

Management of a Small Home Interview Travel Survey Using a Microcomputer

RONALD EASH

In this paper, how a microcomputer and general-purpose business software were used in a small telephone survey of household travel behavior is discussed. Principal applications covered are (a) selecting sample households; (b) mailing to sample households; (c) scheduling interviews; (d) checking the status of each interview; (e) developing the completed interview data sets; and (f) reporting survey results. Staffing requirements and expenditures to complete the survey are also presented. This form of computing is well suited to data collection projects of this size.

The Chicago Area Transportation Study participated in a project to evaluate a commuter rail service extension in northeastern Illinois. The agency's responsibilities were to develop and calibrate a mode choice model to estimate ridership for the proposed commuter rail service. The resulting model includes choice of suburban station, choice of access mode to the suburban station, and choice of central business district (CBD) station to final destination egress mode—the sequence of decisions typically faced by commuter rail riders.

No data set with the detail needed to calibrate this type of model was available, so some data collection had to be included in the project. The use of a current database for the study area also strengthened the study's recommendations.

In this paper, a telephone survey of approximately 200 households to collect data on daily household travel behavior is discussed. This survey was managed with the aid of a microcomputer and commercially available microcomputer software. A small personal computer (PC) was used for (a) selecting sample households; (b) preparing personal form letters and mailing labels; (c) scheduling interviews; (d) tracking the status of each household's interview; (e) compiling data from the completed interviews; and (f) reporting survey results.

THE SOO LINE COMMUTER RAIL FEASIBILITY STUDY

Figure 1 shows the study area for the project, which covers 12 townships, each approximately 36 mi² in area. The track for the proposed service extension is owned by the Soo Line Railroad, which discontinued passenger service on this line in the 1950s.

The lead agency for the Soo Line commuter rail feasibility study was the Northeast Illinois Rail Corporation, now known as METRA. The consulting team assembled for the study was

Chicago Area Transportation Study, 300 W. Adams Street, Chicago, Ill. 60606.

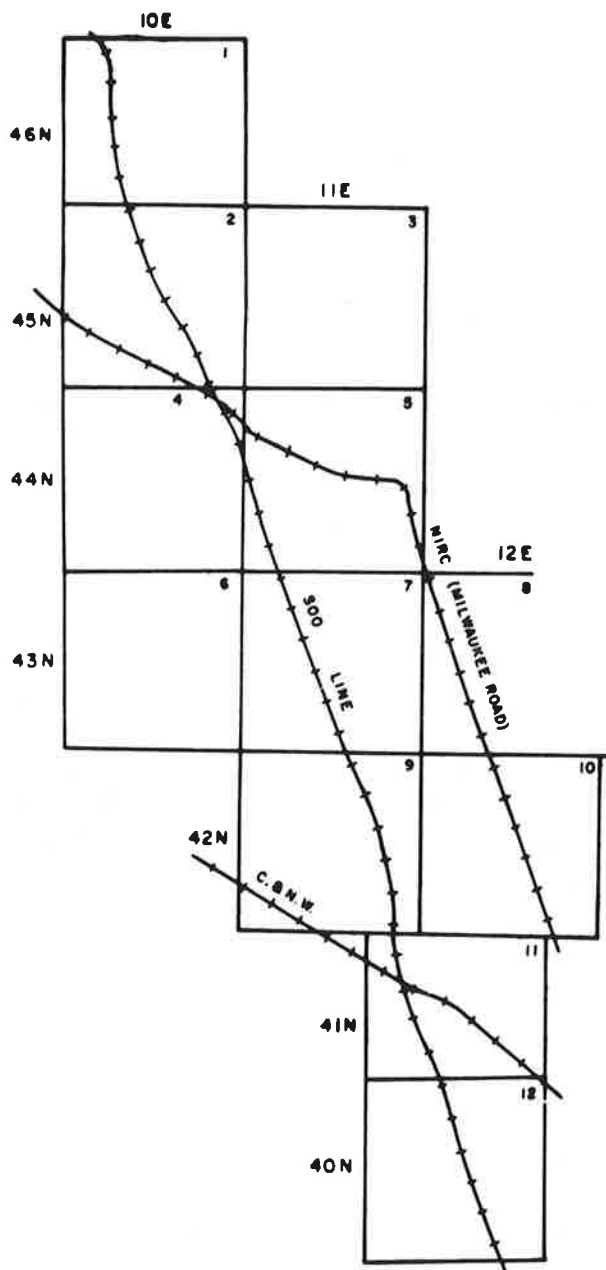


FIGURE 1 The Soo Line Corridor study area.

headed by R.L. Banks and Associates. JHK and Associates and CATS were subcontractors to R.L. Banks and responsible for the ridership estimates and model development.

OVERVIEW OF THE SURVEY

For model calibration, the trip data set obtained from the surveyed households needed to have coverage in two dimensions. Trips had to be distributed across the study area townships and all major travel modes. Minimum numbers of responding households were required for each township and work trip travel mode, so selection of households to contact depended on the interviews completed while the survey was under way.

After a household was selected, a contact letter was sent to the household slightly in advance of the designated survey day. Accompanying this letter was a form that household members could use to keep track of their survey day travel. The evening following the survey day, the household was contacted by telephone. The person who answered was asked to report their household's travel during the previous day using the completed trip reporting form to recall trips. The interviewer would record the household's trips in a standard format, resolve inconsistencies in the reporting of trips, and try to aid in the recall of trips.

SELECTION OF HOUSEHOLDS TO BE CONTACTED

Households were selected for sampling from three sources depending on the probable mode chosen for household work trips. Sample households were obtained by (a) random sampling of households within townships; (b) random sampling of households on monthly ticket lists for existing commuter rail services in the study area; and (c) on-board solicitation of volunteers on several feeder and express bus lines in the corridor. The approximate numbers of these three types of households in study area townships were determined from the 1980 census Urban Transportation Planning Package (1).

The variance in the estimate of a population characteristic obtained from a stratified sample is minimized by proportionally sampling each stratum according to the product of the fraction of population in the stratum times the standard deviation of the characteristic within the stratum (2). For households selected randomly inside townships, the required number of sample households in a township was computed from the number of households in the township and the variance in

township journey-to-work travel times reported in the census. Township household income as reported in the census was the population characteristic used to determine the sample size in the two categories of households with transit work trips.

In the case of the randomly selected households, it seemed desirable to have large samples in townships with varied travel behavior, which is partly measured by the variance in work travel times. Transit-oriented households were sampled to increase the sample sizes in townships with a variety of households. For this purpose, income seemed as good a measure of household variety as any other characteristic available from the census. Table 1 presents the desired household sample sizes assuming 210 (rounding produces a total of 212) completed interviews.

APPLICATION OF THE MICROCOMPUTER TO SAMPLE SELECTION

Randomly sampled households within townships were selected in the following way. Random X-Y coordinates were generated to identify points within air photographs. The nearest intersection to the randomly generated map coordinates was then located. One of the approaches into the intersection was chosen at random, and the range of addresses on this intersection approach determined between the selected intersection and the adjacent intersection. An address in this range was randomly generated and the nearest address that belonged to a private household selected from a reverse (address to telephone number) telephone directory (3). The directory name and address were used for correspondence with the household.

Several short BASIC programs were written to generate map coordinates, select the intersection approach leg, and determine an address within an address range. In situations where there were multiple telephone listings for the same address, indicating more than one household at that address, selection of the household to contact was again at random.

Households from the monthly commuter rail ticket lists were chosen using randomly generated page numbers and page positions in the monthly ticket holder printouts. The printout page numbers and page positions were also selected with a short BASIC program.

TABLE 1 DESIRED HOUSEHOLD SAMPLE BY LOCATION AND WORK MODE

Selection of Households	Township												Total
	1	2	3	4	5	6	7	8	9	10	11	12	
1. Randomly by Township	2	5	3	2	6	3	6	12	22	15	26	28	130
2. Rail Monthly Tickets	0	2	0	1	2	1	3	12	13	10	12	5	61
3. Bus Riders	2	1	1	0	1	1	0	5	4	6	0	0	21
Total	4	8	4	3	9	5	9	29	39	31	38	33	212

MANAGEMENT OF THE INTERVIEWS

Figure 2 is the format of the data file for mailings of survey materials and the scheduling of interviews. This file was prepared with the microcomputer program PFS:FILE (4), an inexpensive (the full retail price was \$125) database management program with limited features. This is one of a number of such programs that can be purchased in this price range with comparable features.

The file format shown in Figure 2 is designed the same way one might prepare a set of index cards, whereby each card keeps track of one contacted household. The format in Figure 2 is for households randomly selected by township. The format for households from the monthly ticket lists and bus riders is identical except for a field that identifies the source of the household.

The name, address, and telephone number of the household in this file are copied from the reverse telephone directory listing. The state and last name of the household, which would appear to be redundant, are included so that the file can be used for mailing labels and personalized form letters.

Fields for scheduling the telephone interview dates are listed in the format underneath the address information. Interviewers are identified in the file along with whether the interviewer could contact the household and, if contacted, whether the household cooperated in the survey. This information allowed for some limited evaluation of interviewer performance. At the end of the format is the final disposition of the household. Is the household included in the data set? unreachable after three call-backs? not a household? or not cooperative?

Household travel was surveyed only on midweek days, Tuesday through Thursday. Phone interviews were usually scheduled for Monday, Wednesday, and Thursday evenings after the dinner hour. In order to complete the mailing of survey forms, households had to be assigned a survey day about 1 week in advance of the interview.

The scheduling of households depended on (a) the availability of telephone interviewers to work the evening following the survey day; (b) the running tabulation of successfully completed interviews by township and mode; (c) the number of households to be recalled (households that had received the mailing but could not be contacted on their assigned survey night); and (d) previous experience with refusal rates. After an evening of interviewing, the file of contacted households was updated the following morning.

MAILINGS TO SAMPLE HOUSEHOLDS

A personalized form letter and a form for tabulation of household trips were mailed to each contacted household. The mailing was scheduled so that the form and introductory letter arrived no more than a day or so in advance of the survey date. Mailing labels were generated from the survey household file discussed, again using the PFS:FILE program. The survey date field was used to pull the addresses of scheduled households for the mailings from all households in the file.

Form letters were printed for each household using the word processing program PFS:WRITE (5), which was compatible with the contact household file. The inside address and salutation were copied from the contact household file and filled out as the letters were printed. The selection of which form letters to print was controlled through the survey date field in the household data file. Labels for the interview forms used by the phone interviewers were also created from the contacted household file.

RESOURCES REQUIRED FOR THE INTERVIEWING

Due to the small size of the survey, only two or three phone interviewers were needed each evening. The interviews were carried out over a 3-week period in the early fall of 1985. During the 3-hour interview period (6:30 to 9:30 p.m.), an

RANDOM SAMPLE HOUSEHOLDS

```

Township Number:      Household Number:
=====
Directory Name:
Address Name (If Different):
Address1:
Address2:
City:                  State:      Zip:
Telephone Number:     Last Name:
=====
Scheduled Survey Date:      Scheduled Survey Call Date:
1st Call Back:      2nd Call Back:      3rd Call Back:
=====
Scheduled Interviewer:
Completed Interview:      Unable to Reach:      No Longer a Residence:
=====
Call Back Interviewer:
Completed Interview:      Unable to Reach:      No Longer a Residence:
=====
Interview in Sample:

```

FIGURE 2 Data file format for contacted households.

experienced interviewer could complete about 10 interviews. Allowing for refused interviews and available call-back interviews, approximately 40 to 50 households received mailings for each survey day.

With the exception of the interview staff and a small amount of staff time to choose the contact households, only one staff member was required to run the survey. A single staff person familiar with database management and word processing software can easily handle the survey mailings and scheduling of interviews for this size of survey. Coordination of the mailings, interview scheduling, and completed interviews—critical for

this type of survey—are difficult when the mailings and data processing require several agency staff.

COMPILATION OF SURVEY DATA SETS

Three data sets for model calibration were prepared from the surveyed households. Data were tabulated by household, by person for each adult member of a household, and by trip for all household trips that could reasonably be completed by commuter rail. All surveyed households are combined in these three data sets.

```

HOUSEHOLD DATA SET:
=====;
Random Sample Households Township Number:
Sample Transit Households Township Number:      Transit Source:
=====;
Household Number:      Address1:
City:                  Zip:
Home Q-Sec:
-----;
Composition:

Adults:      Children:      Enfants:      Female Head:
Elderly:      Handicapped:      Drivers:
-----;
Employment:

Emp. Full-Time:      Emp. Part-Time:      Retired:
Student Full-Time:      Student Part-Time:
-----;
Vehicles:
Automobiles:      Trucks/Vans:

```

FIGURE 3 File format for the household data set.

```

PERSONAL CHARACTERISTICS FILE:
=====;
Random Sample Households Township Number:
Sample Transit Households Township Number:      Transit Source:
Household Number:
=====;
Adults:      Person Number:
Sex1:      Relationship1:      Employment:      Drive:      Elder:      Hand:
-----;
Children:      Child Number:
Sex2:      Age:      Relationship2:
-----;
Work Trip:      Residence:
Travel Time:      City:
Principal Mode:      Zip:
Work City:      Home Q-Sec:
Work Intersection:
Work Q-Sec:
-----;
Commuter Rail:      Line:      Station:
Access Mode:      Parking Cost:      Transit Fare:

Soo Line Trips

Work Trips/Month:      Nonwork Trips/Month:

```

FIGURE 4 File format for the personal data set.

```

=====
DRIVE ACCESS FILE:
=====
Random Sample Households Township Number:
Sample Transit Households Township Number:      Transit Source:
Household Number:      Person Number:
=====
Closest Station:      RR Line1:      RR Station1:
Distance1:      Parking Fee1:      Train Freq1:
-----
Next Closest Station:      RR Line2:      RR Station2:
Distance2:      Parking Fee2:      Train Freq2:
Time Savings2:      Cost Savings2:
-----
Lower Cost Station:      RR Line3:      RR Station3:
Distance3:      Parking Fee3:      Train Freq3:
Time Savings3:      Cost Savings3:
-----
Bus Access Station:      RR Line4:      RR Station4:
Distance4:      Bus Fare:      Bus Freq:
Stop Distance:
-----

Walk Access Station:      RR Line5:      RR Station5:
Distance5:      Train Freq5:
-----
Kiss and Ride Station:      RR Line6:      RR Station6:
-----
Auto Passenger Station:      RR Line7:      RR Station7:
-----
Home Q-Sec:
Work Q-Sec:
=====

LINE-HAUL FILE:
=====
Auto:
Line-Haul Time1:
Destination Parking Fee:
-----
Other Transit:      Stop Distance2:
Line-Haul Time2:      Other Freq:
Line-Haul Fare2:
Walk Distance to Destination2:
-----
Commuter Rail:
Line-Haul Time3:
Line-Haul Fare3:
Bus Distance to Destination:      Added Transit Fare:
Bus Stop to Destination:
Walk Distance to Destination:
-----

GENERAL TRIP CHARACTERISTICS:
=====
Work Trip:
Trip to CBD:
=====

```

FIGURE 5 File format for the trip data set.

These files were again assembled using the program PFS:FILE. No attempt was made to encode the data directly into the microcomputer files during an interview. Interviewers recorded the results of the interview on an interview report form and the data files were then transcribed from these forms.

Figure 3 shows the format for the household data set. This data set describes the household, its composition, employment characteristics, and vehicle ownership. There are a total of 198 records (a segment of the file containing the items in the figure), one for each completed interview.

The format for the personal data set is shown in Figure 4. There is one record for each adult household member and the file describes the individuals and their work trips. It contains 421 records from completed interviews.

The last data set format in Figure 5 is the trip data set. This file includes all reported trips that conceivably could ride existing commuter rail lines in the study area. Generally, an acceptable trip for this file is long enough and directionally oriented so that a significant part of the trip can use available commuter rail services. There are 152 trips in this file, somewhat less than one per household.

The trip data set has three sections, for commuters who board at suburban stations, line-haul travel, and general trip characteristics. The first section of the file summarizes how the individual reaches a suburban commuter rail station when the

line-haul mode choice is commuter rail. The line-haul portion of the file describes the line-haul trip segment and notes how travelers reach their final destination from the point where they depart a line-haul travel mode. This point is usually a parking garage or lot, a rail transit station, bus stop, or downtown commuter rail station. Two other trip characteristics, work or nonwork and CBD destination or non-CBD destination, are noted in the last general trip characteristics section.

The model calibration requires that additional hypothetical trips be entered into the trip file. For example, the complete trip file for model calibration has to include the trip characteristics of current commuter rail trips hypothetically diverted to the automobile driver mode. The cost estimates that follow do not include the costs of estimating the characteristics of these alternative mode trips.

STAFFING AND COSTS OF THE SURVEY

Three hundred seventy-eight households were contacted, and 198 interviews were successfully completed. There were also a few partially completed interviews that produced useful data for one or more of the files.

The survey covered four bimonthly agency accounting periods in September and October 1985. The staff-person-hours presented in Table 2 are tabulated from these four accounting

TABLE 2 RESOURCES EXPENDED ON THE HOME INTERVIEW SURVEY

1. Staff Person-Hours

Work Item	Per Contacted Household	Per Completed Interview
1. Selection of Contact Households	0.24	0.45
2. Mailing of Survey Forms	0.15	0.29
3. Phone Interviews	0.28	0.53
4. Compilation of Survey Data Sets	<u>0.16</u>	<u>0.31</u>
Total	0.83	1.58

2. Approximate Manpower Cost (Average Wage Rate Plus Overhead at 120 percent)

Work Item	Per Contacted Household	Per Completed Interview
1. Selection of Contact Households	\$7.00	\$13.00
2. Mailing of Survey Forms	\$4.40	\$8.50
3. Phone Interviews	\$8.20	\$15.60
4. Compilation of Survey Data Sets	<u>\$4.70</u>	<u>\$9.10</u>
Total	\$24.30	\$46.20

periods. Staff-person-hours are further subdivided into four survey work elements identified by sections in this paper. Slightly more than 1.5 staff-person-hours were required to complete an interview. This amount covers selection of the household through the entry of the household, person, and trip data into the microcomputer files, but does not cover any analysis of the data or model calibration activities.

These staffing requirements were converted to dollar estimates using \$26,000, the average CATS annual salary at the time of the survey, and an agency overhead rate of 120 percent. These figures are presented in the bottom half of Table 2. For this survey, the cost of data collection is just under \$50 for each successfully completed household interview.

The bar chart in Figure 6 shows the staff-person-hours broken down by work element and accounting period. This chart shows how expenditures for these four work elements vary over the time required to complete the survey. Staff-person-hours for sample selection and mailing of survey forms largely precede the actual interviews, whereas the bulk of the expenditures for compiling the data into the microcomputer data sets follows the interviews.

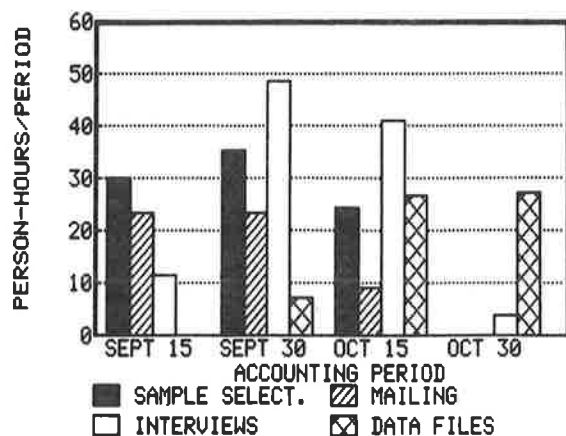


FIGURE 6 Survey staff-person-hours by accounting period.

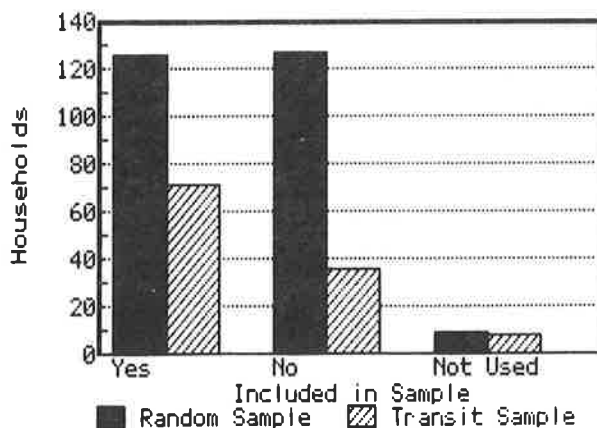


FIGURE 7 Business graphics display of survey data—survey households.

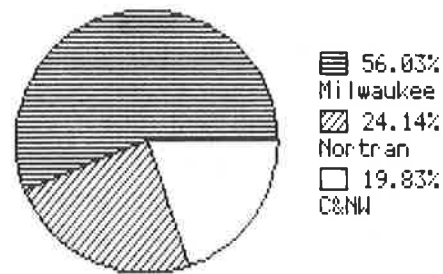


FIGURE 8 Business graphics display of survey data—survey transit households sample source.

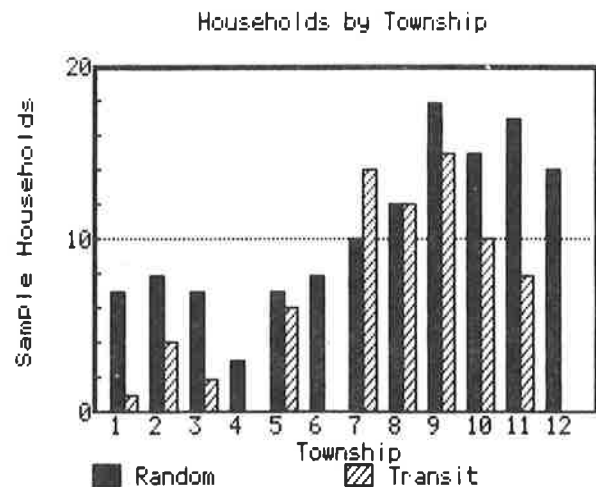


FIGURE 9 Business graphics display of survey data—households by township.

DISPLAY OF SURVEY DATA

Figures 7–9 provides several examples of the use of a business graphics microcomputer program, in this case PFS:GRAPH (6), to display the survey data. The graphs in Figures 7 and 9 summarize the sampled households in the survey. The bar graph in Figure 7 shows the number of responding and refusing households according to how the household was selected to be contacted. The pie chart in Figure 8 is the number of sample households obtained from the monthly tickets and bus ridership (the NORTAN wedge in the pie chart). Successful interviews by townships are shown in the bar chart of Figure 9.

SUMMARY AND CONCLUSIONS

This paper has demonstrated how the use of a microcomputer and general-purpose business microcomputer software can assist typical data collection in a metropolitan planning organization. The survey involved a small number of professional staff. The manager of the project directly managed the survey and performed the microcomputer applications presented in the paper. He always had access to the microcomputer because it was located in his office. The project manager had several years' experience in mainframe computing, a year or so of experience with a microcomputer, and some familiarity with the business software and its capabilities before undertaking

the project. The priority was to complete the survey, not to learn about personal computing.

The following conclusions can be drawn from this experience:

1. There is no need for specialized software or computing skills to support reasonably sized data collection efforts for many transportation planning applications, given the capabilities of general-purpose business microcomputer programs and the ease with which this software can be used.
2. The microcomputer greatly enhanced the project manager's ability to respond to situations that arose during the survey, particularly in the scheduling of households to contact as interviews were completed.
3. The microcomputer permitted a great deal of control over the survey because it reduced the number of administrative and data collection staff needed for the survey. The survey was not subject to delays due to other agency data collection and data processing activities.
4. The completion of the survey within a tight time limit

was to a large extent made possible by the availability of the microcomputer. Access to the machine, data files, and software were unrestricted and work could continue outside normal business hours when it was necessary to catch up to schedule.

REFERENCES

1. *Urban Transportation Planning Package, 1980 Census: Technical Documentation for Summary Tape*. Journey-to-Work and Migration Branch, Population Division, Bureau of the Census, U.S. Department of Commerce, Washington, D.C., 1983.
2. M. Ben-Akiva and S. R. Lerman. *Discrete Choice Analysis: Theory and Application to Predict Travel Demand*. M.I.T. Press, Cambridge, Mass., 1985.
3. *Chicago North Suburban Street Address Directory*. Reuben H. Donnelley Corp., Chicago, Ill., 1985.
4. *PFS: File: User's Manual*. Software Publishing Corp., Mountain View, Calif., 1984.
5. *PFS: Write: User's Manual*. Software Publishing Corp., Mountain View, Calif., 1984.
6. *PFS: Graph: User's Manual*. Software Publishing Corp., Mountain View, Calif. 1983.