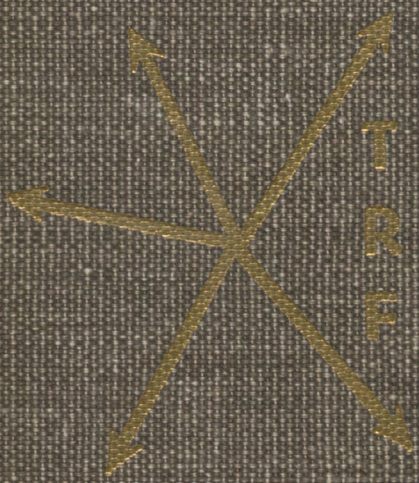
PROCEEDINGS —

Twenty-second Annual Meeting

Volume XXII • Number 1 1981



TRANSPORTATION RESEARCH FORUM

PROCEEDINGS —

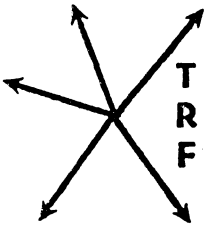
Twenty-second Annual Meeting

Theme:

"Opportunities and Challenges in the New Environment of Transportation"

November 4-5-6, 1981
Golden Gateway Holiday Inn
San Francisco, California

Volume XXII • Number 1 1981



TRANSPORTATION RESEARCH FORUM

An Assessment of Empirical Choice Set Formation

by Katherine K. O'Neil and David O. Nelson**

**Charles River Associates Incorporated, Boston, Massachusetts.*

INTRODUCTION

THIS PAPER presents some recent findings on the topic of choice sets formation. It describes the philosophy of empirically derived choice sets, the Baltimore Disaggregate Data Set (BDDS) approach to choice set formation, findings from the BDDS, and recommendations for further research and improvement in empirical choice set data collection.

The formation of choice sets of individual travel behavior is of particular interest to travel demand modelers. Some researchers indicate that the inclusion or exclusion of alternatives from the disaggregate demand model estimation choice set can have significant impacts on the model results and resulting policy recommendations. Most disaggregate demand models estimated to date have relied upon the analyst to assign choice sets to individuals and develop level of service (LOS) values for alternatives not chosen. The BDDS included a novel experiment in developing choice sets for disaggregate demand modeling. In the BDDS detailed trip report, respondents were queried about alternative modes and destinations for a selected round trip. LOS data were gathered and reported for alternatives actually chosen during the previous six months. We have termed this approach "empirical choice set formation."

The reporting process was designed to generate empirical choice sets for estimation of disaggregate demand models of mode and destination choice. The paper assesses the reporting process and presents CRA's expedience using empirical choice sets drawn from the BDDS for modeling work trip mode choice. Finally, recommendations for improving choice set data collection efforts are discussed.

THE ROLE OF CHOICE SETS IN DISAGGREGATE DEMAND MODELING

In modeling choice behavior, an immediate problem arises in defining the sets of alternatives to be considered by different individuals or groups of people. The choice set in disaggregate demand modeling refers to the range of alternatives facing an individual. In this paper we discuss choice set formation for work trip mode choice models. Why are choice sets important? Meyer has found that model parameters which describe the relationship between predicted utilities and observed choices may be driven as much by variations in choice sets among individuals (which are not fully accounted for in the model) as by variations in preferences (which are accounted for) (Meyer, 1980). The definition of the choice set facing an individual thus may have a significant impact on the stability and significance of model parameters.

Choice sets can vary among individuals for a number of reasons. Ansah (1974) notes that the alternatives faced by an individual depend on the location of the individual relative to the total set of alternatives at a given time of day, his socioeconomic class, the individual's attitudes and beliefs, and his familiarity with each alternative. For example, if the individual is located at a fringe suburban area with no transit, his choice set will only include various automobile alternatives for most trip purposes; or an individual may be unfamiliar with the availability of transit or unaware of the opportunities for ridesharing, in which case these alternatives would not be a part of his choice process. Deterministic choice sets are seldom sensitive to the variability in choice arrays actually facing an individual.

In some probabilistic choice models based on utility maximization, each individual confronted by a choice is considered to have the same choice set available. It is assumed that all individuals in a given market segment (controlling for drivers licenses, automobile availability, etc.) have the same choice set. The underlying theory is that of a perfectly discriminating rational man endowed with complete information. This may be an unacceptable starting point for the analysis of travel behavior (Williams and Ortuzar, 1979). In reality, individuals have limited information concerning the set of alternatives from which they are to choose. Travel demand modeling should evolve to account for the reality of imperfect information and irrational behavior.

Lerman (1975) notes that "The underlying theory of the logit model requires that the choice set for each observation consist of feasible alternatives. This implies that in order to properly estimate a mode choice model, one must know which of the set of possible alternatives are actually available to the individual." Along these lines, at least one researcher has suggested the possibility of a two-stage modeling process wherein first, the analyst models the alternatives in the individual's choice set and second, models the choice among those alternatives (Meyer, 1980). This paper is not so ambitious as to model choice formation, but does not report on some data that may be useful in preparing such a two-stage model. This paper describes an attempt to define empirically the alternatives available to the individual through the use of a technique where the respondent was requested to describe all alternatives to a given trip that he actually used during the previous six months. The alternatives actually used define the feasible alternatives deemed to be available to the individual.

The theory behind the use of empirically derived choice sets is to capture only those alternatives which the individual considered feasible, for which his probability of use was greater than zero. If the individual did not use additional alternatives, he probably did not consider them to be feasible and therefore they were not in his choice set.

WHAT HAVE OTHER RESEARCHERS DONE?

While the transportation research community has developed analytic tools to model the behavior of individuals faced with a set of alternatives, we have few tools for identifying what the alternatives are in the first place. Many model estimation efforts to develop disaggregate mode choice models have assumed that each individual in the estimation data set can choose from a full array of modal alternatives. LOS values for each alternative often have been determined using engineering estimates (CRA, 1976).

While many researchers have assumed that each individual faces the full range of alternatives, Lerman (1975) has taken a different tack. His procedure has been to eliminate infeasible alternatives from the individual's choice set. For example, workers living in fringe areas without transit service are not allowed to have transit as a choice option. This process is termed screening the alternative set, and is an improvement over assuming a full range of alternatives. Lerman indicates that failure to screen the choice set will result in estimates that are biased and inconsistent, and will therefore produce unreliable forecasts of future conditions. He has also found that screening out relevant alternatives (nonzero probability) does not result in inconsistent coefficient estimates.[1] Thus, if an error is to be made in screening, it is best to eliminate relevant options from the choice set rather than include irrelevant ones.

This paper will report on an alternative approach to determining the modal choice sets of respondents using their past travel behavior as representative of the individual's perceived choice set.

EMPIRICAL CHOICE SET FORMATION WITH THE BDDS

The Baltimore Disaggregate Data Set (BDDS) collected for the Federal Highway Administration by Charles River Associates provides interesting data on individual choice set formation. It included a novel experiment in developing choice sets for disaggregate demand modeling. Respondents were queried about mode and destination alternatives for a randomly selected round trip. LOS data were gathered and reported for alternatives actually chosen during the previous six months. We have termed this approach "empirical choice set formation."

This paper describes findings on the empirical approach. In general, this approach does not appear to be an efficient way to collect data on individual travel choice sets. The reasons why it is not efficient are illuminating. First, there is comparatively little variability in travel behavior. Individuals generally travel by the same mode to the same places for a given trip purpose. This is especially true for work trips. Second, when travel behavior does vary, it often varies in nonstandard ways (e.g., the alternative to driving is not simply transit but a combination of shared ride and transit). Our discussion of this process starts with a brief description of the BDDS alternatives identification process.

BDDS ALTERNATIVE GENERATION PROCESS

The BDDS was a home survey of approximately 1,000 households in and around Baltimore, Maryland. A one-day trip diary was collected for every household member. In each interview, one household member was randomly selected to be the "Primary Respondent." One trip from the primary respondent's diary was randomly selected for detailed reporting. From this detailed report the interviewer asked questions to generate

alternatives to the detailed trip. This process for choice set formation is described below.

DETAILED TRIP REPORT

The Detailed Trip Report (DTR) investigated a single randomly selected round trip in great detail. First, the interviewer asked about all other one-way trips associated with this trip necessary to chronicle a complete round trip" from home or work/school. Complete round trips were always recorded in the DTR.

The unit of analysis for the DTR was the link. A **link** was considered any part of a trip wherein one mode was used. When mode changed, one link ended and another began. Every one-way trip contained at least one link and could be composed of many more. For instance, the walk to a bus stop from home would be one link, the ride on the bus another link, the walk afterward, a third link, etc. Because the DTR trip was required both to leave and to return to home or work/school, the reported round trip could have been a chain of trips or a complex tour. A **trip chain** occurred whenever one or more purposes were accomplished by stopping at several locations enroute before returning to the home or work origin.

In practice, the trip selection process yielded three basic types of trips for DTR collection:

- The **home-based work-/school-trips** including all links on the round trip between home and work/school, including stops for shopping and other purposes made enroute between home and work/school;
- The **work-based trips** including all links of a round trip for any purpose and starting and ending at work/school; and
- The **home-based other trips** including all links of a round trip for any nonwork/-school purpose.

LOS data important for mode choice modeling was gathered for each link.

IDENTIFICATION OF ALTERNATIVES

The alternative identification process consisted of a hierarchy of questions about the detailed trip designed to identify alternative modes for all links, alternative trip configurations for trip chains, and alternative destinations for nonwork/school trips. All alternatives that the primary respondent had actually used in the last six months were selected for detailed reporting using the Alternative Trip Report (ATR).

The interrogatory sequence for identifying alternatives for work-/school-trips is shown in Figure 1. If the DTR trip were a simple home-based work/school trip, the interviewer inquired about alternative modes used to make this trip. He recorded all identified alternatives and marked those used during the previous six months (V-5b). If the DTR were part of a chain, the line of questioning was more complex. The interviewer first sought to identify alternatives that maintained the same trip sequence but used alternative modes (V-3c). For discretionary trips in the chain, he asked about the possibility of serving these trip purposes with a simple home-based trip (V-3e). If the chain included more than one work/school location, the interviewer asked about alternative chains using different modes between the two work/school locations (V-3g). He also explored the possibility of traveling to the second work/school location directly from home (V-3i). Each of these different alternatives used during the last six months was marked for the ATR. The interviewer also identified alternative destinations for discretionary links (V-3f) and explored the possibility of transit alternatives for respondents with auto-dominated travel patterns (V-7), but these alternatives could not be selected for the ATR.

A similar procedure was followed for nonwork/school trips. (For more information, see CRA [1980]).

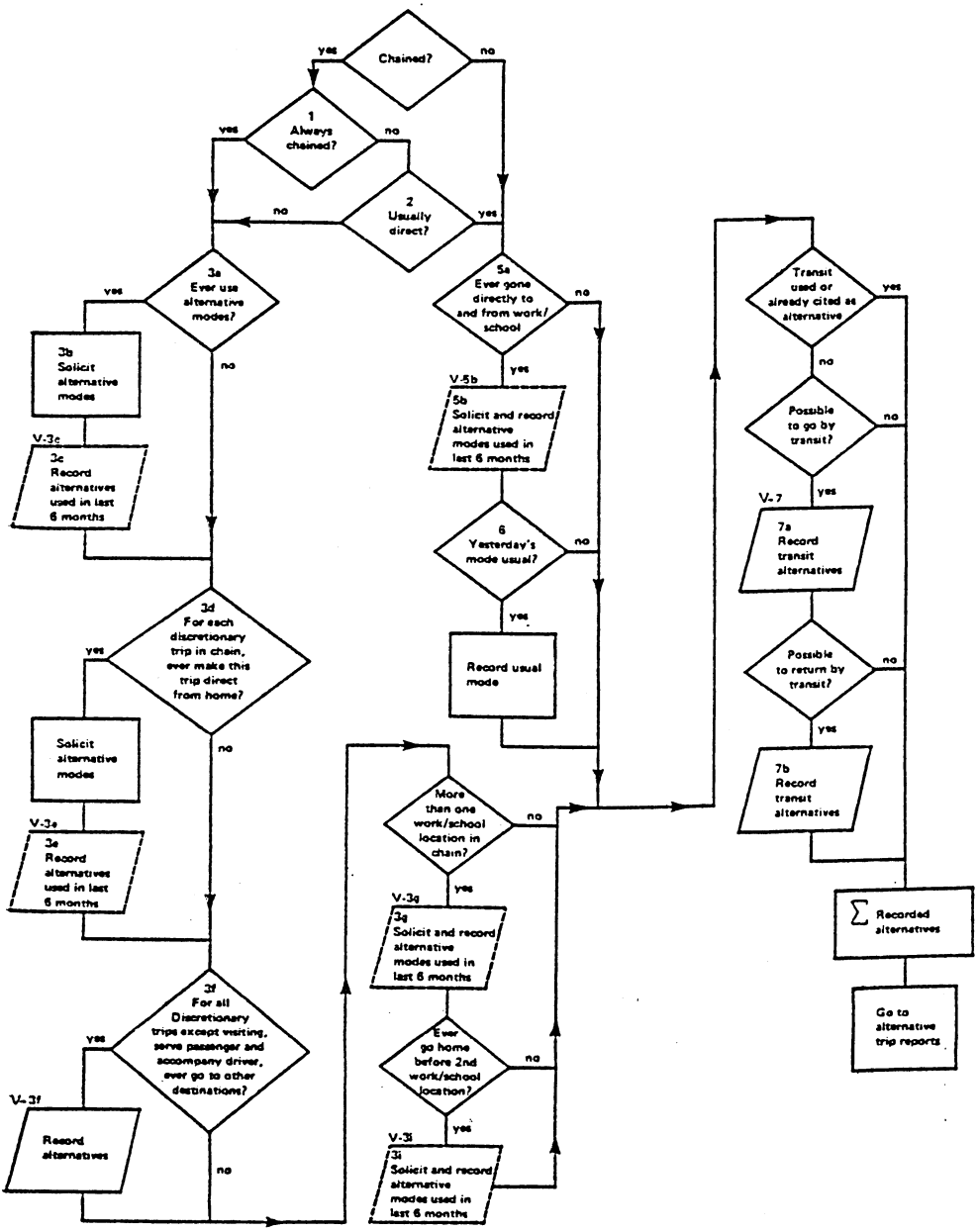
ALTERNATIVE TRIP REPORTS

A separate ATR form was completed for each alternative trip identified during the alternative identification process. The ATR is virtually identical to the DTR. Complete round trips are reported. The unit of analysis for the ATR is the link. Thus, the BDDS detailed and alternative trip reporting process was designed to generate empirically based choice sets for estimation of disaggregate demand models of mode and destination choice. However, the results of this process, as shown below, were less than encouraging.

BDDS CHOICE SET RESULTS

The alternatives generation process, described above, was designed to yield data on mode and destination choices for individuals as determined by their travel patterns for randomly selected trips over the previous six months. However, for a number of reasons this process fell short of the objective of providing a rich set of modeling data on travel

IDENTIFICATION OF ALTERNATIVES FOR HOME-BASED WORK-/SCHOOL-TRIPS (BDDS)



SOURCE: Charles River Associates Incorporated, November 1980.

FIGURE 1

choices. These reasons are described below.

Many problems can be traced to the complexity of urban life and the ways in which travel behavior reflects competing demands on the traveler. First, commuters tend always to make their work trips by the same mode. Second, when they exhibit variability in their mode choices, travelers often use multiple modes, such as shared ride for one link and transit for another. Third, many travel alternatives to simple trips serving one trip purpose are chained trips serving multiple purposes. Fourth, the random trip selection process gathered some data on many types of trips, but seldom gathered enough data on any particular trip purpose to facilitate the development of statistical models. Finally, attrition from the data set, due to households either not traveling on the survey day or not reporting alternatives to the selected trip, was severe. The following sections present results from the BDDS which illustrate these problems.

In many cases the primary respondent did not travel on the survey day. Of the 966 households interviewed, 135 took no trips on the travel day. Of the remaining 831 households, CRA only found ATRs for 389 households. Table 1 shows the contents of the detailed link file data on alternative trips. More than one alternative could be recorded for a single respondent, explaining how 389 households could have 779 one-way trips.

Despite an allowance for nine classes of alternatives, most alternatives reported were of three types. As can be seen from Table 1, most of the alternative trips were alternative modes for simple work/school and nonwork trips (or simple alternatives to chains), and alternative destinations for discretionary trips. There are insufficient cases in the data set to model the other types of alternatives.

Although the study design allowed for a wide variety of modes, only three were represented with sufficient frequency to allow modeling. Table 2 shows the mode for each of the links in the detailed link file. Most of the links were by auto, bus, or walking. There are not sufficient cases with the other modes chosen or as alternatives to enable them to be included in the choice set for mode choice models.

The procedures for selecting the primary respondent and the detailed trip and for identifying the alternatives have possibly adverse ramifications for mix of trips and alternatives found in the data set. Since a primary respondent was randomly selected first, and only then was a trip randomly selected from his trip summary, trips by persons who made few trips on the travel day were more likely to be selected for detailed reporting than were trips by persons who made many trips.

This can be made clearer by a simple example. Suppose the household contains

TABLE 1

NUMBER OF ALTERNATIVE TRIPS REPORTED IN THE BDDS BY TYPE AND NUMBER OF LINKS

Alternative Types	Number of Links 2	3	4	5	6	7	8	Total
I. Alternative mode for work chain	--	--	--	--	--	--	--	--
II. Alternative mode for work chain -- some links removed	6	--	--	--	--	--	--	6
V. Simple worktrip -- alternative mode*	85	7	19	1	5	--	--	117
VI. Alternative modes for nonwork chain	2	9	2	--	1	--	--	14
VII. Alternative Order of Links in nonwork chain	3	2	--	--	--	--	--	5
VIII. Alternative modes simple nonwork	195	5	12	--	4	1	--	217
IX. Alternative destinations	382	12	15	1	7	--	1	418
								779

*This also may be a simple alternative to a complex tour.

Source: Charles River Associates, 1980.

TABLE 2

MODE FOR LINKS IN BDDS DETAILED LINK FILE

	Number	Percent
Auto driver	1,736	40.8
Auto passenger	859	20.2
Walk	997	23.4
Bus	505	11.9
Taxi	71	1.7
Bike	49	1.2
School bus	19	0.4
Motorcycle	19	0.4
Boat	2	0.0
TOTAL	4,257	99.9*

*Rounding error.

Source: Charles River Associates, 1980.

only two members, John and Mary. John only went to market on the travel day whereas Mary went to work and to her welding class. John and Mary have equal probabilities of being selected as primary respondents (0.50). If John is selected as the primary respondent, there is a 100 percent chance his market trip will be selected for detailed reporting. In Mary's case each of her trips has only a 0.50 chance of selection. Consequently, the marginal probability of her trip to welding class being selected for reporting is 0.25, whereas John's market trip has a 0.50 chance of selection at the outset of the household interview. The full ramifications of this selection bias for trips by less frequent travelers in the household has not been explored. We believe it may reduce the representativeness of the sample but does not preclude the estimation of consistent disaggregate demand model parameters.

A comparable selection bias arises from the trip selection process once the primary respondent was chosen. Since an individual one-way trip for any purpose was selected as the basis for compiling a "complete round trip" for detailed reporting, the probability that chains or tours would be selected for detailed reporting was enhanced by a factor equal to the number of one-way trips in the chains.

Related to the problem of overrepresentation of chained trips is the exclusion of chained alternatives to simple trips, resulting from a simplification in the alternative identification process. The alternative identification process broke complex chains into simple alternatives but did not identify complex alternatives to simple trips. Consequently, the data set contains a certain number of "apples and oranges" cases, where the trip taken served two or more purposes but the trip alternative served only one. Moreover, because no chained alternatives to simple trips were gathered, it is not possible to model the decision to make chained trips.

In order to estimate models of single-purpose round trips, CRA collapsed the links in the detailed and alternative trips into single data vectors representing the entire round trip. The mode and purpose variables were collapsed into dummy variables to preserve information to select specific alternative types. This raised a new problem (multiple modes) with the occasional cases where the detailed trip or alternative used a relatively unorthodox mix of modes, such as auto passenger to work and bus back home. Since the state of the art offers little guidance to the modeler in these cases, CRA also eliminated these alternatives from analysis. Similarly, CRA also eliminated alternatives using infrequently used modes such as boat, bicycle, and taxi, since insufficient cases were available to create a data base with these alternatives.

For a model of work mode choice with three alternatives, auto, walk, and transit, CRA edited the contents of the detailed link file to develop a data set containing worktrip DTRs and one or more ATRs that met the simple criteria required by the simplified assumptions of traditional travel choice models. These screening criteria included:

- No mode switching within round trips (multiple modes);
- No multiple purpose trips (chains);
- No round trips that do not end at the first origin; and
- One or more alternatives to the selected round trip.

With respect to modal alternatives, only the chosen trip and one alternative to that trip were sufficient for inclusion in the final sample. The final data set that met these criteria contained only 30 usable observations from an original data set of 966 interviews. Clearly, this particular empirical approach to choice set formation is not data-efficient. (In order to estimate a work trip mode choice model from the BDDS detailed trip file, we were forced to include observations with no alternative trip records and generate choice set data for these records.)

CONCLUSIONS

Two competing general conclusions can be drawn from this experiment. The first conclusion would be that clearly the alternatives generation process used in collecting the BDDS was flawed and contributed to the attrition of travel data. The second, more radical conclusion would be that the simplifying assumptions made in most disaggregate demand models are so abstracted from the reality of individual travel patterns that it is difficult to find individuals in the real world whose true choice process conforms to that supposed by modelers. While it is tempting to accept the latter possibility, the explanation of the problems may be found in the unworkability of using the interview process to generate choice sets in the manner employed in the BDDS.

The Baltimore Disaggregate Data Set was a bold experiment in transportation planning data collection. In the context of an experiment it was a considerable success. Many valuable lessons have been and will be learned from the data source. However, such an ambitious approach to data gathering is not recommended for the calibration or adjustment of planners' applied choice models. It is simply too complex. Rather, for applied purposes the data should be collected with a short and simple questionnaire. Accurate systems data should be used for LOS measures.

The empirical choice set formation process seems to have been too complicated for respondents. However, it may have yielded more useful data if it had been structured differently. Specific recommendations for improvements in empirical choice set formation procedures can be developed based on the BDDS experience. It may be helpful to the interviewer, respondent and analyst to simplify the alternatives generation process by eliminating some questions which yield few positive responses. The length and complexity of the alternatives generation questions sequence may have substantially added to confusion and misreporting without a commensurate increase in useful data obtained. By trying to sample too many purposes, circumstances, and trip configurations the sample was spread too thin. When sampling over a variety of market segments, the probability is high that only a limited number of respondents in any single segment will be in a position to trade between modes and/or destinations. Rather, the empirical choice formation process should focus on one or two purposes of particular interest to the researcher. It will be recalled from Table 1 that over 95 percent of all alternative trips fell into three of nine possible trip purpose categories. Also, it could help to simplify the empirical choice set formation process for mode choice analysis purposes by restricting the search through modal alternatives to a more limited set of modes. Over 95 percent of links in the data set were by auto, bus, or walk.

Sample attrition from the alternative trip generation process is a salient BDDS problem for disaggregate demand modeling using empirical choice set data. DTRs were obtained for trips randomly selected from four trip categories. Dividing the sample in this way obviously limits the number of trips available for modeling any one purpose. In addition, 389 of 966 households reported an alternative trip. This is disappointing. The questionnaire systematically probes for trip alternatives, but ATRs were only completed if the primary respondent reported using an alternative mode or destination in the previous six months. This limit of six months undoubtedly contributed to sample attrition.[2] Perhaps any alternative that the respondent deems feasible should be reported. Some researchers have suggested all alternative modes should be reported, regardless of whether they have been used or are deemed feasible by the respondent. (This may be too extreme, and could yield biased and inconsistent coefficient estimates.)

Some other problems with the BDDS empirical choice set data could perhaps be corrected with an improved approach to trip sampling. Empirical choice set formation procedures for trip selection which do not lead to the overrepresentation of trips by low mobility individuals should be devised. This selection bias may have contributed substantially to the lack of variability in the data since it may be that lower mobility individuals have fewer mode choice alternatives and exercise fewer choices in their travel behavior. Similarly, the trip selection and alternatives generation process should be structured so that complex tours are not overrepresented in the data set. In addition, it may be valuable to solicit complex alternatives to simple trips so that analysts can investigate the decision to make complex tours versus simple trips. Finally, experience with the BDDS suggests that complex trips serving multiple purposes and/or using multiple modes are quite common. This finding may imply that many of the simplifying assumptions used to formulate disaggregate behavior models are unrealistic and that a more

comprehensive approach to understanding traveller's choice may be required.

FOOTNOTES

1 This would be expected given the independence of irrelevant alternatives property of logit techniques and other share models.

2 Whether using a long time period would on balance have produced better alternative data is a design question. Respondents resent being pressed for details that are hard to recall or strike them as having a hypothetical nature.

REFERENCES

J. A. Ansah. 1974. "Destination Choice Modelling and the Disaggregate Analysis of Urban Travel Behavior." Final Report, Purdue University, Indiana, August 28.

J. A. Ansah. 1977. "Destination Choice Set Definition in Travel Behavior Modelling." *Transportation Research* II, 2 (April): 127-140.

Charles River Associates. 1976. "Disaggregate Travel Demand Models, Project 8-13: Phase I Report." Prepared for NCHRP National Research Council. Cambridge, Mass.: CRA, February.

Charles River Associates. 1980. *Disaggregate Travel Demand Models.* Draft Final Report prepared for National Cooperative Highway Research Program, Transportation Research Board, National Research Council, November.

Steven R. Lerman. 1975. "A Disaggregate Behavioral Model of Urban Mobility Decisions." Unpublished doctoral dissertation, MIT, Department of Civil Engineering, June.

Robert Meyer. 1979. "Utility, Uncertainty, and Spatial Adaptation: A Behavioral Theory of Destination Choice Set Formation." August. Prepared for presentation at Transportation Research Board, 1980.

H. C. W. L. Williams and J. D. Ortuzar. 1979. "Behavioral Travel Theories, Model Specification, and the Response Error Problem." Working Paper 116, University of Leeds, School of Geography, July.

ACKNOWLEDGEMENTS

This paper presents some findings from a study of disaggregate travel demand models conducted by Charles River Associates Incorporated, for the National Cooperative Highway Research Program (NCHRP Project 8-13). At CRA the project was directed by Dr. William Tye. The NCHRP project representative was R. Ian Kingham. The authors would like to acknowledge the invaluable technical contributions of Alexander Bardow, and the review and comments of Dr. Timothy Tardiff.

This study was conducted under National Cooperative Highway Research Program Project 8-13. The opinions and findings expressed or implied in this paper are those of the authors. They are not necessarily those of the Transportation Research Board, the National Academy of Sciences, the Federal Highway Administration, the American Association of State Highway and Transportation Officials, nor of the individual states participating in the National Cooperative Highway Research Program.