receiving the same ratings on each mode for a given statement. Various other more subtle patterns may also indicate that a respondent opted not to make individual and, at least partly, independent judgments on each statement. These were looked for together with illogical or improbable responses to other quantitative and qualitative questions. This scrutiny, which took place while the numerical and statistical results were being developed, pointed initially to the superiority of the long form, which was subsequently confirmed by

Table 2. Patterns of missing data for take-home form.

Question	Missing Answers: Table[a] No.	Table[a] Percent	Short Form[b] No.	Short Form[b] Percent	Long Form[c] No.	Long Form[c] Percent
Attitude						
1A. Satisfied with service	7	5.8	4	4.8	5	5.2
1B. Drivers polite	3	2.5	2	2.2	3	3.1
1C. Wait is problem	3	2.5	6	7.1	6	6.2
1D. Schedules difficult	7	5.8	6	7.7	6	6.2
1E. Relax in bus	5	4.2	3	3.6	5	5.2
1F. Bus on time	6	5.0	2	2.4	2	2.1
1G. Weather is problem	3	2.5	7	8.3	3	3.1
1H. Routes go where want	3	2.5	2	2.4	6	6.2
1I. Crime is problem	10	8.3	9	10.7	12	12.4
1J. Maps difficult	9	7.5	NA	NA	8	8.2
1K. News unfair	27	22.5	43	51.2	14	14.4
1L. Bus getting better	11	9.2	5	6.0	7	7.2
1M. Bus company runs trains	19	15.8	11	13.1	9	9.3
Avg, attitude	8.7	7.2	8.3	9.9	6.6	6.8
Marketing						
2. First idea	7	5.8	10	11.9	9	9.3
2. Second idea	10	8.3	11	13.1	10	10.3
2. Third idea	12	10.0	15	17.9	13	13.4
3. Read newspaper	15	12.5	9	10.7	8	8.2
4. Listen radio	18	15.0	10	11.9	7	7.2
5. Watch television	14	11.7	4	4.8	3	3.1
Avg. marketing	12.7	10.6	9.8	11.7	8.3	8.6
6. Frequency use bus	9	7.5	3	3.6	3	3.1
Bus trip parameters						
1A. Origin land use	2	1.7	4	4.8	4	4.1
1C. Access mode	4	3.3	3	3.6	2	2.1
2A. Destination land use	2	1.7	2	2.4	1	1.0
2C. Egress mode	15	12.5	19	22.6	5	5.2
2E. Rather arrive other time	21	17.5	15	17.9	14	14.4
3. Frequency make trip	49	40.8	8	9.5	8	8.2
Avg. bus trip parameters	15.5	12.9	8.5	10.1	5.7	5.8
Bus trip times and costs						
4. Time walking	61	50.8	15	17.9	9	9.3
5. Time waiting	55	45.8	10	11.9	8	8.2
6. Time in vehicles	60	50.0	29	34.5	26	27.0
7. Time looking for parking	99	82.5	51	60.7	46	47.4
8. Pay for parking	100	83.3	51	60.7	48	49.5
9. Cost of ride	88	73.3	35	41.7	35	36.1
Avg, bus trip times and costs	77.2	64.3	31.8	37.9	28.7	29.6
Alternative modes[d]						
Cross out mode used	38	31.7	17	20.2	13	13.4
Named Means A	65	54.2	26	31.0	24	24.7
Named Means B	76	63.3	43	51.2	35	36.1
Named Means C	86	71.7	48	57.1	41	42.3
Avg. alternative modes	66.3	55.2	33.5	39.9	28.3	29.1
Means A times and costs						
1. Time walking	88	73.3	32	38.1	22	29.1
2. Time waiting	90	75.0	NA	NA	31	32.0
3. Time in vehicles	78	65.0	NA	NA	30	30.9
4. Time looking for parking	100	83.3	50	59.5	40	41.2
5. Pay for parking	105	87.5	55	65.5	42	43.3
6. Cost of trip	101	84.2	44	52.4	41	42.3
Avg, Means A times and costs	93.7	78.1	45.3	53.9	34.3	35.4
Means B times and costs						
1. Time walking	90	75.0	46	54.8	36	37.1
2. Time waiting	101	84.2	54	64.3	41	42.3
3. Time in vehicles	83	69.2	54	64.3	45	46.4
4. Time looking for parking	98	81.7	54	64.3	46	47.4
5. Pay for parking	106	88.3	58	69.0	49	50.5
6. Cost of trip	104	86.7	56	66.7	49	50.5
Avg, Means B times and costs	97.0	80.8	53.7	63.9	44.3	45.7
Means C times and costs						
1. Time walking	95	79.2	55	65.5	46	47.4
2. Time waiting	105	87.5	57	67.9	49	50.2
3. Time in vehicles	95	79.2	56	66.7	51	52.6
4. Time looking for parking	110	91.7	61	72.6	32	33.0
5. Pay for parking	113	94.2	61	72.6	54	55.7
6. Cost of trip	109	90.8	58	69.0	50	51.5
Avg, Means C times and costs	104.5	87.1	58.0	69.0	47.0	48.4
Mode preference and captivity						
B. First preferred mode	24	20.0	21	25.0	53	54.6
B. Second preferred mode	30	25.0	27	32.1	57	58.8
B. Third preferred mode	42	35.0	31	36.9	59	60.8
C. Other modes might use	38	31.7	29	34.5	55	56.7
C. Bus only way	66	55.0	43	51.2	75	77.7
Avg, mode preference and captivity	40.0	33.3	30.2	36.0	59.8	61.6

Table 2. Continued.

	Missing Answers					
	Table[a]		Short Form[b]		Long Form[c]	
Question	No.	Percent	No.	Percent	No.	Percent
Socioeconomic						
1. Education	20	16.7	11	13.1	41	42.3
2. Months in Miami	54	45.0	42	50.0	65	67.0
3. Length of residence	19	15.8	14	16.7	49	50.5
4. Ethnic group	16	13.3	10	11.9	45	46.4
5. Age of respondent	12	10.0	20	24.0	46	47.4
5. Sex of respondent	10	8.3	18	21.4	46	47.4
5. Driver's license	18	15.0	20	23.8	45	46.4
6. Number of automobiles	6	5.0	16	19.0	48	49.5
7. Personal income	35	29.2	35	41.7	51	52.6
8. Household income	51	42.5	41	48.8	61	62.9
Avg, socioeconomic	24.1	20.8	22.7	27.0	49.7	51.2
Made comments	86	71.7	72	85.7	85	87.6
Overall avg	46.8	39.0	26.8	31.9	29.1	30.0

Notes: NA = not available. For exact wording and context of each question, see questionnaire, Figures 3-6.
[a]No. distributed, 719; no. of responses, 120; response rate, 16.7 percent.
[b]No. distributed, 719; no. of responses, 84; response rate, 11.7 percent.
[c]No. distributed, 719; no. of responses, 97; response rate, 13.5 percent.
[d]For the table form, the respondent identified these variables by filling out the rows of the matrix.

the quantitative analysis. In addition, it suggested some useful rewordings of both questions and answers and some format changes.

All these factors then pointed toward a decision to use the long form for the main survey. Two factors, however, indicated the need to make a significant modification by eliminating Travel Means C from the survey form. First, it may be noted that for all three forms (Table 2), as one looks from the questions about times and costs for the bus trip through these same questions for Travel Means A, B, and C, the percentage of missing information increases. On the long form, for instance, the percentages increase from 29.6 to 35.4 to 45.7 to 48.4. Additional evidence of this "dropping out" of respondents who evidently tired of answering the same set of questions over and over again is shown in Table 3. The percentage of missing information on the perceptual questions increases from 16.4 to 29.0 to 38.8 to 48.4 as one proceeds from This Bus Trip to Travel Means C.

The second reason for removing Travel Means C from the final version of the questionnaire was the shadow effect of the length of the modal-split questions on the completeness of the questions that followed the modal-split section. Note that for the questions about mode preference and captivity and the socioeconomic questions, the percentage of missing information on the long form is substantially greater than that for the short or table form. Evidently, when respondents tired of the modal-split questions, they did not look to see what came next but were probably sufficiently deterred by the length of the questionnaire that they simply placed it in the envelope for mailing. In fact, this effect was so severe that the missing information on the long form is of the order of twice the percentage on the short and table forms. Some slight effect is seen also in the lower percentage of respondents who wrote comments on the long form. An interesting, but not unexpected, sidelight is the large percentage of persons not responding to the income questions.

A third reason for eliminating Travel Means C was the feeling that doing so might encourage higher response rates to Travel Means A and B. That is, the respondent who, for example, worked his or her way through the questions about the bus trip and Travel Means A might have had a negative reaction to filling out the questions twice more. By reducing the repetition from four times to three, it was hoped to persuade more respondents to persevere and complete the form.

In sum, then, although the table form resulted in a significantly higher response rate, the long form was completed best by the respondents. The length of the long form, however, did result in some negative effects: a drop-off rate in answering the modal-split questions and a lower likelihood of completion of the questions following the modal-split questions. For these reasons, the final decision was to use the long form modified by the elimination of Travel Means C.

Thus, some very positive and, in the long run, cost-saving measures were learned from the rather extensive pilot study of the on-board and take-home forms. More important, a small in-house pretest on secretarial staff of the the table form had failed to uncover the full extent of the problem revealed in the pilot study. Had a decision been made to pretest just the table form on the pilot study, the problem would have been discovered and another pilot study would have been necessary to test the long and the short forms. Even worse, had a decision been made on the basis of an in-house pretest to use the table form, the expensive main survey might have failed to generate data of sufficient quality to support the modeling effort.

Midwest Regional Travel Survey

Additional advantages of performing an extensive pilot study of the survey instruments are shown by experiences on the Midwest survey. Two alternative forms and two alternative survey mechanisms were tested. The two issues to be decided involved the procedure for querying occupation and which of the two surveys (the home-interview attitude survey or the travel logs) should precede the other.

Conventional wisdom in survey research (1,9,10) indicates that asking respondents for occupational information should be done as an open-ended question with a sufficient degree of probing until the interviewer is satisfied that he or she has obtained enough information to permit a coder to categorize the respondent correctly. Three problems exist with this procedure. First, it relies on the ability of the interviewers to probe successfully. Second, the person coding the answers does not have access to the respondent (except with the trouble of a phone

Table 3. Patterns of missing data for mode-specific perceptual questions (take-home long form).

	Missing Answers[b]							
	This Bus Trip		Travel Means A		Travel Means B		Travel Means C	
Question[a]	No.	Percent	No.	Percent	No.	Percent	No.	Percent
1. Too hot or cold	12	13.4	27	27.8	36	37.1	44	45.4
2. Wait 5+ min	10	10.3	26	26.8	35	36.1	45	46.4
3. Get there on time	12	12.4	28	28.9	35	36.1	44	45.2
4. Expensive to CBD	16	16.5	29	29.9	37	38.1	46	47.4
5. Travel with strangers	16	16.5	30	30.9	36	37.1	48	49.5
6. Not allowed to read/write	18	18.6	27	27.8	36	37.1	47	48.5
7. Uncomfortable seats	15	15.5	26	26.8	38	39.2	46	47.4
8. Walk under 10 min	16	16.5	25	25.8	41	42.3	46	47.4
9. Time varies	18	18.6	30	30.9	39	40.2	49	50.5
10. Breaks down	15	15.5	28	28.9	37	38.1	50	51.5
11. Travel in privacy	18	18.6	28	28.9	39	40.2	47	48.5
12. Noisy, bumpy ride	18	18.6	27	27.8	38	39.2	47	48.5
13. Traffic accident	18	18.6	31	32.0	38	39.2	50	51.5
14. Transfers needed	17	17.5	27	27.8	37	38.1	46	47.4
15. Expensive	15	15.5	28	28.9	37	38.1	47	48.5
16. Security	14	14.4	28	28.9	35	36.1	46	47.4
17. Smoking allowed	23	23.7	32	33.0	47	48.5	52	53.6
18. Availability	14	14.4	29	29.9	36	37.1	45	46.4
Overall avg	15.9	16.4	28.1	29.0	37.6	38.8	46.9	48.4

[a] For the exact wording and context of each question, see questionnaire, Figures 3-6. [b] Based on the 97 returned take-home long questionnaires.

Figure 8. Response cards for job and occupation.

JOB CATEGORIES

A. MANUFACTURING OF TRANSPORTATION EQUIPMENT
B. OTHER MANUFACTURING
C. AGRICULTURE, FORESTRY AND FISHERY
D. MINING
E. BUSINESS SERVICES AND REPAIR SERVICES
F. PROFESSIONAL AND RELATED SERVICES
G. WHOLESALE OR RETAIL TRADE
H. FINANCE, REAL ESTATE OR INSURANCE
I. TRANSPORTATION, COMMUNICATIONS, UTILITIES
J. CONSTRUCTION
K. ENTERTAINMENT OR RECREATION SERVICES
L. GOVERNMENT
M. OTHER (Please Describe)

OCCUPATION TYPES

A. PROFESSIONAL OR TECHNICAL
B. FARMER OR FARM MANAGER
C. FARM LABORER OR FARM FOREMAN
D. OTHER LABORER
E. MANAGER, OFFICIAL, OWNER OF A BUSINESS
F. CLERICAL AND SIMILAR WORKERS
G. SALES
H. CRAFTSMAN OR FOREMAN AND SIMILAR WORKERS
I. EQUIPMENT OPERATOR OR MOTOR VEHICLE OPERATOR
J. PRIVATE HOUSEHOLD WORKER (MAID, BUTLER, ETC.)
K. OTHER SERVICE WORKER
L. MILITARY
M. OTHER (Please Describe)

call). Third, asking a respondent for both job type (agriculture, business, government, etc.) and work type (professional, manager, clerical, sales, etc.) as open-ended questions can lead to confusion as to the meaning of the questions.

Thus, a second procedure also was pretested in the pilot study. Response cards (Figure 8) were handed to the respondent with answers to each of the occupation questions. The respondent was then asked to classify himself or herself with some degree of assistance from the interviewer. Interviewers were instructed to make liberal use of the "Other" category when necessary. Note that the categories employed are those used by the U.S. Census Bureau, with some minor wording modifications. One advantage of this procedure is that the respondent is providing his or her perception of his or her occupation. Another advantage is that, because census categories are used, the main survey can be checked against the census for response bias.

At a debriefing session of the interviewers for the pilot study, the interviewers were unanimous in the opinion that the response cards should be used. Both the interviewers and interviewees were reported to have an easier time getting to what the interviewers described as more realistic answers when they employed the cards. Thus, the second procedure was adopted for the main survey.

As mentioned above, the Midwest survey consisted of an attitude survey of one randomly selected respondent and travel logs for each household member older than five years. Two possibilities existed for performing the survey:

Procedure 1: Distribute the travel logs, make an appointment to pick up the travel logs, and then do the attitude survey when picking up the travel logs (travel log first, interview after); or

Procedure 2: Do the attitude survey, distribute the travel logs, and make an appointment to pick up the travel logs (interview first, travel logs after).

Procedure 1 had the following advantages. Because the attitude survey was of very limited utility unless the travel logs were completed and a high percentage of refusals to complete the travel logs was expected, time would not be spent on the attitude survey unless the travel logs were complete. It also would permit the interviewer to probe more easily for completion and correct interpretation of the travel logs. Procedure 2, on the other hand, would permit some rapport between the interviewer and the interviewee to develop during the course of the interview. It might then be expected to be easier to convince the household to take and complete the travel logs.

Both procedures were pretested in the pilot study in which 138 households were contacted. There were 41 nonresponses, including 17 outright refusals, 1 termination, and 23 "no answers." Of the remaining 97 households, half were given travel logs first (Procedure 1); half, interviews first (Procedure 2). As shown below, Procedure 2 was clearly superior.

Response	Percent
Procedure 1	
Refusal of travel log	53
Refusal of interview	5
Completion rate	42
Procedure 2	
Refusal of interview	27
Refusal of travel log	4
Completion rate	69

When presented with the travel logs first, 53 percent of respondents refused to take them compared with a 4 percent refusal rate when the interview was done first. Evidently it is necessary to build up rapport prior to asking respondents to participate in something that, on the surface, appears to be a difficult task. Note also that, in both procedures, once respondents had complied with whatever form was presented first, very low refusal rates (4 and 5 percent) were experienced for the other form.

CONCLUSION

The benefits of testing alternative survey forms when logical arguments concerning the advantages and disadvantages of each form can be offered have been discussed. Two pilot studies designed by us--an on-board survey in Dade County and a regional travel survey in the Midwest--have been used as examples.

This paper has concentrated on one specific aspect of designing surveys and undertaking pilot studies to illuminate and inform the design process. This aspect, frequently ignored in past transportation surveys, is to test alternative designs of questions, survey instruments, or administration procedures of the survey. In the case studies illustrated, a combination of qualitative judgments and scrutiny of returned survey forms and numerical comparisons and tests was used to seek distinctions in effectiveness of the alternatives tested. In the case of the Dade on-board survey, a sufficient sample size was obtained to permit a number of statistical tests of difference between designs. This was useful to support the qualitative judgments but is not essential to the success of the strategy.

In general, pilot studies are constrained to very small samples. Considerable care and attention must be paid to the sampling for useful results to be obtained from such samples; they must be selected carefully and randomly from the same population from which the final sample will be drawn, alternative instruments or procedures must be distributed completely randomly, and all aspects of the survey must be conducted as closely as possible to the expected design of the final survey. Provided that this is done, the small sample will still provide very useful information, even if it is too small to allow statistical comparisons such as those used in the Dade County case study. A good rule of thumb seems to be to aim for a minimum of 50 responses for each alternative tested. If little or no difference, qualitatively and quantitatively, is found between such subsamples, the selection among the tested alternatives is probably not of major significance to the survey results. If large differences are found (even if they cannot be tested statistically), a good basis is provided to select one alternative over another.

In the illustrated case studies, it is apparent that without a pilot study, decisions on the alternative designs and procedures would be likely to have led to significantly lower response rates, less complete responses, or higher cost surveys than was the case after the pilot study results were used. However, the prohibitive effect of current OMB regulations on conducting such pilot studies for surveys covered by these regulations must be noted and should be a matter of major concern to those committed to improving the quality and usefulness of transportation data collection.

ACKNOWLEDGMENT

We would like to thank the staffs of the Dade County Office of Transportation Administration (OTA) and the Kaiser Transit Group (KTG) for their assistance during various phases of this research. Special thanks are due Gary Spivack and Michael Brown of OTA and Howard Eisenstadt of KTG. We also wish to acknowledge the assistance of the Southeastern Michigan Transportation Authority (SEMTA). The views expressed in the paper are ours and do not necessarily reflect the views of OTA, KTG, or SEMTA.

REFERENCES

1. E.R. Babbie. Survey Research Methods. Wadsworth Publishing Co., Inc., Belmont, CA, 1973.
2. D.A. Dillman. Mail and Telephone Surveys: The Total Design Method. New York, Wiley, 1978.
3. I.M. Sheskin, G.S. Spivack, and P.R. Stopher. Dade County's On-Board Bus Survey. Transit Journal, Spring 1981, pp. 15-28.
4. Transit Development Program. Office of Transportation Administration, Dade County, Miami, FL, Feb. 1980.
5. Preliminary Draft of Countywide Attitude Survey. Dade County Metro Transit Agency, Miami, FL, 1980.
6. I.M. Sheskin and P.R. Stopher. The Dual Survey Mechanism as a Device for Gauging Nonresponse Bias. Presented at Annual Meeting of Southeastern Division of Association of American Geographers, Blacksburg, VA, Nov. 1980.
7. I.M. Sheskin and P.R. Stopher. Transportation and Energy--Some Urban Perceptions. Proc., Transportation Research Forum, Vol. 22, No. 1, 1981, pp. 332-342.
8. P.R. Stopher and A.H. Meyburg. Survey Sampling and Multivariate Analysis for Social Scientists and Engineers. Lexington Books, Lexington, MA, 1979.
9. C.H. Backstrom and G.D. Hursh. Survey Research. Northwestern Univ. Press, Evanston, IL, 1963.
10. C.J. Dixon and B. Leach. Questionnaires and Interviews in Geographic Research. Study Group in Quantitative Methods, Institute of British Geographers, Univ. of East Anglia, Norwich, Concepts and Techniques in Modern Geography No. 18, 1979.

Vehicle Origin Survey

LARRY D. CRABTREE AND GARY KRAUSE

The vehicle origin survey (VOS) is an effective and economical method of determining the origin (home address) of motorists by using vehicle license plate numbers recorded at selected locations. The license plate numbers are matched against the national registration files of R. L. Polk and Company and based on the vehicle owner's address, various geographical codes (including zip code, census tract, and block group) are applied to establish the origin of the vehicles surveyed. Applications cover a broad spectrum including transit planning (park-and-ride and nonwork bus route needs), transportation planning (commuter traffic and origin/destination trip tables for airports, employment centers, stadiums, etc.), and environmental engineering (gas conservation and air/noise quality). In short, VOS can be of assistance to any local government or private activity or business that relies on or is related to the automobile (including light trucks) as a means of transportation. The process of selecting survey locations, the collection time periods, and the size of the sample are defined by the user and depend on the scope, extent, and intent of the survey. Data collection is a straightforward process that can be provided either by the user or by an outside collection agency. Quality control is the key element with emphasis on recording the license numbers accurately and legibly (on forms) or dictating clearly on voice tape. Survey outputs are in the form of statistical tables and computer tapes (geocoded to census geography), which can be supplemented by graphic presentations and computer dot mapping overlaying local street maps.

The majority of personal transportation needs are provided by passenger cars and light trucks. By recording vehicle license plate numbers gathered at any location (intersection, destination, etc.), the residence (origin) of the vehicle owners can be established.

This is a straightforward approach that has been proven effective but because of methodological complexity and cost has been somewhat restricted in use. The vehicle origin survey (VOS) overcomes these difficulties and offers an efficient and economical way to obtain the benefits of this technique. License plate numbers are gathered, transferred onto magnetic tape, and matched against the R. L. Polk and Company nationwide motor vehicle registration files. Matched output is provided on computer tape and summary statistical reports. Computer-generated maps can be prepared that identify the geographic location of the registered owner. The following items concerning the owner and the vehicle are provided:

1. Geography of owner's residence--county, postal town, zip code, census tract, and block group; and
2. Vehicle information--model year, fuel type, number of cylinders, and cubic-inch displacement.

In addition to the standard geographic codes listed above, other geographical indicators could be provided (e.g., traffic zones, municipalities).

Agreements with various states preclude the use of name and address of the registered owner; therefore, this information can be provided to the public sector only if written approval is granted by the appropriate state motor vehicle authorities.

SURVEY APPLICATIONS

The VOS has been used to provide essential data for various planning programs. These include the following:

1. Park-and-ride lots--Surveys were conducted in the Detroit area by Southeastern Michigan Transportation Authority (SEMTA), which identified the areas where commuter bus and commuter train riders originated. The results also indicated mileage (as the crow flies) from residence to parking location, residence location overlap between adjacent stations, and information concerning the need to extend certain routes (Figures 1 and 2).
2. Airport use--A survey of motor vehicles parked at the Greater Cincinnati Airport established the residence distribution of airline passengers throughout the metropolitan area. Figures 3 and 4 specify the origin of these vehicles by distance (1-mile increments) from the airport and also the relative vehicle density by distance from the airport. The number of vehicles from the area covered by each 1-mile concentric ring is divided by the square miles in that geographical area to determine the vehicle density per square mile.
3. Commuter parking--Data gathered at a major Cincinnati downtown commuter parking facility (Riverfront Stadium) indicated that 50 percent of the commuter vehicles were from 11 zip-code areas, 75 percent from 30 zip-code areas, and all 613 vehicles in the survey covered almost 100 zip codes (Figures 5 and 6).
4. Commuter traffic--Commuter entrances to downtown Cincinnati were surveyed at peak hours to determine the origin of vehicles at the various entry points. Figure 7 summarizes the results by census tracts. Figure 8 charts vehicle residence location for the Central and Seventh entranceway and indicates that the majority of vehicle owners using this entrance lives within a radius of 5-10 miles.
5. Bus-route planning--SEMTA is using the nonwork trip data gathered at major regional shopping centers to assist in planning nonpeak bus routes to better utilize equipment and provide a public transportation alternative for shoppers. Figures 9 and 10 illustrate the type of data used in this survey.

VOS, particularly when coupled with follow-on surveys, can also be effectively applied to a variety of other transportation studies. Technical vehicle information (vehicle type, model year, cylinders, cubic-inch displacement, etc.) included in the output could prove valuable in estimating fuel consumption and air and noise quality.

Since VOS data are coded at various geographical levels, the output is being used effectively in conjunction with other data bases. SEMTA has incorporated VOS with demographic data, employer data, and home interviews in their total planning efforts.

SURVEY STRUCTURE

The survey method is dictated by the location(s) and the purpose of the survey. Survey location, days of the week, time of day, number of days, etc., as decided on for two of the applications described above were as follows:

1. Park and ride--Gather all license plate numbers of vehicles using parking lots of park-and-ride facility. Observe vehicles and record license numbers for vehicles parking and for vehicles dropping off individuals. Observe vehicles to exclude (or separately identify) non-park-and-ride use of lots.
2. Commuter traffic--Record license plate numbers of vehicles passing the location from 7:30 to 9:00 a. m. on a typical weekday.

Figure 1. Vehicle origin survey: Jeffries and Middlebelt park-and-ride lot.

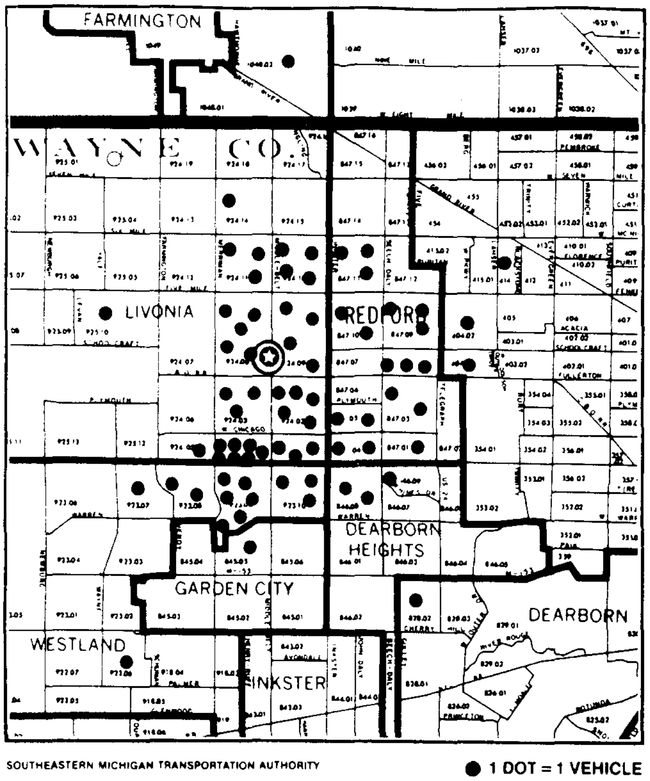


Figure 2. SEMTA park-and-ride survey.

LOCATION: FOREST CITY - JEFFRIES FREEWAY AND MIDDLEBELT

DATE	TIME	PLATE	ZIP	TRACT	COUNTY	********TOWN**
102479	920AM	LR8163	48073	1C21.00	26125	ROYAL CAK
102479	920AM	JGJ838	48024	1C48.02	26125	FARMINGTON
102479	920AM	PXD231	48060	6106.00	26125	PORT HURON
102479	920AM	NSW783	48205	64.00	26163	DETROIT
102479	920AM	OYS629	48210	122.00	26163	DETROIT
102479	920AM	DXJ761	48223	404.01	26163	DETRCIT
102479	920AM	MHL535	48223	404.02	26163	DETROIT
102479	920AM	NSZ880	48223	414.00	26163	DETRCIT
102479	920AM	MBJ565	48205	707.03	26163	DETRCIT
102479	920AM	MFJ568	48128	828.02	26163	DEARBCRN
102479	920AM	MBZ378	48135	845.05	26163	GARDEN CITY
102479	920AM	NTM833	48127	846.08	26163	DEARBCRN HTS
102479	920AM	OYK136	48127	846.08	26163	DEARBCRN HTS
102479	920AM	NTF351	48127	846.09	26163	DEARBCRN HTS
102479	920AM	RCX660	48239	847.01	26163	DETRCIT
102479	920AM	MFM398	48239	847.03	26163	DETRCIT
102479	920AM	MFR669	48239	847.03	26163	DETROIT
102479	920AM	NSY776	48239	847.04	26163	DETROIT
102479	920AM	NT8893	48239	847.04	26163	DETRCIT
102479	920AM	DK8766	48239	847.05	26163	DETRCIT
102479	920AM	PCF925	48239	847.05	26163	DETRCIT
102479	920AM	MHL461	48239	847.08	26163	DETROIT
102479	920AM	TJK625	48239	847.08	26163	DETROIT
102479	920AM	PKX853	48239	847.08	26163	DETROIT
102479	920AM	JCM925	48239	847.09	26163	DETRCIT
102479	920AM	NJC371	48239	847.09	26163	DETROIT
102479	920AM	MLX911	48239	847.09	26163	DETRCIT
102479	920AM	OJP443	48239	847.10	26163	DETROIT
102479	920AM	HHZ642	48239			
102479	920AM	CJM179	48239			
102479	920AM	DKF647	48239			
102479	920AM	CJS673	48239			
102479	920AM	DNZ325	48185			
102479	920AM	MMG188	48185			
102479	920AM	NST823	48185			
102479	920AM	RBZ158	48185			
102479	920AM	RNV257	48185			
102479	920AM	CHV789	48185			
102479	920AM	PXG510	48185	923.09	26163	WESTLAND
102479	920AM	DKH935	48185	923.09	26163	WESTLAND
102479	920AM	PCK808	48185	923.09	26163	WESTLAND
102479	920AM	OGF444	48185	923.09	26163	WESTLAND
102479	920AM	TML198	48185	923.10	26163	WESTLAND
102479	920AM	NTC816	48185	923.10	26163	WESTLAND
102479	920AM	NSY721	48185	923.10	26163	WESTLAND
102479	920AM	RSS428	48150	924.01	26163	LIVCNIA
102479	920AM	NST390	48150	924.01	26163	LIVCNIA
102479	920AM	ROP166	48150	924.01	26163	LIVCNIA
102479	920AM	NTC973	48150	924.02	26163	LIVCNIA
102479	920AM	NSV761	48150	924.02	26163	LIVCNIA

This is a partial listing of the data used to conduct a Vehicle Origin Survey for one of several SEMTA Park & Ride locations. The dot map on the opposite page illustrates by census tract the area this location serves. When evaluated in conjunction with surveys for other locations and with other data bases (demographic employment interview) the need for & effectiveness of a transit system can be determined.

PREPARED BY URBAN SCIENCE APPLICATIONS, INC.
DETROIT, MICHIGAN JAN 21, 1980

SAMPLE SIZE

A VOS can be conducted by gathering all vehicle license plate numbers or relying on a representative sample. Samples will be more effective when the total number of license plates to be recorded would otherwise be very large. The total number of license plates to be recorded depends on both the extent of the geographic areas to be covered and the level of geographic detail required.

1. Regional application--When vehicles are expected to originate from an entire standard metropolitan statistical area, or a major portion of one, a survey size equivalent to approximately 1 percent of total households (but not less than 2500 license plates) is required if reasonably adequate counts are to be expected at the census-tract level. If reliable measures by time of day and/or day of week are also desired, larger samples may be required. Increases in sample size will be necessary if data are gathered for several different survey locations.

2. Local application--Surveys with as few as 300 observations have proven effective when a low-volume location with a more localized draw is involved.

MATCH RATES

On average, approximately 75 percent of the license plate numbers are matched in the R. L. Polk and Company files. The primary factors for a 25 percent nonmatch rate are as follows:

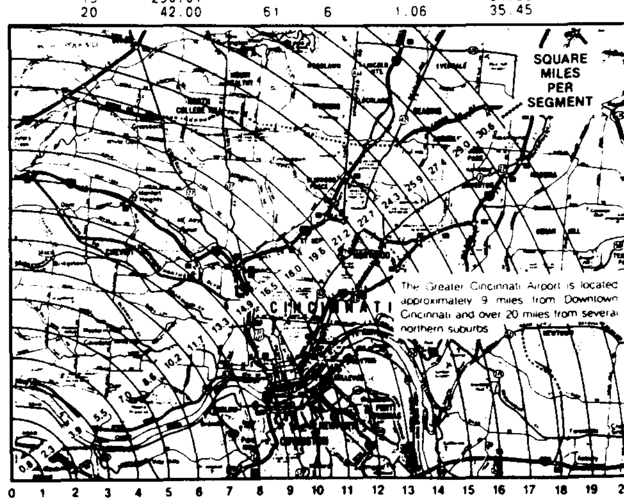
1. A portion of the vehicles originates outside the geographical area included in the study (for cost effectiveness the portion of the registration file to be searched is predefined),

Figure 3. Vehicle origin survey: September 1980, Cincinnati metropolitan area.

SAMPLE CONTAINS 567 OBSERVATIONS WITH 178 DIFFERENT CENSUS TRACTS

RANK	CENSUS TRACT	COUNTY	COUNT	% OF TOTAL	ACCUM. %
1	109.00	17	29	5.11	5.11
2	[illegible]	61	22	3.88	8.99
3	221.01	61	12	2.12	11.11
4	239.00	61	11	1.94	13.05
5	244.00	61	11	1.94	14.99
6	215.02	61	11	1.94	16.93
7	319.00				18.69
8	221.02				20.28
9	215.03				21.69
10	249.00				23.10
11	240.00				[illegible]
12	215.01				25.75
[illegible]	214.02				26.98
14	235.00				28.22
15	226.00				29.45
16	212.00				30.69
17	207.03				31.92
18	236.00				33.16
19	250.01				34.39
20	42.00	61	6	1.06	35.45

License plate numbers gathered in the parking lots (including satellite lots) at the Greater Cincinnati Airport were processed and 567 of the license plate observations were census tract coded. These 567 vehicles originated from 178 different census tracts and the tract with the greatest number was 109.00 (29 vehicles or 5.11% of the total). The cumulative percent (Accum. %) indicates that 25% of airport usage originated from 12 census tracts.



2. There are errors in the collection or preparation of data, and

3. New license plate numbers are not yet in the registration files (the vehicle registration files are updated from one to three times per year depending on the state involved).

An analysis of the unmatched records found in vari-

Figure 4. Vehicle origin survey: Greater Cincinnati Airport.

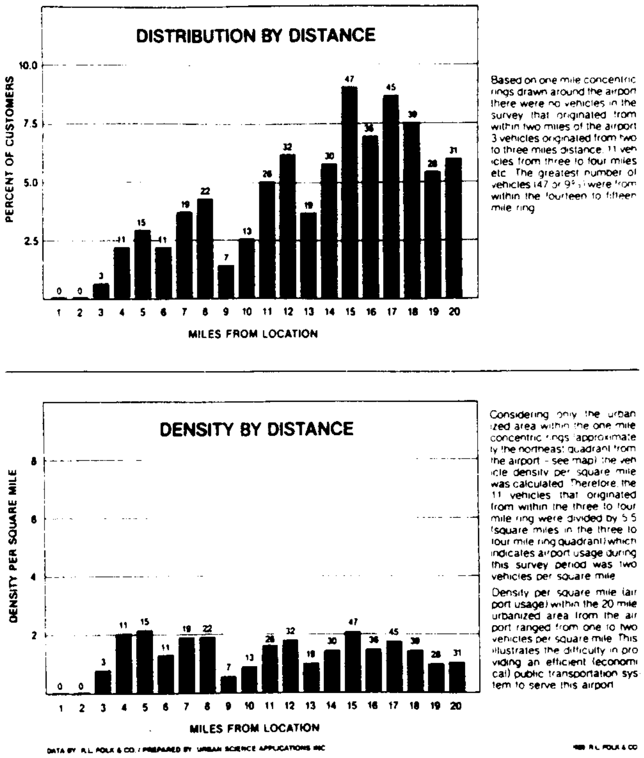


Figure 5. Vehicle origin survey: September 1980, Riverfront Stadium.

TOTAL SURVEY CONTAINS 613 OBSERVATIONS WITH 96 DIFFERENT ZIP CODES

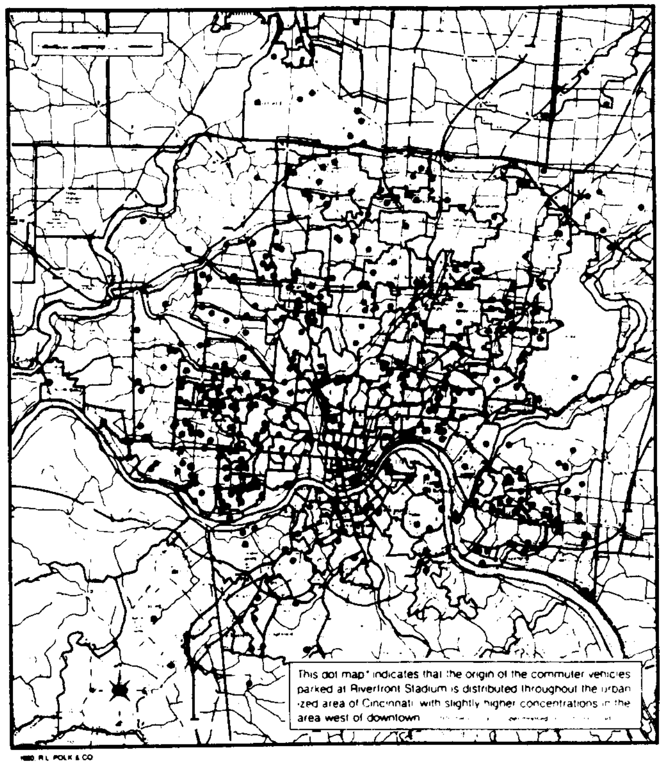
RANK	ZIP CODE	COUNT	% OF TOTAL	ACCUM. %
1	45238	75	12.23	12.23
2	[illegible]	[illegible]	[illegible]	20.39
3	45230	42	6.85	27.24
4	45239	30	4.89	32.14
5	45242	23	3.75	35.89
6	45231	19	3.10	38.99
7	45208	15	2.45	41.44
8	45244	15	2.45	43.88
9	45224	15	2.45	46.33
10	45215	14	2.28	[illegible]
11	41017	14	2.28	50.90
12	45220	13	2.12	[illegible]
13	45240	13	2.12	55.14
14	41075	11	1.79	56.93
15	45236	10	1.63	58.56
16	45014	10	1.63	60.20
17	45243	9	1.47	61.66
18	45245	9	1.47	63.13
19	45204	9	1.47	64.60
20	45237	8	1.31	65.91
21	45213	7	1.14	67.05
22	45223	7	1.14	68.19
23	45246	7	1.14	69.33
24	45140	7	1.14	70.47
25	45227	6	0.98	71.45
26	45212	6	0.98	72.43
27	45209	6	0.98	73.41
28	45002	5	0.82	[illegible]
29	45206	5	0.82	75.04
30	45205	5	0.82	[illegible]
31	45229	5	0.82	76.67
32	41011	5	0.82	77.49
33	45241	5	0.82	78.30
34	45225	5	0.82	79.12
35	45233	5	0.82	79.93
36	45216			
37	45219			
38	45226			
39	45069			
40	45102			
41	41042			
42	41015			
43	45013			
44	45030			
45	45150			
46	45217	3	0.49	86.79
47	45202	3	0.49	87.28
48	45001	3	0.49	87.76
49	45015	3	0.49	88.25
50	45036	3	0.49	88.74

Riverfront Stadium and adjacent areas are used as commuter parking lots during normal working hours. A survey of these lots during a work day (when special events were not taking place) resulted in 613 vehicles originating from 96 different zip codes.

Zip code 45238 had the greatest number of vehicles (75 or 12.23% of the total). The last column (cumulative %) indicates 50.90% of the vehicles originated from only 11 zip code areas, and 75.04% from 29 zip codes.

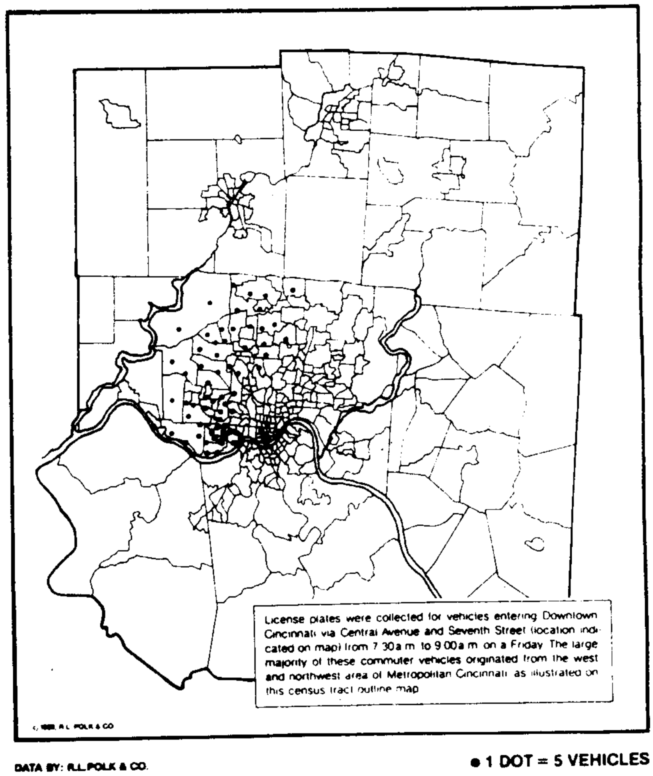
PREPARED BY URBAN SCIENCE APPLICATIONS, INC.
DATA BY: R.L. POLK & CO NOV 25, 1980

Figure 6. Vehicle origin survey: commuter parking, Riverfront Stadium area.



DATA BY: R.L. POLK & CO. / PREPARED BY: URBAN SCIENCE APPLICATIONS INC. ● 1 DOT = 1 VEHICLE

Figure 7. Vehicle origin survey: commuting traffic, Central and Seventh, Cincinnati metropolitan area.



DATA BY: R.L.POLK & CO. ● 1 DOT = 5 VEHICLES

Figure 8. Vehicle origin survey: distribution by distance, Cincinnati metropolitan area.

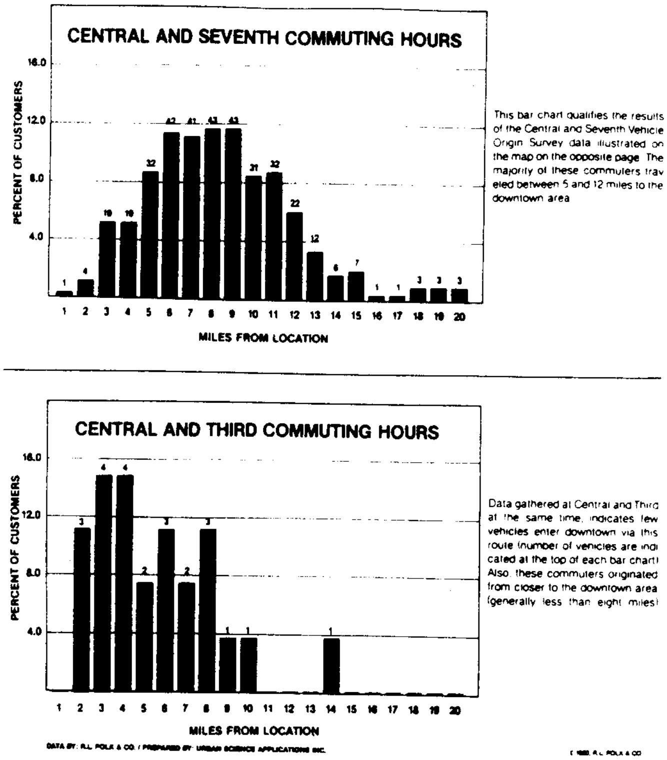
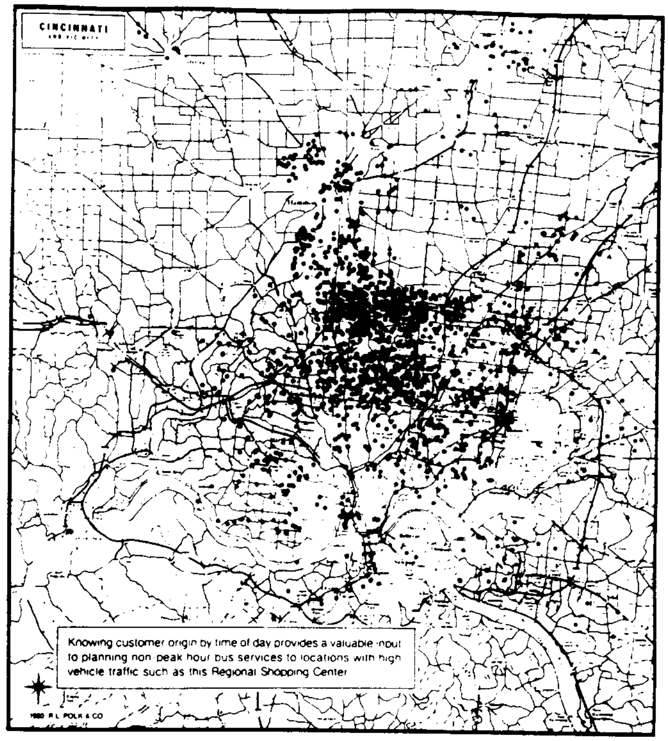


Figure 10. Vehicle origin survey: Tricounty Regional Mall, Cincinnati metropolitan area.



DATA BY: R.L. POLK & CO. / PREPARED BY: URBAN SCIENCE APPLICATIONS INC. • 1 DOT = 1 VEHICLE

Figure 9. Vehicle origin survey: Regional Mall, Cincinnati metropolitan area.

TOTAL SURVEY CONTAINS 2041 OBSERVATIONS WITH 134 DIFFERENT ZIP CODES

RANK	ZIP CODE	COUNT	% OF TOTAL	ACCUM. %
1	45240	190	9.31	9.31
2	45231	153	7.50	16.81
3	45014	147	7.20	24.01
4	45246	143	7.01	31.01
5	45215	137	6.71	37.73
6	45069	124	6.08	43.80
7	45241	88	4.31	48.11
8	45239	76	3.72	51.84
9	45011	72	3.53	55.36
10	45242	62	3.04	58.40
11	45013	52	2.55	60.95
12	45218	50	2.45	63.40
13	45140	44	2.16	65.56
14	45040	43	2.11	67.66
15	45042	42	2.06	69.72
16	45236	36	1.76	71.48
17	45224	34	1.67	73.15
18	45211	32	1.57	74.72
19	45237	29	1.42	76.14
20	45238	27	1.32	77.46
21	45230	22	1.08	78.54
22	45015	20	0.98	79.52
23	45247	19	0.93	80.45
24	45150	18	0.88	81.33
25	45212	18	0.88	82.21
26	45216	18	0.88	83.10
27	45223	16	0.78	83.88
28	45220	15	0.73	84.62
29	45243	14	0.69	85.30
30	45213	13	0.64	85.94
31	45056	13	0.64	86.57
32	45227	13	0.64	87.21
33	45208	12	0.59	87.80
34	45229	11	0.54	88.34
35	45209	10	0.49	88.83
36	45030	9	0.44	89.27
37	45005	9	0.44	89.71
38	45232	9	0.44	90.15
39	45036	8	0.39	90.54
40	41011	6	0.29	90.84
41	45039	6	0.29	91.13
42	45122	6	0.29	91.43
43	45157	5	0.24	91.67
44	45050	5	0.24	91.92
45	41017	5	0.24	92.16
46	45205	4	0.20	92.36
47	45414	4	0.20	92.55
48	45370	4	0.20	92.75
49	45219	4	0.20	92.94
50	45133	4	0.20	93.14

PREPARED BY URBAN SCIENCE APPLICATIONS, INC.
DATA BY: R.L. POLK & CO. OCT 17, 1980

ous surveys has been carried out; the results are as follows:

Factor	Percent
Outside survey area	10
Not in state file (input error)	8
In state file, not yet in R. L. Polk and Company file	7
	25

Further analysis of the 7 percent not yet in Polk files indicated that these were generally distributed in the same manner as the matched license plates.

The match rate is used as a quality control measure to evaluate surveys from market to market and from time period to time period. Also, by assigning codes to data collectors and keypunch operators, the accuracy of recording and preparing data can be measured by comparing individuals' match rates to the norm.

DATA-GATHERING TECHNIQUE

Gathering data is more logistical than technical. Users can easily collect their own data or employ outside data collection. The technique for gathering the numbers, like the sample size, is dictated by the survey location and purpose. A commonsense approach is required:

1. Stationary point--Moving traffic requires a vantage point that gives a clear view of all vehicles (from the rear only in one-license-plate states). The collector must be close enough (preferably within 50 ft) for visual observation of the license numbers on vehicles in all traffic lanes. This can be accomplished by standing on a sidewalk, on the shoulder of the road, or even on an overpass. If a low profile is desired, the collector

can sit in a parked vehicle adjacent to the observation point. This same technique can be followed to record license plates as cars enter downtown parking garages.

2. Walking--An area such as a commercial strip that is congested with vehicles scattered throughout requires the data collector to walk. It is too difficult and often dangerous to drive under these conditions and effectively collect the data.

3. Moving vehicle--Large parking lots (e.g., regional shopping centers and commuter parking lots) that hold great numbers of vehicles can efficiently be surveyed from a moving vehicle. Traffic is usually light and the lot can be driven slowly to ensure accurate collection of data.

SURVEY METHODOLOGY

Two methods of recording data have been used effectively to date. These are forms and voice tape. Each has advantages and disadvantages that, in part, depend on the location as well as the individual collector.

1. Forms
   a. Use when small number of vehicles per site expected or small sample collected
   b. Maximum of 300 license plate numbers per hour can be recorded
   c. Requires less training and skill than voice recording
   d. Forms easily controlled and audited
   e. Writing, however, must be legible so as not to confuse 8 and B, 2 and Z, etc.
2. Voice tape
   a. Use when large sample required and vehicles concentrated
   b. 500 plate numbers per hour easily recorded
   c. More effective when collecting from moving vehicle
   d. Diction and enunciation very important (use words instead of letters--"Able" or "Apple" for A, etc., and say "Stop" after each license plate number)
   e. More equipment, greater expense, and sometimes technical problems
   f. Data preparation personnel must be trained to keypunch accurately from voice tape
   g. Greater potential for error

Cameras and hand-held keyboard entry directly to tape are also available but have not been used in VOS and thus cannot be evaluated at this time.

CONCLUSION

Experience to date has demonstrated that the VOS can provide a cost-efficient, highly useful data input to the overall transportation planning process, which can be further enhanced by follow-on surveys of motor vehicle owners. In addition, the ability to computer-map motor vehicle origin by census tract (or other small area) provides the professional and nonprofessional alike with an immediately understandable picture of the commuter and nonwork trip "marketplace" by specific destination. And when coupled with total vehicle ownership by census tract, demographics, etc., this service provides other measures such as "market penetration" relationship between public and private transportation use at the small geographic area level.

The service includes output tapes, statistical reports, and computer mapping. It is important to restate that name and address of registered owners are not available to the commercial or private sector and only available to the public sector when written approval is granted by the appropriate state motor vehicle authorities.

Analysis of Employee Residential Locations for Transit Planning

RAI PARVATANENI AND TIMOTHY LAMBERT

The development of a data base that describes the residential locations of employees working in the Detroit central business district (CBD) and adjoining major activity centers is described. The data base helped to conduct immediate and short-term transit service planning functions of the Southeastern Michigan Transportation Authority in the Detroit metropolitan area. This data-base development was undertaken because of the limitations of the existing sources that describe the work-related travel. Data describing the employee residential locations of selected major employers were gathered from personnel departments. The employers provided either an address list of their employees or summaries by zip-code locations. The residential locational descriptions of 33 555 employees for the CBD and 34 583 employees for the adjoining activity center represented sample rates of 31 and 52 percent of the total employment. An expansion methodology was developed and deployed to project sample data to the total employment population for 1980. Further, 1985 residential location projections were made by using the base-year data and regional population and employment-growth factors. The base-year location data at census-tract level for each employer or groups of employers and summaries for the total employment became valuable information in instituting peak-period route services; existing services were modified and route-effectiveness measures were developed. The base-year and 1985 data were also used in short-term transit service planning.

The Southeastern Michigan Transportation Authority (SEMTA) plans, constructs, and operates public transportation facilities and services. Although the authority's area of jurisdiction covers the seven counties of southeastern Michigan, SEMTA primarily serves suburban to downtown Detroit commuter travelers and travel demands between suburban communities. Under a purchase-of-service agreement, SEMTA is also responsible for Detroit services operated by the City of Detroit Department of Transportation (DDOT).

Although SEMTA was created in 1967, the authority's operations actually began in 1971, with the first of several purchases of private carriers. Over the years, SEMTA ridership has steadily increased. Ridership since 1974 has increased at an annual rate of 13 percent from 7.1 million to more than 13.4 million annual passengers. Because of the trend toward ridership increases, SEMTA will have to carefully monitor and plan for future service improvements that will accommodate the potential ridership.

Currently, a significant portion of the travel market for SEMTA services is those traveling for work purposes. A recently conducted transit user survey indicates that approximately 90 percent of peak-period SEMTA users and 50 percent of DDOT users belong to this group. But only 25-30 percent of the total downtown-oriented commuting travelers use the transit service; therefore, potential exists to enlarge the transit market by attracting automobile users to public transportation.

To better serve commuter travel and increase transit ridership by diverting automobile users to public transportation, the current commuter travel behavior should be better understood. A review of existing information on commuter travel behavior showed many limitations for use in service planning.

Although the U.S. Census Bureau, through the decennial and annual housing surveys (1,2), provides information on the residential and employment ends of journey-to-work travel, there are several drawbacks for use in service planning. Primarily, the release of the census data, often three to four years from the survey date, makes these data less useful in route planning, which requires a more current data base.

Second, census data are gathered on a small-sample basis, which yields aggregate travel movements in the region. Although these data are at the analysis zonal level when released, they do not focus sufficiently on the trip end. That is, information on the commuter's work location is not specific to a particular establishment; rather, it is limited to respective analysis zones in the employment center. Although this allows trip patterns to be identified, marketing efforts at specific, high-potential employers are not possible. The Institute of Transportation Engineers Committee 6A-12 (3) examined the applications and limitations of the 1980 census data and recommended additional data-collection activities, including the employer surveys to supplement the census data.

Similarly, the home-based work travel data developed from the application of traditional travel-demand modeling chain would not provide the detail needed in service planning. The Transportation and Land Use Study (TALUS) (4), conducted in 1965, represents travel patterns now obsolete due to significant regional urban sprawl, varied energy supplies, and demographic changes during the past 15 years.

Because of these limitations and the recognized need to supplement the 1980 U.S. Census results, a data base describing the residential locations of employees in selected activity centers was developed to enhance the understanding of work travel that affects the design of transit services. This paper describes the procedures for collecting representative residential location data and a methodology for expanding the sample data to the total employment population in selected employment centers. The collection of the sample location data base was focused on major employers in the Detroit metropolitan region because of the large number of work trips generated by these firms. The application of the expansion methodology resulted in the estimated census-tract level residential locations of all employees working in the Detroit central business district (CBD) and adjacent central functions area (CFA). This distribution represents the total potential work-travel market for the delivery of public transportation.

This paper also describes a methodology to forecast 1985 residential locations of CBD/CFA employees by using the 1980 base-year data and regional population and employment-growth factors. Further, it provides a summary of the varied applications of the data in both immediate and short-term transit planning.

STUDY AREA

Although public transportation is provided in all seven counties, the primary service area is the City of Detroit and the adjoining three-county area (Figure 1). Currently, the transit service outside this area mainly serves the elderly and the handicapped and, to some extent, internal travel within a few satellite cities.

As in most large U.S. cities, the maximum peak-period travel is oriented toward the most densely business-populated area of the region. This area, shown in Figure 2, covers two activity centers, namely, the Detroit CBD bounded by the freeways and the CFA adjoining the CBD on the north side. The employment densities are 125 547 employees/mile^2 and 24 015 employees/mile^2, respectively. This paper describes the collection of the employee residential location data through contacts with employers in these areas and the analysis of those data to develop the potential transit demand for the travel made from various points in the tricounty area to the two activity centers.

COLLECTION AND PROCESSING OF EMPLOYEE RESIDENTIAL LOCATION DATA

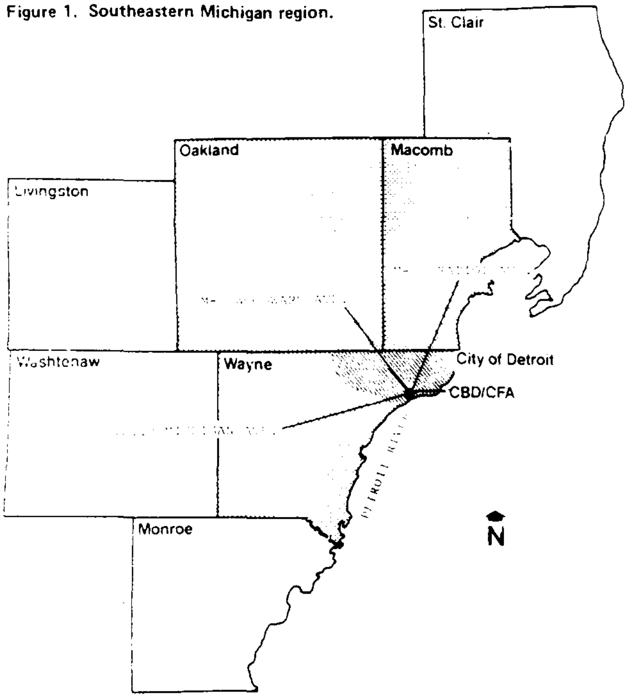
Employee residential data were collected from major employers to produce a representative sample of various industrial employment categories. The approach for data collection consisted of a planning/marketing staff team that provided the employer with the following information in an arranged meeting and requested the residence location of their employees:

1. Description of existing transit services to the employment location,
2. Planned service improvements,
3. Company's opportunity to participate in planning, and
4. Company's opportunity to market public transit to its workers.

The interest exhibited by the employers in reaction to the team's marketing approach accelerated the rate of data collection and added to SEMTA's credibility in the business community. Requests were made for employees' home address lists (names deleted) with the street address, the name of the city/township, and the zip code. Data generation usually posed no problem, since most employer personnel files were computerized. However, some employers did not provide specific address data for reasons of confidentiality but did provide aggregated summaries of the number of employee residences in each zip-code area. A record of the information gathered for each employer is maintained on standardized forms. These contain the following information:

1. Business name and address,
2. Business contact person,
3. SEMTA contact person,
4. Contact date,
5. Level of aggregation (i.e., addresses versus zip codes), and
6. Format of data (i.e., hard-copy printout and/or magnetic tape).

Figure 1. Southeastern Michigan region.



For those employers who supplied address information and not the zip-code summaries, the data were summarized by zip code through manual tabulation. The reason for these tabulations of address data is that the data could be readily used in the service planning; a delay of about six months, which is the normal time needed for complete geoprocessing (i.e., assigning addresses to census tracts) of the data and preparation of summaries, would have made the data inadequate for use. To perform automated geoprocessing in a cost-effective way, it was better to accumulate several employer data sets.

The address information of six CBD employers was geoprocessed by using the U.S. Census ADMATCH computer program to allow analysis at a finer level of geography. The geoprocessing consists of associating the census tract number (1970 tract geography) with each of the address records by using the DIME Geographic Base File (GBF) and the ADMATCH programs. This resulted in an address-to-census tract match rate of approximately 75 percent.

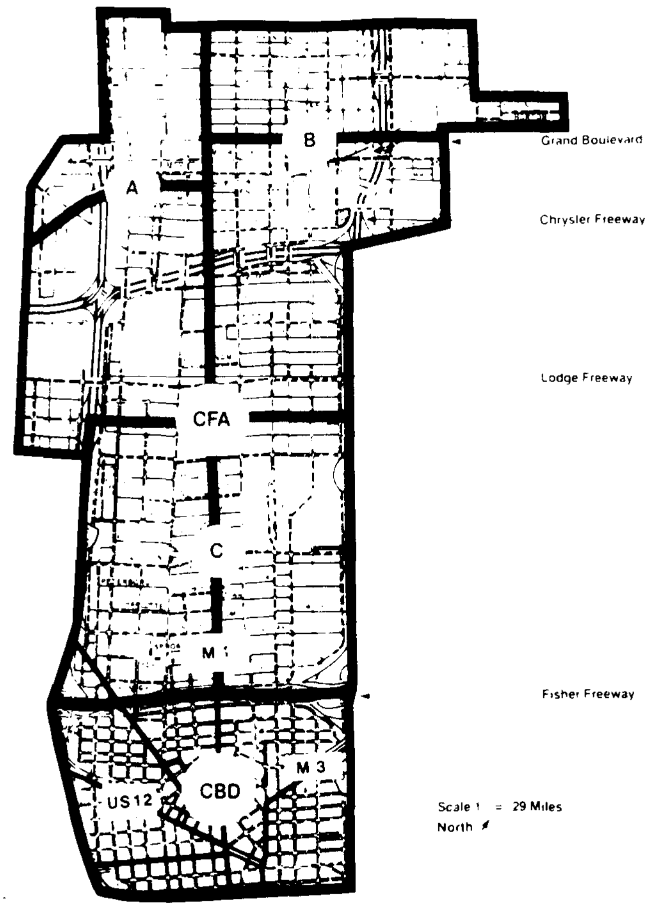
Records were unmatched if the address was outside the immediate tricounty area, address input was misspelled, or the address had an inexact or new street identifier. The remainder of the unmatched records were manually geoprocessed to identify the census tracts. Then summaries of residence locations by 1970 census tracts were derived from the geoprocessed data.

The analysis reported in this paper used the employee residential location data for 6 employers at census-tract level, which accounted for 15 000 employees, and for another 10 employers at zip-code level, which accounted for another 17 000 employees. This represented a sample of more than 32 000 workers (30 percent) of 106 715 employees in the CBD. For the CFA, the data from 10 employers at zip-code level were used, which accounted for 34 583 employees (52 percent) of a total of 66 042.

EMPLOYEE DATA EXPANSION FOR DETROIT CBD

To identify the potential travel-demand areas, the sample employee home location data were systematically expanded to the total CBD employee population.

Figure 2. CBD and CFA activity-center boundaries.

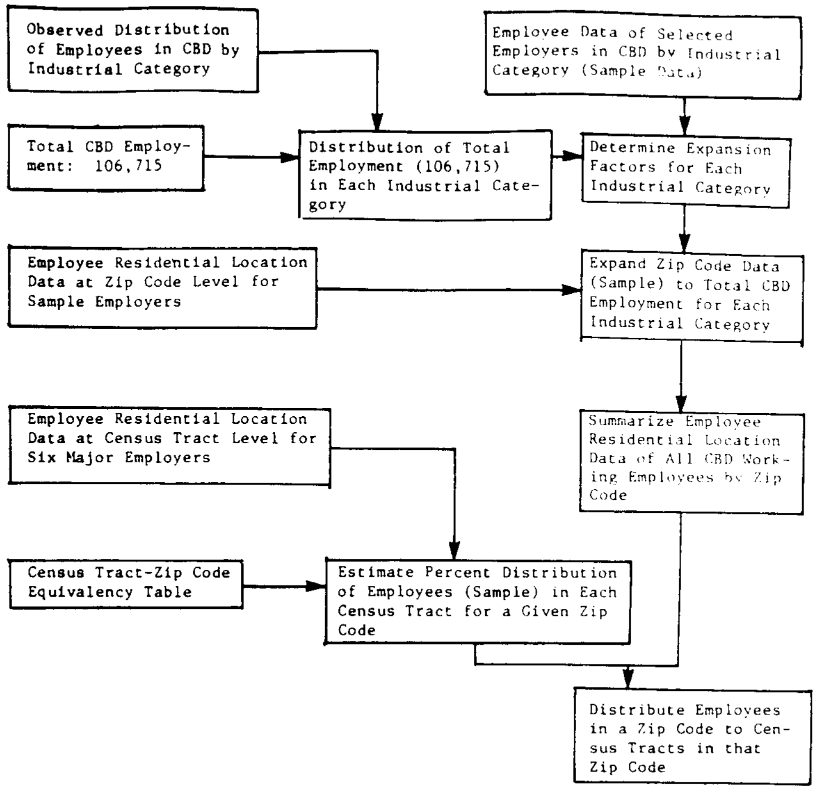


The purpose of this process was to develop total demand set at a disaggregate level (census tract) so that the data could be used directly in transit planning. Because the CBD and the CFA had distinctly different sample rates, separate methods were developed and used for expanding sample data to the total residential location distributions.

The expansion process is shown in the flowchart presented in Figure 3. The flowchart depicts the zip-code level expansion process from the sample data to the total CBD employment. The flowchart then shows the process by which these zip-code level data representing the total CBD employment were distributed to the census-tract level. The census-tract distribution was based on the distribution of the address data of the previously mentioned six CBD employers. The assumptions used in this process were as follows:

1. That major CBD employers in a single industrial category show similar employee home location distributions and that, conversely, the distributions would differ between types of industries; and
2. That the distribution of employee living patterns among various census tracts within a given zip code is not in the same proportion to the number of households in the census tracts or the number of persons in the tracts. (This hypothesis was validated by comparing the distribution of employee residential locations of selected employees who had supplied their address data against the distribution of households.)

Figure 3. Process for deriving residential locations of all CBD employees.



Derivation of CBD Expansion Factors

A direct expansion of the sample data to the total employment was not appropriate because the sample data were an inaccurate representation of the total employment in the CBD and the respective proportions for each type of industry. This meant that an expansion within similar industrial categories was more appropriate.

The industrial classifications used in developing regional small-area forecasts were considered appropriate in this analysis. Some adjustments to these classifications consisted of further groupings of regional categories when there was insignificant sample size in any category. The final industry categories used were the following:

1. Manufacturing (automobile and other);
2. Transportation, communications, utilities;
3. Wholesale and retail trade;
4. Finance, banks, insurance;
5. Public administration; and
6. Natural resources, construction, business and professional services.

All major employers who provided location data were classified in one of the categories presented in Table 1. To maintain confidentiality, the names of employers are not shown.

Table 1 also lists the total number of employees from the sample in each category. The data from one automotive employer were excluded from Table 1 because this employer recently moved from the suburbs to the Detroit CBD and its residential locations were atypical.

Table 1 also estimates the total number of employees working in the CBD by industrial category. These estimates were based on the control total of 106 715 employees, distributed in each category, based on employment data from the Michigan Employment Security Commission (MESC).

The next column in Table 1 indicates the total number of employees, excluding the one automotive company and all federal government employees. A total of 5674 federal employees were estimated to be working in the CBD. The residential location data from these employers were not available, but home locations are most likely distributed throughout the region because, unlike local government employees, no residency requirements exist for federal employees. The last column in Table 1 lists the expansion factor to project sample data to the control totals within each industry category.

Expansion Process at Zip-Code Level

The actual expansion from the sample data to the control totals within each category was performed at zip-code level; the expansion factors are those shown in Table 1. When more than one employer was listed in a single category, the numbers of employees within each zip code were combined. The process is illustrated in Table 2 for the industry category Finance, Banks, Insurance for the three zip codes 48015, 48026, and 48043.

The actual process was completed with a standard computer package, which allowed for an automated expansion. The process was performed with all the zip codes; this resulted in the expanded data set, which was then adjusted to include the location data of the automotive company and the federal employees.

Distribution of Zip-Code Data to Census Tracts

The expanded residential data were further disaggregated to the census tracts based on the observed location distribution of the six CBD employers. These data gave the percentage distribution of employee residences among the census tracts within each of the zip codes. In order to derive the total regional distribution, a zip-code census-tract equivalency table was used.

Table 1. Derivation of CBD expansion factors.

Industry Category	Sample Employer	No. of Employees in Sample	Observed Distribution[a] (%)	CBD Employees by Category	CBD Employees Without Automotive Company and Federal Employees	Expansion Factor
1. Manufacturing (automobile and other)	Automotive company	504	5.57	5 944	4 184[b]	8.302
2. Transportation, communications, utilities	Utility 1, utility 2, utility 3	7826	10.3	10 992	10 992	1.405
3. Wholesale and retail trade	Retail store	3705	13.4	14 300	14 300	3.859
4. Finance, banks, insurance	Bank 1, bank 2, health insurance company	8236	23.7	25 291	25 291	3.071
5. Public administration	City government, county government	8230	17.5	18 675	13 001[c]	1.580
6. Natural resources, construction, business and professional services	Accounting company 1, accounting company 2, accounting company 3, hotel, engineering consulting company	3887	29.46	31 438	31 438	8.087

[a] Observed distribution was derived based on Michigan Employment Security Commission (MESC) data.
[b] Total employment adjusted to exclude recently relocated automotive company (1760 employees).
[c] Total employment adjusted to exclude federal government employees (5674 employees).

Table 2. Zip-code level expansion process (CBD) for one industrial category.

Zip Code	No. of Employees in Sample: Bank 1	Health Insurance Company	Bank 2	Total No. of Employees in Sample	Expansion Factor	After Expansion
.	.	.	.	.	.	.
48015	0	7	1	8	3.071	25
48026	3	30	9	42	3.071	129
48043	13	181	29	223	3.071	685
.	.	.	.	.	.	.

The data manipulation for deriving the percentage distribution of residence locations for each zip code in the region would have been a tedious process. However, computerized techniques made this distribution possible. The result of this process was the estimated residential location data at census-tract level for all CBD employees (102 723). The final data format is shown below:

1970 Census-Tract Number	No. of Employees Who Live in This Census Tract and Work in Detroit CBD
1001.00	50
1001.01	111
1001.02	69
.	.
.	.
.	.
7115.0	22

The total employment as contained in the file was less than the control total of 106 715 for two reasons. First, there are employees who work in the CBD but live in Canada and outside the region. Second, the regional DIME/GBF file includes only the tricounty area; those working outside the tricounty area were not included.

EMPLOYEE DATA EXPANSION FOR CFA

The CFA expansion process differed from that used for the CBD for two reasons. First, sample employee residential location data collected from employers accounted for 34 583 of the 66 042 total employees in the CFA. Since the sample consisted of more than 50 percent of the total, an assumption was made that the sample was representative of the total employment population and that errors due to simple expansion would be minimal.

Second, the CFA sample data gathered were not geoprocessed and were only available at the zip-code level; hence, within zip-code areas, the distribution of CBD employee residential locations by census tract was assumed to hold for the CFA employees also. During the follow-up analysis, this assumption will be tested after CFA address data have been geoprocessed. The expansion process performed on the data is described below.

Derivation of Expansion Factors and Expansion for Zip Codes

Because the CFA included a large geographic area, it was divided into three analysis districts to classify data down to a level suitable for service planning. The expansion process was performed separately on each of the CFA districts designated A, B, and C (see Figure 2). The sample details for each of these districts are presented in Table 3. Also shown in the table are the expansion factors for Districts A and C, where

Expansion factor = (control employment totals)/(sample employment totals).

As shown in Table 3, the expansion factor for District A is 1.388 and for District C, 2.295. For these two districts the expansion was performed at the zip-code level from the sample to the totals in the CBD.

Since the sample employment total for District B was very small compared with the total, a direct expansion as above was not considered appropriate. The distribution of the entire CFA sample by zip codes was derived and used for District B. This process is illustrated in Table 4.

Thus, for each of the three CFA districts, the residential location of all employees by zip codes in the region was derived. All the data manipulations were performed by using a standard computer package program.

Distribution of Expanded CFA Data from Zip Codes to Census Tracts

To further distribute the zip-code data to corresponding census tracts, data from the Detroit CBD

Table 3. Derivation of CFA expansion factors.

District	Sample Employer	Sample Employment	Total Sample Employment by District	Adjusted Employment Totals	Expansion Factor
A	University staff	16 024	23 827	33 068	1.388
	Motor company	5 069			
	Computer company	2 005			
	Hospital 1	729			
B	Hospital 2	1 660	1 795	12 412	.[a]
	Art institute	135			
C	Hospital 3	3 723	8 961	20 562	2.295
	Hospital 4	1 496			
	Hospital 5	1 279			
	University (medical staff)	2 463			
Total			34 583	66 042	

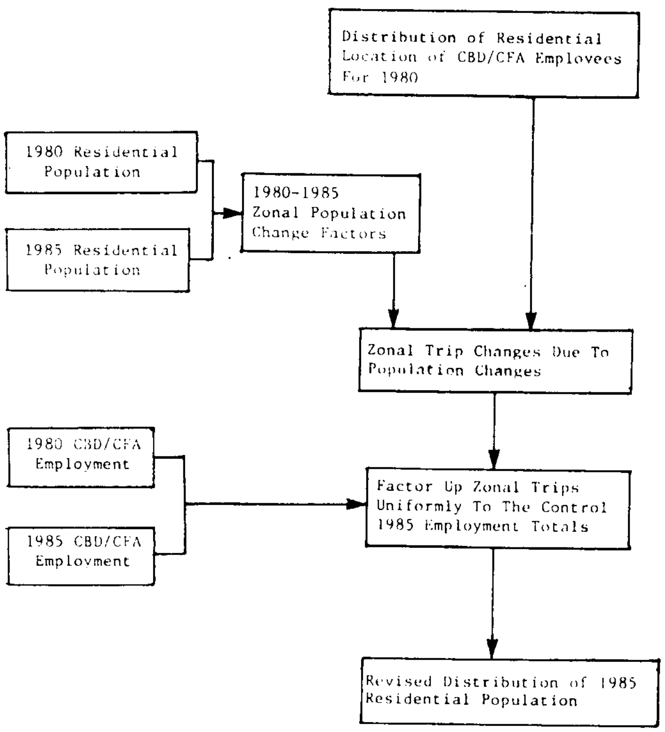
[a]Because the expansion factor was greater than 6, a different method was used for District B as explained in the text.

Table 4. Zip-code level expansion for CFA (District B).

Zip Code	Total Sample Employment for All 10 Employers[a]	Distribution of Total[b] (%)	District B Employment Distributed[c]
.	.	.	.
.	.	.	.
48015	607	0.017	221
48026	449	0.0130	162
.	.	.	.
.	.	.	.

[a]Total sample employment = 34 583.
[b]Column 3 = column 2 divided by 34 583.
[c]Column 4 = column 3 times 12 412 (total District B employment).

Figure 4. Process for forecasting 1985 residential locations.



were used. This process consisted of determining the percent distribution of employment from the six major employer data sets for each of the census tracts in any given zip code and then distributing the total employment of that zip code to the census tracts based on the derived distribution. This process then yielded the employee residential location data at census-tract level for the CFA.

EMPLOYEE RESIDENTIAL LOCATION DATA FORECASTS FOR 1985

This section describes the methodology and its application for developing projections of residential locations for employees working in the CBD and the CFA for 1985 based on 1980 data.

The basic factors considered in the 1985 projection process were the following:

1. 1980 base employee residential location data files for the CBD and the CFA,
2. Employment growth rates for the CBD and the CFA from 1980 to 1985, and
3. Changes in population from 1980 to 1985.

The process used for projecting the 1985 data is illustrated in the flowchart (Figure 4). Based on the population shifts from 1980 to 1985, it was assumed that the residential locations of CBD and CFA working employees will exhibit shifts similar to that of the entire population. The adopted regional 1980 and 1985 small-area forecasts of population (number of people) were used to adjust the 1980 employee residential location data to the population shifts. The 1985 employment forecasts for the CBD and the CFA were determined to be 123 789 and 70 731, respectively. The adjusted 1980 employee residential location data were projected to reflect 1985 employment totals.

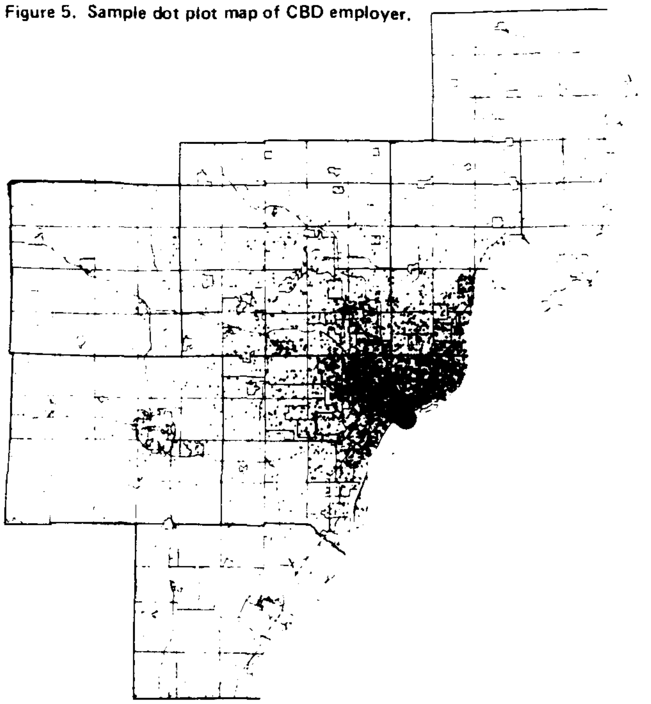
Preparation of Base Data

A computer file was created with employment and residential information for 1980. The file holds data for 1446 regional analysis units, and since regional population forecasts are based on these analysis units, the 1980 census-tract residential location data were converted to analysis-zone geography. A census-tract and analysis-zone equivalency file was used to do this.

The input data file variables in this process are as follows:

1. Analysis unit number,
2. 1980 CBD working employees who live in this zone,
3. 1980 CFA employees from District A who live in this zone,
4. 1980 CFA employees from District B who live in this zone,
5. 1980 CFA employees from District C who live in this zone,
6. 1980 population for this zone, and
7. Projected 1985 population for this zone.

Figure 5. Sample dot plot map of CBD employer.



Adjustments of Residential Data to Population Shifts

The basic assumption is that the employee residential locations would move in the same direction as general population shifts. If a zone experiences a reduction in population from 1980 to 1985, the employees who live in that zone and work in the CBD or the CFA will likely be reduced proportionately. Similarly, areas with increased population in 1985 will reflect an increase in employee residential locations. The adjustment factor was derived by dividing projected 1985 population by the 1980 population for each of the zones (total of 1446) in the file. Thus,

Adjustment factor (i) = [1985 population for analysis unit (i)]/[1980 population for analysis unit (i)],

where i ranges from 1 to 1446.

The adjustment factor was then applied to the 1980 residential location data. Multiplication of the 1980 data by the adjustment factor for a given zone reveals the adjusted number of employees in that zone. There is only one adjustment factor for each zone.

Adjustment of Location Data to 1985 Control Employment Totals

The residential location data due to adjustments for the population shifts resulted in a data set with fewer employees in recognition of the decreasing regional population forecasts. The total employment observed in the adjusted data files was 98 175 for the CBD and 30 848 for District A, 11 636 for District B, and 19 337 for District C of the CFA. In spite of the decreases in the overall population for the study area, the employment for CBD and CFA is expected to increase from 1980 to 1985.

Based on the predicted regional employment-growth trends (5), the 1985 control employment was determined to be 123 789 for the CBD and 35 416 for District A, 13 293 for District B, and 22 022 for District C of the CFA. However, as observed in the expansion process for 1980, it was assumed that 3.7 and 2.2 percent of the CBD and the CFA employees will live outside the study area. Discounting for this factor, the CBD and the CFA control employment totals were 119 209 for the CBD and 34 634 for District A, 13 000 for District B, and 21 538 for District C of the CFA.

Expansion factors (EFs) were derived from the revised control totals and the adjusted employee data:

EF for CBD = 119 209/98 175 = 1.214.

EF for District A = 34 634/30 848 = 1.123.

EF for District B = 13 000/11 636 = 1.117.

EF for District C = 21 538/19 337 = 1.114.

The adjusted 1980 file for 1980-1985 differential population was factored up uniformly, based on the above expansion factors, which resulted in 1985 employee residential location data projections for each of the analysis units.

APPLICATIONS OF EMPLOYEE RESIDENTIAL LOCATION DATA

SEMTA has used residential location data in many service-planning and corridor-analysis projects. Perhaps the most important result is that the previous planning data base and the transit demand estimation methodology, which employed population density of potential service areas as an indicator of certain types of travel habits, have been replaced. This new data base is much more efficient because it represents population densities of actual travelers with a known destination.

Graphic Displays

Graphic displays effectively illustrate the residential location patterns and thus the travel patterns with CBD and CFA orientation. They simply and quickly lend a sense of the overall distribution pattern of the CBD and CFA employee populations. Two types of graphic displays have been extensively used in the transit planning. The first is a rough, inexpensive display of the number of employees per zip-code area. The number of employees in a zip-code area is represented by color coding; different colors indicate the varying density of employee residences per acre. The advantage of this method is that a single business or a group of businesses located in the same place or block can be studied for trends in residential patterns. However, this display is limited in that it does not effectively illustrate trip origin densities when the size of the zip-code areas varies.

One way to overcome this limitation is to develop a computerized dot-plotting program. This program uses the digitized census-tract or zip-code boundary coordinate files and randomly places a dot in the appropriate area for each employee (see Figure 5 for an example--a reduced version of the actual map, which is usually at a scale of 1:250 000). The dot-plotting and the color-coding techniques are two examples of graphic tools being employed in various industries. The variety of industries using these graphic capabilities has been well documented (6).

Although the dot-plotting method does give a better representation of the actual number of employees in each unit of analysis, it is also limited. In some areas, absolute numbers cannot be gleaned from the display when the computer overplots several dots in one place. This problem can be solved by using larger scale maps and smaller dots;

even with these limitations the display serves to illustrate general travel density trends. These displays are also useful in visually analyzing new and existing routes. The residential location data and the graphic displays have contributed to the development of several route-planning activities scheduled for execution within six months. The data have been used frequently to modify specific routes to increase ridership. Two examples of such efforts are presented below.

Route Planning

A potential market on a major employment site was identified and served by establishing an express route. A major utility located on the western side of the CBD had no express route to the eastern suburbs. The residential location data analysis indicated that this company had a concentration of employees in the express service area. An extension of routing in the work-destination end of the trip resulted in a significant ridership increase on the route. Boarding counts at the utility site confirmed that this increase came primarily from that workforce.

SEMTA also rerouted a portion of a CBD local service by employing residential location data and displays. Originally, the service had a single-route configuration in the home (residential) end of this trip. Once the expanded data were derived, SEMTA staff observed that by branching (i.e., deviating route segments from the main routing), substantial numbers of potential riders would gain access to transit. The service area was widened by the route deviations, which put transit service closer to CBD and CFA employees' neighborhoods.

Corridor Analysis

Employee residential location data are also adaptable to corridor service-planning projects, which extend to midrange (i.e., five-year) time periods. SEMTA reviewed the level of transit service provided to different areas of the region, projected the potential demand from the data set, and determined that the western suburbs were underserved, particularly by the park-and-ride commuter routes. SEMTA employed the Interactive Graphic Transit Design System (IGTDS) to analyze service to a single corridor. IGTDS is a set of computer programs developed by General Motors Corporation to assist planners in designing and evaluating transit alternatives by using computer graphics and analysis. With the location data demand set, alternative park-and-ride routes were tested, which yielded potential routes and park-and-ride lot locations.

Another example of an automated corridor analysis that uses the data is a feasibility analysis of a commuter rail system along another corridor in the region's northeast area. The modal-split models used the data to estimate ridership by various transportation modes. The modeling process consisted of validating the primary and submodal-split models by applying them to another corridor currently served by the commuter rail service for base 1980 conditions.

Model results were matched against the actual ridership on various transit modes, including commuter rail, to validate the model coefficients. After this step, they were applied to the study corridor and produced ridership estimates close to the actual ridership. These models were then applied to simulate the projected 1985 travel demand on each mode. Based on the analysis results, it seems that sufficient demand will be present in 1985 to support the proposed rail line.

Additional benefits derived from the team approach of data collection include the following:

1. Placement of a sales-ticket booth at a major employment center,
2. Establishment of sales agents at major employment sites, and
3. Enhancement of SEMTA's credibility in the regional business community.

The third point is especially important, since SEMTA pursues joint development projects and seeks to expand the employer base to employment centers not in the CBD or the CFA. Finally, nontransit benefits also resulted from the residential location data. Other government and private agencies have used these data to examine the regional demographic and economic trends.

CONCLUSIONS

An up-to-date data base describing home-to-work travel demand is necessary and useful to conduct transit service-planning activities efficiently. Major employers helped to gather residential location data for their employees and to develop a travel-demand data base. The team data-collection efforts proved to be quite effective, as illustrated by the positive response from the employers.

The data gathered are the most current information at the level of geography needed in service planning. The data summaries and display techniques developed in this study provide insight and understanding of actual travel demand on a route and corridor basis. These summaries and displays have already been used for route modifications in the downtown area to effectively serve the employment center. A unique expansion methodology was developed to estimate total travel demand from sample data. Expanded employee residential location data at census-tract level and densities assisted in planning on the home end of the work travel. These data have also been used in service improvements and for the design of new express routes. Further, the base-year data and 1985 projections were used for short-term corridor planning.

The methodologies described in this paper are unique and do not involve hypothetical modeling theories. Rather, they provide a reliable, accurate, and up-to-date data base. To supplement the dynamic planning activities, the data base can be continuously monitored with little effort. In fact, because the data base has been widely and successfully used, the authority plans to expand its services to include other employment centers in the region. Efforts are also under way to provide a nonwork travel data base. The end product will be a comprehensive demand set that reflects current travel habits and will result in more sensitive and efficient transit planning.

ACKNOWLEDGMENT

The opinions, findings, and conclusions expressed in this paper are ours and not necessarily those of the the Urban Mass Transportation Administration of the U.S. Department of Transportation.

REFERENCES

1. Journey To Work. Bureau of the Census, U.S. Department of Commerce, Rept. PC 2-6D, 1973.
2. Selected Characteristics of Travel to Work in 20 Metropolitan Areas: 1977. Bureau of the Census, U.S. Department of Commerce, Rept. P-23, No. 105, Jan. 1981.

3. ITE Committee 6A-12. Preparation for the 1980 Census. ITE Journal, March 1979, pp. 41-47.
4. Profile of Southeastern Michigan TALUS Data. Detroit Regional Transportation and Land Use Study Committee, Detroit, MI, Aug. 1968.
5. Review of 1974 Version of Small-Area Forecast for Use in Alternative Analysis. Southeast Michigan Council of Governments, Detroit, MI, Oct. 1976.
6. H. Takeuchi and A.H. Schmidt. New Promise of Computer Graphics. Harvard Business Review, Jan.-Feb. 1980, pp. 122-131.

Method for Determining and Reducing Nonresponse Bias

PETER R. STOPHER AND IRA M. SHESKIN

Nonresponse bias is of continuing concern in participatory surveys of human subjects. It has led frequently to the adoption of expensive interview surveys in place of cheaper self-administered surveys because of relative response rates. Nonresponse bias has been estimated from comparison of early and late returns in self-administered surveys, from comparison of socioeconomic and demographic variables between the survey and census data, from special efforts to contact a sample of nonrespondents, and by assuming extreme values for nonrespondents. None of these methods is totally effective, whereas the relative economy of self-administered surveys has grown and suggests a reexamination of the value of such surveys. A method is outlined by using two survey mechanisms, including a conventional self-administered procedure, where the joint mechanism retains most of the economies but adds information on nonresponse and provides a means to increase response levels of the self-administered segment. Results from two transportation surveys are described and nonresponse biases and response levels are discussed.

One of the first decisions in any survey design is to select the mechanism by which the survey will be performed. Input to this decision includes specification of the purpose of the survey, definition of the sampling frame, determination of desired confidence levels (and thus sample size), labor availability, time and budget constraints, types of questions that need to be asked, likelihood of obtaining accurate answers, length of the survey, and expected response rate (1). Each survey effort is to some extent unique and thus the choice among the face-to-face interview, the mail questionnaire, the telephone interview, and a number of other alternatives must be made for each survey by using a careful balancing procedure that considers the various advantages and disadvantages of each method.

One of the most important of these factors is the expected response rate because of the effects both on costs and on the unknown bias that a low response rate may introduce. More often than not, if respondents are placed in direct contact with an interviewer, the response rate is assumed to be high, generally on the quite strong grounds that refusal is less acceptable to a personal request than it might be to any impersonal approach such as a mail survey. In contrast, significantly lower response rates are assumed to occur when no personal request is involved or when the request is only to accept a survey form and not to answer specific questions. However, as Dillman (1) points out, this supposed significant advantage in response rate may be due, to some extent, to the manner in which response rates are calculated for the mail survey versus the face-to-face interview survey.

Irrespective of the survey mechanism, nonresponse occurs. It can be classified into two forms: genuine and nongenuine nonresponse (2). Genuine nonresponse is not the concern of this paper. This is defined as the nonresponse occasioned by selecting sampling units that are subsequently found to no longer be a part of the survey population (e.g., vacant or demolished houses, addresses that do not exist). In contrast, nongenuine nonresponse is defined as that nonresponse which occurs by the voluntary action of a sampled respondent not to participate in the survey. Genuine nonresponse is not of serious concern because it can be assumed generally to be a random or quasi-random occurrence that adds no significant bias to the survey data and that can be corrected largely by expanding the sample appropriately to cover its expected or encountered level (3). Nongenuine nonresponse is a documented source of bias for a number of reasons (4). It has been shown in a number of instances that those who do not respond to a survey possess generally a characteristic of direct relevance to survey measurements. For example, in surveys of travel habits and needs (an area well known to us), nonrespondents are most likely to be drawn from two segments of the population: those who travel very extensively and who therefore would be subject to much longer questioning on travel habits for a period such as 24 h and those who travel very little or not at all and who doubt the relevance of the survey to them or of themselves to the survey (5,6). This facet alone is a major cause of nonresponse bias. Others, which do not need elaboration here, include educational and income bias to written questionnaires and life-style biases associated with the state of being at home for the survey (1).

As a general rule, it can be assumed that the potential existence of and the extent of nonresponse bias caused by nongenuine nonresponse is correlated with the size of the nonresponse rate. Although it appears that little scientific evidence exists to support this hypothesis (particularly given the paucity of studies of nonresponse itself, let alone the biases and their relationship to rate), this assumption carries a fairly substantial weight of circumstantial common sense. For the purposes of this paper, it will be accepted as a reasonable postulate and not subject to further question.

Given, then, the parallel factors of an expected relationship between nonresponse bias and the common assertion that personally conducted surveys have higher response rates than impersonally conducted ones, it is not surprising that the majority of human surveys have tended to be carried out by means of direct interviewing in preference to most other methods of survey.

This paper raises three parallel concerns that derive from this state of affairs. First, some problems concerned with the calculation of response rates on face-to-face interview surveys versus mail surveys are discussed. Second, given the tremendous differences in unit costs of personal interviews versus self-administered surveys, it is becoming increasingly worthwhile to seek relatively economical ways to improve self-administered surveys so that better response rates can be obtained and stronger advantages developed for these significantly more economical procedures. We believe that the dual survey mechanism (DSM) described herein is a valuable procedure for improving self-administered surveys. Third, given that nonresponse occurs, some procedure is needed that will provide a means to estimate the extent and shape of nonresponse bias. In this respect, we argue that traditional methods of measuring nonresponse bias (comparison between sample survey and census figures, interviewing by, say, face-to-face interview a subsample of nonrespondents to a mail survey, comparison of early and late returns, and assumption of extreme values for nonrespondents) have significant disadvantages that lead to their not being used in many practical fields of survey research.

The utility of the suggested alternative procedure--the DSM--for determining and reducing nonresponse bias is illustrated by using the results of three travel-behavior surveys, one conducted in Dade County (Miami), Florida, and two in Washtenaw County (Ann Arbor), Michigan.

RESPONSE RATES FOR MAIL AND INTERVIEW SURVEYS

Two options are available for participatory surveys--personal interview or some form of self-administered survey. To a large extent, conventional wisdom in transportation data collection (and in other fields) has been to use face-to-face interviews. This has been based on the notions that response rates are higher, that data are less subject to both error and bias, and that certain items of interest in transportation surveys cannot be collected by using a self-administered survey. The major acknowledged disadvantages of interview surveys are the length of time required to collect the data (particularly for on-board vehicle surveys) and the cost, which currently ranges from about $35 to more than $500 per interview in transportation applications. A major advantage of self-administered surveys is their cost, which may range from as little as $1 for each complete response to a high of about $30.

Recent research in West Germany (2), however, suggests that the response rates claimed for interview surveys may be inflated. Although response rates often are cited as being 90-95 percent or higher, such rates are generally misleading because they are calculated on a different basis than are the response rates of self-administered surveys, which yields an automatically higher figure for the interview survey. For example, for a mail-out, self-administered survey, response rates are calculated as the proportion of those surveys mailed out that were returned as usable responses. Frequently, the proportion of mailed-out surveys not delivered or delivered to an address that was temporarily or permanently vacant will not be known. Conversely, interview-survey response rates usually are based on the total number of completed interviews plus terminations and refusals. Often not computed into such response rates are the number of "no answers," failed requests for calls back, under construction, no such address, and the like, which would be made up from a back-up sample and would be discounted prior to computing a response rate. Therefore, comparable response rates between these two survey types generally have not been reported. Such comparable rates would show interview surveys to achieve a much lower response rate than usually has been reported. As an example, in the 1980 Southeast Michigan Regional Travel Survey, which was an at-home interview survey of 2706 households, the calculated response rate was 85 percent. If "no answers," failed requests for calls back, under construction, no such address, and the like are added in, the response rate drops to 65 percent (7). These ideas are in agreement with Dillman (1, p. 50), who points out that "in face-to-face and telephone interviews a refusal is not considered as such until a contact is made. In mail studies, the opposite is assumed, that is, a nonresponse is a refusal until proven otherwise." Also, researchers often fail to report the way in which the response rate was calculated.

TRADITIONAL SOLUTIONS TO NONRESPONSE BIAS PROBLEM

There exists no solution to the nonresponse bias problem that can guarantee absolutely that [R], the set of respondents, is a random sample of [S], the set of selected individuals in the sample (8). One common procedure is to assume that [S] is a random sample of [P], the population, and then to test for significant differences between [R] and [P] on a set of known variables for [P]. Thus, by using a series of one-sample significance tests, one could, for example, test to see whether the mean income (adjusted for inflation) of [R] is significantly different from the mean income reported for the study area by the census. If no significant differences are found, it would be reasonable to assume that the incomes of [R] and [NR], the set of nonrespondents, are not significantly different and that no response bias exists with respect to income. If, on the other hand, [R] is found to be biased toward upper-income categories (as is likely to be the case), it becomes possible to weight the answers of those of lower income who did respond to produce [R*], the set of respondents with answers weighted to reflect more accurately the distribution of incomes in the study area. This procedure could be performed for various variables and different surveys weighted with different factors to reflect known distributions more accurately.

Although the above procedure may be effective in some cases for adjusting for nonresponse bias, a number of significant problems exist:

1. Although many demographic variables are available in the census, many important variables for which one might want to check for nonresponse bias may not be available from a census. Suppose a survey queries attitudes about energy costs in a given county. If a greater percentage of automobile users than bus users answers the survey, one may want to weigh the results from the bus users to reflect modal split in the county more accurately. This is only possible if the number of bus and automobile users in the county is known from the census.
2. Demographic variables may be available in the census but may be significantly dated.
3. The census data may be inaccurate because the census also is likely to suffer nonresponse from the same groups of people as a sample survey.
4. If the population from which the sample is being drawn is a subpopulation (such as the users of a given facility), it is highly unlikely that a census exists of such users.

An excellent example of the application of this first procedure is provided by Young and Willmott (9) in their 1970 study of family sociology in London. Census data for 1971 were available to them only in terms of the sex variable, and no response bias was shown. Age, marital status, and occupation comparisons had to be made with the older Sample

Census 1966. They show their sample to be somewhat underrepresentative of the young and the single. This may have been due to the greater difficulty of finding such people at home, or to an inadequate sampling frame, or to a change in the demographic structure of the population over the four-year period since the census. As expected, comparison of the occupation variable showed some tendency for those in professional and managerial positions to be more responsive to the survey.

A second technique used to judge nonresponse bias is to select a random sample of nonrespondents at the completion of the survey and to make special and persistent efforts to gain some brief information from this sample. Thus, if the original survey mechanism was a mail survey, a brief home interview or telephone survey might be devised on nonrespondents and tests for significant differences between [R] and [NR] performed. This technique also has a number of disadvantages:

1. A significant cost is added to the survey.
2. The amount of time needed to complete the survey is extended because the survey of nonrespondents cannot begin until all nonrespondents are identified.
3. Not all nonrespondents to the first survey will cooperate with the second survey.
4. Suppose one is looking for differences between respondents' attitudes and behavior. If such differences are found, the question arises whether these differences really exist or are caused by the different survey mechnisms used for [R] and [NR].
5. If the follow-up survey occurs after the original survey, it may be that attitudes and/or behavior may have been changed by some outside factor. Thus, it is possible that the original set of respondents [R] completed a survey on energy prior to, say, an oil embargo, whereas those respondents in [NR] are being queried after the oil embargo.

By using this methodology, Goudy (10), in a sample of the general public in rural communities in northern Iowa, raised the response rate from 79 to 93 percent by following a mail survey with a face-to-face interview of nonrespondents to the mail survey. Although the additional interviews resulted in only slight changes in the demographic characteristics of the respondents, the changes were in the expected direction. The proportion of respondents with less than 11 years of school increased from 31 to 33 percent and the proportion with income below $6000 went from 24 to 26 percent.

A third traditional method for dealing with nonresponse bias is to compare early responses with late responses (11). The assumption inherent in such a comparison is that respondents who mail in their questionnaires very late or who answer only after some follow-up effort (such as a reminder postcard) are similar to nonrespondents.

A number of articles have appeared employing this method of comparing early and late respondents to travel surveys. In a travel survey by Wright (12), two reminder letters were mailed to nonrespondents and followed, if necessary, by a personal visit. Significant differences were found between early and late responses in age, sex, occupation, length of residence, and ownership of dwelling unit. No significant differences were found in education, household size, location of the household, and relationship of the respondents to the head of the household.

Waltz and Grecco (13) also compared early and late respondents. Respondents differed significantly by sex, education, occupation, length of residence, and ownership of dwelling unit. No significant differences were found for age, city of residence, marital status, household size, and type of dwelling unit. They also compared respondents and nonrespondents who were shown to differ significantly on length of residence, ownership of dwelling unit, and type of dwelling unit.

Galin (14) also compared both early and late respondents and respondents and nonrespondents as part of a data-collection effort for the Australian Road Research Board. Postcards with eight questions were handed to drivers at a cordon line. The vehicle type (car, truck, van) and the sex of the driver were noted. No significant differences were found for these two variables between those who did mail back the postcard and those who did not. When early and late respondents were compared, no significant differences were found in trip purpose, trip length, vehicle type, age, number of years driving, and sex.

Finally, Kanuk and Berenson (15), in a comprehensive 1975 literature review of mail surveys and response rates, concluded that research efforts to determine the difference between respondents and nonrespondents have focused on demographic, socioeconomic, and, to a lesser extent, personality variables. The only widespread finding is that respondents tend to be better educated than nonrespondents and thus have greater facility in writing.

A variation of this technique for a telephone survey has been suggested by O'Neil (16). He compared those who responded to the survey on first contact with those who answered only after having refused on the first attempt. The "resistor" group, for example, were shown to be more likely from blue-collar occupations and lower in income and education, although O'Neil judges the differences to be unimportant. One very significant drawback to this third traditional procedure, whether for mail or telephone survey use, is that it is based on the unproven and somewhat dubious assumption that those who respond to a survey late or only after some follow-up effort are similar in characteristics to nonrespondents.

Finally, Cochran (17) suggests a procedure that assumes extreme values for nonrespondents. Unfortunately, as shown by Fuller (18) and Wayne (19) under a variety of conditions, the calculated confidence intervals are almost always far too wide to permit meaningful inferences from the data. In sum, all traditional methods for dealing with nonresponse bias have been shown to have significant disadvantages.

DSM AS APPROACH TO NONRESPONSE BIAS

Three goals are implicit in the selection of a survey mechanism, as described in the preceding sections of this paper: lack of bias, economy, and knowledge about the characteristics of inevitable nonrespondents. No single survey mechanism succeeds in achieving all three. Beginning from the premise that the home-based, personal interview is the most effective way to minimize nonresponse and its associated bias but that such a mechanism is rapidly becoming far too uneconomical for many applications, we sought to develop a mechanism that would provide significant economies at a much smaller loss to response and bias.

The mechanism developed is the coupling of a short, relatively inexpensive form of personal contact as a prior approach to a longer, self-administered survey. At least two versions of this mechanism have been developed: (a) an intercept survey in which there is a personal request to complete and hand back a short survey form and a following take-home/mail-back survey and (b) a brief telephone interview followed by a mail-out/mail-back survey. These designs seek several common goals:

1. Through several mechanisms, to increase the response rate to the self-administered survey;
2. To provide the means to execute follow-up on the mail survey (which is often missing in a take-home/mail-back survey) as a means to build the response rate;
3. To provide some useful information on those who respond to the personal contact but refuse to respond to the mail survey, which thus provides a partial measure of nonresponse; and
4. To use the initial contact in several different ways to define more precisely and clearly to potential respondents the situational context for the self-administered survey.

The DSM is much less expensive than home-based personal interviews but is not limited, as are some of its obvious single-mechanism alternatives, in the length of the survey that can be executed nor in the contextural situation. (Clearly, if the survey purposes can be fulfilled by a 5-min intercept or telephone survey, there can be no possible value from a DSM; the use of the DSM is where a longer survey is needed to satisfy the measurement requirements.)

The first and fourth points above merit some elaboration. Increases in response rate should arise from several aspects of the procedure. First, people are generally more likely to respond to a brief intercept survey (e.g., an on-board bus, plane, and train survey) or to a 5- to 10-min telephone interview than to a significantly longer survey of almost any type. For a number of people, this will create a seeming obligation to agree to and to complete the subsequent longer survey--the standard marketing device of compliance with a small request leading to compliance with a subsequent longer request (20). Second, an intercept survey frequently creates circumstances that induce an enhanced response from such effects as peer pressure and the appearance that the survey is neither long nor difficult to do. Similarly, refusal rates for short telephone surveys are usually very low, particularly if the approach is from or on behalf of a public-service agency.

The context-setting capability of the initial contact is also extremely valuable. In an intercept survey with subsequent self-administered survey, the self-administered survey may ask questions about the activity that was intercepted. If the self-administered survey is completed some time after the activity occurred, recall may be a problem. The occurrence of the intercept survey, however, may serve to help fix the specific occasion in the minds of respondents. When the initial mechanism is a telephone survey, the telephone contact can be used to specify a particular day or activity that should be the subject of the self-administered section. This provides a control of situational context that is usually lacking in self-administered surveys.

The remainder of this paper discusses two case studies of the use of the DSM and its benefits in specific contexts. Case Study 1 is the Dade County On-Board Transit Survey (21), which was an intercept-and-mail DSM. Case Study 2 is a similar survey for Washtenaw County, Michigan, which included both an intercept-and-mail DSM and a telephone-and-mail DSM.

CASE STUDY 1: DADE COUNTY ON-BOARD TRANSIT SURVEY

Dade County, Florida, is involved currently in the construction of a rail rapid transit system, a downtown people mover, and a revised and expanded bus network expected to cost a total of about $1 billion and all scheduled for completion by 1984. The Dade County Transit Development Program, 1980-1985, calls for survey work to elicit information concerning the manner in which the bus system is being used currently. Such information is to be employed in adjusting the bus system to user needs as new vehicles are purchased, in designing the feeder bus network into the rapid transit stations, and in updating available modal-split models (22).

Because the desire was to sample only those individuals who ride the bus and such individuals comprise a small percentage of the county's population, the only possible cost-effective means of reaching bus riders was an intercept survey. The short length of time for which many riders are on a bus, the obvious difficulties of conducting an interview under such circumstances, and the fact that an interviewer needs to select respondents dictated the use of a self-administered form.

Four competing forces presented themselves: (a) the volume of information needed from each rider was extensive and filled 10 pages of legal-size paper; (b) the longer the form, the lower the response rate is likely to be; (c) persons on short bus rides could not be asked to fill out long forms while riding; and (d) some respondents (particularly the large number of elderly in Dade County) would experience physical discomfort from trying to read and write on a moving bus.

Thus, a DSM was developed that contained five parts:

1. An instruction page;
2. Form a, designed to be completed and returned on board the bus, although designed so that it could be mailed back instead if the respondent so desired (the on-board form);
3. Form b, designed to be completed at home and mailed back (the take-home form);
4. An envelope for the return of the take-home form; and
5. A cover letter from the Dade County Transportation Coordinator, designed to lend credibility and encourage response.

The Dade County intercept-and-mail DSM was designed specifically to accomplish the following:

1. The on-board form was designed to gain response from the type of person who would give 3-5 min but would certainly not go to the trouble of carrying home a survey form, spending 45 min to fill it out, and then remembering to mail it back.
2. The on-board form also could take advantage of people's feelings about being good citizens by way of a "demonstration effect." That is, suppose forms are handed to 30 persons on board a bus. If even some minimum number sit down and begin to fill out the form, the chances are good that others will follow the lead to avoid feeling guilty and being viewed by fellow passengers in a negative way for not cooperating. Persons who would be reluctant respondents also will be encouraged to cooperate when they see that the survey does in fact take only a few minutes.
3. The fact that respondents were handed the form while boarding the bus and were thus a captive audience also helped to encourage response. Unlike a personal interview at home or a telephone interview, where the interviewer may be interrupting the interviewee involved in some activity, most bus riders usually do little with the time they are on the bus. The survey could thus act as an interesting diversion.
4. As mentioned above, certain questions are best answered while a respondent is performing a given activity because loss of information can be

expected if time is permitted to lapse. For example, "How long did you wait at the bus stop for this bus?" is answered most accurately (in terms of the respondent's perception of waiting time) immediately following the wait. Thus, a major problem of travel surveys, that respondents have difficulty remembering trip details or even that a trip was made, is averted.

5. As mentioned above, most intercept surveys that involve only a take-home, mail-back form have no possibility of a follow-up, because the addresses of those taking the forms are unknown. The Dade County DSM, by asking for the address of the respondent on the on-board form (for the purpose of sending them a free bus pass incentive), permitted a follow-up to proceed for those not returning the take-home form.

6. The Dade County DSM also took advantage of the idea that compliance with a small request can be effective in encouraging compliance with a longer request. It is also more likely that respondents will remember to fill out the take-home form given that they have already spent some time that day on the survey. Also, some people have an aversion to leaving a job only half done.

7. An effective device used in the Dade County DSM was to promise respondents to both the on-board and the take-home forms that a free bus pass would be forthcoming.

8. An additional advantage of the on-board form is that, although the form was designed to be self-administered, those who had trouble filling out the form could seek assistance from fellow passengers or from the survey worker.

9. One rather unexpected benefit of the take-home form was that in addition to the 181 persons who returned both the on-board and the take-home forms, 120 persons who did not complete the on-board form did fill out the take-home form. A number of factors may have contributed to this outcome: (a) some persons are discomforted by reading and writing in a moving vehicle; (b) some respondents were on the bus for too short a time to fill out the on-board form; and (c) survey workers reported that many elderly persons did not bring their reading glasses. In all cases, these nonrespondents perceived the take-home form as more important or, in spite of a business-reply panel on the on-board form, might have assumed that the on-board form could not be mailed in.

10. Perhaps the most important benefit of the on-board form is that it permits the evaluation of possible response biases in the take-home form. For the pilot survey in Dade County, 632 persons answered the on-board form, whereas only 181 of these (29 percent) answered the take-home form. Thus, if the on-board form did not exist, the response rate would have been about one third. More important, dividing the 632 persons into the [R] and [NR] groups depending on whether or not they had responded to the mail-back permits the identification of age, sex, driver's license, transit captivity, and geographical location (via zip code) bias.

As Table 1 shows, such biases were apparently not significant in this survey. For both the [R] and [NR] groups, about 48 percent are female, about 60 percent have driver's licenses, and about 43 percent are captive to transit. The percentages of respondents in each of the age categories are strikingly similar; there are two exceptions. Those less than 12 years old did not return the take-home form and those older than 70 were more likely to complete the take-home form.

An interesting but not unexpected observation is that, in all cases, item nonresponse on the on-board form is significantly higher for the [NR] group than for the [R] group. For example, 6.7 percent of the [R] group did not answer the question about sex, whereas 12.2 percent of the [NR] group left this question blank.

Other variables that appear on the on-board survey may also be used to check for possible biases in the attitude questions on the take-home form. If attitudes toward transit are influenced by the way one uses (or is forced to use) the system (a reasonable assumption), then if a disproportionate share of the [R] group are forced to use transfers, wait longer for the bus, etc., than those in the [NR] group and if these negative service aspects are reflected in a more negative response to attitude questions on the take-home form, the results from the attitude questions would be more negative than would be the case if everyone completed the take-home form. No significant difference between [R] and [NR] for average waiting time and the need to transfer was found, as shown in Table 2.

Although it can be argued successfully that not all nonresponse bias on either the take-home (mail) or the on-board (intercept) survey form can be identified because there are people who will not respond to either form, a successful argument has been made that the nonresponse rate is decreased considerably by the on-board form.

CASE STUDY 2: WASHTENAW COUNTY ON-BOARD TRANSIT SURVEY AND TELEPHONE MAIL SURVEY

Washtenaw County, Michigan, has been considering expansion of its bus system into more rural areas as well as various funding options for the system. To garner information on the feasibility of various plans, a survey similar in structure to that used in Dade County was designed that contained both an on-board and a take-home form and accrued all the same types of benefits described for the Dade County survey. The response rate on the on-board form was 88 percent; on the take-home, 38 percent.

Of the 1171 respondents to the on-board form, 44 percent (510) sent in a mail survey. Thus, [NR] constituted 661 individuals (56 percent). Variables of age, automobile ownership, transit captivity, length of residence, sex, and driver's license were available for checking for nonresponse bias.

A result that confirmed the Dade County results was that in all cases those who had not returned the mail survey were also less likely to complete the questions on the on-board form fully. In the [NR] group, 7.1 percent left the age question blank (versus 2.4 percent for the [R] group); 7.6 percent of the [NR] group omitted automobile ownership (2.2 percent of the [R] group); 9 percent of the [NR] group omitted transit captivity (2.2 percent); 5.3 percent of the [NR] group left the length-of-residence question blank (3.5 percent); 7 percent omitted the question about sex (2.6 percent); and 6.0 percent omitted the driver's-license question (1.2 percent).

By using either chi-square or Kolmogorov-Smirnov tests (as appropriate), no significant differences were found between [R] and [NR] for age, automobile ownership, transit captivity, and length of residence. On the other hand, respondents to the mail survey were more likely to be female (chi-square = 16.4) and to not have a driver's license (chi-square = 39.7). Thus, analysis of the mail-back survey might be weighted to reflect more accurately the characteristics of the bus-riding public as revealed by the on-board survey, on which the response rate was more than twice that on the mail-back survey.

As a second part to this overall survey effort, another DSM was used, which consisted of a brief (5-min) telephone interview of a random, stratified sample (by using random-digit dialing), followed by a mail-out, mail-back survey. From eligible telephone contacts, a response rate of 80 percent was obtained, whereas the mail-back survey achieved a 56 percent response rate.

The results bear considerable similarity in the existence of nonresponse bias to the results of the on-board bus sample. A total of 2468 usable responses was obtained from the mail-back survey. As before, nonrespondents to the mail-back survey had a significantly higher rate of nonresponse to telephone survey questions on all questions. Some significant differences were found in the stratum response rates, although not in the major urbanized areas of the county. Apart from that, respondents to the mail-back survey were more likely to hold a driver's license and consequently to use cars as driver or passenger on a frequent basis and were more likely to have lived in the area for a long time, to own one or more automobiles, to be female, and to be older than the nonrespondent. Of particular importance here is that nonresponse bias appears related (as for the on-board survey) to variables related to the survey issues (i.e., transportation and mobility) and the funding of transportation investment. Again, this suggests the need to weight the survey results to reflect population characteristics more accurately if results are to be used to represent the county population.

Table 1. Response bias in mail-back survey revealed by on-board survey, Dade County.

	Mail-Back Respondents [R]		Mail-Back Nonrespondents [NR]	
Variable	Frequency	Percentage	Frequency	Percentage
Sex (χ^2 = 4.05)[a]				
Male	88	48.6	186	41.3
Female	81	44.7	209	46.4
No response	12	6.7	55	12.2
Total	181		450	
Age (D = 0.075)[a]				
<12	0	0.0	10	2.2
12-17	20	11.0	50	11.1
18-34	54	29.8	139	30.9
35-54	57	31.5	140	31.1
55-69	18	9.9	46	10.2
>70	18	9.9	13	2.9
No response	14	7.7	52	11.6
Total	181		450	
Driver's license (χ^2 = 1.56)[a]				
No	61	33.7	152	33.8
Yes	103	56.9	223	49.6
No response	17	9.4	75	16.7
Total	181		450	
Captivity status (χ^2 = 0.019)[a]				
Could be driver or passenger	41	22.7	102	16.9
Could be passenger	20	11.0	49	16.2
Could be driver	33	18.2	81	15.8
Could not go by car	69	38.1	171	36.0
No response	18	9.9	44	15.1
Total	181		450	

Notes: Raw chi-square values are reported. D = Kolmogorov-Smirnov statistic.
[a]Not significantly greater than 0 at the 0.001 confidence level.

CONCLUSION

This paper has first made the point that the response rates on interview surveys have often been overestimated, whereas the rates for mail surveys are often underestimated. Given the enormous cost savings of a self-administered mail survey, it would seem worthwhile to develop methods to improve such procedures. Certainly, Dillman's (1) total design method deserves significant attention in this respect. The DSM procedure described above has been shown to have significant advantages. The idea can be extended to virtually any survey effort. Researchers need not always think of a mail survey or a telephone survey or a face-to-face survey but rather the proper mix of these methods, which allows the researcher to take advantage of the benefits of each survey mechanism and to avoid as many disadvantages as possible.

One of the most important benefits of the DSM is its ability to increase response rates and thus decrease nonresponse bias. More important, the DSM facilitates the determination of the existence of nonresponse bias and provides a procedure for correcting for it. The traditional solutions to the nonresponse bias problem--to perform one-sample statistical tests on variables available in the census, to make a special effort to gain cooperation of a sample of those who have refused to cooperate, to compare early and late returns, or to assume extreme values for nonrespondents--although useful, have been shown to have some significant drawbacks. As an alternative, the DSM employed in the Dade County and Washtenaw County on-board transit surveys has been shown to be beneficial in ameliorating nonresponse bias somewhat by improving response rates and eliciting some information from those who will only take the time to respond to a brief survey form.

ACKNOWLEDGMENT

The research on which this paper is based was financed in part through funds from the U.S. Department of Transportation by support of local Unified Planning Work Programs. We would also like to thank the staffs of the Dade County Transportation Administration and the Ann Arbor Ypsilanti Urban Area

Table 2. Response bias in service variables from Dade County on-board survey.

	Take-Home Respondents [R]			Take-Home Nonrespondents [NR]		
Service Variable	Min	Frequency	Percentage	Min	Frequency	Percentage
Waiting time	χ = 18.20 S = 15.25 Median = 15[a]			χ = 17.32 S = 14.96 Median = 15[a]		
Need to transfer (χ^2 = 0.0003)						
Yes		66	36.5		164	33.3
No		107	59.1		266	56.4
No response		8	4.4		19	10.2
Total		181			450	

[a]Both distributions are highly skewed to the right because zero is a lower bound on waiting time. The rather high standard deviations are due, then, to a number of reasonably high waiting times experienced by a reasonably small number of persons.

Transportation Study Committee, under whose auspices this work was executed. We are responsible, however, for the facts and accuracy of the data presented here. The contents reflect our views and are not necessarily those of any of the participating agencies.

REFERENCES

1. D.A. Dillman. Mail and Telephone Surveys: The Total Design Method. Wiley, New York, 1978.
2. W. Wermuth. Uncertainty of Household Travel Surveys in Methodological and Measurement Contexts. In New Horizons in Travel-Behavior Research (P.R. Stopher, H. Meyburg, and W. Brög, eds.), Heath, Lexington, MA, 1981.
3. P.R. Stopher and A. Meyburg. Survey Sampling and Multivariate Analysis for Social Scientists and Engineers. Heath, Lexington, MA, 1979.
4. W. Brög and A.H. Meyburg. Consideration of Nonresponse Effects in Large-Scale Mobility Surveys. TRB, Transportation Research Record 807, 1981, pp. 39-46.
5. A.H. Meyburg and W. Brög. Validity Problem in Empirical Analyses of Non-Home-Activity Patterns. TRB, Transportation Research Record 807, 1981, pp. 45-50.
6. W. Brög and A.H. Meyburg. The Nonresponse Problem in Travel Surveys--An Empirical Investigation. TRB, Transportation Research Record 775, 1980, pp. 34-38.
7. Southeast Michigan Regional Travel Survey: Final Report 5. Schimpeler-Corradino Associates, Coral Gables, FL, June 1981.
8. K.A. Brownlee. A Note on the Effects of Nonresponse on Surveys. American Statistical Association Journal, 1957, pp. 29-32.
9. M. Young and P. Willmott. The Symmetrical Family. Panther Books, New York, 1973.
10. W.J. Goudy. Interim Nonresponse to a Mail Questionnaire: Impacts on Variable Relationships. Iowa Agriculture and Home Economics Experiment Station, Ames, Journal Paper j-8456, 1976.
11. M.N. Donald. Implications of Nonresponse for the Interpretation of Mail Questionnaire Data. Public Opinion Quarterly, Vol. 24, 1961, pp. 99-114.
12. M. Wright. Response and Non-Response to a Postal Questionnaire. Greater London Council, Res. Memorandum RM-548, 1978.
13. E.W. Waltz and W.L. Grecco. Evaluation of a Mailed Planning Survey. HRB, Highway Research Record 472, 1973, pp. 92-107.
14. D. Galin. How Reliable Are Mail-Surveys and Response Rates: A Literature Review. Journal of Marketing Research, 1975, pp. 440-453.
15. L. Kanuk and C. Berenson. Mail Surveys and Response Rates: A Literature Review. Journal of Marketing Research, 1975, pp. 454-473.
16. M.J. O'Neil. Estimating the Nonresponse Bias Due to Refusals in Telephone Surveys. Public Opinion Quarterly, Vol. 42, Summer 1979, pp. 218-232.
17. W.G. Cochran. Sampling Techniques. Wiley, New York, 1963.
18. C.H. Fuller. Weighting to Adjust for Survey Nonresponse. Public Opinion Quarterly, Vol. 38, 1975, pp. 239-246.
19. I. Wayne. Nonresponse, Sample Size, and the Association of Resources. Public Opinion Quarterly, Vol. 39, 1976, pp. 557-562.
20. D.J. Bem. Self-Perception Theory. In Advances in Experimental Social Psychology (L. Berkowitz, ed.), Academic Press, New York, 1972.
21. I.M. Sheskin, G.S. Spivack, and P.R. Stopher. The Dade County On-Board Transit Survey. Transit Journal, 1981, pp. 15-28.
22. Transit Development Program 1980-1985. Dade County Office of Transportation Administration, Miami, FL, March 1980.

Small-Sample Home-Interview Travel Surveys: Application and Suggested Modifications

PETER R. STOPHER

A method was put forward three years ago for estimating the sample sizes needed for travel surveys from information contained in earlier household surveys. The method showed that very small samples (of the order of 1000-3000 households) could be used to update trip rates and the succeeding steps of travel forecasting by using the information on standard deviations contained in 1950 and 1960 data. Despite the potentially far-reaching impacts of this method, little use appears to have been made of it. An application of the method is described that shows that, in a region of more than 1.6 million households, a sample of 2600 households was estimated as being sufficient to achieve measurement of trip rates to within ±5 percent sampling error with 90 percent confidence. After the survey had been executed, measured trip-rate variances and sample distribution were compared with those used for sample-size estimation from 1965 data. Although variances and distributions were found to have changed quite substantially, the sample was found to have produced trip-rate estimates that were within or no more than ±1.5 percent beyond the specified design sampling error. Second, it was found that the method originally put forward does not provide efficient or intuitively appealing samples for the common case of stratified trip-generation relationships. For this case, a procedure is put forward to specify the required levels of error in each stratum in such a way that account is taken of the magnitude of the trip rate and the size of the stratum. It is shown that this procedure is more efficient and that it yields more intuitively appealing sample distributions than the assumption implied by the earlier procedure of an identical percentage error for each stratum.

Many of the large urban areas of the United States are continuing in the 1980s to do transportation planning by using forecasting procedures calibrated on data collected in the 1960s. These data were generally collected by means of a random or systematic sample of households; the sampling rate was from 1 to 5 percent of the regional population. In urban areas of 100 000 population and more, this might have involved anywhere from a few thousand to 20 000 or 30 000 households in the sample. Because of the high cost of such surveys, few have been conducted since about 1972, and it is unlikely that funding will exist in the foreseeable future for such major surveys. Currently, the cost of a household interview such as that used in the 1960s data collection is anywhere from about $60 to $200; some instances of specialized data collection run well in excess of even $200. Such unit costs translate into survey costs of, perhaps, $200 000 for a small urban area of 100 000 population to several million dollars for urban areas such as New York, Chicago, and Los Angeles.

Given the age of the current primary data bases, the realities of urban growth and change of the past two decades, shifts in economic trends and patterns, and the emergence of higher fuel costs and potentially uncertain fuel supply, it is not surprising that many urban areas are concerned now to generate a new planning data base and provide the means to update or rebuild their travel-forecasting procedures. Given the tremendous costs of repeating the 1960s data collection and the dwindling of available planning funds in real dollars, the interest of many planners has turned to small samples, where "small" connotes absolute sample sizes of less than 5000 households, irrespective of urban-area size.

A major impetus was given to this direction by the work of Smith (1), which showed how to use the information collected in earlier surveys to design an efficient sample of very small size for updating travel-forecasting procedures. Smith's method uses the standard deviations obtainable from the 1960s data to compute coefficients of variation for relevant travel measures and then to compute the sample sizes needed to achieve a prescribed accuracy at specified confidence limits in new measurement of those variables. Smith showed that, for a particular scheme of trip-generation estimation, a sample size below 1000 households would achieve an accuracy of ±5 percent with 90 percent confidence for the estimation of trip rates. He then showed that this same sample size would be more than adequate to calibrate a gravity model of trip distribution and a modal-split model. Despite the significance of these findings, there appear to have been few attempts to utilize Smith's procedure since it was published. This paper reports on one such application of the formula and shows comparisons between the computations of error and sample size made from the original 1960s data and those from the new data. Although some changes in values were found, it is notable, as shown in subsequent sections of this paper, that these varying values would not have affected the sample sizes materially. The paper also describes a problem encountered with Smith's procedure and proposes a modification that should prove more useful in the future.

PRACTICAL SAMPLE

The critical variable for sample-size determination was defined to be the household tripmaking rate. The existing trip-generation forecasting procedure consists of four linear-regression equations with the independent variables of family life cycle, income, household size, and automobile availability; stratifications to four equations are on the basis of area type. Area type was defined in terms of a combination of employment density and residential density, such that the first area type comprises zones with a high density of employment, whereas the second, third, and fourth are zones of low employment density and residential density that is high in area type 2 and declines successively to area type 4.

The decision was made to seek the same accuracy level in each area type by specifying that trip rates in each area type be estimated to within ±5 percent with 90 percent confidence. While the original trip-generation modeling from 1965 data had been done by using regression, the data were reanalyzed as rates by using the Automatic Interaction Detection (AID) procedure to select subgroups within each area type by the other independent variables. AID is essentially a clustering procedure that was used to cluster households by sociodemographic characteristics within area types. Clustering was based on the tripmaking of the households. A total of 35 clusters were identified, as given in Table 1. The 1965 average trip rates for the four area types, the populations of the four area types, and the translation of these figures into total trips are given in Table 2. By using the trip rates of Table 2, it can be seen that the trip rates in area type 1 were to be estimated to some value equivalent to 1.78-1.96 with 90 percent confidence, between 3.71 and 4.11 in area type 2, and so forth.

Table 1. AID groups identified by area type.

		Characteristic			
Area Type	Subgroup	Income Group	Automobiles	Household Size	Life-Cycle Group
1	14	1	0	All	3
	15	1	0	All	1,2,4,5
	18	1	1,2+	1,2	All
	19	1	1,2+	3+	All
	4	2,...,5	All	1	All
	10	2,3	All	2	All
	16	2	All	3+	All
	17	3	All	3+	All
	12	4,5	All	2,3+	2,3,4
	13	4,5	All	2,3+	1,5
2	18	1	0	1,2	All
	19	1	1,2	1,2	All
	16	2-5	All	1	All
	17	2-5	All	2	All
	6	All	0	3+	All
	8	All	1	3+	3,4
	9	All	1	3+	1,2,5
	10	1,2,3	2+	3+	All
	14	4,5	2+	3+	1-4
	15	4,5	2+	3+	5
3	10	All	All	1	All
	12	All	All	2	3,4
	13	All	All	2	1,2,5
	8	All	0,1	3+	3,4
	9	All	0,1	3+	1,2,5
	6	All	2+	3+	1-4
	7	All	2+	3+	5
4	12	1	All	1,2	All
	14	2-5	All	1	All
	15	2-5	All	2	All
	8	All	0,1	3+	2,3,4
	9	All	2+	3+	2,3,4
	6	All	0,1	3+	1,5
	10	1,2,3,4	2+	3+	1,5
	11	5	2+	3+	1,5

Notes: Symbols used in this table are defined as follows. Income group: 1 = < $4000/year, 2 = $4000 - 5999/year, 3 = $6000 - 7999/year, 4 = $8000 - 9999/year, 5 = > $10 000/year; automobiles: 0 = no automobile available, 1 = one automobile available, 2+ = two or more automobiles available; household size: 1 = one-person household, 2 = two-person household, 3 = three-person household or more; life-cycle groups: 1 = head of household <35 years, no children <18; 2 = head of household 35-65 years, no children <18; 3 = head of household 65 years or more, no children <18; 5 = head of household any age, youngest child 6-18 years.

Table 2. Trip rates and total trips by area type.

Area Type	Trip Rate	Households (1980 estimate)	Total Trips (1980 estimate)
1	1.87	84 484	157 985
2	3.91	191 886	751 157
3	5.21	1 034 090	5 389 574
4	5.19	344 023	1 784 929
Total		1 654 483	8 083 645

Smith's procedure (1) was applied within each area type and to household subgroups defined by the AID analysis. The computations for this are given in Table 3. It should be noted that, unlike the recommendations made in Smith's paper (1), a coefficient of variation (CV) of 1.0 was not assumed, but individual CVs were calculated throughout. In fact, the CVs are found to exhibit considerable variation; they range from 0.227 to 1.477, but most values are below 0.8. The procedure requires that a sample size be computed on the basis of the required accuracy at the specified confidence level by estimating a pooled CV over the identified subgroups. Subsequently, the sample size may be readjusted on the basis of the subsample size in the critical cell, where this is defined as the cell that has the largest CV. Application of the sampling procedure generates a sample size for each cell based on its contribution to the overall CV. To draw the sample in this manner, however, would require information on the cell membership of every household in the population, which is clearly not likely to be available. Rather, the sample is likely to be drawn at random, in this case from all households in an area type. Given data on the frequency with which households occurred originally in the sample within each cell, an expected sample distribution can be computed. This will usually be different from the sample distribution based on the contribution to the overall CV. This shows clearly in Table 3 when the columns "Allocated Sample" and "Expected Sample" are compared.

The initial sample sizes computed from the procedure are 610, 450, 343, and 404 households, respectively, for the four area types, which gives a total sample requirement of 1807 households. If one imposes the requirement that the critical cell (indicated by footnote a in Table 3) must be sampled at the design sample size, then the expected sample should be increased by the ratio of the allocated to expected sample for the critical cell in each area type (1). This produces the values shown in the column "Full Random Sample" and produces samples of 1157, 660, 481, and 524 for the four area types, respectively, and a total sample of 2822 households.

Although this completed the sample-size computation from a statistical standpoint, it was not considered to have defined an acceptable sample on the basis of other needs of the sampling procedure. The study region consists of multiple jurisdictions for which various planning and policy actions are expected to be done by the metropolitan planning organization (MPO). For planning based on this survey to be acceptable to the various jurisdictions, there is a need for the sample to be reasonably proportionately distributed over the jurisdictions. The expected distribution of the sample by jurisdiction (by using the eight primary jurisdictional levels) and four area types is given in Table 4 together with the percentage of the sample in each entry of the table. Table 5 notes the percentage of the population in each cell. A comparison of these two tables shows that the sample distribution is disproportionately heavy in area type 1 and jurisdiction 1.

From the politics of MPO planning, this is not acceptable. Therefore, several changes were made to the sample sizes based on the statistical sample and jurisdictional concerns.

The first adjustment made was to reduce the size of the sample for area type 1. The required sample here almost doubled in size when the critical cell was considered, although this cell generates very few of the regionwide trips. It was determined that accepting the expected sample of 24 households would increase the error at 90 percent confidence from ±5 percent to ±7.5 percent. This was felt to be acceptable in light of the very large increase needed otherwise in this sample size and its implication for the entire sample distribution.

The second adjustment was based on the selected method of sampling. Smith's procedure is based on the assumption that a simple random sample is selected. The sampling procedure used in this case, however, was a three-stage sampling procedure by using zones, blocks, and households as the sampling units for the three stages. Multistage sampling provides considerable gains in sampling accuracy and inexpensiveness when a full enumeration of the final sampling units does not exist but increases the sampling error over that of simple random sampling of the final-stage units (2,3). To calculate the sampling error for the multistage procedure, it would be necessary to know the standard deviations of trip rates by zone and by block. This information was not available and could not be computed readily at the time of sampling, so precise sampling errors could not be computed. To allow for the increased error, an across-the-board arbitrary increase of 10 percent was applied to the sample sizes. Given the importance of area type 3, by virtue of both its trip rate and the proportion of households, it was decided to add a further 95 households to this sample, distributed proportionately over all jurisdictions. This brought the total sample to 620 in area type 3. Finally, 50 households were removed from area type 2 and added to area type 4 to be distributed over all jurisdictions except 1. These sample-size changes were decided on as being politically or judgmentally desirable and were not based on statistical analysis. A summary of these changes is given in Table 6, and Table 7 gives the final designed sample.

RESULTS OF SURVEY SAMPLE EXECUTION

In execution, a total of 2706 interviews were conducted, of which 2446 were considered to be sufficiently complete for analysis, including data on the independent variables for trip-rate analysis. Some of the 2446 sampled households had an estimated income based on data on area type, available vehicles, and number of workers. The distribution of the achieved sample by area type and jurisdiction is shown in Table 8. Comparisons of this table with Table 7 show that a fairly good approximation to the design sample was achieved, with the exception of area types 1 and 2 in jurisdiction 1. The samples in these localities proved to be quite problematical due to urban renewal and localities of high unemployment.

Table 9 gives the computations of sample size given the trip rates and their standard deviations as actually measured in the survey. The sample sizes attained were in all cases close to or in excess of those required for ±5 percent error at 90 percent confidence, despite the changes in critical cells and the general shifts in CVs.

On the basis of this use of the procedure for sample estimation, after the elapse of more than 15 years, it appears that the sample sizes estimated are perfectly adequate and sufficiently robust to provide acceptable accuracy, even where trip-rate measures have not been very stable. Furthermore, even though quite small sample sizes are generated, these are proved adequate to measure trip rates to the required level of accuracy. Through this method, a major cost saving is realized. In 1965, the TALUS survey sampled 4 percent of the region's households. With the increased region and population, which totaled more than 1.65 million households in 1980, the same sampling rate would have required a sample of 66 000 households. At $100 per interview, the survey would have cost $6 600 000, $6 300 000 more than this survey. A problem does arise, however, in that specifying the level of accuracy in terms of trip rates appears to be inadequate with respect to accuracy of trip estimation. This issue is discussed at greater length in the remainder of this paper.

APPROPRIATE SPECIFICATION OF ACCURACY

In the case study described in this paper, the regional population of households was stratified first into four area types. Subsequently, the same percentage error was specified for each area type, notwithstanding the major differences in subpopulation size and the variations in trip rates present across the area types. This process led to domination of the sample by area type 1, even though this area type produces only 1.95 percent of regional trips. To achieve the required 5 percent accuracy in this stratum, 1157 sample households were needed out of a statistically computed total sample of 2822 households for the four strata. In other words, 41 percent of the sample was required to measure 1.95 percent of the regional trip total. This situation arises for several reasons. First, because there are few households in this area type, the simple random sample from 1965 located few households in this stratum (4.5 percent or 523 out of a sample of

Table 3. Sample-size calculations for all four area types.

Cell	CV_i	F_i	F_iCV_i	W_i	Allocated Sample	Expected Sample	Full Random Sample
Area type 1							
14	0.426	0.171	0.0728	0.097	59	104	204
15	0.678	0.237	0.1606	0.214	130	145	284
18	0.787	0.079	0.0622	0.083	51	48	94
19	0.743	0.044	0.0327	0.044	27	27	53
10	0.828	0.114	0.0944	0.126	77	70	137
16	0.884	0.118	0.1043	0.139	85	72	141
4	0.527	0.081	0.0427	0.057	35	49	96
17	0.865	0.054	0.0467	0.062	38	33	65
12	1.200	0.064	0.0768	0.102	62	39	76
13[a]	1.477[a]	0.039	0.0576	0.077	47[a]	24[a]	47
			0.7508		610	610	1197
Area type 2							
18	0.271	0.099	0.0268	0.042	19	45	66
19	0.460	0.074	0.0342	0.053	24	33	48
16	0.283	0.057	0.0161	0.025	11	26	38
17	0.549	0.215	0.1181	0.183	82	97	142
6	0.526	0.050	0.0264	0.041	18	23	34
8	0.772	0.153	0.1179	0.183	82	69	101
9	0.863	0.138	0.1191	0.185	83	62	91
10	0.833	0.066	0.0554	0.086	39	30	44
14	0.842	0.081	0.0681	0.106	48	36	53
15[a]	0.944[a]	0.066	0.0627	0.097	44[a]	30[a]	44
			0.6446		450	450	660
Area type 3							
10	0.268	0.057	0.0154	0.027	9	20	28
12	0.460	0.043	0.0195	0.035	12	15	21
13	0.458	0.178	0.0814	0.145	50	61	86
8	0.593	0.199	0.1182	0.210	72	68	95
9	0.227	0.120	0.0274	0.049	17	41	58
6	0.705	0.209	0.1471	0.261	90	72	101
7[a]	0.795[a]	0.194	0.1541	0.274	94[a]	67[a]	94
			0.5631		343	343	481
Area type 4							
12	0.357	0.089	0.0317	0.052	21	36	47
14	0.320	0.028	0.0090	0.015	6	11	14
15	0.401	0.191	0.0764	0.125	50	77	100
8	0.578	0.173	0.0998	0.163	66	70	91
9	0.726	0.209	0.1515	0.248	100	84	109
6[a]	0.788	0.092	0.0728	0.119	48[a]	37[a]	48
10	0.764	0.082	0.0628	0.103	41	33	43
11	0.787	0.137	0.1074	0.176	71	55	71
			0.6114		404	404	524

[a]Critical cell.

Table 4. Initial sample distribution by jurisdiction and area type.

	Area Type									
	1		2		3		4		Total	
Jurisdiction	No.	Percent	No.	Percent	No.	Percent	No.	Percent	No.	Percent
1	661	23.4	566	20.0	110	3.9	2	0.1	1339	47.0
2	199	7.0	80	2.8	140	5.0	89	3.2	508	18.0
3	132	4.7	9	0.3	104	3.7	192	6.8	437	15.5
4	51	1.8	0	0	89	3.2	59	2.1	199	7.0
5	95	3.4	7	0.2	24	0.8	45	1.6	171	6.1
6	12	0.4	0	0	4	0.1	53	1.9	69	2.4
7	8	0.3	0	0	8	0.3	44	1.6	60	2.1
8	0	0	0	0	2		40	1.4	42	1.5
Total	1158	41.0	662	23.4	481	17.1	524	18.5	2825	100.0

Table 5. Percentages of population by area type and jurisdiction.

Jurisdiction	Area Type 1	2	3	4	Total
1	2.9	9.9	14.3	0.1	27.2
2	0.9	1.4	18.3	3.5	24.1
3	0.6	0.2	13.6	7.6	22.0
4	0.2	0	11.5	2.3	14.0
5	0.4	0.1	3.1	1.8	5.4
6	[a]	0	0.5	2.1	2.6
7	[a]	0	1.1	1.7	2.8
8	0	0	0.3	1.6	1.9
Total	5.1	11.6	62.5	20.8	100.0

[a] Less than 0.1 percent.

Table 6. Adjustments to statistical sample by area type.

Area Type	Adjusted Sample from Smith's Procedure	Reduction in Area Type 1 Sample	Adjustment for Multistage Sampling	Adjustment for Area Type 2,3,4 Samples
1	1157	610	680	680
2	660	660	725	675
3	481	481	525	620
4	524	524	575	625
Total	2822	2275	2505	2600

Table 7. Distribution of final sample.

Jurisdiction	Area Type 1	2	3	4	Total
1	388	578	141	2	1109
2	117	82	181	106	486
3	78	9	135	229	451
4	30	0	114	71	215
5	56	7	31	54	148
6	7	0	5	63	75
7	5	0	11	53	69
8	0	0	3	48	51
Total	681	676	621	626	2604

Table 8. Distribution of executed sample.

Jurisdiction	Area Type 1	2	3	4	Total
1	348	465	139	0	952
2	98	73	170	90	431
3	73	23	156	219	471
4	33	0	166	72	271
5	41	13	38	37	129
6	11	0	0	61	72
7	13	0	16	50	79
8	0	0	0	41	41
Total	617	574	685	570	2446

Table 9. Calculations of sample sizes based on survey results.

Cell	CV_i	F_i	$F_x CV$	W_i	Optimal Sample	Expected Sample	Full Sample	Executed Sample	Design Sample
Area type 1									
14	0.652	0.238	0.155	0.181	144	188	296	147	117
15	1.313	0.066	0.087	0.102	80	52	82	41	161
18	1.061	0.149	0.158	0.185	146	118	186	92	54
19	1.352	0.107	0.145	0.170	134[a]	85[a]	134[a]	66[a]	30
10	0.585	0.089	0.052	0.061	48	70	110	55	78
16	0.720	0.109	0.079	0.092	73	86	136	67	80
4	1.245	0.070	0.087	0.102	81	55	87	43	55
17	0.318	0.028	0.009	0.010	8	22	35	17	37
12	1.112	0.044	0.049	0.057	45	35	55	27	44
13	0.341	0.100	0.034	0.040	31	79	125	62	27+
			0.855		790	790	1246	617	683
Area type 2									
18	1.038	0.099	0.103	0.141	82	58	97	57	67
19	1.217	0.064	0.078	0.106	62[a]	37[a]	62[a]	37[a]	50
16	0.742	0.206	0.153	0.208	122	121	203	118	39
17	1.066	0.037	0.039	0.053	31	22	37	21	145
6	0.651	0.054	0.035	0.048	28	32	54	31	34
8	0.875	0.061	0.053	0.072	42	36	60	35	103
9	0.622	0.099	0.062	0.084	49	58	97	57	93
10	0.899	0.117	0.104	0.141	83	68	114	67	45
14	0.397	0.171	0.068	0.093	54	100	168	98	55
15	0.430	0.092	0.040	0.054	32	54	90	53	45[a]
			0.735		585	586	982	574	676
Area type 3									
10	0.825	0.162	0.134	0.218	89	66	95	113	35
12	0.807	0.215	0.174	0.283	116	88	126	150	27
13	0.773	0.079	0.061	0.099	41	32	46	55	111
8	0.857	0.099	0.085	0.138	57[a]	40[a]	57[a]	69[a]	124
9	0.427	0.262	0.112	0.182	74	107	153	183	75
6	0.410	0.052	0.021	0.034	14	21	30	36	130
7	0.245	0.113	0.028	0.046	19	46	66	79	120
			0.615		409	410	573	685	621
Area type 4									
12	0.726	0.205	0.149	0.253	95	77	118	117	56
14	0.912	0.179	0.163	0.277	103[a]	67[a]	103[a]	102[a]	18
15	0.520	0.107	0.056	0.095	36	40	61	61	120
8	0.656	0.084	0.055	0.093	35	32	49	49	108
9	0.747	0.037	0.028	0.048	18	14	21	21	131
6	0.402	0.249	0.100	0.170	64	93	143	143	58[a]
10	0.289	0.073	0.021	0.036	13	27	42	42	51
11	0.252	0.068	0.017	0.029	11	26	40	40	86
			0.589		375	376	577	570	626

[a] Critical cell.

Table 10. Constant-magnitude error by area type.

Area Type	Avg Trip Rate	90 Percent Confidence Limit of Sampling-Error Value	Percent
1	1.87	±0.244	±13.0
2	3.91	±0.244	±6.24
3	5.21	±0.244	±4.68
4	5.19	±0.244	±4.70

Table 11. Z^2/E^2 and sample size by stratum.

Area Type	Percent Error Required	Z^2/E^2	$C\bar{V}_i$	n
1	13.0	160.1	0.7508	90
2	6.24	695.0	0.6446	289
3	4.68	1235.5	0.5631	392
4	4.70	1225.0	0.6114	458
Total	--	--	--	1229

Table 12. Comparisons of sample and population.

Area Type	Percent of Population	Percent of Percentage Sample	Percent of Absolute Sample
1	5.1	41.0	10.1
2	11.6	23.4	24.3
3	62.5	17.0	31.5
4	20.8	18.5	34.1

11 512 usable household records). Thus, the means and standard deviations were estimated from very small samples. (All but one of the 13 cells had samples less than 90 and that one exception had a sample size of 123. Among the other area types, with a total of 33 cells, the smallest sample size was 47 and the next was 138. The remaining cells ranged from 150 to 1125 households.)

Second, area type 1 was defined only in terms of high employment density. The zones occur mainly in central business districts (CBDs) and outlying business districts (OBDs) and exhibit wide variations in residential characteristics. Trip rates varied by cell from 0.27 to 5.15; the mean was 1.87. Variations in the other trip rates were generally markedly smaller. Thus, area type 1 households constitute a diverse group of households in terms of tripmaking and are inaccurately measured because of the small sample size. Third, although the initial sample size estimation is close to the sample sizes of the other area types (610 compared with samples between 343 and 450), one cell--the critical cell--in area type 1 has a very small frequency of occurrence but a large CV. It serves to double the sample size to 1157. This also should be seen in the context that this cell is responsible for 0.08 percent of the region's tripmaking.

The basic problem identified by this case study is that the sample trip rates bear no relation to the planning units of measurement, for which sampling is really designed. Given that trip rates are the units that will be estimated and about which standard deviations and means are known from previous surveys, the primary issue becomes one of how to weight the trip rates so that the samples drawn are in reasonable relation to the impact of the rates on estimation of travel volumes. This problem arises only under the circumstances that some form of stratification or segmentation takes place and samples are to be estimated independently for each stratum or segment. Complication is added by the fact that a unique set of sampling rates cannot be obtained from the equation for the sampling error for a stratified sample with variable sampling fraction, unless some relationship is prespecified between the sampling rates or stratum sampling errors.

First, consider the effects of the stratification used in this case. If one applies the estimating procedure for a stratified sample with variable sampling fraction, the sampling error for the regionwide average trip rate can be computed. The estimation is made from the following:

$$\text{s.e.}\,(\hat{\bar{y}}) = (\sum_i g_i{}^2 n_i s_i{}^2/\hat{N}^2)^{1/2} \quad (1)$$

where

$\text{s.e.}\,(\hat{\bar{y}})$ = sampling error of $\hat{\bar{y}}$,
g_i = expansion factor for stratum i,
n_i = sample size in stratum i,
s_i = standard error of $\bar{y}_i$, and
$\hat{N}$ = estimated total population = $\Sigma_i g_i n_i$.

The estimated standard errors of the stratum trip rates are ±2.113, ±1.083, ±1.198, and ±1.220, respectively. By using the original sample sizes from Smith's procedure shown in Table 4, the sampling error is ±0.0364.

The weighted average trip rate is 4.886, so the error at 90 percent confidence is ±1.22 percent. By using the adjusted samples shown in Table 7, this sampling error reduces to ±0.0323 and a 90 percent confidence bound of ±1.09 percent. Clearly, by specifying ±5 percent error in each stratum, the error over all strata is much less than ±5 percent, as expected. It is interesting to note that the reduced sample in area type 1 is outweighed by the increases in area types 3 and 4.

As a means to define more appropriate sample sizes for a stratified sample, consider specifying an error on the weighted average trip rate. If one specifies a requirement of ±5 percent error on the average trip rate of 4.886, this represents an error of ±0.244. Now, suppose that this error in the rates is specified for each stratum. This means that, irrespective of stratum, any given household will have the same probability of a misprediction of given magnitude. The reason for choosing this definition of error is that it means that tripmaking by each household is estimated to the same absolute level of accuracy. Thus, in looking at any group of trips, such as those in a corridor, on a specific facility, or those in a subarea, all of the trips in the group will have been estimated to the same level of accuracy, irrespective of the type of household that generated the trips. It implies also that one is less interested in household trip rates per se but is more interested in numbers of trips by some grouping geographically or modally.

In this case study, the effect of this is to specify the trip rates and 90 percent confidence limits on error (see Table 10). This is markedly different from the constant percentage error, which at 5 percent generates absolute errors of ±0.094, ±0.196, ±0.261, and ±0.260, respectively. Again, the implications of this are that with many more households in area types 3 and 4 than in 1 and 2, the absolute error in trips will be higher than with the specification shown in Table 9.

Applying these new sampling errors produces different values of Z^2/E^2 for each stratum (see Table 11). It can be seen that the sample sizes are markedly different from those obtained from the constant percentage error sample. If these individual samples are then allocated across the cells of each area type as before, an increase in sample size is required for the critical cell of each area type, which increases the four samples to 176, 424, 550, and 594, respectively, and requires a total of 1744 households. This is noticeably smaller than the 2822 generated from the percentage sample. Consider also the percentages of the sample and population in each area type for this procedure compared with the previous one, as given in Table 12. The new sample's percentages bear a more logical relationship to the population than those of the original sample.

Because all sample sizes were increased to produce the minimum required sample in the critical cell, the final sample of 1744 households will produce a smaller error than the specified ±5 percent of the weighted average trip rate. By using the estimation for a stratified random sample with variable sampling fraction, the sampling error is found to be ±0.035, which produces a 90 percent confidence limit on the error of ±1.18 percent. This is slightly less than the ±1.22 percent error obtained from the 2822 sample. An interesting comparison can be obtained to the achieved sample of 2446 with its distribution among the area types. This sample provides a sampling error of ±0.0313, which is ±1.05 percent at 90 percent confidence. Because of the changed distribution imposed in design and further shifted in execution, this sample produced a smaller error on overall trip rates than the statistically designed sample based on a ±5 percent error. The greater efficiency of the absolute-value-based sample is shown by increasing that sample of 1744 households to 2446 with the same proportionate distribution as in the 1744 sample. In that case, the error on the overall weighted trip rate is ±0.0296, which gives a 90 percent confidence limit of ±1.00 percent. This shows that the absolute-value sample is more efficient than the percentage-based sample as well as being more reasonable on the basis of prediction of trip volumes. Similarly, increasing the sample size to 2822 reduces the sampling error yet further to ±0.93 percent at 90 percent confidence.

CONCLUSION

The sample-size estimation procedure developed by Smith (1) has been shown to produce an adequate sample for updating trip-generation rates from previous years' surveys. Despite changes in the distribution of households over the relevant cells, the sample produced trip-rate estimates that were within ±1.5 percent of the required 90 percent confidence limit on sampling error, even though the executed sample was about 6.5 percent short of the design sample and more significant shortfalls of 10 and 15 percent occurred in area types 1 and 2. The method appears robust enough to be able to handle the realities of real-world survey execution and changes in population distribution over the elapse of 15 years.

The case study used here also shows that this procedure for sample estimation may need to be used as only the initial estimate of sample size and distribution. Political and jurisdictional realities are likely to require that the sample sizes be changed and augmented to satisfy other requirements than purely statistical ones. Nevertheless, judicious changes should not threaten the statistical reliability of the sample, if these changes are made with the goals of the sampling clearly in mind.

Finally, this case study shows that the sample is likely to be estimated inappropriately if the trip-generation procedure is based on stratification and sample sizes estimated independently in each stratum. In this case, independent estimation can lead to domination of the sample by a stratum of households that has a low trip rate and that may represent a very small proportion of regional households. In this case, this was found to happen, so that a stratum containing 5.1 percent of regional households and producing 1.95 percent of regional trips was estimated to require 41 percent of the sample. The need was identified, therefore, to determine a more rational basis for specifying the permissible sampling error than the direct extension of Smith's procedure, which leads to specifying a constant percentage error for all strata.

The proposed modification for stratified sampling is to estimate the permissible error as an absolute number (fraction) of trips per household and then calculate this as a fraction of the mean trip rate in each stratum. This procedure has been shown to generate a smaller sample requirement than that by using a constant percentage error and to provide a distribution of the sample by stratum that is intuitively more appealing. In this case, the low trip-rate stratum requires 10 percent of the sample instead of 41 percent, which seems much more reasonable for the stratum's contribution to trip totals and to probable error. Furthermore, the resulting sample in this case is smaller and has a smaller overall error than the sample generated from a constant relative error. Comparing the overall weighted trip-rate error between the absolute-error method and the relative-error method, one finds that the absolute-error method reduces the sampling error by almost 25 percent or reduces the required sample size by 38 percent.

ACKNOWLEDGMENT

I would like to thank staff at the Southeastern Michigan Transportation Authority and the Southeastern Michigan Council of Governments for the use of data from the 1965 Detroit Regional Transportation and Land Use Study (TALUS) and the 1980 regional travel survey that provided the case study material for this paper.

REFERENCES

1. M.E. Smith. Design of Small-Sample Household-Interview Travel Surveys. TRB, Transportation Research Record 701, 1979, pp. 29-35.
2. F. Yates. Sampling Methods for Censuses and Surveys. Charles Griffin and Sons, London, 1965.
3. P.R. Stopher and A.H. Meyburg. Survey Sampling and Multivariate Analysis for Social Scientists and Engineers. Heath, Lexington, MA, 1979.