TECHNICAL REPORT STANDARD PAGE

1. Report No. 337	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle **Cost-Effective Data Collection in Louisiana**	5. Report Date: September 2002 6. Performing Organization Code	
7. Author(s): Chester G. Wilmot Peter R. Stopher	8. Performing Organization Report No. 337	
9. Performing Organization Name and Address: Department of Civil and Environmental Engg. Louisiana State University Baton Rouge, LA 70803-6405	10. Work Unit No. 11. Contract or Grant No.: LTRC No. 98-4SS	
12. Sponsoring Agency Name and Address: Louisiana Transportation Research Center 4101 Gourrier Avenue Baton Rouge, LA 70808	13. Type of Report and Period Covered: Final Report, 06/01/97 - 12/31/99 14. Sponsoring Agency Code	
15. Supplementary Notes		

16. Abstract

The purpose of this research was to identify cost-effective methods to accumulate data for metropolitan transportation planning in Louisiana. The research was directed at making maximum use of existing data sources, investigating the transferability of data to metropolitan areas in Louisiana, considering the use of small local samples to update transferred or outdated local data, and investigating the potential of Global Positioning System (GPS) equipment to collect travel-related information.

Data considered for transfer are aggregate relationships, distributions, and proportions of the data from one or more external, primary, or secondary data sources. Transferability to the Baton Rouge metropolitan area was tested by comparing aggregate measures of data from the National Personal Transportation Survey (NPTS) of 1995 for metropolitan areas, with populations between ½-1 million, and from the North Central Texas Council of Governments (NCTCOG) survey of 1996, with values obtained from the Baton Rouge Personal Transportation Survey of 1997. Some aggregate measures were found to be transferable to Baton Rouge while others were not. The possibility of using current, local data to update transfer data was investigated by collecting data from a sample of 108 households in Baton Rouge in 1998 and using it to update aggregate measures from NPTS 95 and NCTCOG. While the sample was too small to provide consistently improved transferability to the transferred data, simulated results with a sample of 450 households suggested that data updating can provide aggregate measures of transferred data that are comparable with locally-collected data.

Three research thrusts were conducted to investigate the potential of improving the efficiency for collecting local data. First, stratified sampling of households to improve the homogeneity of data in strata was investigated. Second, a new form of time-use diary using a day-planner type format was developed and tested. Third, Global Positioning System (GPS) devices were used to record auto travel. All three tactics appeared to improve the efficiency of collecting local data.

17. Key Words Data, Transferability, Planning, Modeling, GPS, Panel, Survey.		18. Distribution Statement Unrestricted. This document is available through NTIS, Springfield, VA 21161.	
19. Security Classif. (of this report) Not applicable	20. Security Classif. (of this page) Not applicable	21. No. of Pages 180	22. Price

COST-EFFECTIVE DATA COLLECTION IN LOUISIANA

by

Chester G. Wilmot, Ph.D.

Louisiana Transportation Research Center and
Department of Civil and Environmental Engineering
Louisiana State University
Baton Rouge, Louisiana 70803

and

Peter R. Stopher, Ph.D.

Department of Civil and Environmental Engineering
Louisiana State University
Baton Rouge, Louisiana 70803

LTRC PROJECT NO. 98-4SS
STATE PROJECT NO. 736-99-0504

conducted for

Louisiana Department of Transportation and Development
Louisiana Transportation Research Center

September 2002

ABSTRACT

The purpose of this research was to identify cost-effective methods to accumulate data for metropolitan transportation planning in Louisiana. The research was directed at making maximum use of existing data sources, investigating the transferability of data to metropolitan areas in Louisiana, considering the use of small local samples to update transferred or outdated local data, and investigating the potential of Global Positioning System (GPS) equipment to collect travel-related information.

Data considered for transfer are aggregate relationships, distributions, and proportions of the data from one or more external, primary, or secondary data sources. Transferability to the Baton Rouge metropolitan area was tested by comparing aggregate measures of data from the National Personal Transportation Survey (NPTS) of 1995 for metropolitan areas with populations between ½ to million and from the North Central Texas Council of Governments (NCTCOG) survey of 1996 with values obtained from the Baton Rouge Personal Transportation Survey of 1997. Some aggregate measures were found to be transferable to Baton Rouge while others were not. The possibility of using current, local data to update transfer data was investigated by collecting data from a sample of 108 households in Baton Rouge in 1998 and using it to update aggregate measures from NPTS 95 and NCTCOG. While the sample was too small to provide consistently improved transferability for the transferred data, simulated results with a sample of 450 households suggested that data updating can provide aggregate measures of transferred data that are comparable with locally-collected data.

Three research efforts were conducted to investigate the potential of improving the efficiency of collecting local data. First, stratified sampling of households to improve the homogeneity of data in strata was investigated. Second, a new form of time-use diary using a day-planner type format was developed and tested. Third GPS devices were used to record auto travel. All three tactics appeared to improve the efficiency of collecting local data.

ACKNOWLEDGMENTS

The research reported in this document was funded by the Louisiana Transportation Research Center. The study was conducted under the supervision of the following Project Review Committee:

Harold R. Paul, LTRC
Arthur D. Rogers, LTRC
James R. Joffrion, DOTD
Daniel J. Broussard, DOTD
Eric I. Kalivoda, Ph.D., DOTD
Virgil W. Page, FHWA
Babak N. Naghavi, Ph.D., LTRC
Huey P. Dugas

The assistance of the Project Review Committee in the execution of the study is gratefully acknowledged.

Surveys conducted as part of this study were executed by the Louisiana Population Data Center of Louisiana State University. Trisha Dudula and Shivaprasad Shivananjappa assisted in distributing questionnaires and processing the returns. Their contribution is acknowledged with appreciation. Prashant Bachu and Kalyan Koppineedi assisted in the execution of the GPS survey and in the development of programs to evaluate and process GPS track file data for inclusion in a Geographic Information System. Their contribution to this portion of the study is acknowledged with gratitude.

IMPLEMENTATION STATEMENT

The results of this research are intended as input to transportation planning in the metropolitan areas of Louisiana. Methods are described that transportation planners may employ to make greater use of existing data, collect new data more efficiently, and combine existing and new data most effectively. The material provides suggestions to accumulate transportation data but does not prescribe a specific procedure; planners in each metropolitan area must decide on appropriate procedures for their area. The viewpoint adopted by the authors is that maintaining a flexible application environment encourages innovation and development.

The suggestions included in this report are not restricted to Louisiana or to metropolitan areas. The procedures have general applications. However, the findings are tentative and further development and testing are needed to confirm the usefulness of the proposed procedures.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Background

Data are essential to transportation planning; the collection of data is, therefore, a significant activity of the transportation planner. It is also an expensive activity, making careful design and planning of survey instruments and procedures essential to reducing costs and increasing the effectiveness of the data to be collected. To understand what surveys and sampling requirements might be needed by the transportation planner, it is necessary to first review the nature of the data needs for transportation planning. At the outset, it is useful to distinguish between a census and a sample. A census involves measurement or enumeration of every member of a subject population. (The word population is used here to define the universe of the units of interest to the study, which may be people, vehicles, buildings, etc.). A sample involves choosing a subset of the population from the universe. This sample may be small or large, depending on many factors. However, the intent is to always draw a sample from the population that is considered to be representative of the entire population no matter how small or large the sample.

Data are needed for three main purposes: description of the present situation, input to development and use of transportation planning models, and monitoring the effects of the implementation of policies, strategies, and investments. Principal among the data needs for transportation are data on the supply of each of the transportation systems and the use being made of it, i.e., the demand for transportation. In addition, there may be a need for qualitative information relating to perceptions of the transport system, preferences and opinions about the system, etc. *[1]*.

The first purpose, description, entails the need to know how the transportation system is currently being used, where problems arise, and the severity, location, duration, and timing of problems. There are many sources of anecdotal information that can suggest where the system is failing. Transportation within the urban area is an issue discussed in everyday conversation, and most people consider themselves experts on at least some part of the system. As a result, transportation experts will frequently be recipients of much gratuitous information on transport problems. However, it is very difficult to have a good idea of just how the system operates on a daily basis. In part, this is because the road system of most cities is a very large and quite a complex phenomenon. A large city may have many thousands of kilometers of roads. On the other hand, it is because many failures in the system are non-recurring failures such as an accident in a particular location or a temporary breakdown in the functioning of a traffic signal. Generally, it is important to separate the nonrecurring incidents that cause a breakdown in the functioning of part of the system from the recurring problems that indicate some permanent deficiency in the system.

The transportation system is, thus, very difficult to measure or assess. Most urban areas maintain a traffic-counting program that provides knowledge of the traffic volumes at selected points throughout the area. However, these counts are usually small in number compared to the length of the roadway system. Measurement of the hour-by-hour performance of the entire system would involve an enormous study at extremely high cost, which has probably never been attempted in any urban area in the world. In place of the complete measurement of the system, a sample of measurements were taken. This information was then used to provide a description of the entire urban system and its functioning. Thus, the first requirement from the data is to provide a basis for describing how the system currently functions.

The second purpose of data is as input to the development of forecasting models either without any changes being made to the system, or in response to various possible changes that could be made. The various different computer models have different data needs although the broad categories of data change relatively little from model to model.

The third purpose of data is to monitor the results of certain actions taken with respect to the transportation system. For example, a roadway widening may be undertaken as a result of measurements that have determined a shortfall in the capacity of the corridor under study. After completing the widening project, it is appropriate to determine if the project has had a beneficial effect on travel through the corridor or has changed travel patterns in some way. This requires the measurement of the performance of the widened facility, other parallel facilities, and facilities that feed traffic to the widened roadway. Also, there may be a desire to determine how the traveling public perceive the widened roadway in terms of satisfaction with the result or opinions about the performance of the corridor since the widening.

Because qualitative data needs are usually defined specifically to each study in which they are collected, it is not possible to provide a list of typical data items required in this category. However, there are fairly standard categories of quantitative data required. These can be subdivided into the categories of supply and demand data as follows:

1. Supply Data
    - Capacity (possibly a function of the number of lanes of roadway, number of public transport vehicles, etc.)
    - Design speed
    - Type of service provided (arterial roadway versus collector/distributor versus freeway versus local road; express bus or train service, local service, skip-stop service, etc.)
    - Use restrictions (e.g., turn prohibitions, parking permitted or prohibited, operation only in the peak, etc.)

2. Demand Data
    - Volumes of use by time of day, means of travel, and specific location
    - Current actual speed, both peak and off-peak
    - Costs and times experienced by users by time of day or by origin-destination locations
    - Attributes of users that relate to levels of use and methods of use, e.g.,
        - Income
        - Age
        - Car ownership
        - Driver's license status
        - Household size
        - Working status, etc.

These various data needs cannot all be collected by a single survey procedure and instrument. Rather, transportation data collection normally involves a number of different data-collection activities each using different survey methods, instruments, and sampling procedures. However, this project focuses on the collection of demand data, particularly relating to the specific trips undertaken during a prescribed period of time and the attributes of those making the trips. This is often referred to as Household or Personal Travel Surveys.

The Personal Travel Survey involves collecting data from a small sample of households (approximately 3,000). The data collected are descriptors of the household itself (such as the type of dwelling unit, the number of people in the household, the number of vehicles available, the income of the household, etc.), descriptors of each person living in the household (e.g., educational level, worker or student status, age, gender, relationship to the respondent, etc.), and details of traveling done by household members for a 24-hour weekday period. On the average, in the late 1990s, the cost of such a survey was estimated to be around $130 for each completed household. Thus, a survey of 3,000 households will cost close to $400,000.

As a result of this magnitude of cost, most metropolitan areas that conduct surveys do so only once every ten or twenty years. When these areas decide it is time to collect data, it is usually a major effort to put together the money required. It is often assembled from a variety of jurisdictions within the metropolitan area and usually exhausts most planning funds for the year or two during which the survey is to be conducted. On the other hand, a number of metropolitan areas do not or rarely collect data of this type at all.

In Louisiana, New Orleans and Baton Rouge each conducted a personal travel survey in the early 1960s. Since that time, no new data were collected in Baton Rouge until 1997 when an add-on to the Nationwide Personal Transportation Survey (NPTS) was purchased to obtain data on 1,395 households in the metropolitan area. New Orleans has not completed a new survey since 1965, although one began in late 1999, which will continue into the early part of 2000. Slidell conducted a survey in late 1997, using postcard solicitation and a mail-out, mail-back format. Approximately 800 households provided responses. The other metropolitan areas in the state have never collected any data of this type. As a result, these metropolitan areas have had to depend largely on borrowing models from elsewhere in the country and using limited local data such as traffic counts to calibrate (i.e., adjust) and validate the models for the local situation. New Orleans has continued to use models that were estimated on the 1965 data and has updated those models from time to time using other available local data. Baton Rouge no longer has access to the 1960s data and has relied on borrowed models. Slidell has not used the data from 1997 for any modeling purposes to date.

Many metropolitan areas have used models borrowed from elsewhere. There has been a substantial willingness when local data either do not exist or are found to be deficient, to borrow one or more of the models for the four-step, travel-demand process from another location. Sometimes the effort is undertaken to update the models to the new area in the borrowing process, but there are a number of cases where this is not done. While borrowing models is done almost routinely now by a number of MPOs and other agencies, borrowing data is virtually unknown. Most MPO staffs are of the opinion that the travel in any given metropolitan area is sufficiently unique, making borrowed data have little or no value.

However, there is no scientific basis for this position. At the same time, there is no information available to determine if the results of borrowing data and updating them to the local area would produce a different set of models compared to borrowing models fitted in a specific area and updating them to the local one. A review of the literature uncovers no treätment of the topic of borrowing data.

The main focus of this research is to develop methods or procedures for travel surveys that would be more cost effective than current methods in order to make it more feasible for smaller metropolitan areas to collect data and to determine to what extent the data can be borrowed from other localities. Several different efforts have been pursued in this research to seek methods that may satisfy this overall objective. A further purpose of this research project is to explore the potentials for using one or more panels in Louisiana as a mechanism to provide updating capabilities for any existing cross-sectional surveys, to provide measurements of the dynamics of travel behavior changes, and to provide a mechanism to uncover likely responses to various current policies and strategies in transportation.

Data Needs

The principal focus of personal travel surveys is to serve the needs of transportation planning models. Within this focus, there are at least two principal needs that must be fulfilled. First, data may be needed for model estimation. This involves the development of new models of travel demand for a region, usually by using standard model structures. For this activity, the data needs are the most intensive, because the largest samples and a balance of data between alternative destinations, modes, and routes are needed to be able to obtain reliable estimates of model coefficients and parameters. Other data, besides what may be obtained from the personal travel survey, are also required. Principle among these data needs are the inputs required to construct highway and transit networks from which travel time and related data are derived as an input to the models.

Second, data may be needed to update or validate existing models. In this case, the data needs are substantially less intensive because it is not necessary to completely re-estimate the coefficients and parameters. Rather, updating and validation can usually be achieved with significantly smaller data sets in which not all of the data items required for estimation will necessarily be present. For example, updating or validating trip generation models may require selecting a few of the cells from the cross-classification table of the trip production model and sampling sufficient households to determine a current value for the trip rates. These rates would be compared to existing model rates, and a decision would be made on updating the rates with the new data as well as applying a factor to other rates in the cross-classification scheme that were not measured at this time. Generally, updating can be accomplished with a data set of about 1,000 to 1,500 completed households. Similarly, validation may be achieved by using a small sample of households to determine how well the models predict their travel choices. A non-random, even non-representative sample may be selected for this purpose, and sample sizes will generally be quite small, approximately on the order of a few hundred households. These sample sizes are in contrast to the 3,000 or more households required for model estimation.

Data Requirements of Recent Legislation

Recent legislation such as the Clean Air Act Amendments of 1990 (CAAA), the Intermodal Surface Transportation Efficiency Act of 1991 (ISTEA), and the Transportation Efficiency Act for the 21st Century of 1997 (TEA-21) has all mandated certain changes in how planning is to be completed, which have ramifications for data collection *[2]*. For example, the CAAA made the following changes in data needs *[3, p. 14]*:

- "Demographics – smaller/variable zone systems, GIS-based, wider range of data (e.g., employment categories, housing/rental prices, crime rates).
- "Networks – greater detail (down to arterials because emissions are estimated for the entire network), consistent with zone system scale, GIS-based, reflect economies of scale.
- "Facility Performance – need improved speed/flow relationships, validation data.
- "Conventional Home Interview Data – needed for model development/refinement, detailed spatial emissions analyses (in San Diego, such data revealed orders of magnitude differences in emissions projections).
- "License plate surveys/cordon counts – for off-model flows.
- "Longitudinal surveys/panels – to evaluate response to TCMs and land-use dynamics."

Most of these requirements were further strengthened by the provisions of ISTEA that have been carried through into TEA-21. In a speech from the same conference in which the above listing occurred, Arthur B. Sosslau summarized a number of data collection issues from both the CAAA and the ISTEA *[3, p. 8]*: more cost-effective data collection methods are needed, with greater accuracy than prior survey methods; data are also required on good movements; research is needed on the data collection requirements for intermodal planning; consistency in data collection efforts is necessary, together with "...replicable information from multiple sources..."; research is needed on the data requirements for estimating land-use impacts and changes from increasing facility capacity; and data are needed for assessing the impacts of TCMs. These all represent new data requirements stemming from the CAAA and the ISTEA.

Data Requirements of New Transportation Planning Procedures

One of the important issues arising out of recent legislation is the inadequacy of much of the current transportation planning modeling. Two efforts of interest to this project are occurring. The first of these is an upgrade of existing transportation planning models, and the second is the development of a potentially new procedure for transportation planning models. These model improvements and changes mandate transformation in the data requirements to use upgraded models or the new procedure for modeling. In the updating efforts, one of the major changes required is the introduction of some form of time-of-day model or factors into the modeling process *[2]*. This requires more intensive data travel timing in the network. Another effort is to model trip chains rather than trips, which requires accurate information on the sequencing of trips. There are, again, implications for network detail and accuracy, which do not impact the collection of household travel data directly. However, there is also a need to introduce travel costs into some model steps, requiring better information on the costs of travel, parking, and related activities with better information on household and personal incomes. Automobile ownership modeling is suggested as another step that should be included, which may require an additional data collection on the vehicles available to households and the circumstances of their acquisition. Again, the need for data to support transportation-system-sensitive land-use models is brought out with the concomitant requirement to obtain better data on land use, its impacts on transportation, and the impacts of transportation on land use. Finally, the models need to consider non-motorized modes (walk and bicycle) including a requirement of detailed data on the use of these modes. This last requirement also specifies a need for even more detailed networks in order that walking and bicycling can be modeled effectively.

The second direction being pursued is the development of an entirely new modeling strategy, the TRansportation ANalysis and SIMulation System (TRANSIMS), which is a transportation system simulation model currently being developed for the U.S. DOT. The specific data needs of TRANSIMS are not yet known since the model is not fully operational. However, initial indications are that the model will have extensive data needs well beyond those of current models both in detail and in scope. However, work is also progressing on developing methods and techniques that diminish the input data requirements of TRANSIMS because of the extent of current requirements *[4]*. Specifics of either the full model data input needs or of those that will be required under the diminished requirements have not yet been publicized.

Summary of Data Needs

The data needs for transportation planning can be summarized briefly. The detailed items that are required will depend on the specific models that are to be run. However, the major requirements are:

- Highway networks, including data such as speeds, travel times, capacities, turning characteristics, facility type, area type, etc.

- Transit networks, including data such as speeds, travel times, headways by period of the day, vehicle capacities, etc.

- Household demographics, including income, vehicle ownership or availability, household size, ownership type, housing type, home address, etc.

- Personal demographics such as driver-license status, age, gender, highest level of education attained, employment status and location, student status and school location, etc.

- Travel data for a period of at least 24 hours, including mode(s) used, purpose of travel, number of accompanying family members, costs of travel, time of day, duration of travel, origin and destination locations, etc.

Data Sources

Introduction

The data available to satisfy the needs described in the previous section are generally available from three different sources. Cost-effective data collection requires that data from each source be used.

The first source of transportation planning data is data collected for a specific purpose other than transportation planning but is useful for transportation planning. Generically, these data bases are referred to as *secondary data* by transportation planners, because their use in transportation planning is secondary to the prime purpose for which the data were collected. A good example of a popular secondary data source is the census data collected every ten years to enumerate the population that is also used for a variety of other purposes. Secondary data are usually provided by public agencies, but private companies sometimes do provide data in niche areas not catered to by public sources. Data from public agencies are usually free while those from private companies must be purchased, sometimes at considerable cost. A good source of public data is the Bureau of Transportation Statistics of the Federal Highway Administration (http://www.bts.gov). They also publish a comprehensive directory of public and private secondary data sources in the U.S., Canada, and Mexico *[5]*. The directory describes each data set, its source, main features, sponsoring agency, form, and availability. More secondary data sources are available to transportation planners today than ever before because old databases have been maintained, and several new databases have been established. Examples of some new databases are data from the Commodity Flow Survey and the American Travel Survey first conducted in 1993 and 1995, respectively, by the Bureau of Transportation Statistics.

Second, data may be collected specifically for the purpose of providing information needed in a particular application. These data are referred to as *primary data* because they are collected for the primary purpose for which they are used. Primary data are associated with a particular study area at a particular time and have been the main source of transportation planning data in past transportation studies.

Third, data may be obtained by transferring data from elsewhere into a particular application. This is referred to as *transferred data*. The degree of success with which this is accomplished is referred to as the transferability of the data. Transferability is sometimes described in terms of geographic, temporal, or cultural transferability when information is transferred spatially, over time, or among different cultures, respectively. However, any difference between the conditions when information is collected and when it is used involves the issue of transferability. Changes in development, activity, land use, socio-economic structure, values, attitudes, and travel behavior are fundamentally responsible for changes in transferability. These may or may not be related to spatial, temporal, or cultural differences. Factors affecting transferability in transportation planning and the measure of their impact are not well understood.

Secondary Data Sources

General Census Data. Probably the most important source of secondary data for transportation planners is the decennial census. A wide variety of information is provided, and a list of the information available to the public on computer files from the Bureau of the Census is attached as Appendix A. The data can be accessed through their website at http://www.census.gov or by way of the Louisiana Population Data Center at LSU at http://www.lapop.lsu.edu. Other useful state data centers are listed in Appendix B.

Additional information on updates or projections is also available on census data. The Louisiana Population Data Center at LSU and other state data centers listed in Appendix B list some of this data. For example, 1997 estimates of population by parish and employment by industry type in Louisiana are available together with projections to the year 2020 of population in each parish by gender, race, and age group.

Most of the information in the census data that is useful for transportation planning is obtained in the so-called "long form" of the census. This form is issued to approximately one in six households and contains information on automobile ownership, household income, number of workers, and details on the journey to work *[6, p. 2]*. However, a recent initiative of the Bureau of the Census is aimed at replacing the long form with a continuous random sampling of approximately three million households annually to procure the same information. Starting on a trial basis in 1996, the process has gradually been expanded, and results from samples drawn at 31 test locations in 2000 will be compared with results obtained from the 2000 census long-form for the same areas. This continuous sample survey, called the American Community Survey (ACS), is expected to entirely replace the long form of the census in 2010. A full description of the ACS is included in Appendix C. While the ACS has not been conducted in Louisiana yet, full implementation of ACS throughout the country is scheduled for 2003. The ACS promises to provide year-by-year updates on all census data but does not replace the full enumeration of the population and the collection of limited additional data in the "short form" of the traditional decennial census.

Several data files have been prepared from the census data that have particular application to metropolitan transportation planning. Some of the more important of these are the Census Transportation Planning Package (CTPP) and the Public Use Microdata Samples (PUMS) databases. These databases are described more fully below.

CTPP. Since the 1970 census, special data files have been prepared by the Bureau of the Census from the long form of the census for use in transportation planning in metropolitan areas. These data files were initially referred to as the Urban Transportation Planning Package (UTPP), but since the 1990 census, they have been called the Census Transportation Planning Package (CTPP). They consist of socio-economic, demographic, and journey-to-work data aggregated to Traffic Analysis Zone (TAZ) level in urban areas and parish level for rural areas. The data are provided in an Urban Element for metropolitan areas and a Statewide Element for other areas. The data are provided on CD-ROM for each of eight divisions in the country. Louisiana is within the "west south central" division including Arkansas, Louisiana, Oklahoma, and Texas. The data are distributed free of charge by the Bureau of Transportation Statistics.

CTPP data are widely used, and it is reported that their level of use is increasing with each census *[7, p. 22]*. The disadvantage of the data is that they are only collected every ten years and that they are usually only released two to three years after collection. New developments with the American Community Survey (ACS) described above will likely eliminate these disadvantages and make the CTPP data even more useful.

PUMS. Public Use Microdata Samples (PUMS) are individual records of a sample of households from the long form of the census data with names, addresses, and other personal information removed to ensure anonymity. Data are provided at the person and household level for a sample of long-form households in geographic areas called Public Use Microdata Areas (PUMAs). A PUMA is an area with population in excess of 100,000 so that individual respondents cannot be identified by their characteristics. PUMS files are available in two sizes: a one percent and five percent sample of households (a three percent sample of the elderly is also available). A description of the PUMS data is provided in Appendix D.

The usefulness of the PUMS data lies in the disaggregate level at which the data are provided. In other words, information is provided at person and household levels of respondent characteristics and journey-to-work travel. Personal characteristics such as age, gender, and occupation or household characteristics such as household size, vehicle ownership, and household income can be acquired directly from the data. This is particularly useful when the structure of the population is needed in terms of multiple characteristics, as when seeking the proportion of the population who simultaneously fit into different categories of household size and vehicle ownership.

PUMS data have the same disadvantage as CTPP data because they are only prepared every decade. However, updates are prepared as part of the American Community Survey and as the ACS becomes more universally applied, more updated PUMS data files will be available.

BEA. The Bureau of Economic Analyisis (BEA) of the Department of Commerce has historical and forecast data at parish levels for population, employment, and income. Its main value is in its projected data, which is for all counties (and parishes) in the nation for the years 2000, 2005, 2010, 2020, and 2040. For transportation planning, the projected values must be broken down from parish level to traffic analysis zone level, but the BEA projections are good aggregate estimates that can serve as constraining totals to which disaggregate estimates must comply.

BEA data are available on CD-ROM from the Bureau's order desk (1-800-704-0415) or are available either in microfiche or printed form from Louisiana State University's Middleton Library, which is one of the federal depository libraries for Louisiana. County (or parish) projections were last made in 1990. The projections made in 1995 were only made at Metropolitan Statistical Area (MSA), Bureau of Economic Analysis Zone, or the state level. Recent projections, which extend up to the year 2045, are no longer at the county level as before but are at the coarser level of MSA, BEA zone, or state level. BEA zones are economic areas geographically designated by 183 zones nationwide. Louisiana is divided into 6 BEA zones, although all BEA zones do not necessarily conform to state boundaries. A diagram showing BEA zones is included in Appendix E. A description of the 1995 BEA data that are distributed by CD-ROM and entitled the Regional Economic Information System (REIS) is included in Appendix F. REIS contains historic and projected data with the projected data included in a subdirectory PROJECT.

NPTS. The National Personal Transportation Survey (NPTS) is a nationwide travel survey conducted every five to seven years among randomly selected households throughout the nation. First conducted in 1969, it was repeated in 1976, 1983, 1990, and 1995. The size of the 1995 NPTS data sample is 42,033 households, consisting of 95,360 persons making a recorded number of 409,025 daily trips. The data can be downloaded from the NPTS website (http://www-cta.ornl.gov/npts/1995/Doc/index.shtml) free of charge.

The data are presented in six files, each presenting the data in terms of a different analysis unit. Household information is presented in a household file, personal information in a person file, and details of individual trips shorter than 100 miles and occurring on a specific survey day in a day trip file. Information on longer trips that occurred any time in a specified two-week period is recorded in a travel period file, while detail on segments of trips in the day trip file are recorded in a segment file. Information on vehicles is provided in a separate vehicle file. The files form a relational data base through the cross-referencing of household and person identification numbers in each file.

NPTS data provide a very good means of tracking change in travel characteristics and travel behavior over time, although changes in survey methodology have reduced the validity with which this can be done. The data are not a good source of information for individual geographic areas because sample sizes in individual areas are small. However, it is a very valuable data source when the influence of geographic areas are not expected to be significant, or their influence can be adequately captured by variables within the data. The NPTS data are a valuable source of transfer information.

Commercial Data Bases. A variety of companies have data that can be useful to transportation planners, but some of the most prominent are Dunn and Bradstreet on the provision of employment data, Reebie and Associates for the provision of freight transportation, and Data Resources Incorporated (DRI) on income and material costs. DRI is the primary source of projections of the cost of labor, material, and equipment in industry.

Primary Data Sources

Following a period of almost 30 years in which data had not been collected in a metropolitan area in Louisiana to support a transportation study, several travel surveys have been conducted in the past few years in urban areas in Louisiana. These surveys provide good information in the area in which they were collected, but they also provide potential transfer information for other metropolitan areas in the state. In addition, several travel surveys have been conducted in metropolitan areas in the U.S. in the past few years, which can also serve as potential transfer information to metropolitan areas in Louisiana. Some of the more important local and out-of-state travel surveys are reviewed below.

Baton Rouge Personal Transportation Survey (BRPTS). As an add-on to the NPTS survey of 1995, the Louisiana Department of Transportation and Development commissioned the same company that conducted the NPTS 95 survey (Research Triangle Institute) to survey additional households in the Baton Rouge area with exactly the same instrument they used in the NPTS 95 survey. An additional 1,395 households were surveyed in 1997 in this exercise. The data file contains information on 3,068 persons and 13,194 day trips and are presented in exactly the same format as that of the 1995 NPTS data.

Southern St. Tammany Parish Survey. The consulting firm Urban Systems Inc. conducted a mail-out, mail-back survey in southern St. Tammany parish in 1997. Data were collected on weekday travel (4 a.m. to 4 a.m., Monday to Friday). Households were recruited by sending out postcards to the approximately 56,000 households in the area. Approximately 2,500 households agreed to participate, but completed questionnaires only were received from approximately 800 of those households.

New Orleans Survey. A survey initiated in the latter half of 1999 for the Regional Planning Commission of New Orleans is being conducted by the same consulting firm that conducted the southern St. Tammany parish survey. The survey was in progress at the time of the writing of this report and is scheduled for completion in the first half of the year 2000. Recruitment is by Random Digit Dialing (RDD) with recruitment rates varying between 20 and 25 percent of those contacted. However, initial response rates from those that agreed to participate are low, and it is expected that the final response rates may be similar to those experienced in southern St. Tammany parish survey. However, considering that the population of the New Orleans metropolitan area is considerably larger than that of the southern St. Tammany parish, the data file obtained from this survey is likely to be in the order of 5,000 households, which will constitute a valuable source of transfer data.

NCTCOG Survey. An activity diary survey was conducted in the Dallas-Ft.Worth area in 1996 by the North Central Texas Council of Governments (NCTCOG). Its recent vintage and proximity to Louisiana make it a good candidate for possible transfer data. The data consists of responses from 3,996 households providing information on 44,244 trips.

Other Recent Travel Surveys. Several other transportation surveys have been conducted in the last five or six years that could potentially serve as transfer data to metropolitan areas in Louisiana. These include the Portland, Oregon, two-day activity diary survey conducted in 1994 of 8,900 households, the Salt Lake City survey conducted in 1993 of approximately 2,000 households, and the Detroit survey conducted in 1994 of approximately 7,000 households. However, a much larger number of primary data sets than those mentioned above are available. Stopher and Metcalf reported in their synthesis of travel survey methods in 1996 that they had identified 55 state and metropolitan surveys conducted since 1990 *[8, p. 8]*.

Data Collection

Evolution of the Data Collection Process

Over the past decade, there have been a significant number of changes developed in the formats and procedures used for household travel surveys *[9], [10]*. In the 1980s, the standard household travel survey was a trip-based diary of varying format. However, most diaries asked about the next trip undertaken by the respondent, and then collected various items of data on that trip, including the purpose of the trip and the address of the origin and the destination of the trip. Some of these diaries were developed as a series of sheets of paper, some as tabular forms, and others as booklets. Another common feature was that, without exception in the United States, the surveys were carried out using some form of telephone and mail contact, but face-to-face interviews were not considered to be practical, affordable, or safe to conduct in urban areas across the country. In contrast, surveys in other countries such as in the United Kingdom and Australia continue to use face-to-face interviews.

In 1990, the first activity diary was developed in the United States. This diary was in booklet form at and used essentially the same questions as the trip or travel diary, but the order of the questions were changed *[11]*. The purpose of this change was to alter the focus of respondents to their activities instead of focusing on travel. The focus on travel seemed to lead to a loss of information because people did not consider some of their very short trips and spur-of-the-moment trips important enough to report. Anecdotal evidence suggested that trip rates were higher from the activity diaries than from trip or travel diaries. Several versions of activity diaries were developed in the next few years and used in such areas as southern California, Salt Lake City, Detroit, and elsewhere. In general, the activity diary was typified by working from the perspective of asking "What did you do next?", but also in treating travel as being distinct from an activity. In other words, this type of diary tended to define an activity as something done at a place and travel as something connecting activities in different places but not as an activity.

In the mid-1990s, some activity diaries were developed into time-use diaries. There has been some confusion about this particular nomenclature. Some use the term to denote a format in which travel is no longer defined as a connection between activities but is treated as another type of activity. Others have defined the time-use diary as one that collects data about in-home activities and activities done outside the home in detail. In this report, the true time use diary is defined as one that does both, i.e., collects data on in-home activities (in more detail than just work at home and other) and also treats travel as another type of activity.

The general thrust of these developments has been to try to correct deficiencies in other survey formats, particularly the under-reporting of travel, and also to enrich the total data base by collecting data on in-home activities that may be traded off with out-of-home activities and by collecting transit trip data in detail. The latter can be achieved by defining that each activity involved in traveling by transit represents another activity. Thus, each walk, wait, travel on a vehicle, etc. is considered a separate activity, and data are requested on locations of waits and other attributes of the activities involving movement *[12]*.

A consistent feature of the surveys conducted in the past 40 years has been cross-sectional surveys, collecting large or moderate samples of households once every ten or more years for which the samples are drawn independently of one another. Depending on the purposes for which the data are used, these cross-sectional surveys may not represent the optimal use of resources. In many fields of endeavor, where there are continuing changes occurring, an alternative survey methodology is used. This is known as the panel survey *[13]*. More correctly, one should consider these types of surveys as surveys with overlapping samples that are conducted with a relatively frequent periodicity, e.g., once per year. Only one transportation panel has been conducted in the U.S. for a significant period of time, which is the Puget Sound Panel Survey. Although there have been a number of other efforts to set up panels, most of them have collected no more than one repeat of the data (a second wave)

[14]. There have also been transportation panels in other countries, most notably in the Netherlands where the Dutch National Mobility Panel was established in the mid 1980s *[15]*. The merits of panels have been explored by various researchers *[16]*, *[17]*, *[18]*. Principally, the idea behind the use of a panel was to first undertake a comprehensive cross-sectional survey with a relatively large sample of households (e.g., on the order of 3,000 to 5,000). From within the sample, a panel was selected and then surveyed every six to 12 months. The changes detected in the panel were then applied to the cross-sectional data to "update" the data on a frequent basis. Theoretically, the cross-sectional survey should have never gone out of date through a process of this type. In practice, there comes a point where the cross-sectional data are sufficiently old, and a new cross-sectional data set must be collected.

In contrast to the decennial or bi-decennial collection of a major cross-section, there are at least three advantages that a panel offers. The first is the ability to continually update the cross-sectional data so that after five, ten, or fifteen years from collection, the cross-sectional survey still remains largely relevant and sufficiently up-to-date to be usable for many of its original purposes. Without a panel for updating, the data become no longer relevant and need to be discarded. Second, the panel data alone can be used to track the dynamics of change. For example, in the Puget Sound Panel Survey, among the data measured was the choice of travel mode for the work trip. In the Seattle area, it had been observed over some years that transit ridership was fairly stable in total ridership numbers. This implied that a segment of the population rode buses, while the remainder of the population drove cars. The panel revealed that bus ridership was not stable, and most riders rode the bus for about one year then used a car. In the meantime, these converted transit riders were replaced by new transit riders on a continuous basis. This information clearly suggests quite different strategies for building bus ridership than the cross-sectional data suggests. Hence, there are clear policy advantages to be obtained from panels. Third, the panel can also be used to obtain information on immediate issues and policies in transportation. This is possible because there exists a group of households whose travel and demographic characteristics are already known and with whom an abbreviated survey is already being undertaken. In addition, survey-specific, attitudinal, or stated response questions can provide rich information for policy and decision makers.

Problems with Current Data Collection Procedures

Not withstanding the developments of the past several years in household travel surveys, at least two problems remain in the collection of such data. First, the diary format is often poorly comprehended by respondents and considerable effort is often required in data repair, or the decision is made to collect the data by Computer-Aided Telephone Interviewing (CATI) as a means to avoid the extensive data repair that may be necessary *[10]*. This problem is not a function of the specific physical design of the diary as a booklet in tabular format, or as a series of sheets of paper, but it seems to reside in other issues relating to the design of the survey such as the specificity required from people about their travel, which many are unable or unwilling to provide, and question wording and presentation. The second significant problem that is not addressed by the time-use diary and is possibly even exacerbated by it is that of the identification of the geographic locations where non-travel activities take place. Here the problem is that the majority of people do not know an address that is sufficiently complete for geocoding many of the locations they typically visit. Most people can provide a complete address for home; many can also do so for their workplaces. Some parents will know the street address of the schools that their children attend, but most will know, at best, the name of the school and possibly the street on which it is located. Addresses of shops, banks, post offices, friends' homes, restaurants, etc. are generally not known.

Geographic locations are extremely important for transportation analysis. Geographic Information Systems (GIS) as a data analysis, display technique and a platform on which to run travel-demand models to increase the precision of information on locations. Many transportation analysts would like to be able to code all trip ends to a latitude and longitude, rather than a traffic analysis zone. However, imprecise data on locations makes even coding to traffic analysis zones often an inaccurate and time-consuming process. More than one consultant working on preparing travel survey data has defined the geocoding activity as a "black hole," that sucks up all available budget and labor resources and provides no light on the subject, whatsoever. While there are several approaches to solving this problem, the research conducted in this project addresses one method in particular, namely the use of Global Positioning System (GPS) devices to record time and position of travelers.

New Developments

As a result of the new legislation reviewed in the section on data needs above, there have been new developments in the area of data collection. Within the Travel Model Improvement Program (TMIP) of U.S. DOT, within which TRANSIMS is being developed, there is a track devoted to improving data collection methods. As part of that effort, several steps have been taken towards the improvement of data collection methods. First, a new version of the

FHWA *Origin-Destination Surveys Guide* has been produced called the *Travel Survey Manual [19]*. This manual sets out more up-to-date procedures and methods for conducting household and personal travel surveys and the accompanying surveys such as screenline and cordon surveys. Another activity sponsored under this track was a conference on "Household Travel Surveys: New Concepts and Research Needs," which was tasked with defining a research program for Track D of TMIP *[20]*. As a result of this conference, the proof-of-concept project on the use of a GPS for tracking spatial travel locations recommendation was implemented as reported in the *Lexington Area Travel Data Collection Test [21]*. Another topic that came out of that conference that is currently being researched is the potential of using the Internet as an alternative method of response for those households that wish to use it. Publication of results of this work has not yet occurred. Most of the other recommendations have yet to be implemented.

New developments can probably be summarized into a number of categories as follows:

- Development of new formats for travel diaries, particularly in the direction of activity and time-use diaries
- Addition of new data elements to surveys, particularly such personal descriptors as education level, race and ethnicity, detailed work status, and personal income
- Preliminary experimentation with the use of GPS devices to track travel
- Use of GIS to record travel survey data and to assist in the geocoding process

There is a reluctance by many public agencies to embrace new and less tried methods. Consequently, a considerable amount of survey work continues to use methods originally developed in the 1970s and 1980s, often with significant negative impact on the quality of the data collected.

Strategic Approach to Cost-effective Data Collection

A National Cooperative Highway Research Program (NCHRP) project was completed in 1997. Its objectives were the identification of data needs for statewide and metropolitan transportation planning, the availability of current data, evaluating the cost-effectiveness of data collection activities, and investigating the potential of data integration and consistency among planning organizations. The study developed a strategic approach that can be applied by state and MPO officials to facilitate data needs assessment, data organization, data-collection priorities, and data dissemination. Study results are summarized in NCHRP

Report 401 *[22]*. A final report and a supplemental document on data collection practices and sources were also produced *[23]*, *[24]*.

The process proposed in the study involves first establishing a data task force of representatives from the state, MPO, local authority, transportation agencies, and other stakeholders such as air quality agencies to manage and coordinate data collection. Data needs are identified by using classic strategic business planning principles, and each agency in an area is urged to conduct exercises to identify its data needs. The study then suggests a comprehensive framework in which the data from all the different agencies are organized (see Appendix G). This allows data needs to be identified collectively for an area rather than at individual agency levels. The study recommends that redundancy of data among agencies be reviewed and a small data committee be established to determine the relative importance of different data sets and the cost of collecting them. They recommend a rough evaluation to identify the most cost-effective methods of collecting important data. Less important data may be collected less frequently or use methods that are cheaper and, possibly, are less accurate. Data sharing among agencies is encouraged, and the establishment of shared data information systems such as Geographic Information Systems (GIS) is suggested as a practical means of creating and maintaining a common data base in an area. Issues of data consistency and standardization are central to the practical implementation of the system proposed in this study.

The study suggests several principles leading to cost-effective data collection that can be applied in this study. These include:

- use of existing (i.e. secondary) data sources to the greatest extent possible
- sharing of data among agencies in the same area
- identifying relative importance of individual data items and identifying the most cost-effective data collection method commensurate with the accuracy justified by the importance of the item.
- establishing consistency and standardization of data items.
- establishing an information system that permits ready access to the data by all users.

In addition to these the use of transferred data may be appropriate and helpful sometimes. The transferred data can be updated with local data in those situations where local data are collected. If local data collection procedures are standardized, more opportunities to transfer data that are consistent and of recent vintage will be presented.

The relevance of the Faucett study to the research conducted in this study is evident, although this Faucett process is more strategic than that foreseen for this project. Their proposed process requires considerable effort in establishing and maintaining a database that is accessible and useful to all users in the area. The process does promise a more cost-effective database because it benefits from lack of redundancy, emphasis on useful and important data, standardization, and sharing of information, but it requires a major investment in effort to establish and maintain.

Transferability of Transportation Planning Data

Transferability has been a subject of research in transportation planning since soon after the introduction of the system-wide transportation planning process in the mid 1950's *[25], [26]*. The initial issue was whether models calibrated on data in the base year remained valid when they were used to forecast travel in the future. Later, the matter of being able to use a model calibrated in one area and apply it in another was studied. The matter of geographic transferability of models was investigated for the potential savings in time and money that could be realized by not having to collect data in each area in which an application was required.

In studying the question of the transferability of transportation planning models, investigations of both temporal and geographic transferability produced mixed results. Some models displayed high levels of transferability while others did not. However, one finding was consistent: the transferability of models was improved when local data were used to update transferred model parameters *[27], [28]*. Given this experience with models, it is anticipated that the transferability of data will be equally enhanced by using local data to update transferred data.

The transferability of transportation planning data has not been studied in the past, although use has been made of the expected general stability of some transportation data items for quick estimates or validation of other data. For example, the Institute of Transportation Engineers (ITE) has produced a database of urban travel characteristics from a number of studies conducted in the United States in recent years *[29]*. Similarly, the Federal Transit

Administration has produced a database of the characteristics of urban transportation transit systems from a number of transit systems around the U.S. *[30]*. These databases provide typical values of a wide range of travel and system characteristics as well as the socio-economic structure of travelers. These values can be used for coarse-level estimation of transit ridership or response to policy changes, or they can be used to test the reasonableness of values obtained in other surveys. Another example of the use of standard values is the trip generation rates produced by the Institute of Transportation Engineers' in which average trip generation rates for a wide range of land uses are provided *[31]*. These rates are widely used in estimating trip generation for individual land uses.

When the question of the transferability of transportation planning data is reviewed, the nature of the data considered for transfer must be defined. Traditionally, data collected for a transportation study consist of detailed information on a sample of households and the travel they conducted on a certain day. The data includes records of individual trips between specific locations in the study area. Clearly, data of this type are not transferable. However, simple relationships within the data that portray basic travel behavior may be transferable. For example, travel behavior such as trips per day, trip length distribution, time of day of travel, and mode choice may be transferable between two locations if they are made among households with similar characteristics operating within similar travel environments. Thus, similar households operating in similar travel environments are expected to exhibit similar travel behavior irrespective of their location.

The concept that travel behavior is a function of the characteristics of the traveler and the transportation system (i.e., the travel environment) is not new; it was postulated from the very beginning of the development of the transportation planning process *[32]*. The difficulty has been in attaining sufficient characteristics of the traveler and travel environment to identify their close relationship to travel behavior. In transportation planning models, this has depended on including sufficient relevant variables in the model specification and having a model format that allows accurate emulation of observed travel behavior. To the extent that models have achieved this goal, transferability has been improved as demonstrated by those studies that have shown that improved model specification leads to improved transferability *[33], [34]*. The same principle is expected to apply to the transfer of data where more accurate data are more transferable. Greater accuracy of data is achieved by making the data more representative of the population they represent. Data considered for transfer consist of average relationships in the data such as trip rates, trip length distributions, time of day patterns, and overall mode share proportions. Accuracy of these relationships is enhanced when the population is segmented so that the relationships are as homogeneous as possible within each segment. If travel behavior is dictated by the characteristics of the traveler and his/her travel environment, then segmentation leading to homogeneous travel behavior segments will occur in the characteristics of the traveler and the travel environment.

It is interesting to note that while data used to estimate transportation planning models is usually provided at the individual household, person, and trip level, it is not needed in that much detail to calibrate many of the models used in travel demand estimation. For example, a cross-classification trip generation model is estimated by observing the average trip rate in each category of the model without having to know the detail of individual trips. Similarly, a gravity trip distribution model is calibrated on total productions and attractions by zone and the distribution of trips by travel time interval. Thus, a gravity model can be calibrated without knowledge of individual trip interchanges, since only zonal trip ends and the proportion of all trips by travel time intervals is required for calibration. The fact that trip generation and trip distribution models are usually estimated for different trip purposes, separately, does not affect the observation.

For a modal split, disaggregate data are needed to calibrate a discrete choice model, but aggregate data can be used to modify the model. For example, the alternative specific constants in a multinomial logit model can be adjusted to ensure that the model reproduces local aggregate mode shares by subtracting $\ln(P_i/Q_i)$ from the alternative specific constant of a transferred model when P_i and Q_i are the mode share for alternative i in the transfer and local settings, respectively *[35, p. 237]*. If more local information is available such as aggregate mode shares for two or more subpopulations or subregions of an area, then local alternative specific constants α_i and a scale factor λ can be estimated from the set of nonlinear simultaneous equations of the form shown in equation (1) below for each share *i* in subpopulation or subregion *m [36, p. 27]*. The parameters can be estimated within a logit model estimation package.

$$S_{i,m} = \frac{e^{\alpha_i + \lambda\beta x_{i,m}}}{\sum_{j=1}^{J} e^{\alpha_i + \lambda\beta x_{i,m}}} \qquad (1)$$

where,

$S_{i,m}$ = share of alternative i in subpopulation or subregion m.

x_i = independent variable values for alternative i.

J = total number of alternatives.

It is true that not all travel demand models commonly used in transportation planning function on aggregate data. For example, regression models of trip production are typically calibrated on household-level data. However, trip attraction regression models are usually zonal-level models using variables such as total employment, total floor area by land use type, and average residential density in each zone as attraction variables in the model. In addition, regression models can also be made to function on aggregate data by establishing dummy variables that represent groups of households. In fact, regression models provide the opportunity to combine disaggregate and aggregate data into a single model if some disaggregate data are available, but only aggregate data is available for the other variables in the model. A regression model with only dummy variables is equivalent to a cross-classification model.

The fact that many travel demand models can be calibrated on aggregate data such as trip rates, trip length distributions, and mode share proportions is fortuitous since these are the same characteristics of travel behavior that are potentially transferable and can be obtained from existing data sources. Thus, to the extent that data can be successfully transferred, a means of relieving the need for local data collection will be identified.

OBJECTIVE

The objectives of this study are:

- To investigate the potential of using transportation planning data collected elsewhere or at another time and to satisfy the data needs of current transportation planning activities in metropolitan areas of Louisiana. One objective is to investigate the geographic and temporal transferability of data used in metropolitan transportation planning in Louisiana. The use transferred data reduces the demand for current data collection, leading to cost and time savings.

- To investigate the potential to use small panels to collect detailed data that can be used for continuous updating of various data sources. Past studies of transportation model transferability have shown that transferred information can be dramatically improved in a local setting if small samples of local data are used to update transferred information. Therefore, it is the purpose of this study to investigate whether a panel of approximately 200 to 500 households from across the state, surveyed on a regular basis, may provide adequate geographic and temporal updating capabilities to transferred data used in individual MPOs in Louisiana.

- To extend and undertake further testing of the potential to use a GPS-PDA instrument for collecting travel-related information, particularly under circumstances where a fully-rectified GIS of the street system does not exist and to obtain several improvements to technology and user interface.

SCOPE

The research in this study was directed at developing cost-effective methods of data collection for transportation planning of personal travel in metropolitan areas of Louisiana. Freight transportation and inter-city travel were not included in the investigation.

Data requirements for transportation planning can be broken into those required to estimate and operate a model in the current environment and those that portray future demographic and transportation system conditions, serving as input to the model when predicting future travel. The preparation of forecast data is not considered in this analysis.

The methods investigated in this study represent some of the ways in which data can be collected more cost-effectively, but they do not represent all the potential means of achieving that goal. In addition, the investigations conducted in this study have produced tentative findings that should be confirmed with further study.

METHODOLOGY

Approach

The overall approach adopted in this study, to provide cost-effective data collection for metropolitan transportation planning, was accomplished by pursuing three consistent and complementary research efforts. First, the investigation was directed at making maximum use of existing data. Existing data included both secondary and primary data. Second, since existing data were unlikely to satisfy all transportation planning data needs, research was directed at collecting local data as efficiently as possible. Third, research was directed at combining the existing and new data as effectively as possible.

The first research effort of making maximum use of existing data was pursued by identifying primary and secondary data sources, describing the nature of the data in each source, analyzing data from some of these sources, and evaluating the accuracy of the data by comparing their values to those of a local survey.

The second research effort, the efficient collection of local data, was pursued by investigating how stratified sampling could be used to permit lower overall sampling rates, how a new design of the time-use diary could lead to greater accuracy and ease of reporting among respondents, and how a Global Positioning System (GPS) device placed in each vehicle of a household could lead to a more accurate and complete recording of travel. The new time-use diary was employed to collect data from a stratified sample of respondents as a demonstration of the establishment of a panel that could be surveyed on a regular basis. The panel could be used to maintain an up-to-date database of local data.

The third research effort, the effective combination of existing and new data, was pursued by investigating how the two data sources could best complement or enhance each other. For example, one of the features of existing cross-sectional data sets is that they became outdated quickly. Using a regularly-surveyed panel for local data allows transferred data to be brought up to date and adjusted to fit local conditions. The procedure combining information from two or more sources, giving appropriate weight to each, is the subject of discussion in the next section.

Data Updating

A past means of updating the parameters of transportation planning models that has been used is Bayesian updating *[35]*. It assumes that model parameters estimated of one sample belong to the same population as estimates from a similarly specified model of another sample. The Bayesian updating process permits estimates from the two sources to be combined to provide new estimates that incorporate the influence of both. The process is particularly useful when local samples are too small to provide reliable estimates on their own or local data are outdated. There is no requirement that one of the samples be a local sample but only that estimates from the different sources belong to the same distribution and, obviously, that the area where the updated model is to be used is well represented by the areas from which the contributing samples come.

While Bayesian updating has been used to update models it may also be used to update data. A particularly convenient form of Bayesian updating involves the use of conjugate priors. A conjugate prior is a prior distributed in such a manner that when it is combined with the likelihood function of local data, a posterior distribution is produced with the same functional form as the prior *[38, p. 412]*. It is convenient because it allows the resulting posterior distribution, in this case the updated data, to be expressed in simple terms of the prior distribution, the transfer data, and local data. A popular conjugate prior is normally a distributed prior combined with distributed local data. This produces a posterior distribution which is normally distributed with parameters θ and σ^2, derived from the characteristics of the prior and local data as follows:

$$\theta_{\text{updated}} = \frac{\dfrac{\theta_{\text{transfer}}}{\sigma^2_{\text{transfer}}} + \dfrac{\theta_{\text{local}}}{\sigma^2_{\text{local}}}}{\dfrac{1}{\sigma^2_{\text{transfer}}} + \dfrac{1}{\sigma^2_{\text{local}}}} \quad \textbf{(2)}$$

and

$$\sigma^2_{\text{updated}} = \frac{1}{\dfrac{1}{\sigma^2_{\text{transfer}}} + \dfrac{1}{\sigma^2_{\text{local}}}} \quad \textbf{(3)}$$

where,

θ = mean of the data item.

σ^2 = variance of the data item.

It is interesting to note that applying the above updating process recursively to more than two data sources is mathematically equivalent to adding additional terms to the expression. In other words, an updated data value can be obtained from multiple sources provided the data can be assumed as coming from the same normal distribution in each source. In this case, the expression for the updated data value for data from N sources would be:

$$\theta_{\text{updated}} = \frac{\sum_{n=1}^{N} \frac{\theta_n}{\sigma_n^2}}{\sum_{n=1}^{N} \frac{1}{\sigma_n^2}} \tag{4}$$

and

$$\sigma^2_{\text{updated}} = \frac{1}{\sum_{n=1}^{N} \frac{1}{\sigma_n^2}} \tag{5}$$

In equations (2) and (4) the data item values from the contributory sources are weighted by the inverse of their variance. Intuitively, this is appealing since data values known with less certainty (i.e. have larger variance) contribute less to the updated estimate of the data item than those known with greater certainty. An analyst may elect to manually alter the variance of a data item from a particular source to alter the weight of that source if the automatically assigned weight is considered inappropriate.

Definition of a Panel

The term "panel" is used in survey research to denote a survey that is conducted on the same respondents on two or more occasions. However, it is more correct to use the term "overlapping samples" to describe this type of sampling *[13]*. In general, samples on two or more occasions may be drawn in one of two distinct ways: as independent (non-overlapping) samples or as overlapping samples. This concept can be most easily understood by considering the following diagram (figure 1), which shows a non-overlapping sample and three types of overlapping samples.

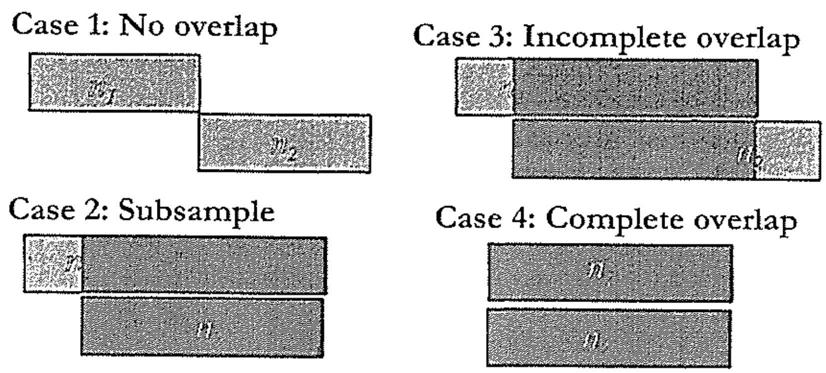


Figure 1

Illustration of overlapping and non-overlapping samples

In figure 1, n_1 denotes the sample drawn on the first occasion, and n_2 denotes the sample drawn on the second occasion. In conventional household travel surveys, samples are drawn independently every ten to twenty years. This represents the situation exhibited in case 1 where the n_1 samples drawn on the first occasion are independent and different from those drawn on the second occasion, n_2. If there are any common households or persons between the first and second occasion, it is by coincidence, not design. The samples are normally considered completely separate and independent; this is shown by the lighter color in each of the two samples.

In case 2, a sample of n_1 households was drawn on the first occasion. On the second occasion, a smaller sample of n_2 was drawn, where every household in the second sample was included in the first sample, but some households from the first occasion were not included on the second occasion. This defines the second occasion as a subsample of the first. This is one of the forms of the overlapping sample. In case 3, the samples overlap so that some of the households sampled on the first occasion were not sampled on the second while some households were sampled on the second occasion and not included in the first sample. This defines an incomplete overlap. A special part of case 3 occurred when n_1 and n_2 were chosen to be equal. In this case, there was a common component of the sample, although the subsample of the first occasion, that is not included in the second, was *replaced* by an equal number of households for the second occasion. Case 4 illustrated a situation in which there was a complete overlap of the samples from the two occasions. Case 4 is what is most often thought of by the term *panel*.

In reality, creating case 4 for a survey of human populations is extremely difficult, even impossible, unless the time span between the first and second occasion is very short, the survey is very simple, and all members of the sample on the first occasion have a very strong commitment to participate in the second occasion. In reality, several events may occur to members of the sample on the first occasion that will prevent inclusion of some of them on the second. First, some members of the sample in the first occasion may die in the time between the two occasions. Second, if the survey is specific to a geographic location (e.g., a metropolitan area), some members of the sample in the first occasion may move away before the second occasion. Third, some of the sample from the first occasion may simply change their minds about participating in the second occasion. Fourth, in the event that the survey is of a household, the following circumstances may occur: a household may break up by death, divorce, or children leaving home. If any of these events occur, then the desire to create a panel survey is forced into becoming either a case 2 or a case 3 sample. Case 2 defines a situation in which some members of the sample from the first occasion have been lost to the survey by the time of the second occasion, and the sample panel is allowed to become smaller. Case 3 defines a situation in which the sample loss from the first occasion is made up by adding new samples on the second occasion either to maintain the panel size, enlarge the original panel, or operate with a somewhat smaller panel.

In general, no matter what the reason for loss of sample between the two occasions, the loss of members of the panel from one occasion to another is termed panel *attrition*. Of course, it follows that the longer the period between the two occasions, the greater the attrition is likely to be. Also, if the survey is to be repeated on more than two occasions, then the larger the number of occasions, the greater the attrition will be. In fact, a further element may be added to the attrition, which is fatigue of panel members arising from repetitive surveying.

From the above cases of overlapping samples, different terms can be used for different cases. Case 4 remains as before as a true panel, but could also be defined as the ideal case of a panel without attrition. Case 2 is then a panel with attrition, while case 3 represents a panel with attrition and replenishment or replacement. Furthermore, if in case 3 there is an attempt to add additional members to the panel in the second or subsequent wave of the survey beyond what is required to replace those panel members lost to attrition of various types, then this case may become a case of a panel with replacement and enrichment.

Surveys that are conducted at several points in time are sometimes referred to as *time-series surveys* or *longitudinal surveys [1, p. 34-35]*. It is useful to make a distinction between these two as does Hensher *[39]*. Time-series data are values of a single variable over time. Thus, daily traffic volumes observed at a permanent traffic counting station or annual transit ridership on a mass transit system could be represented as time series data. Longitudinal data include multiple variables in which the independent variables and their causal relationships to the dependent variables may change over time. Longitudinal data includes panel data for transportation planning. A further distinction is that time-series data are collected in order to track growth or change over time, while longitudinal data are collected to assist in understanding the underlying causes of the process being surveyed.

Reasons to Consider Using a Panel

Probably the most common type of panel recognized is a panel used in a medical study. Panels have been used to study the effects of suspected toxic or carcinogenic substances, to study the effects of lifestyle changes on health and well-being, and to study the effects of prescription medications. Such studies are frequently described in the news media. In these cases, the reasons for using panels instead of cross-sectional samples seems to be self-evident – the scientists want to study the effects of something that causes change in the long term but may not be observable in the short term. Therefore, the obvious technique is to sample a group of individuals who are either exposed to the suspect toxic or carcinogenic compound, are dosed with the medication, or are willing to undergo the lifestyle change under investigation. However, the real reason for using a panel is more scientific than this suggests.

The sampling errors of using a panel are smaller than those of repetitive cross-sections when the subject of interest changes over time. This can be shown mathematically. To see the mathematical proof of this, it is necessary to first discuss sampling errors and to develop an understanding of the sampling error properties of different methods of repetitive samples.

Sampling Error. Sampling error is a property of all samples, but it disappears when a census is taken. Here the strict definition of a sample is used as a part of the total population, while a census, on the other hand, is a survey of all members of a population. Sampling error must always exist because it is a function of the extent to which individual idiosyncracies in the population are either omitted from the sample or occur in the sample more frequently than in the total population. Sampling error also varies with the method of sampling used. In this exposition, it is assumed that samples are drawn by the method of *simple random sampling*, although all of the following results can be readily applied to any other form of sampling such as proportionate sampling, disproportionate sampling, etc.

In a simple random sample, the sampling error for the mean, measured from the sample, is given in equation 6.

$$sampling\ error\ of\ mean = \frac{\sigma\sqrt{(1-f)}}{\sqrt{n}} \tag{6}$$

where,

$\sigma =$ standard deviation of the measure, usually estimated from the sample by s

$n =$ sample size

$f =$ the sampling fraction, or n/N, where N is the population size

The term $(1-f)$ disappears when N is very large and/or when n is very small. It is known as the finite population correction factor. If all other factors are kept equal, it can also be seen that as the sample size becomes large, the sampling error becomes smaller.

Now, consider the sampling error for a difference in means between two samples. Suppose that the variable of interest is the average car ownership, denoted by $\bar{x}$, and that the subscripts *1* and *2* are used to denote the first and second occasions of the survey, respectively. The sampling error of the difference in the mean car ownership between two surveys is given in general by equation 7.

$$\text{s.e.}(\overline{x}_2 - \overline{x}_1) = \sqrt{(\text{s.e.}\overline{x}_2)^2 + (\text{s.e.}\overline{x}_1)^2 - 2\,\text{cov}(\overline{x}_1, \overline{x}_2)} \qquad (7)$$

Suppose that the two samples are independent samples. In this case, the covariance term can be assumed to be zero, and the sampling error becomes that shown in equation 8.

$$\text{s.e.}(\overline{x}_2 - \overline{x}_1) = \sqrt{(\text{s.e.}\overline{x}_2)^2 + (\text{s.e.}\overline{x}_1)^2} \qquad (8)$$

In other words, this means that the sampling error of the difference in the values is equal to the square root of the sum of the squared sampling errors from each of the two surveys. If the two surveys had identical sampling errors, this would indicate that the sampling error of the difference would equal approximately 1.4 times the sampling error from either survey. Suppose average car ownership were measured as 1.95 in the first survey, 2.35 in the second survey, and that the sampling errors, identical in both surveys, were ±0.25. The difference in car ownership between the two surveys is 0.4, and the sampling error of the difference is ±0.35. With 95 percent confidence, the difference is between -0.29 and +1.09. This means that it is not certain if car ownership changed between the two occasions.

Now, suppose that car ownership is measured using overlapping samples. The least error will occur when the overlapping samples are a panel with no attrition, i.e., case 4. In this case, the covariance term in equation 2 is maximized. If the sampling error of car ownership is approximately the same on each occasion and there is a high correlation, R, between car ownership on the first occasion and car ownership on the second occasion (meaning that most households increase car ownership in like proportions if they were the same car ownership level on the first occasion), then the covariance will be almost equal to the square of the sampling error from either occasion. In such a case, the sampling error is approximately zero, as is shown in equation 9.

$$\text{s.e.}(\overline{x}_2 - \overline{x}_1) = \sqrt{2(\text{s.e.}\overline{x})^2 - 2R(\text{s.e.}\overline{x})^2} = \text{s.e.}(\overline{x})\sqrt{2(1-R)} \qquad (9)$$

If the same values of car ownership were observed in these two surveys and the correlation between the two was 0.95, then the sampling error of the difference would be ±0.08. In this case, there be a 95 percent confidence that the difference lay between 0.25 and 0.55. There would be a 95 percent confidence that car ownership had increased and that the increase was between one-quarter and one-half of a car per household. This is clearly a far superior result than that obtained from the non-overlapping samples.

In the case of an overlapping sample of the case 3 type, the amount of the covariance that is subtracted is based on the size of the common segment of the sample. The greater the overlap, the more covariance subtracted. In a case 2 situation, however, the result is a little different. In this case the sampling error is obtained from equation 10, in which n_c denotes the common subsample size.

$$\text{s.e.}(\bar{x}_2 - \bar{x}_1) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_c} - 2\frac{s_{12}}{n_1}} \qquad \textbf{(10)}$$

The sampling error may not, in this case, benefit from the additional sample that is included in the first survey. This will depend on the relative magnitudes of n_1, n_c, s_1, s_2, and s_{12} and the correlation between the initial sample and the subsample on the second occasion. Indeed, there are conditions under which the sampling error of the difference is actually *increased* as a result of the additional sample in the first survey *[13, p. 460]*.

In summary, the statistical benefit of overlapping samples and panels can generally be seen as reducing the uncertainty in the measurement of the difference in any variable between the two occasions compared to non-overlapping samples. Clearly, however, the extent of the overlap is key, and attrition of a panel needs to be minimized in order to achieve maximum benefit from this statistical characteristic. It is also important to note that, while all of the above analysis and discussion has focused on two occasions, the same benefits will arise from multiple occasions, and the above formulas can be applied to each pair of occasions that are to be compared.

Other Benefits of Overlapping Samples. There is a second benefit to be obtained from overlapping samples on multiple occasions. In situations in which dynamic processes underlie any of the measures of interest to the analyst, the existence of these dynamic processes will become evident only when overlapping samples are used. An illustration of this fact is found in findings from the Puget Sound Transportation Panel regarding transit riders, as aforementioned. A series of cross-sectional surveys in the Seattle region had indicated for some time that transit ridership was fairly constant in terms of total numbers of riders. The cross-sectional surveys led to the assumption that in Seattle there existed a population sub-group of transit riders, and a much larger population subgroup of people who did not ride transit. However, the Puget Sound Transportation Panel found that transit usage was continually changing. A significant proportion of transit users were younger adults who started out as transit users, acquired a car after a year or two, and then switched to single-occupant auto *[40]*. As fast as these transit users moved to auto use, they were replaced by new young adults as transit riders. This could be revealed only by a panel. If the policy of the region was to seek methods to increase transit ridership, then very different policies would be called for from the information obtained in the repetitive cross-sections rather than from the panel. In the first case, policies would be developed to try to lure car users from their cars into transit. In the second case, however, the policies would concentrate on retaining transit users after one to two years and seeking ways to discourage them from switching to auto.

The ability of a panel survey to uncover dynamic processes of this type is clearly a very important attribute of the panel, but it is also not the only additional advantage that can be obtained. Recalling the statistical benefits of overlapping samples, it is also clear that better information of a given accuracy can be obtained from a smaller sample when overlapping samples are used than where independent cross-sectional samples are used. One of the major problems with survey data for transportation planning is that it rapidly becomes out-of-date. There is a need to collect a new data set in order to be able to maintain models with up-to-date information. This is an expensive process, with the costs of each completed household being as much as \$125 to \$150. Assuming that each time a household travel survey is completed, approximately 3,000 households should be sampled, this translates to a cost of around \$400,000 for each cross-sectional survey. In contrast, Murakami and Ulberg quote a comparable contractual cost of about \$170 per household for four waves of the Puget Sound Transportation Panel *[40, p. 166]*. This means that four waves (four annual repeat surveys of the same households) cost between 13 and 36 percent more than a single cross-sectional household survey. Total costs of the panel are approximately twice this amount. However, this includes agency costs such as project management and agency contributions to design, implementation, coding, geocoding, and analysis that are not reflected in the average cost of cross-sectional surveys.

These considerations lead to a further possible use of a panel – to update an existing cross-sectional survey and prolong its useful life. For example, suppose that a cross-sectional survey of 5,000 households were to be undertaken at a certain point in time, for example, in the year 2001. Under the normal procedures, this survey would need to be repeated by 2011, because the original data would be outdated by this time. Thus, a cost of approximately $700,000 (in constant 2001 dollars) would be incurred to maintain the data base. It should also be noted that since the data are usually not available for modeling and other analysis until eighteen months to two years after collection, the data are likely to have been more outdated than desired for at least the last three or four years of each decade. Now suppose that 500 of the households in this sample are impaneled. Suppose further, that this panel is surveyed annually beginning in the year following the initial cross-sectional survey at an average cost of $45 per survey. After ten years, the cost of the panel would be $225,000 for ten waves. The use of these 500 panel measurements each year to update the other 4,500 households in the original cross-sectional data set would result in the sample still being considered reasonably up-to-date at the end of ten years. To compare costs, suppose that the panel were to be ended after ten years and the updated household data were then used for a further ten years. Total costs at the end of twenty years would be $925,000 compared to $1,400,000 for two cross-sectional surveys. Clearly, if updating can be performed from a panel, then the panel offers considerable potentials for cost savings over standard methods of surveying for planning purposes. This would be an obvious contribution to cost-effective data collection methods.

Potential Problems of Panels

While the preceding section has outlined the major advantages to be obtained from the design and use of a panel, there are also problems with panels. The first and most obvious problem is the continued care and maintenance of a panel. A cross-sectional survey requires a one-time allocation of funds, albeit a very large allocation, after which no further funds are needed for the survey for ten to twenty years. However, use of a panel requires an annual budget outlay, albeit very small in comparison, which must be maintained each year or else the panel is likely to be lost. The attrition mentioned earlier in this report becomes a severe problem if a panel is not used for a period of two or more years, after which time it is possible that so many of the original panel members have been lost that it becomes necessary to start with a fresh panel. In the case of the Puget Sound Transportation Panel, a period of two years did elapse between the first and second waves, but several contacts were made with the panel members during that time *[40, p. 176]*. Even so, attrition between the first and second waves was markedly larger than between subsequent waves, where surveys were conducted on an annual basis. Thus, to maintain a panel, it is necessary for the government agency that is sponsoring the work to maintain an annual budget item that is at least sufficient in making periodic contacts with panel members and that should, ideally, allow for a survey at least once each year. It is interesting to note that, in the Puget Sound Transportation Panel, there are still 329 households from the original 1989 panel that are in the panel as it enters into the tenth wave in 2000. The original sample in 1989 was 1713 households, so attrition over 10 waves and 12 years has totaled 81 percent. However, it should also be noted that 34 percent of the attrition occurred over the first three waves and four years, so only a 47 percent of attrition has occurred over seven more waves and eight more years.

The second major disadvantage of a panel is response bias *[41]*. Response bias arises from two principal causes: attrition and conditioning. Attrition introduces bias when it does not occur randomly across the panel members. Experience seems to indicate that usually it is a non-random process. This can have serious impacts in introducing biases into the results that the preceding calculations on sampling errors do not take into account. For example, suppose that a panel begins with 500 households. Suppose that 100 of these households leave the panel between wave 1 and wave 2. Suppose that, of the remaining 400 households who are surveyed in both the first and second wave, 15 non-car-owning households acquire a car, while the remaining 25 non-car-owning households remain without a car. One might conclude from this that 37.5 percent of non-car-owning households have switched to car-owning. However, suppose that there are 10 non-car-owning households in the 100 households that were in wave 1 and not in wave 2. At the extremes, either all 10 of these might have become car-owning households, or none of them may have changed. In the first case, this would lead to a true percentage of 50 percent shifting from non-car-owning to car-owning, while the latter case would lead to the true percentage of 30 percent. This range of change is much wider than the sampling normally expected. Without knowing what the panel drop-outs actually did, one is left with a potential for serious bias, particularly if attrition is related to measures of interest to the survey. (This is, of course, similar to the problem of nonresponse in a cross-sectional survey, particularly where the reason for a household not responding is related to the purposes of the survey. As in that case, there are no good solutions to the bias problem.)

It is possible to build a model of the attrition process, as noted by Horowitz, but this also seriously complicates the analysis and still leaves the probability that some amount of bias is not dealt with *[41]*, *[42]*. A number of models have been attempted to provide these corrections, e.g., Meurs and Ridder *[43]*, Pendyala and Kitamura *[44]*, and Brownstone and Chu *[42]*. However, as noted by Horowitz, the conclusion of Brownstone and Chu should be emphasized, that "...no *ex post* econometric technique can substitute for minimizing attrition in the first place..."*[41], [42]*. These aspects of bias are present in any panel in which non-random attrition takes place and would be equally likely to occur in panels for medical and other purposes as for those in transportation.

It is also important to note that initial nonresponse to the panel survey is perhaps even more important than in a cross-sectional survey, again particularly when the reasons for nonresponse are related to the goals of the survey. For example, if those who travel very little are more likely not to respond (on the assumption that their travel is too little to be of importance to the survey) or if those who travel a great deal are more likely not to respond (because of the large amount of work involved in completing the diary), then the panel starts out as a biased panel. Attrition is now likely to exacerbate the situation. However, correcting for initial nonresponse generally cannot be done with some type of model because of the lack of information on nonrespondents.

The second source of bias in panels is due to conditioning. This is either a behavior change occurring in members of the panel because of their membership in the panel or is a behavior change made by panel members as a result of the tasks involved in being a part of the panel. For example, being questioned once a year about commuting habits may sensitize panel members to issues relating to commuting, possibly with the effect of making it more likely that they will choose socially-desirable modes for commuting such as carpools, vanpools, or transit. On the other hand, the task in which panel households are involved, in completing one or more days of diary travel, may lead household members to minimize travel on diary days in order to reduce the amount of work involved in responding to the survey. This may be a temporary and short-lived change in behavior, but it is nevertheless one that could lead to significant error in the conclusions drawn from the survey.

Probably the principal conclusion to be drawn about conditioning biases is that their existence restricts the useful life of a panel and may also, in some circumstances, require special considerations in the design of a panel. For example, the California Air Resources Board recently requested proposals for a study on the effectiveness of clean air strategies that rely on publicity about a potential for ozone exceedance and, on the following day, requested that commuters use transit, carpools, or other means instead of driving alone. It was noted, however, that it would be difficult to impanel commuters and ask them if their behavior changed as a result of such publicity. The panel survey itself would alert them to the expected response and would be likely to cause panel members to respond to the publicity more if they were not members of the panel. A survey design was required, instead, that would not use a potentially conditioned panel. Information is not available as to how this has been achieved.

Establishment of a Panel

The initial recruitment of a panel is usually accomplished within the recruitment for a cross-sectional household travel survey. As part of the recruitment or follow-up calls, respondents are asked if they are willing to be contacted again with a similar survey. Those responding affirmatively constitute a pool from which a panel can be drawn. It is usually necessary to determine willingness to participate further from a significantly larger sample than is needed for the panel itself in order to allow a panel to be established that has the desired demographics and other characteristics. If the panel is to be established other than within the conduct of a cross-sectional survey, then an early question in the recruitment call must determine if the household or person is willing to participate in an ongoing study. Again, it is necessary to obtain a significantly larger sample of contacts than desired for the panel.

In the LSU diary survey, the question on helping in future surveys was posed in the reminder call following the diary day. The question asked:

> "I have one last question to ask you. In about six months, we will be asking some of the households who took part in this study to help us with another step in the study. Would you be willing to help us again at that time?"

In the pilot survey, 43 out of 90 households recruited were contacted on the second reminder. Of these 43 households, 18 (42 percent) sent back completed surveys, thirty of the 43 households that were successfully contacted on the second reminder responded to the question. (It is not clear why the other 13 did not respond to the question.) Of the 30 households, 20 said yes to this question, seven indicated that they would consider it if contacted in 6 months, one household refused outright, and two households gave what was interpreted as a "soft" refusal. (A hard refusal occurred if the respondent indicated that he or she should not be contacted. If the respondent said no without qualifying the no then it was considered a soft refusal.) There were three households that responded to the survey, were contacted on the second reminder, but did not respond to the question on panel membership. There were six households that responded positively to the panel question but failed to return any survey materials. Of the 20 households that answered "yes," 14 (70 percent) returned completed diaries and could therefore be considered candidates for a panel. In fact, 90 households were successfully recruited and should have been contacted in this reminder call. Of the 47 households that were not contacted, responses were obtained from 16 (34 percent). It would appear from these statistics that being contacted on the second reminder had a small effect on the responses. However, the fact that only half of those who responded to the survey were asked about panel membership is a design flaw that was not noted at the time of the pilot survey.

In the main survey, a similar result occurred. In this case, 128 households out of the 275 recruited households were contacted successfully for the second reminder. Of these, 34 households completed surveys, representing 27 percent of those contacted in the second reminder (compared to over 40 percent in the pilot survey). Another 36 households who were not contacted in the second reminder completed and returned diaries. This represents just under 25 percent, showing that the second reminder appears to have had little effect on the overall response rate (even less so than in the pilot survey). Of the 128 households contacted in the second reminder, there is again an unexplained 60 households that never answered the question on panel membership. Of the 68 that did answer, 29 indicated they would be willing to participate, and 33 indicated a maybe (contact me again at the time of the survey). There were no hard refusals, and there were six soft refusals. What is notable here is the much larger number of possible and the smaller proportion of outright affirmatives. Of the 29 who indicated "yes" to the panel, only 16 (55 percent) sent back completed diaries. Similarly, 16 (48 percent) of the 33 who indicated "maybe" returned completed diaries. Only two households that were contacted on this reminder call but did not answer the question on panel membership returned surveys. None of those who indicated "no" to the panel returned surveys, as was also the case in the pilot survey.

From this exercise, combining the pilot and main surveys, there are 47 households that indicated a definite or possible affirmative to being in a panel. There are a further 57 households that were never asked the question on panel membership that did return completed surveys in either the main or pilot survey. This indicates a maximum potential panel of 104 households. The main change that should be made is in asking the question about panel participation. There is good reason to keep this question until after diaries have been completed, because the respondents now know what is involved in the survey. It could be assumed that affirmative responses would have been much higher if the question had been asked in the recruitment call at which time respondents did not know what tasks would be involved. Even when diaries have been sent out and households had supposedly completed the diaries, those that indicated willingness to participate in a panel did not always complete and return diaries in this survey. Hence, between 55 and 70 percent of those indicating willingness were potential panel members on the basis of completing the initial survey.

Incentives. It has been found that response rates to surveys increase markedly when incentives are offered *[14], [45], [46]*. There are also concerns that incentives may introduce an additional bias in the responses as noted by Richardson et al. and others *[47]*. However, the prevailing practice seems to be to offer incentives, on the assumption that the added response outweighs the concerns with potential biases. In fact, if, as has been suggested by some, the biases may tend to draw responses from low income households, this is not necessarily a problem because most household travel surveys have under-representation of this group of households *[47, p. 92]*.

Problems may arise when offering incentives with respect to permitted uses of the funds for the survey. In some instances, there are state or local prohibitions on the use of money for incentives of this type. However, this can often be circumvented through some offering of something other than money. In the 1990 Activity Diary survey in Boston, it was not legal for the state to send money to survey participants as an incentive *[11]*. Instead, tickets for the state lottery were sent, each ticket being worth $1. There was one ticket sent for each travel diary to be completed. This was at least partially responsible for raising the pilot survey response rate of 26 percent (without an incentive) to the main survey response rate of 42 percent (with the incentive provided). Other changes in the instrument and the survey procedures probably also contributed to the increase, since the prevailing opinion seems to be that incentives can increase response rates on the order of 4 to 10 percent.

Generally, the most effective incentives are relatively small. Experiments with different levels of incentive have concluded, currently, that an incentive of $1 to $2 per diary is sufficient in both cross-sectional and panel surveys. Higher incentives, such as $10 per household, have not been found to be as effective. However, for the record, it should be noted that the Lexington experiment on use of GPS devices used an incentive of $50 per household, partly to persuade households to spend the effort of mounting the GPS units and partly to ensure that the devices were returned. Experimentation has also determined that providing incentives prior to the survey task is more effective than those promised to be sent after completion of the task. This also avoids the difficulty of establishing when the survey task has been completed well enough for an incentive to be sent to the household. Incentives sent with the diaries are designed to create some level of guilt on the part of the recipient that, if he or she accepts the incentive and does nothing about completing the survey, then this is socially unacceptable behavior.

In this research, funds did not permit offering incentives to participating households. It can be expected that incentives would have increased response rates by about 8 percent and that recruitment to a panel would have been more productive. The effects of a larger incentive on a panel or repetitive incentives on a panel have not been investigated in the transportation survey arena.

Maintenance of a Panel

As noted earlier, one of the major problems with panels is attrition. Attrition occurs for a variety of reasons, many of which are unavoidable. Households leave the geographic area of interest, are dissolved through death or divorce, or move and become untraceable even though they may have moved within the geographic area of interest. Attrition also occurs as households become tired of participating, lose interest, or, for other reasons, do not respond. Panel maintenance has to do with two things: avoidance of attrition that can be avoided and replacement of those panel members who are lost to attrition.

Avoidance of Some Attrition. The principal strategy here is to attempt to maintain interest in the panel and contact with panel members. It is generally recommended that panel members should be contacted at least every six months. This serves several purposes. Forwarding addresses are usually kept on file by the Post Office for six months and mail is forwarded to the new address in that period. Similarly, telephone number changes are usually provided for three to six months. Thus, new phone numbers or new addresses can usually be traced if contact is made every six months. Second, such contact can serve to remind panel members of the existence of the panel and to keep their interest in participating. Additionally, it has been found useful in a number of cases to send out a periodic newsletter to provide information on what is being done with the panel data and associated issues and to send birthday and Christmas cards to panel members.

Hence, it is important to set up a program of contacts for a panel in order to reduce attrition. This contact should probably take the form of a biannual newsletter, possibly sent out in March and September, birthday and Christmas greetings sent at the appropriate times for each member of a panel household, and pre-notification letters prior to initiation of a new wave of surveys. It is advisable to have all mailing items printed with directions for the Post Office to not forward mail but to notify of address changes.

Rules also must be established for circumstances where the household moves or changes significantly. This will usually depend on the purpose and scope of the survey. For example, if a panel is set up for the Baton Rouge metropolitan area, any household that moves out of the area to another metropolitan area in Louisiana is automatically out of the panel. However, if the panel were set up for the entire State of Louisiana, such a household move would not disqualify the household from continuing in the panel. Additions to a household through marriage, birth, adoption, or moving in of close family members (e.g., parents, or a child who has been away at school) will not normally cause a disqualification of the household from the panel. Separation or divorce can be treated in two possible ways. Assuming that one or both parties remain in the geographic area, each may be continued as a panel member. Alternatively, since this produces households that are probably in a different socioeconomic grouping than before, the resulting households may be dropped from the panel and a new household substituted. If neither household remains in the region of the survey, then the household should be replaced.

Panel Replacement. Replacement in panel surveys, as has been discussed previously, is necessary to maximize the value of the panel. The method of replacement, however, is subject to a number of different considerations. Replacements can be made in several different ways *[44]*:

- Pure random sampling from the present population of households;
- Endogenous sampling, in which the same sampling strategy is used as in the original sampling and the attempt is to replace each lost household with one that matches as closely as possible to the lost household and was also present in the region at the time of the original panel establishment;

- Exogenous attrition-compensating sampling, in which exogenous variables are used to identify subgroups of households with higher attrition rates that are over-sampled in the refreshment sampling;

- Endogenous attrition-compensating sampling, in which endogenously-defined subgroups of households found to have higher attrition rates are over-sampled in the refreshment sample; and

- Sampling to capture population dynamics, in which new immigrants to the region and other groups that are represented differently in the current population may be included in the refreshment sample.

The specific method used to replace panel members lost to attrition depends significantly on the purpose of the panel. For example, in a panel that is designed to track specific behaviors, such as transit riding or carpool use, endogenous sampling would be the most appropriate method to use. Where the purpose is to provide a picture of the dynamics of change within the original panel and also in the current population, refreshment would most likely follow either the first method of pure random sampling or the last method of sampling for population dynamics. The attrition-compensating samples could be used in either circumstance, if certain subgroups of the population were found to be more likely to leave the panel, in order to provide for better comparability over successive waves with less need for subsequent panel refreshment.

In the event that the original panel was sampled to meet a particular distribution of population subgroups, the replacement sample would normally be drawn to return to approximately the same distribution for each successive wave. For example, suppose the panel is sampled to provide a distributed sample by household size and vehicle availability, for a 350 household panel survey as shown in table 1. Suppose that in recruitment for the second wave, the distribution becomes that shown in table 2.

Table 1
Hypothetical distribution for a panel survey

Vehicles available	Household size				
	1	2	3	4	5+
0	10	10	10	10	10
1	25	25	25	25	25
2	-	25	25	25	25
3+	-	-	25	25	25

Table 2
Second wave distribution for a panel survey

Vehicles available	Household size				
	1	2	3	4	5+
0	8	6	7	9	5
1	20	22	18	19	26
2	-	24	21	24	18
3+	-	-	17	22	24

In this case, the refreshment sample would be drawn to return each of the cells of the matrix in table 2 to the values in table 1. The exception would be that the 5-plus household size with one vehicle has exceeded the original, presumably as a result of shifts into that cell from elsewhere in the matrix and, therefore, would remain unchanged. This strategy would be correct because the original sample was not distributed proportionally to the population. There is, thus no reason to sample proportionally in the refreshment sample. However, two choices still remain. The refreshment sample can be restricted to households that were present in the region at the time that the original panel was drawn, or they could be drawn from all households currently in the region. The former option would keep close to retaining the characteristics of the original panel and would not reflect changes that have taken place in the population in the time elapsed between the panel waves, whereas the latter option would capture population changes that have occurred over the period between panel waves.

Uses of a Panel

The principal uses of a panel have been well summarized by Paaswell *[48, p. 10-11]*, viz:

> *A growing body of literature has shown how transportation panels can be used to anticipate, evaluate, and forecast transportation changes. Panels provide more complete information on the behavioral aspects of transportation planning than do other means. Panels can be used to explain how or why individuals or groups make transportation choices, and can also help planners forecast responses to specific policy initiatives. Because panels can be broadly structured and can respond to change over time, they can represent the extensive set of transportation customers and their evolving concerns.*

Furthermore, panels offer the potential of updating expensive cross-sectional surveys by tracking the dynamic changes occurring within the population. For example, if vehicle miles of travel (VMT) are increasing in the population (absent significant changes in vehicle ownership, licensing of drivers, and other similar changes), then the identification of the increases in the panel VMT can be converted to a factor in applying to VMT estimates in a cross-sectional survey completed at the same time as the first wave of the panel. This use of a panel may provide the possibility of avoiding large, costly cross-sectional surveys on ten-year intervals, and it may allow, instead, the continual updating of a large cross-sectional survey data base on an annual basis. This is similar to the procedure currently used by the U.S. Bureau of the Census to update population, employment, and some population demographics on an ongoing basis between the decennial censuses.

A further use of panels is to provide more extensive measurements of behavior and preferences than is normally possible in a cross-sectional survey. Generally, any survey is limited in length to what respondents will tolerate as the maximum size of the survey-completions task. Currently, cross-sectional household travel surveys are limited to collecting a selection of household and personal characteristics, vehicle characteristics, and descriptions of all travel completed in a 24 or 48-hour period by members of each sampled household. In a panel, most of these data do not need to be collected repetitively over time;

indication of change or confirmation of no change since the last wave of the panel is usually all that is necessary. This means that subsequent waves of a panel can be used to collect additional data that are not normally within the tolerance of the responding public. For example, a subsequent wave might collect data about one week of travel by household members, may request completion of a stated response survey, or may request data on what alternatives were considered for each of the trips conducted by the household in a day.

From the cost-effectiveness perspective of data collection in Louisiana, the use of the panel to update cross-sectional surveys appears to be, by far, the most compelling argument. An investigation of a panel's potential use is explored in the next section of this report.

ANALYSIS

Assembly of Existing Data Sources

Introduction

Existing data were used for several purposes in this study. First, they were used to test transferability by observing the similarity of data among data sets. Since differences may sometimes be due to variations in data definition, data collection procedure, or survey protocol, consistency among tested data sets was important. Second, existing data were used to identify segments that are as homogeneous as possible in travel behavior. Third, existing data the in Baton Rouge metropolitan area were used as a reference to which other data areas could be compared.

NPTS 1995 Data

The National Personal Transportation Survey (NPTS) of 1995 is a large, comprehensive data set of travel data. It consists of 42,033 households of which approximately 25,000 are randomly selected from across the nation. The remainder are additional households surveyed within metropolitan areas who purchased add-in surveys from the agency collecting the NPTS data. The data include detailed information on the household and its members, their travel, and the conditions under which the household functions. Because the original sample was randomly selected from across the nation and there were an estimated 98,990,000 households in the U.S. in 1995, the sampling rate in all but those areas with add-in surveys was approximately 25 out of every 100,000 households. Thus, samples within individual areas are small and generally not suitable for use within metropolitan studies. On the other hand, the large number of observations and the richness of the NPTS data set make it an ideal set to identify the impact that characteristics of travelers and their environment have on travel.

The NPTS 95 data have household and person weights that permit expansion of the sample to represent the population and to adjust for non-representativeness in the sample. Non-representativeness results from non-representative sampling, non-response, and different response rates among respondents. Since the NPTS 95 data sampling frame was a list of household telephone numbers in the U.S., household weights were adjusted to accommodate the different sampling rates resulting from households with more than one telephone number.

Household weights were subsequently adjusted to sum to the estimated number of households in the U.S. in 1995 (98,990,000). This incorporated the expansion factor into the household weight but also adjusted the rates of those included in the sample for those omitted because they had no telephone. The household weights were further adjusted to match marginal controls such as equal weight totals for each month of the year, estimated numbers of households in subregions in the country, ethnic mixes in the country, etc. Person weights were also established in the NPTS 95 data. Person weights differ from household weights because they accommodate non-response within the household, when some persons in a household respond and others do not. Person weights were first adjusted to total 241,675,000, which was the estimated population in the U.S. in 1995, and then were adjusted to population totals in similar subdivisions of the population as those used in adjusting the household weights. The detail of how weights were established for both households and persons in the NPTS 95 data can be found in the documentation distributed with the NPTS 95 data. An excerpt is included in Appendix H.

It is important to note that household weights and person weights are different and cannot be used interchangeably. Household weights must be used to weigh household variables such as household size, household income, number of vehicles in the household, and number of workers in the household. On the other hand, person weights accommodate the fact that persons of a particular age, gender, worker status, or other characteristic may be less likely to respond to the survey than others. Thus, individual household members are weighted differently to make the sample representative of the population. Because trips are reported at the person level, person weights are also applicable in weighting trip characteristics in the data.

Since the focus of this research is on data for metropolitan transportation planning, which usually models weekday travel, the NPTS 95 data was screened to include only urban households who were surveyed on a weekday. This reduced the data to 19,764 useable households.

Household information in the NPTS 95 data includes household size, household income, number of drivers, vehicle ownership, stage-in-the-family-life-cycle of the household, household location, home ownership, type of dwelling, number of workers, and other household information. Person information includes age, gender, worker status, relationship within the household, race, education level, and information on distance to work, mode, and transit availability for the journey to work. The travel information from NPTS 95 used in this study included starting times, ending times, distance, mode, cost, driver, waiting and walking times for transit, and trip purpose for all trips made during a 24-hour period by each respondent 5 years of age and older. Conditions in which each household operates are described in terms of population density, median house price, median household income, and percent home ownership in the census tract or block in which the household is located. In addition, the presence or absence of different forms of transit, distance to transit, and qualitative information on the standard of transit for travel from home are also included in the data.

Baton Rouge Personal Transportation Survey

The Baton Rouge Personal Transportation Survey (BRPTS), conducted in 1997, is identical in format to the NPTS 95 data but the sample was drawn entirely from the Baton Rouge metropolitan area. The data file consists of 1,395 households of which 679 reside in an urban environment and were surveyed on a weekday.

In this study, the BRPTS data was used as control data against data transferred into the Baton Rouge area from elsewhere. Like the NPTS 95 data, the BRPTS data has weighting factors that compensate for non-response and biases. Subsequently, the data can be made representative of the population through application of the weighting factors applied appropriately at household and person level. Because of the limited size of BRPTS data, sampling error is an issue when the sample is segmented.

PUMS 1990 Data

The PUMS 1990 data are useful for the disaggregate information that can be acquired from them. Since they are census data from the "long form," they also contain information on the journey to work. In this study, the PUMS 1990 data were used to determine the proportion of the population in different household size/vehicle ownership categories in Baton Rouge and to determine the journey-to-work characteristics of households within these categories.

When census data are used to estimate the travel characteristics of the population, adjustments must be applied to the data in order to get accurate estimates. Appropriate adjustment factors have been calculated to account for absenteeism from work on the survey day, reporting of "normal" mode rather than actual mode, multiple jobs, and trip chaining *[49]*. These adjustment factors were applied to the PUMS 1990 data in this study.

NCTCOG Data

The North Central Texas Council of Government's data (NCTCOG) was selected for analysis in this study because of its recent vintage (1996) care with which it was prepared and the proximity of Dallas-Ft. Worth to Louisiana. The data were selected as a representative primary data set that could demonstrate the transfer of data.

The data contain a variable which indicates whether all in the household completed the survey. Only those households in which all members completed the survey were included in the analysis in this study. The final data file consisted of 3,996 households which recorded 38,401 daily trips.

Characteristics of a Suitable Panel

Identification of Homogeneous Travel Groups

The purpose of establishing homogeneous travel groups was to identify segments of the population that displayed minimum variance in their travel behavior. Sampling within such segmented groups would place the lowest demands on sample size within each group and would allow for more accurate estimation of travel characteristics with a given total sample size. The panel used for updating purposes was sampled according to the established segmentation scheme.

Two methods of identifying homogeneous travel groups were employed in this study. First, Analysis of Variance was used. Analysis of Variance identifies and determines the relative contribution of individual variables to the variance observed in a single dependent variable (Univariate Analysis of Variance) or simultaneously in multiple dependent variables (Multivariate Analysis of Variance). In this study, only Univariate Analysis of Variance was used because seeking homogeneity among multiple variables can limit the homogeneity obtained on any one variable, and trip rates were considered the most important travel characteristic being investigated. The second form of investigation used was tree-building classification systems. These systems employ algorithms that search the data for the best way that it can be segmented to reduce the variance in a single variable. The procedures conduct an exhaustive search of all possible divisions of data and present the results in the form of an inverted tree starting with all observations in the trunk, branching into ever-increasing subdivisions of the data as the tree attains more branches. The final segments are described by the end branches. A convenient source of these tree-building classification systems is an add-on to the Statistical Package for Social Sciences (SPSS) called Answertree. Answertree has four alternative segmentation procedures that an analyst may choose from. All perform a tree-building activity, although some permit only dichotomous divisions of data at each level and some limit the types of variables that can be considered. The Classification and Regression Tree (C&RT) method permits only binary splitting of data but is capable of handling continuous, ordinal, and nominal variables *[50]*. The C&RT method was used in this analysis.

The major operational difference between Analysis of Variance (ANOVA) and the tree-building systems is that in ANOVA the analyst must submit the variables on which segmentation is to be tested. If the variables are continuous, it must prescribe intervals by which the continuous variables can be altered into a manageable number of discrete values. Thus, ANOVA requires participation and judgement from the analyst in suggesting segmentation schemes to be tested. Only those segmentation schemes submitted are evaluated, and there is no guarantee that the optimum segmentation scheme has been included among those evaluated. In contrast, the tree-building classification systems seek out the variables and the boundaries that produce the most distinct segments. They employ an exhaustive search of all divisions and evaluate them on numerical criteria. Because the process is so numerically-based, analyst participation or intervention in interpretation, evaluation, and manipulation of the final decision tree (pruning) is required.

The NPTS 95 data of all urban residents surveyed on a weekday was used for the identification of homogeneous travel groups. The dependent variable was trips per household per day for all trip purposes combined and for each of the trip purposes of home-based work (HBW), home-based other (HBO), and non-home based (NHB) travel. Using ANOVA, a large variety of variables describing the household and its environment were tested. Through a process of trial and error, the most significant factors were found to be household size, number of vehicles in the household, and family-life-cycle. In the NPTS 1995 data, family-life-cycle was 10 discrete categories of households based on the number and work status of adults and the number and age of children in the household. The results of an ANOVA run on urban households surveyed on a weekday in the NPTS 1995 survey are shown in table 3.

Table 3

ANOVA of trips of all purposes per household per day

Tests of Between-Subjects Effects

Dependent Variable: Count of household trips (0-72)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	480921.335[a]	264	1821.672	51.112	.000
Intercept	20863.917	1	20863.917	585.398	.000
LIF_CYC	1111.031	9	123.448	3.464	.000
HHVEHCNT	3033.671	10	303.367	8.512	.000
HHSIZE	7156.514	9	795.168	22.311	.000
LIF_CYC * HHVEHCNT	3457.707	45	76.838	2.156	.000
LIF_CYC * HHSIZE	5253.989	35	150.114	4.212	.000
HHVEHCNT * HHSIZE	6898.408	47	146.775	4.118	.000
LIF_CYC * HHVEHCNT * HHSIZE	7472.815	97	77.039	2.162	.000
Error	694954.951	19499	35.641		
Total	3045421.000	19764			
Corrected Total	1175876.286	19763			

a. R Squared = .409 (Adjusted R Squared = .401)

The analysis shown in table 3 divides the data into 10 life-cycles, 11 vehicle ownerships (0-10 vehicles per household), and 10 household size groups (1-10 persons per household), making a total of 1,100 subdivisions of the data. This segmentation captures approximately 41 percent of the variation in the data (R^2=0.409) with household size being the most influential variable, followed by vehicle count and family life-cycle, respectively. The results in table 3 show, by the significance of the products of the variables in the table, significant second and third-order interaction among the variables. Interaction is the joint impact of two or more variables on a dependent variable distinct from the impact of the individual variables on the dependent variable.

Repeating the analysis with trips distinguished by trip purpose produces the results shown in tables 4-6. For home-based-work (HBW) trips, number of workers is the main descriptor of the number of trips. For home-based-other (HBO) and non-home-based (NHB) trips, a similar grouping scheme to that used for all trip purposes was found to capture the most variation in data values. However, the amount of variation in the data captured by the segmentation scheme was particularly low (14.7 percent) for NHB trips, as shown in table 6.

Table 4

ANOVA of HBW trips per household per day

Tests of Between-Subjects Effects

Dependent Variable: Count of household HBW trips(0-19)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	23546.627[a]	257	91.621	53.101	.000
Intercept	898.238	1	898.238	520.591	.000
LIF_CYC	47.414	9	5.268	3.053	.001
HHVEHCNT	96.764	10	9.676	5.608	.000
WRKCOUNT	689.893	7	98.556	57.120	.000
LIF_CYC * HHVEHCNT	110.616	52	2.127	1.233	.121
LIF_CYC * WRKCOUNT	183.090	33	5.548	3.215	.000
HHVEHCNT * WRKCOUNT	179.615	34	5.283	3.062	.000
LIF_CYC * HHVEHCNT * WRKCOUNT	340.260	107	3.180	1.843	.000
Error	33656.055	19506	1.725		
Total	105226.000	19764			
Corrected Total	57202.682	19763			

a. R Squared = .412 (Adjusted R Squared = .404)

It may be expected that more than the 41 percent of the variation in HBW trip rates should be captured by the segmentation scheme since each worker normally makes two work trips per day. Thus, there should be a close correlation between the number of workers and number of work trips. However, one of the problems of current trip purpose classification procedures is that multi-purpose trips often obscure the prime purpose of a trip. For example, a person dropping a child off at school on their way to work and going by a grocery store on their way home from work will record no HBW trips but two HBO and two NHB trips. This anomaly that a worker may go to work and not register a work trip may be overcome by using the concept of trip chains rather than trips. A trip chain is the journey between home and work, irrespective of the number of stops in between. Using chains rather than trips to measure work trips and employing the same segmentation scheme as shown in table 4 increases the R^2 value to 0.49.

Table 5
ANOVA of HBO trips per household per day

Tests of Between-Subjects Effects

Dependent Variable: Count of household HBO trips (0-50)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	180778.3ª	264	684.766	44.841	.000
Intercept	6287.642	1	6287.64	411.735	.000
LIF_CYC	785.863	9	87.318	5.718	.000
HHVEHCNT	384.024	10	38.402	2.515	.005
HHSIZE	3066.488	9	340.721	22.312	.000
LIF_CYC * HHVEHCNT	1738.947	45	38.643	2.530	.000
LIF_CYC * HHSIZE	2393.663	35	68.390	4.478	.000
HHVEHCNT * HHSIZE	2940.134	47	62.556	4.096	.000
LIF_CYC * HHVEHCNT * HHSIZE	3691.548	97	38.055	2.492	.000
Error	297770.9	19499	15.271		
Total	969923.0	19764			
Corrected Total	478549.2	19763			

a. R Squared = .378 (Adjusted R Squared = .369)

Table 6
ANOVA of NHB trips per household per day

Tests of Between-Subjects Effects

Dependent Variable: Count of household NHB trips (0-41)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	39298.300ª	264	148.857	12.771	.000
Intercept	1433.055	1	1433.055	122.946	.000
HHSIZE	520.490	9	57.832	4.962	.000
HHVEHCNT	706.841	10	70.684	6.064	.000
LIF_CYC	217.700	9	24.189	2.075	.028
HHSIZE * HHVEHCNT	1120.268	47	23.835	2.045	.000
HHSIZE * LIF_CYC	919.804	35	26.280	2.255	.000
HHVEHCNT * LIF_CYC	704.342	45	15.652	1.343	.062
HHSIZE * HHVEHCNT * LIF_CYC	1956.645	97	20.172	1.731	.000
Error	227279.8	19499	11.656		
Total	448139.0	19764			
Corrected Total	266578.1	19763			

a. R Squared = .147 (Adjusted R Squared = .136)

Using the Classification and Regression Tree (C&RT) segmentation method produces similar results to those achieved with ANOVA. For the all-trip-purpose trip rates, the C&RT procedure identified household size, family life cycle, and number of vehicles in the household as the most influential factors in identifying homogeneous segments in household trip rates. The segments, their mean trip rates (in trips/household/day), and the standard deviation of the mean trip rate are shown in table 7.

Table 7
Segments of homogeneous all purpose household trip rates

<table>
<tr><th>House-
hold size</th><th>Family life cycle</th><th># of
vehs.</th><th>Mean
trip rate</th><th>Std.
dev.</th></tr>
<tr><td rowspan="4">≤2</td><td rowspan="2">Single adult, no children
Single adult, retired, no children
Single adult, youngest child age 0-5</td><td>0</td><td>3.17</td><td>2.80</td></tr>
<tr><td>≥1</td><td>4.73</td><td>2.97</td></tr>
<tr><td rowspan="2">Single adult, youngest child age 6-15
Single adult, youngest child age 15-21
Two or more adults, no children
Two or more adults, retired, no children</td><td>0</td><td>5.50</td><td>4.24</td></tr>
<tr><td>≥1</td><td>8.09</td><td>4.83</td></tr>
<tr><td rowspan="2">3-4</td><td>Two or more adults, youngest child age 0-5
Two or more adults, no children
Single adult, youngest child age 16-21
Single adult, youngest child age 0-5
Single adult, no children</td><td>≥0</td><td>10.90</td><td>6.63</td></tr>
<tr><td>Single adult, youngest child age 6-15
Two or more adults, youngest child age 16-21
Two or more adults, youngest child age 61-5</td><td>≥0</td><td>15.12</td><td>8.14</td></tr>
<tr><td rowspan="2">>4</td><td>Single adult, no children
Single adult, youngest child age 0-5
Single adult, youngest child age 6-15
Single adult, youngest child age 16-21
Two or more adults, youngest child age 0-5
Two or more adults, youngest child age 16-21
Two or more adults, retired, no children</td><td>≥0</td><td>16.61</td><td>10.29</td></tr>
<tr><td>Two or more adults, no children
Two or more adults, youngest child age 6-15</td><td>≥0</td><td>22.07</td><td>11.37</td></tr>
</table>

vehs. = vehicles
std. dev. = standard deviation

Segmentation methods such as C&RT have a statistic titled "risk factor," which with a continuous target variable such as the trip rate used in this analysis, is the variance of the observations about the segment means. For the all-purpose household trip rates shown in table 7, the risk factor was 38.5. Considering that the variance of the all-purpose household trip rates without segmentation is 59.5, the segmentation process has captured (1-38.5/59.5) or 35.3 percent of the variance in the data. This is a similar value to that obtained with ANOVA, although the number of categories in the ANOVA analysis was considerably greater because a category was established for each combination of the values of household size, number of vehicles, and family life cycle (i.e. 11x10x10 or 1,100 categories). Thus, the segmentation process was able to capture, in the eight segments shown in table 7, a large proportion of the variation captured in a more comprehensive segmentation scheme with the same variables in ANOVA.

Applying the C&RT segmentation process to the HBO and HBW trip rates in the sample produced the segmentation scheme shown in tables 8 and 9. For the HBO trip rates, the variance was 24.2 unsegmented and 16.1 segmented. Thus, the segmentation acounted for approximately 33 percent of the variation in the data. For HBW trip rates, the variance was 2.9 unsegmented and 1.8 segmented, accounting for approximately 38 percent of the variation in the data.

Table 8
Segments of homogeneous home-based-other household trip rates

<table>
<tr><th>House-
hold size</th><th>Family life cycle</th><th>Mean trip rate</th><th>Standard deviation</th></tr>
<tr><td>≤2</td><td rowspan="2">Single adult, youngest child age 6-15
Two or more adults, youngest child age 6-15
Two or more adults, youngest child age 16-21
Two or more adults, retired, no children</td><td>4.95</td><td>3.57</td></tr>
<tr><td>3</td><td>6.70</td><td>4.43</td></tr>
<tr><td rowspan="2">≤3</td><td>Single adult, no children
Single adult, retired, no children</td><td>2.05</td><td>1.96</td></tr>
<tr><td>Single adult, youngest child age 0-5
Single adult, youngest child age 16-21
Two or more adults, youngest child age 0-5
Two or more adults, no children</td><td>3.44</td><td>3.03</td></tr>
<tr><td rowspan="2">4</td><td>Single adult, no children
Single adult, youngest child age 16-21
Two or more adults, youngest child age 0-5
Two or more adults, youngest child age 16-21
Two or more adults, no children
Two or more adults, retired, no children</td><td>6.35</td><td>4.62</td></tr>
<tr><td>Single adult, youngest child age 6-15
Two or more adults, youngest child age 6-15</td><td>9.67</td><td>5.86</td></tr>
<tr><td rowspan="2">>4</td><td>Single adult, no children
Single adult, youngest child age 16-21
Two or more adults, youngest child age 0-5
Two or more adults, youngest child age 16-21
Two or more adults, no children
Two or more adults, retired, no children</td><td>9.48</td><td>6.83</td></tr>
<tr><td>Single adult, youngest child age 6-15
Two or more adults, youngest child age 6-15</td><td>13.16</td><td>7.70</td></tr>
</table>

Table 9
Segments of homogeneous home-based-work household trip rates

# of workers	Family life cycle	Mean trip rate	Standard deviation
1	Single adult, youngest child age 0-5 Single adult, youngest child age 6-15 Single adult, youngest child age 16-21	0.94	1.08
	Single adult, no children Two or more adults, youngest child age 0-5 Two or more adults, youngest child age 6-15 Two or more adults, youngest child age 16-21 Two or more adults, no children Two or more adults, retired, no children	1.31	1.13
2	Single adult, youngest child age 0-5 Single adult, youngest child age 16-21 Single adult, youngest child age 16-21 Two or more adults, youngest child age 0-5	1.96	1.47
	Two or more adults, youngest child age 6-15 Two or more adults, youngest child age 16-21 Two or more adults, no children Two or more adults, retired, no children	2.39	1.63
3	All family life cycles	3.44	2.10
>3	All family life cycles	4.61	2.62

From the ANOVA and C&RT results it is clear that the most significant factors affecting trip rates are household size, family life cycle, number of vehicles, and number of workers. Thus, a segmentation scheme was chosen from among these variables to serve as the basis for stratifying the sample for the panel survey conducted in this study. The segmentation scheme must consist of a limited number of segments so that a reasonable number of observations can be included in each segment without making the total sample too large. In addition, a common segmentation scheme must be chosen for all trip purposes. After review of the results, the population proportions in each segment, and consideration of ease of use, the segmentation scheme shown in table 10 was chosen. The segmentation scheme shown in table 10 captured 31 percent of the variation in the all-purpose trip rate in the data. The table shows the segment number followed by the percentage of the NPTS 95 sample in each segment in parentheses.

Table 10
Chosen segmentation scheme

<table>
<tr><th rowspan="2">Number of vehicles</th><th colspan="5">Household size</th></tr>
<tr><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th></tr>
<tr><td>0</td><td colspan="5">1 (10.7%)</td></tr>
<tr><td>1</td><td rowspan="3">2 (17.0%)</td><td>3 (10.9%)</td><td colspan="3">5 (8.1%)</td></tr>
<tr><td>2</td><td rowspan="2">4 (22.0%)</td><td>6 (7.4%)</td><td colspan="2">8 (12.6%)</td></tr>
<tr><td>3+</td><td>7 (4.6%)</td><td colspan="2">9 (6.7%)</td></tr>
</table>

Simulated panel results

To observe how the travel characteristics estimated from a panel are likely to relate to the travel characteristics of the population it is representing, a simulated panel was created from the NPTS 95 data for Baton Rouge. The resulting travel characteristics were compared with those of the Baton Rouge Personal Transportation Survey of 1997 (BRPTS 97). Since the panel was expected to consist of a total between 200 and 500 households, a simulated panel of 50 households in each of the 9 segments was established to test the ability of a panel to reflect travel characteristics accurately. The panel was established for urban residents of Baton Rouge and weekday travel characteristics were observed. The panel was established by randomly sampling from the NPTS 95 data among households living in urban areas with a population between 500,000 and 1,000,000 and were surveyed on a weekday.

Estimates of overall travel behavior from the simulated panel were made by multiplying average values in each segment of the simulated panel with the proportion of the population in that segment. In this particular application, where the panel was simulated for Baton Rouge, segment proportions were estimated from the NPTS 95 data for MSAs with populations between 500,000 and 1,000,000. These segment proportions were estimated using household trip weights to represent the proportions in the population.

Trip rates, mode shares, and trip length distributions from the simulated panel were compared with observed values from BRPTS surveys. Because both data sets permit the estimation of weighted and unweighted travel characteristics, both weighted and unweighted values have been estimated in the comparisons which follow. However, in a real panel consisting of locally-collected data, it may often be difficult to determine appropriate weights accurately.

Trip rates estimated from the simulated panel and those observed in the BRPTS 97 data are shown in table 11. A statistical test of the similarity of the trip rates from the two data sources can be made using a z-test of the difference between two means *[51]*:

$$z = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}} \qquad \textbf{(11)}$$

where,

$\overline{x}_n$ = mean travel characteristic value from source n, and

σ_n^2 = variance of travel characteristic value from source n

Using the above equation, none of the trip rates in table 11 were found to be significantly different from each other at the 0.05 level of significance. The unweighted trip rates from the two data sources appear more similar than the weighted trip rates, although it must be remembered that the BRPTS values are only sample estimates themselves and, therefore, are subject to error. For example, the observed BRPTS trip rate of 1.9 home-based work trips per day is higher than would normally be expected.

Table 11
Comparison of simulated panel and BRPTS trip rates

Trip purpose	Trips/household/day			
	Simulated panel		BRPTS 95 data	
	Weighted	Unweighted	Weighted	Unweighted
All purposes	10.2	9.6	10.9	9.7
HBW	1.4	1.5	1.9	1.6
HBO	5.5	5.1	5.6	4.9
NHB	3.3	3.0	3.5	3.2

Mode shares by trip purpose are compared between the simulated panel and the BRPTS data in table 12. To test the difference between the shares, the following test statistic suitable for testing the difference between two proportions in large samples is used *[51, p. 255]*:

$$z = \frac{\frac{x_1}{n_1} - \frac{x_2}{n_2}}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} \qquad \text{with } \hat{p} = \frac{x_1 + x_2}{n_1 + n_2} \tag{12}$$

where,

x_k = number in population k choosing the mode under consideration.

n_k = number in population k.

Values that are significantly different at the 0.05 level of significance are shown with an asterisk in table 12. The simulated and observed values are very similar, but there is a general tendency for the simulated panel results to overestimate bus and walk and underestimate auto driver. This reflects Baton Rouge's tendency to be less transit and walk-oriented than other metropolitan areas of similar size.

Table 12
Comparison of simulated panel and BRPTS mode shares

Trip purpose	Mode	Mode share			
		Simulated Panel		BRPTS	
		Wtd.	Unwtd.	Wtd.	Unwtd.
HBW	Auto driver	0.82	0.82*	0.89	0.90
	Auto passenger	0.10	0.09	0.09	0.09
	Bus	0.05	0.04	0.01	0.01
	Walk	0.04	0.05	0.01	0.01
	Taxi	0.00	0.00	0.00	0.00
HBO	Auto driver	0.67	0.65	0.63	0.62
	Auto passenger	0.21	0.23	0.27	0.27
	Bus	0.06	0.06	0.05	0.05
	Walk	0.06	0.06	0.04	0.06
	Taxi	0.00	0.00	0.00	0.00
NHB	Auto driver	0.70	0.69	0.73	0.73
	Auto passenger	0.24	0.22	0.22	0.23
	Bus	0.03	0.03	0.02	0.01
	Walk	0.03	0.05	0.03	0.03
	Taxi	0.00	0.00	0.00	0.00

Trip length distributions from the simulated panel and the BRPTS data are shown in table 13. With the exception of short, home-based-work trips, which the simulated panel suggests should be higher than those measured in the BRPTS survey, the travel time distributions are very similar. Statistical tests were conducted on the differences between the simulated panel and BRPTS values using the test statistic in equation (12). None were found to be significantly different at the 0.05 level of significance.

Table 13
Comparison of simulated panel and BRPTS trip length distributions

Trip purpose	Travel time interval (mins.)	Percentage of trips in time interval			
		Simulated Panel		BRPTS	
		Wtd.	Unwtd.	Wtd.	Unwtd.
HBW	0-15	48	49	36	33
	15-30	36	37	41	44
	30-45	10	10	14	14
	45-60	2	1	5	5
	>60	4	3	4	4
HBO	0-15	59	60	60	61
	15-30	29	28	28	28
	30-45	3	5	6	6
	45-60	2	1	2	2
	>60	7	5	4	3
NHB	0-15	59	65	59	60
	15-30	27	24	29	28
	30-45	6	4	6	7
	45-60	1	1	2	2
	>60	7	5	4	4

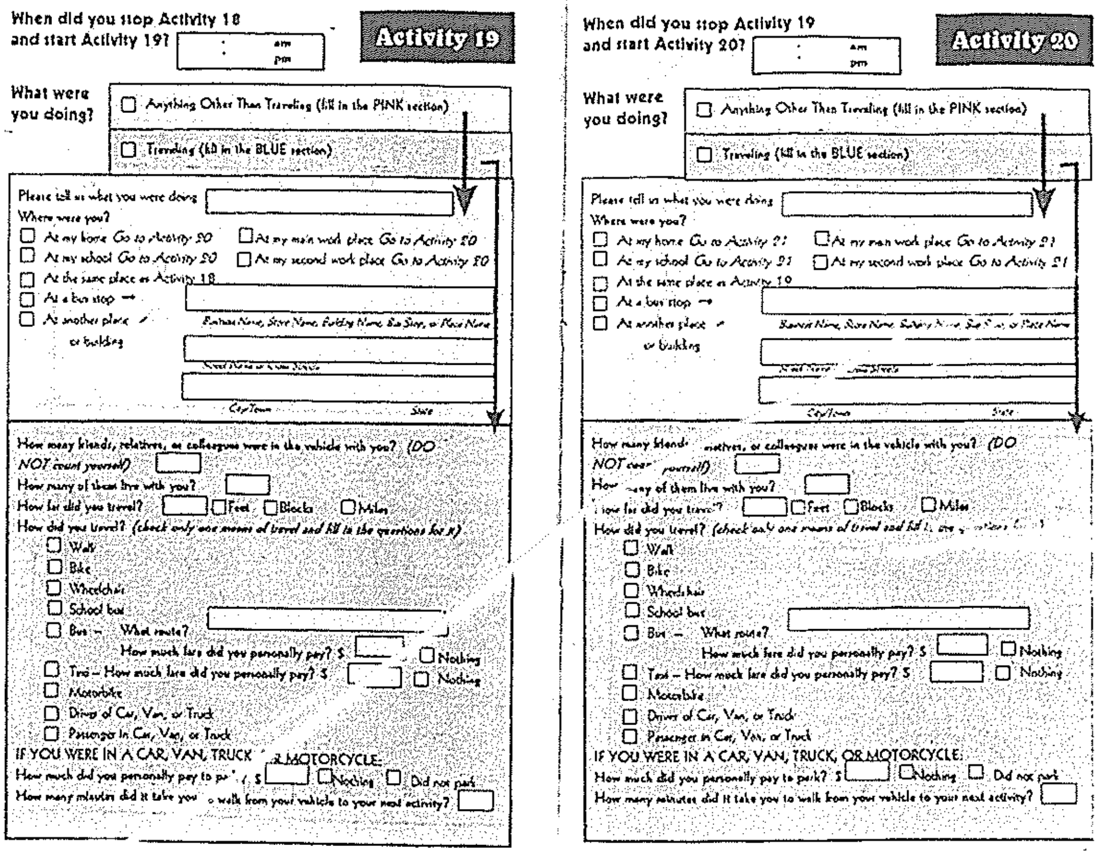
Reviewing the results in tables 11-13, it is apparent that even with a small panel of 450 segmented households, fairly accurate estimates of travel characteristics can be obtained. If the panel is established in a particular area, it will capture many of the peculiarities and idiosyncrasies of that area and, therefore, be more representative than a simulated panel.

Development of Experimental Design of a Time-use Diary

Time-use Diary Design

Initially, the research focused on the development of an improved time-use diary. The goal was to allow ease of use and more coherence with the respondent's needs, while still allowing the analyst to collect a much richer data set and include details of in-home activities. As with most previous activity and time-use surveys, the initial design of this diary asked respondents to record data about consecutively-numbered activities. Activities were defined broadly, activity at a place and activity of traveling between places both were defined as activities, and space was provided to answer questions about each. One significant change that was implemented in this design was to allow people to write in their own description of the activity, rather than trying to have respondents select from a lengthy list of activity codes. An example of two pages from such a diary is shown in figure 2.

It was considered necessary to allow for at least 30 activities to be recorded, and this format produced a bulky diary of approximately 40 pages in length. This caused an overflow of more than 30 activities and some limited instructions and personal data. A different format of diary was also developed in an attempt to reduce the bulk. This was based on the popular personal planner formats that have been available to business people and others for some time. It is also based loosely on designs used in time-use research *[52], [53]*. An example of part of this diary is shown in figure 3. In this format, with the same range of questions as in the first diary, the diary required a total of 16 pages, including the personal data and the covers. No memory jogger was required and no overflow was needed for a larger number of activities because the format accommodated all activities that the individual was willing to report. This version was created partly in response to experience with previous consecutive activity diaries in which there have always been problems requiring diary repair. It has been unclear whether people use the memory jogger or if it helps in filling out the diary.



When did you stop Activity 18 and start Activity 19? : am pm

Activity 19

What were you doing? ☐ Anything Other Than Traveling (fill in the PINK section) ☐ Traveling (fill in the BLUE section)

Please tell us what you were doing

Where were you?
☐ At my home Go to Activity 20 ☐ At my main work place Go to Activity 20
☐ At my school Go to Activity 20 ☐ At my second work place Go to Activity 20
☐ At the same place as Activity 18
☐ At a bus stop →
☐ At another place or building → Business Name, Store Name, Building Name, Bus Stop, or Place Name
Street Name or Cross Streets
City/Town State

How many friends, relatives, or colleagues were in the vehicle with you? (DO NOT count yourself)
How many of them live with you?
How far did you travel? ☐ Feet ☐ Blocks ☐ Miles
How did you travel? (check only one means of travel and fill in the questions for it)
☐ Walk
☐ Bike
☐ Wheelchair
☐ School bus
☐ Bus — What route?
How much fare did you personally pay? $ ☐ Nothing
☐ Taxi — How much fare did you personally pay? $ ☐ Nothing
☐ Motorbike
☐ Driver of Car, Van, or Truck
☐ Passenger in Car, Van, or Truck
IF YOU WERE IN A CAR, VAN, TRUCK, OR MOTORCYCLE:
How much did you personally pay to park? $ ☐ Nothing ☐ Did not park
How many minutes did it take you to walk from your vehicle to your next activity?

When did you stop Activity 19 and start Activity 20? : am pm

Activity 20

What were you doing? ☐ Anything Other Than Traveling (fill in the PINK section) ☐ Traveling (fill in the BLUE section)

Please tell us what you were doing

Where were you?
☐ At my home Go to Activity 21 ☐ At my main work place Go to Activity 21
☐ At my school Go to Activity 21 ☐ At my second work place Go to Activity 21
☐ At the same place as Activity 19
☐ At a bus stop →
☐ At another place or building → Business Name, Store Name, Building Name, Bus Stop, or Place Name
Street Name or Cross Streets
City/Town State

How many friends, relatives, or colleagues were in the vehicle with you? (DO NOT count yourself)
How many of them live with you?
How far did you travel? ☐ Feet ☐ Blocks ☐ Miles
How did you travel? (check only one means of travel and fill in the questions for it)
☐ Walk
☐ Bike
☐ Wheelchair
☐ School bus
☐ Bus — What route?
How much fare did you personally pay? $ ☐ Nothing
☐ Taxi — How much fare did you personally pay? $ ☐ Nothing
☐ Motorbike
☐ Driver of Car, Van, or Truck
☐ Passenger in Car, Van, or Truck
IF YOU WERE IN A CAR, VAN, TRUCK, OR MOTORCYCLE:
How much did you personally pay to park? $ ☐ Nothing ☐ Did not park
How many minutes did it take you to walk from your vehicle to your next activity?

Figure 2
Excerpt from the initial activity diary

In both diaries, a common approach was taken to the issue of collecting data on in-home activities. First, respondents were instructed to write a description of each activity in his or her own words. Some guidance was provided in an example page within the diary. However, respondents were also told that any activities they did not care to report in detail could simply be indicated as "other personal activities." Hopefully, this would cover such personal activities as bathing, personal care, dressing, and other private activities of no interest to transportation planners.

Time-Use Diary – Activity Details

All Activities			At a Place	Travel
Start Time		Please write in what you did	Where were you? Place name, building name, business/store name, street address, city *(Skip this if you were traveling)*	How did you travel?
Hour	Minutes			
3 am				
4 am				
5 am				
6 am				
7 am				
8 am				
9 am				
10 am				
11 am				
12 pm				
1 pm				
2 pm				

Time-Use Diary – Travel Details

Traveling *Skip if you were at one place)*						
Hour	How many friends, relatives were with you?	How many of them live with you?	How Far did you travel?	What Bus Route? *(if applicable)*	How much did you pay for bus/taxi fare or parking?	How many minutes did you walk to your next activity?
3 am						
4 am						
5 am						
6 am						
7 am						
8 am						
9 am						
10 am						
11 am						
12 pm						
1 pm						
2 pm						

Figure 3
Focus group version of day-planner diary

There is a major difference between these two versions of the diary. In the consecutive-activity diary, the respondent is able to check boxes corresponding to the possible responses to questions. For example, to answer where the activity took place, the respondent can check home, school, work, or the same place as the previous activity, among others. Similarly, mode of travel can be selected from a list. Therefore, the bulk of this diary is deceptive in that the respondent has to write relatively little and the task of completion actually proceeds rapidly. In contrast, the day-planner diary requires that every response be written out in the appropriate location. It also requires people to be able to respond on a tabular form, which may be a problem for some respondents.

The focus group. The two diaries were tested in a focus group, assembled from a range of staff at the Louisiana Transportation Research Center. Those participating were all employees of the state Department of Transportation and included clerical and secretarial staff, lab assistants, and professionals. The group, which consisted of fifteen people, was assembled, and the purposes of the focus group were outlined. Each member of the focus group was given two diaries, one of the consecutively-numbered activity time-use diaries and one of the day-planner format diaries. The covers of half of each type of diary were blue, and the covers of the other half were green. The blue diaries were coded for completion on the following Wednesday, while the green ones were coded for the following Thursday. Half of the focus group members were instructed to fill out the consecutively-numbered activities diary first and half to fill out the day-planner diary first. The focus group was reconvened on the day after the last diary day.

The focus group was encouraged to discuss the various aspects of each diary, including the information requested, the format of questions and answers, the provision of instructions, and any other relevant aspects. Among the responses, it was found that most did not read any instructions unless they got into filling out the diary and did not know what to do. It was also found that some of the closed-ended questions seemed unduly limiting to respondents. After extensive discussion, the group was also asked to comment on their preference of the two diary formats. An overwhelming majority of the participants agreed that the day-planner diary was considerably easier to complete, allowed them more freedom to express themselves, and took relatively little time. In addition, most agreed that the bulky, consecutively-numbered diary was actually much quicker to complete than its bulk suggested, but it had a significant flaw. This flaw basically had to do with a respondent forgetting to record one activity and discovering the omission later in the day. To correct this, it was necessary to erase most of the activities already entered and rewrite them, in order to include the omitted activity at the correct sequential location. As a result, most focus group members admitted that they would simply leave out the missing activity, even if it created a significant error in the diary. This problem did not arise in the day-planner diary, where insertion could simply be made at the relevant time of the day.

A number of useful points were made about the diary design. One of these led to changing the day-planner format to spread the 24 hours over three pages instead of two, allowing more space within each hour of the day. Other changes were also made in the layout of the instructions, the content of the instructions, and the layout of the personal information questions.

The pilot survey. Following completion of these changes, a full-scale pilot survey was undertaken with the day-planner format of the diary. The survey was completed by 36 households as a CATI-mail-out, mail-back survey (Computer-Aided Telephone Interview). The survey recruitment was conducted by the Louisiana Population Data Center at LSU, which has the capability to undertake telephone surveys using CATI. The research team developed scripts for recruitment and reminder calls and provided these to the Data Center, which was programmed into CATI software. The specific product used for this work was Sawtooth Corporation's Ci3 CATI software.

Based on the research reported earlier (the section on identifying groups with homogeneous travel characteristics), it was decided that any panel or general-purpose sample should be selected on the basis of a matrix of household size and vehicle ownership/availability. This had proved to be the most restrictive classification that identified households with homogeneous travel characteristics. It also corresponded more often than not to the basis of trip production modeling in most standard cross-classification models. Therefore, for sampling in the pilot survey, it was decided to test the method of sampling from this matrix and to determine whether any sub-groupings would be necessary within the matrix. Accordingly, the matrix shown in figure 4 was used for this survey. The cells for zero-car households were combined, because this has become such a small proportion of the population and made finding sufficient samples by household size extremely difficult. Combination of the vehicle ownership levels for one-person households was justified, because a one-person household was not likely to travel more with two or three cars. Similarly, combining the two and three+ vehicles for two-person households was justified on the same reasoning. No other combinations appeared to be justified for the pilot survey, but during the pilot survey, it was determined desirable to group together each four and five+ person household for vehicle ownership levels one and two. Along with a desire to obtain 100 households and eleven cells in the final matrix, the aim was to recruit nine households in each cell.

<table>
<tr><td rowspan="2">Vehicles available</td><td colspan="5">Household size</td></tr>
<tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5+</td></tr>
<tr><td>0</td><td colspan="5"></td></tr>
<tr><td>1</td><td rowspan="3"></td><td></td><td></td><td colspan="2"></td></tr>
<tr><td>2</td><td rowspan="2"></td><td></td><td colspan="2"></td></tr>
<tr><td>3+</td><td></td><td></td><td></td></tr>
</table>

Figure 4
Pilot survey sampling scheme

Households were defined as complete if no more than one diary (or two diaries in the case of large households) was missing in the return. The survey protocol was designed to run as follows: households were recruited using random-digit dialing and a short CATI interview that established some basic facts about the household, determined membership of the size and vehicle availability cells of the sampling matrix, and elicited the mailing address for the diaries. Households were then mailed diaries for each resident in the household on the diary day, together with a cover letter explaining the survey purpose and a vehicle and household information form. A reply-paid mailing envelope was also included in the package. Three attempts were made to call each household on the evening before the beginning of their diary day, to make sure that they had received the package, remembered the diary day, and to answer any questions they may have about the survey materials. Three attempts were made to call the household on the evening following the diary day to remind the household to send back the completed diaries and to deal with any who had forgotten to complete the diaries by reassigning the diary day. Additional telephone calls and postcard reminders were scheduled over the subsequent three weeks, alternating between postcard and telephone, for those households from which diaries had not yet been received. Figure 5 shows the desired contact regime set up for the pilot survey.

	Day from Recruitment Call												
	0	1-5	6	7	8	9-12	13	14-15	16	17-19	20	21-22	23
Action	Recruit		Pre reminder	Diary Day	Post reminder		Phone reminder		Postcard reminder		Phone reminder		Postcard reminder

Figure 5
Schedule of Contacts for the Pilot Survey

No incentives were used in this survey. However, in an attempt to improve overall response rate, a strategy was employed to attempt to make the telephone part of the survey more "respondent friendly." While this procedure was considered to be rather burdensome to the Population Data Center that performed the calls, it was stipulated that, as far as possible, the same interviewer was to call a given household on each of the contacts made. The script also provided for this interviewer to provide his or her name several times in the course of the calls and to provide the respondent with his or her telephone number for any questions that might arise. This contrasts standard protocols in which a household will most probably be called by a different interviewer on each occasion that a telephone call is placed. The idea behind this approach was to try to make the respondent feel more valued and to make the survey a little less impersonal.

A total of 1,670 telephone numbers were attempted. Of these, 1,418 were ineligible (e.g., business numbers, numbers not in service, group quarters, etc.). This high number is not unexpected. The numbers were generated by the research team and no attempt was made to remove business numbers, numbers used by the telephone company, and major blocks of numbers not in current use. In addition, because Baton Rouge has a large student population, there are a substantial number of phone numbers that are out of use at any given time because of changes in student locations and from nonpayment of bills. The remaining 252 numbers were either eligible or had eligibility unknown (the recording of the dispositions did not allow determination of the actual number of those with eligibility unknown, which was a defect in the Ci3 software). A total of 90 households were successfully recruited, representing a recruitment rate of at least 36 percent, and there were 151 refusals or terminations. As noted earlier, households were recruited into household size and vehicle availability categories. Table 14 shows the results of the recruited households, but does not show the aggregations actually used in the sampling. For example, all zero vehicle households were grouped together, and households of sizes four and five+ were grouped together for one and two vehicles. In the pilot survey, the aim of survey for the nine households in each sampling category was not enforced. The aim was to determine the degree of difficulty in achieving this goal in the main survey. As can be seen from table 14, the sample is not split as evenly as desired. Large samples were found in the one person one-plus vehicle category and in the two person two-plus vehicle category. Non-car-owning households were, as expected, very hard to find, probably because many such households also do not have a telephone.

Table 14
Recruited households, pilot survey

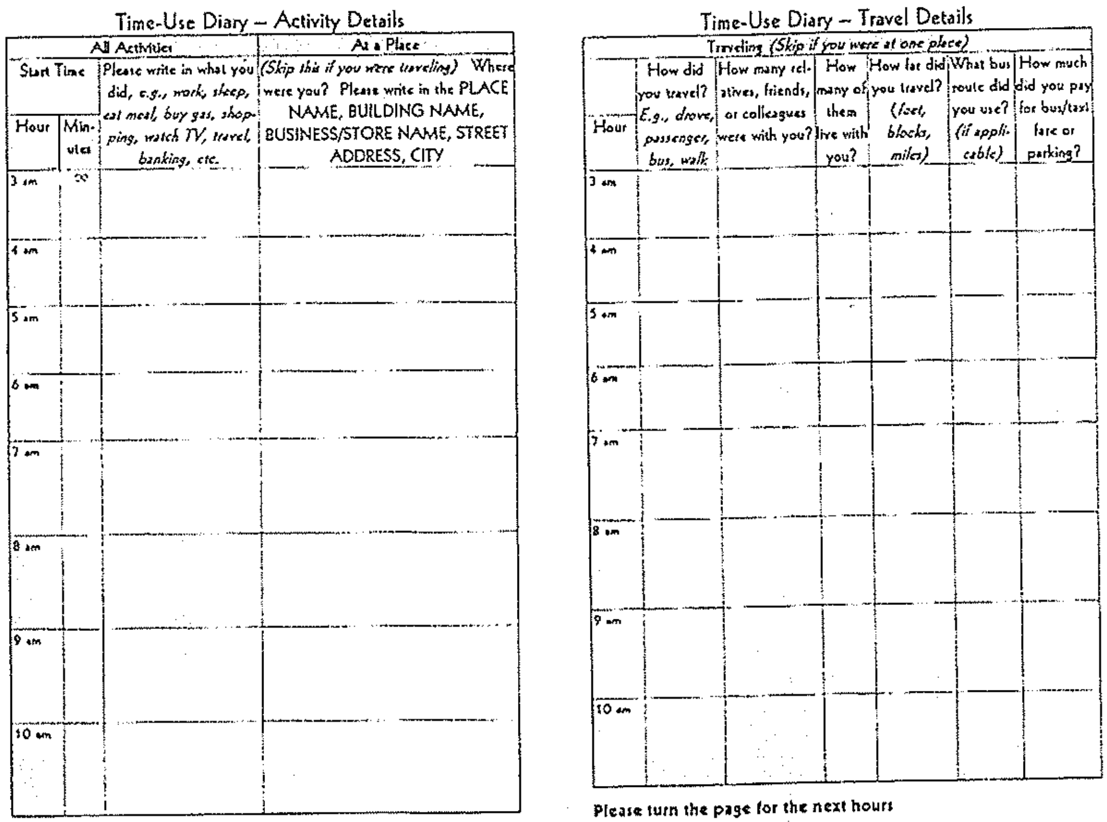
<table>
<tr><th rowspan="2">Number of vehicles</th><th colspan="5">Household size</th></tr>
<tr><th>1</th><th>2</th><th>3</th><th>4</th><th>5+</th></tr>
<tr><td>0</td><td>2</td><td>2</td><td>0</td><td>1</td><td>0</td></tr>
<tr><td>1</td><td rowspan="3">22</td><td>11</td><td>1</td><td>0</td><td>1</td></tr>
<tr><td>2</td><td rowspan="2">20</td><td>5</td><td>6</td><td>3</td></tr>
<tr><td>3+</td><td>7</td><td>6</td><td>3</td></tr>
</table>

The total number of households that returned completed diaries through the mail was 36, or 38 percent of households where survey materials were sent out. Overall, this is a 14 percent response rate (thirty-six percent times thirty-eight percent), which was better than expected for a mail-back survey. Of the 36 households that returned diaries, 34 returned useable diaries. There were 87 people in the 36 households, 81 people in the useable households, but activity information was provided on only 74 of those. Thus, the final data set consisted of data on 34 households and 74 people.

This was an encouraging result. Typical response rates from mail-back surveys with and without incentives have been running in the range of five to 24 percent in the U.S. in recent years, and they are declining, as seen on actual surveys mailed out to respondents *[8]*. Although numerical analysis of the pilot survey was not attempted at this stage, a detailed examination was made of the diaries that had been returned. It found that these diaries provided considerable detail on people's activities on the diary day (including some personal activities of no interest to transportation planners), and that there did not appear to be a serious problem with people understanding the diary completion task. In fact, a check of educational level and age showed that people with very little education and elderly people had both provided complete responses. Examination of the responses also showed that there were many instances where more detail was provided on personal activities than was needed, indicating that those who responded to the survey apparently had not found the questioning on their time use to be intrusive or an invasion of their privacy. Of course, those who did find the survey intrusive or difficult to complete probably did not return surveys. Project funding did not permit any follow up with respondents to find out how they reacted to the surveys.

Some brief summary statistics have been prepared from the pilot survey, to determine numbers of activities and trips. An examination of the diaries indicated that some analytical work on the diaries will increase the numbers of activities and trips reported, because respondents did not always separate travel and an activity. This may suggest some further modifications to the diary format or the use of CATI for data retrieval. Without editing of the data, however, there are 1,102 activities reported by the 74 people, averaging just under 15 activities per person. There are 316 trips reported, averaging 4.27 per person or 9.29 per household. These figures are at the upper end of the rates reported in activity surveys. It should also be noted that seven households had one or more diaries missing, causing this reported household trip rate to be too low. Indeed, if each of those persons with missing diaries had reported the average number of trips per person of those who completed diaries, the household trip rate would increase to 10.18 trips per household.

The Main Survey. Based on the pilot test, the day-planner diary worked well and only superficial changes were made to the forms. No changes were made to the survey protocol. Figures 6 and 7 show the principal changes in the diary instrument between the pilot test and the main survey. These changes were made to add emphasis to activity and travel by printing them in red, and to do the same to the headers of "At a Place" and "Traveling..." Also, the repeat of the hours on the right page was dropped, and the tables were extended closer to the binding, allowing more room in the right page columns.



Time-Use Diary – Activity Details

All Activities			At a Place
Start Time		Please write in what you did, e.g., work, sleep, eat meal, buy gas, shopping, watch TV, travel, banking, etc.	(Skip this if you were traveling) Where were you? Please write in the PLACE NAME, BUILDING NAME, BUSINESS/STORE NAME, STREET ADDRESS, CITY
Hour	Minutes		
3 am			
4 am			
5 am			
6 am			
7 am			
8 am			
9 am			
10 am			

Time-Use Diary – Travel Details

	Traveling (Skip if you were at one place)					
Hour	How did you travel? E.g., drove, passenger, bus, walk	How many relatives, friends, or colleagues were with you?	How many of them live with you?	How far did you travel? (feet, blocks, miles)	What bus route did you use? (if applicable)	How much did you pay for bus/taxi fare or parking?
3 am						
4 am						
5 am						
6 am						
7 am						
8 am						
9 am						
10 am						

Please turn the page for the next hours

Figure 6
Pilot test version of the day-planner time-use diary

The main survey used the same protocol as the pilot survey and aimed to draw a sample of 100 completed households. The survey was conducted at the end of January through the beginning of February 1999.

Time-Use Diary – Activity Details

All Activities			At a Place *(Skip if you were traveling)*
Start Time		Please write in what you did, *e.g., work, sleep, eat meal, buy gas, shopping, watch TV, travel, banking, etc.*	Where were you? Please write in the PLACE NAME, BUILDING NAME, BUSINESS/STORE NAME, STREET ADDRESS, CITY
Hour	Minutes past the hour		
3 am	00		
4 am			
5 am			
6 am			
7 am			
8 am			
9 am			
10 am			

Time-Use Diary – Travel Details

Traveling *(Skip if you were at one place)*					
How did you travel? *E.g., drove, passenger, bus, walk*	How many relatives, friends, or colleagues were with you?	How many of them live with you?	How far did you travel? *(feet, blocks, miles)*	What bus route did you use? *(if applicable)*	How much did you pay for bus/taxi fare or parking?

Please turn the page for the next hours

Figure 7
Main survey version of day-planner time-use diary

Calling extended over a period of two weeks for recruitment, using 4,630 telephone numbers from the RDD sample. Of these 4,630 numbers, 2,926 were determined to be ineligible. Again, these numbers were not screened for business numbers, blocks of numbers not in service, or numbers assigned to telephone company use. A total of 275 recruitment calls were completed, and there were refusals and terminations totaling 678 numbers. This is considered to be a notably high refusal rate and is probably because student workers were used to conduct the telephone interviews as opposed to more seasoned adult interviewers. The 740 telephone numbers that were used ended as busy, no answer, or otherwise did not have eligibility determined, and there were 11 unconverted soft refusals, i.e. refusals that appeared possible to change to compliance if a more effective interviewer were to call back. This results in a response rate for the recruitment, using CASRO[1] standards, of 24 percent. This is a disappointingly low rate, but has nothing to do with the format of the diary itself, because the diary was only seen by those who agreed to participate in the recruitment call.

Although it was desired to use the same procedure of assigning one interviewer to make all calls to a given household, it appears that this may not have been done completely in the main survey partly due to the longer time that this survey took. The same reminder procedure was also employed in this survey, which resulted in calls being made to households over a five-week period. Because of turnover in the students available to act as interviewers, many interviewers that began the survey were not there at the completion. This high rate of turnover may have contributed substantially to the lower response rate.

As for the pilot survey, a sampling scheme was used to stratify households by household size and vehicle availability. In the main survey, the design attempted to obtain between 10 and 30 households in each of the final household categories, which are shown in table 15 together with actual numbers of households recruited. Some cells exceeded the targets because of the time-lag in recruiting and marking targets in the cells.

Table 15
Distribution of main sample over the sampling categories

<table>
<tr><th rowspan="2">Number of vehicles</th><th colspan="5">Household size</th></tr>
<tr><th>1</th><th>2</th><th>3</th><th>4</th><th>5+</th></tr>
<tr><td>0</td><td colspan="5">15</td></tr>
<tr><td>1</td><td rowspan="3">34</td><td>25</td><td colspan="3">22</td></tr>
<tr><td>2</td><td rowspan="2">39</td><td>32</td><td colspan="2">32</td></tr>
<tr><td>3+</td><td>33</td><td colspan="2">42</td></tr>
</table>

1 CASRO is the Council of American Survey Research Organizations. It has developed and recommends use of a method for computing response rates that is described in Appendix I.

Diaries were mailed back by 72 of the respondents or 26 percent of those recruited, representing an overall response rate of six percent. This is a lower response rate than the pilot survey, and there are no clear reasons that have been uncovered for the decline in the response rate. Both the recruitment rate and the mail-back rate are about 10 percent lower in the main survey than in the pilot survey, and there appears to be evidence of problems with the interviewers. This may be an indication of less skilled interviewers or less effective training for the main survey. Of the 72 households returning diaries, 70 provided useable diaries and other information. Detailed analysis of the survey results are found later in this chapter.

Acquisition and Enhancement of GPS Units

Purpose and Background

The purpose of using GPS units is to improve the geographic information requested in a standard personal travel survey. As noted in the introduction, most people have only partial information on many of the places they visit each day, and the level of information they possess is usually insufficient to meet the needs of transportation planners for geographic coding. GPS units have the property of being able to determine a geographic location and a precise time at which the geographic location is recorded. The first known use of GPS instruments to assist in travel survey measurement is the proof-of-concept project performed by Battelle for the Federal Highway Administration *[21]*. The device tested by Battelle consisted of a GPS antenna that was about the size of a computer mouse, with a magnetic base connected by cable to a Personal Data Assistant (PDA). It measured about five inches by three inches by one inch thick. Wiring for the GPS antenna had to be placed around the passenger door with the GPS antenna on the exterior roof of the car. The PDA was placed in the car and connected to the cigarette lighter receptacle for power purposes. The GPS antenna required a two to five minute warm-up period to acquire signals and determine its position.

The proof-of-concept project involved sampling households in Lexington, Kentucky, and providing one GPS-PDA unit to each household. This unit was then placed in the main car used by the household. Each household that agreed to participate was paid an incentive of $50 to use the unit for one week, and to complete a travel diary for one of the days in that week. The unit worked in the following manner: when the driver got in the car to begin traveling, he or she was to answer a series of questions displayed on the screen of the PDA.

These questions related to who was the driver, the purpose of the driver's trip, the number of persons accompanying the driver, the name/relationship of each accompanying person, and the purpose of each accompanying person. During the time taken to respond to these questions, the GPS antenna would be warming up, so that positional information would begin to be recorded by the PDA as soon as the vehicle began moving (although this was not always the case, when the driver was alone or if the driver forgot to enter the data at the start of the trip). The PDA would record positional and time information throughout the trip until the ignition was turned off or the end trip button on the PDA screen was touched. If the end trip button was touched, but the ignition was still on, the PDA would display questions for the beginning of another trip.

The proof-of-concept test was successful. Households were willing to undertake this activity and successfully provided information for one week. The data have been analyzed to determine reasonableness and usefulness, and some limited analysis has been done for the correspondence between GPS-PDA data and the diary data. However, the proof-of-concept test has not provided a unit that can be applied readily in field practice. The purpose of this part of this project was to develop the concept further by finding a more "user-friendly" unit and developing a new protocol for its use so that it can be used by MPOs and other agencies to collect data on travel patterns.

Review of Instruments

In this project, exploration of an alternative GPS procedure has been initiated. The first requirement in this process was to select a technology for the GPS unit. A considerable effort was undertaken to examine the specifications of a variety of GPS antennas and PDAs. The Battelle experiment used a Garmin GPS30 TracPak PC antenna without differential correction and a Sony MagicLink PIC_2000 PDA. The GPS30 had to be mounted on the car roof, because it was unable to operate through a car windshield. It was also found that this GPS antenna had problems "seeing" sufficient satellites in downtown building "canyons" and where tree canopies are fairly dense. This latter shortcoming poses major problems in a city such as Baton Rouge, where tree canopies are extensive. The Sony MagicLink is a PDA with a large screen, no keyboard, and a stylus for use in entering data or responding to questions. It could be programmed to display questions by means of a PCMCIA card, which could also be used to store information recorded from the GPS antenna and the PDA itself. The unit is shown in figure 8 with the antenna at the bottom right of the picture, the Sony MagicLink on the left, the carrying bags in the center, and the cables and cigarette-lighter adapter.

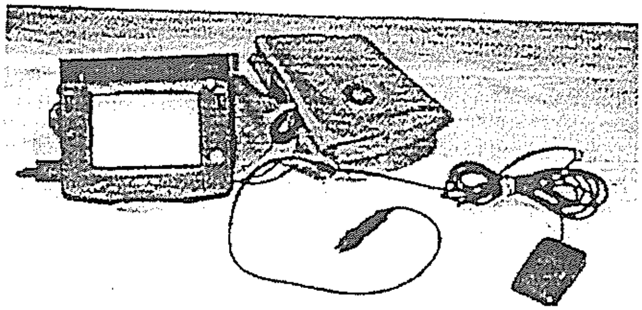


Figure 8
The Battelle GPS equipment *[21]*

There are a number of problems or disadvantages with the Battelle unit. The first is simply the cost of the unit. As purchased by Battelle, the PDA and GPS together with cables and other requirements cost a total of approximately $1300. The second problem was the warm-up time required for the GPS30, which could result in a significant amount of missed data, including making the origin of the trip uncertain. In some instances, it could even lead to missing a trip altogether. Third, the protocol for this equipment required the driver of the vehicle to touch the "Start" symbol on the PDA screen in order to begin recording GPS positional data and to define when the trip actually started. Fourth, it relies on the driver being conscientious in answering all of the questions on the PDA before beginning the trip, including a correct definition of the trip. Suppose that a driver sets off to work and enters his purpose as work on the PDA before leaving home. On the way to work, he suddenly realizes that he is almost out of gas, and stops at a gas station to fill up. The trip is now no longer a work trip but a trip to purchase gasoline, which ends at the gas station. A new trip now begins that is from the gas station to work and has the purpose "other-to-work." Procedures were developed in the software to allow the driver to make this change, but this again depended on the driver responding to the PDA before continuing his drive to work. The objectives of this part of this project were to reduce, as far as possible, these negative aspects of the Battelle equipment and to also take advantage of the latest technology available.

By the time this project began, the Sony MagicLink was no longer in production and a similar PDA was not available. A number of GPS antennas were available with improved warm-up times (as little as 30 seconds) and better ability to "see" the satellites through a windshield and tree canopies. The link to a PDA, however, was more problematic. With the exception of the Palm Pilot®, a large-screen PDA with no keyboard was almost entirely missing from the market. The Palm Pilot did not provide the means to program and had very limited memory at the time that this project began. Interfacing a GPS antenna with the Palm Pilot was also a problem. An extensive internet search was performed to seek both good GPS antennas and palm top computers that could be coupled together and used in a car, powered from the cigarette lighter. The requirements for the palm top computers were that they should have:

- At least 2 megabytes of memory
- Preferably a PCMCIA slot for either additional memory, a program, or the interface with the GPS antenna
- A RS-232 port for the GPS antenna (unless it would connect through a PCMCIA card)
- Backlighting for viewing at night
- A touch screen
- Rechargeable power pack or a working battery life of at least 48 hours
- An adapter to fit the cigarette lighter
- Moderate price (preferably under $300)

No palm top computer was found that met all of these specifications, and most failed to meet at least three. None were designed to interface with a GPS antenna, except the Palm Pilot, which could be used for route finding, but had no mechanism to store track data from the GPS antenna. From this search, it appeared that it would be both somewhat complex and quite expensive to put together a GPS antenna of appropriate capabilities with a palm top computer in order to emulate the device created by Battelle. One of the particular concerns of this project was to develop a less expensive device. The result of the search indicated that each unit might still cost approximately $1,000.

In 1998, some new GPS devices began appearing on the market. These devices were designed to store track-point data[2] to allow a user to find and retain the shortest paths to different locations. These devices were researched further in hope of selecting one of them that recorded only the track-point data and collected the other desired data by some alternative protocol. The search provided a number of potential devices. However, the major limitation found was the storage capacity for track-point data. Only one device seemed to offer the potential GPS antenna features with what was considered an acceptable level of storage. This was the Garmin GPS-III+®. The GPS-III+ consists of an antenna that can be

[2] Track-point data are the data recorded by a GPS of latitude, longitude, and time, at specified sampling frequency, e.g., one second, and that define a track traversed by the device.

rotated into a number of positions, attached to a small unit that has a number of control buttons and a screen to display current location and the immediate background map. The unit is powered by four AA batteries and has an approximate battery life of about 10 hours. However, the device is also provided with a power cable and a cigarette lighter plug. The device is shown in figure 9.

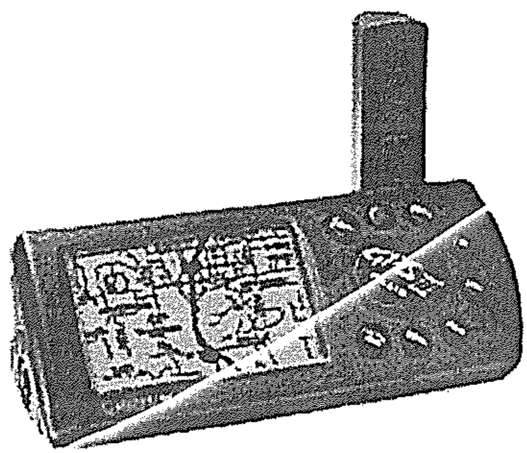


Figure 9
The Garmin GPS III+

The Garmin GPS III+ is approximately five inches long, stands two inches high (not including the antenna), three-and-three-quarters inches high with the antenna vertical, and two-and-one-half inches deep, including the antenna. It weighs nine-and-one-half ounces. The base of the unit is rubberized so that the unit will sit on a car dashboard without slipping, although a bean bag weight is also available to provide greater stability if needed. The unit retails for under $400. The device has a number of options that can be programmed with respect to display, recording, units of measurement, etc. The programming is achieved through the eight small buttons and one large button on the device. The GPS-III+ is able to record 1,900 track points, and its memory can be set either to wrap (so that the earliest recorded points are overwritten by the latest) or to fill to storage capacity and then stop recording.

An important feature of the Garmin GPS III+ relates to its power on and off features. If the unit contains charged batteries, but is plugged into the cigarette lighter socket of the car, when the power to the lighter socket is removed, the GPS III+ displays a message "Powering Off in 30 Seconds – Press Enter to switch to battery power." It then performs a count down and automatically turns off after 30 seconds if the enter button has not been pressed. The device does not have the capability, however, to power on automatically when the ignition is turned on and power is again supplied to the socket. This means that the device requires no action to turn off after a car is parked but must be manually turned on at the beginning of any car trip.

Acquisition and Testing of the Garmin GPS-III+

One of these units was acquired for testing to determine its capabilities and whether it could meet the requirements of the device that was sought for this project. In addition, it was desirable to test a unit to determine how to develop a protocol for its use. The unit was put through a number of tests, and test data were collected with the unit in order to determine its suitability and to understand what information could be obtained from it. Included in the tests were the following:

- Collecting up to three days of data in and around Baton Rouge
- Collecting three days of data that included travel to New Orleans and back
- Collecting data in another part of the country with long distances included in the collection
- Leaving the device turned on in a parked car to determine battery drain
- Testing the collection of data with persons unassociated with the project

The first two experiments were designed to determine how much of the data could be collected and stored. Normally, the device will record position data every 1 second. This was found to be impractical for collecting three days of data. Travel time in three days clearly exceeded 1900 seconds, or approximately 31 minutes, which is the maximum duration of data that could be collected with 1-second interval recording. The interval was then re-set to 0.1 miles. This should allow approximately 190 miles of data to be recorded. It was found that the device does not maintain the 0.1 mile interval but varies around it, overriding it in a manner that is not entirely clear from an examination of the data. An example of track-point data collected with a 0.1 mile interval setting is shown in table 16.

Table 16
Example of track-point data for a short city trip

ID	Longitude	Latitude	Track	Local Date	Local Hour	Local Minute	Local Secs	Time(Sec)	Distance (mi)	Speed (m/hr)
1	-91179162	30409447	1	7/9/99	14	54	14			
2	-91180975	30410105	1	7/9/99	14	55	16	62	0.117	6.797
3	-91182407	30410378	1	7/9/99	14	55	55	39	0.087	8.036
4	-91183129	30410654	1	7/9/99	14	56	29	34	0.047	4.976
5	-91184805	30410775	1	7/9/99	14	56	55	26	0.100	13.858
6	-91185859	30411159	1	7/9/99	14	57	19	24	0.068	10.213
7	-91185746	30412328	1	7/9/99	14	58	7	48	0.081	6.071
8	-91185349	30413106	1	7/9/99	14	58	47	40	0.059	5.279
9	-91183775	30413578	1	7/9/99	14	59	13	26	0.099	13.730
10	-91182353	30414015	1	7/9/99	14	59	42	29	0.090	11.151
11	-91181648	30412634	1	7/9/99	15	0	14	32	0.104	11.713
12	-91180980	30411298	1	7/9/99	15	0	45	31	0.100	11.658
13	-91179271	30411126	1	7/9/99	15	1	12	27	0.102	13.652
14	-91179041	30409544	1	7/9/99	15	1	43	31	0.110	12.776
15	-91178777	30408502	1	7/9/99	15	2	21	38	0.074	6.972

From the "Distance (mi)" column, it can be seen that the intervals actually range from as little as 0.059 miles to as much as 0.117 miles. Some of the discrepancy can be accounted for because the distances calculated in the table are determined as straight-line distances between the successive latitude and longitude values. However, this cannot account for all of the variability in the distance intervals. The total distance for this trip was 1.239 miles and the average interval was 0.088 miles. Another track, collected a few days later with the same instrument and in the same data set was 6.7 miles long and had an average interval of 0.113 miles. A subsequent trip that involved travel from a home in Baton Rouge to the New Orleans airport, covering a total distance of 69.9 miles (odometer reading is 74 miles), gave an average recording interval of 0.201 miles. Apart from this case, most of the longer trips found averaged somewhere in the region of 0.12 miles as the recording interval.

Table 16 shows data after some processing from the GPS III+ device. However, the downloaded data were found to require considerable manipulation to be useful. Table 17 shows the data for table 16 before processing. The first column of data is either an H for "Header" or a T for "Track Data." The data are then recorded with latitude first and longitude second with the compass bearings provided first (N or S for latitude and W or E for longitude), followed by the actual bearing with degrees and minutes concatenated and fractions of minutes following. It was proposed that the data from the GPS devices would be used in TransCAD. For TransCAD to recognize the coordinates, the longitude must be first and the latitude second. Furthermore, there is no compass bearing in a TransCAD latitude and longitude; rather, S latitude and W longitude are negative, while the others are positive. TransCAD requires also that latitude and longitude be expressed in millionths of degrees.

Table 17
Raw track data from the GPS III+

H	LATITUDE	LONGITUDE	DATE	TIME	ALT	;track
T	N3024.56687	W09110.74972	09-JUL-99	19:54:14	-9999	
T	N3024.60630	W09110.85851	09-JUL-99	19:55:16	-9999	
T	N3024.62271	W09110.94444	09-JUL-99	19:55:55	-9999	
T	N3024.63929	W09110.98774	09-JUL-99	19:56:29	-9999	
T	N3024.64653	W09111.08832	09-JUL-99	19:56:55	-9999	
T	N3024.66955	W09111.15156	09-JUL-99	19:57:19	-9999	
T	N3024.73971	W09111.14481	09-JUL-99	19:58:07	-9999	
T	N3024.78638	W09111.12099	09-JUL-99	19:58:47	-9999	
T	N3024.81471	W09111.02652	09-JUL-99	19:59:13	-9999	
T	N3024.84094	W09110.94123	09-JUL-99	19:59:42	-9999	
T	N3024.75806	W09110.89890	09-JUL-99	20:00:14	-9999	
T	N3024.67791	W09110.85883	09-JUL-99	20:00:45	-9999	
T	N3024.66761	W09110.75631	09-JUL-99	20:01:12	-9999	
T	N3024.57266	W09110.74247	09-JUL-99	20:01:43	-9999	
T	N3024.51017	W09110.72666	09-JUL-99	20:02:21	-9999	

The date and time recorded by the GPS III+ are measured in Universal Time (i.e., what used to be known as Greenwich Mean Time), which is five hours ahead of Central Daylight Time and six hours ahead of Central Standard Time. Comparing tables 17 and 16, it can be seen that the times and dates in table 16 are converted to local time and date rather than universal time and date. No information is provided on altitude unless the GPS III+ is initialized at a known altitude. Since altitude is not of interest in this work, the initialization has not been attempted, and the altitude column is filled with a missing data code of -9999. Although the GPS III+ shows a column headed as track, no data are recorded in that column.

Initially, Excel® macros were written to make these conversions and to reorder the data columns to produce the format shown in table 16. This provided the means to begin to examine the data further, to determine the feasibility of displaying the tracks in TransCAD, and to determine what further analysis or processing would be required.

The results of the long-distance tests showed that it is extremely important to ensure that old data are deleted from the device before commencing a new data-collection period. By mistake, the device was not cleared before beginning the three days that included the New Orleans travel resulting in part of the data being lost. The device was also not set on wrap, which would have automatically deleted the earliest recorded tracks as additional track data were collected. Hence, a portion of the long-distance data were lost. The testing of the device in Southern California worked well and produced good track information. It confirmed the ability of the device to collect significant amounts of data over a three-day period, including long trips.

The battery-life test established that the listed battery life was probably correct, which indicated a life of about 6 hours on four AA batteries. This confirmed that it would be necessary for the device to always be allowed to turn off automatically at the end of a trip (i.e., when the car engine was switched off), and that the device would have to be manually restarted at the beginning of each subsequent trip. The only exception would be if the device was placed in a car that has a permanently live cigarette lighter. Some makes and models have this feature.

Testing the collection of data with a person unassociated with the project was enlightening. This showed that there are problems, somewhat as was anticipated, in the person remembering to turn on the device each time the car is started. Occasionally, it was found that the device had been turned on part way into a trip, while other trips were missed completely. This suggested that some sort of reminder sign or other mechanism would be needed to try to overcome this problem.

During the tests, it was noted that the GPS III+ occasionally lost signal, presumably due to obstructions to its view of at least four satellites. When this occurred, the track point file treated the resumption of a signal as though a new track was started. This resulted in a situation in which some trips were made up of two or more tracks. In order to recover the trips, it would be necessary to devise rules that would permit tracks to be joined together when they represented a continuation of the same trip. Conversely, it was also noted that the GPS III+ would not create two tracks if a person stopped for some intermediate purpose (such as at a drive-through bank, fast food drive through, or other stop not requiring the car engine to be turned off). This raised the question as to whether such stops could be identified and distinguished from a traffic stop, that might occur at a stop light or stop sign. Therefore, some additional experimental runs were made in which one of the project staff made such intentional stops.

These additional experimental runs showed that stops could often be identified and distinguished from traffic stops. However, the identification depended on the length of the stop. Should someone stop at a curbside to pick up or drop off a passenger and stop less than 30 seconds, this would be identified as a traffic stop. From experimentation with the various files and as is discussed further in the next section, tracks were joined if the elapsed time between them was less than 400 seconds and if the time between the last two track points of the first track was not more than 30 seconds greater than the time between the penultimate track point and the one before it. In addition, the beginning point of the second track must be at a different location than the ending point of the first track, indicating that the vehicle was traveling during the elapsed time. A track was to be split into two tracks if the time between any pair of successive track points was greater than 150 seconds, or a series of points occurred with effectively no change in position.

Using the track files that had been obtained from these tests, the next task was to determine what additional processing would be necessary to make the track data usable. The desire was to find some way in which the driver of the car could be informed of each track that was found recorded in the device, to allow the driver to be asked about details of the trip such as purpose, other occupants in the vehicle, etc. This led to the next step in the process – the development of visual basic programs to process the track point data.

Visual Basic Programs to Process the Track Point Data

The goal of this task was to produce files that could be used for input into the subsequent data collection tasks and would produce coherent trip records from the track-point data. Early in the process, it was determined that the track point data from the GPS III+ could not be read directly into a GIS software such as TransCAD, but must undergo some transformations to be usable. The initial transformations are described in the preceding section. The transformations must be saved in a database file that can then be read into TransCAD or similar software. Here, each track file is saved as a separate geographic point file. The transformation steps accomplished in this process are:

- To select the input file (track file downloaded from Garmin GPS III Plus)
- To open the selected track file into an Excel work sheet
- To delete the unnecessary header information and unnecessary columns
- To rearrange the columns to make the file compatible with TransCAD
- To convert the longitude and latitude values into a format read by TransCAD
- To convert the UTC time and date to Baton Rouge local time and date
- To insert a column header in the first row of the file to provide a description for each column of data
- To number each track point and each track
- To insert the header at the beginning of each new track
- To compute the time, distance, and speed data between each pair of consecutive track points
- To join two different tracks into one if the criteria set for joining tracks is met
- To mark the cells between two tracks that could be joined based upon user judgement
- To split a track into two if an intermediate stop is observed
- To insert the header at the beginning of each newly formed track
- To renumber the track points and tracks
- To insert an additional column for identification adjacent to address column
- To format the column widths to make the file suitable for printing

Each track point must be numbered with a unique number for TransCAD to be able to read the file and convert it to a geographic file. The tracks are numbered for analytical convenience. The headers are needed for each new track, so that each track can be saved eventually as a separate file. All tracks recorded on the GPS III+ are initially in a single file with at least one embedded blank line between two successive tracks and a header row. For TransCAD to be able to plot the tracks as separate entities, it is necessary to save each separate track as a unique file. In addition, is a need for visual inspection of the data. For this reason, it is desirable to print out the transformed track point file to allow visual inspection for joining and splitting tracks where the criteria are not completely fulfilled. An example of data from two consecutive tracks, as downloaded from the GPS III+, is shown in table 18.

Table 18
Example of the beginning of a raw track file from the GPS III+

H	S	OFTWARE	E	& VERSION							
I	P	CX5 2.09									
H	R	DATUM		ID	X DA					DF	
M	G	WGS 84		12	1 +0.00000	0e	+	0		0	0
H	C	OORDINA	Y	STEM							
U	L	AT LON DM									
H	L	ATITUDE	L	ONGITUDE	DATE	TI	M	E			ALT
T	N	3024.222	W	9110.676	24-Aug-99	13	:	12	:	50	-9999
T	N	3024.225	W	9110.683	24-Aug-99	13	:	13	:	14	-9999
H	L	ATITUDE	L	ONGITUDE	DATE	TI	M	E			ALT
T	N	3024.054	W	9110.503	24-Aug-99	16	:	29	:	14	-9999
T	N	3024.113	W	9110.441	24-Aug-99	16	:	31	:	9	-9999
T	N	3024.185	W	9110.364	24-Aug-99	16	:	31	:	18	-9999
T	N	3024.187	W	9110.304	24-Aug-99	16	:	31	:	41	-9999
T	N	3024.225	W	9110.357	24-Aug-99	16	:	38	:	24	-9999
T	N	3024.222	W	9110.442	24-Aug-99	16	:	38	:	39	-9999
T	N	3024.188	W	9110.531	24-Aug-99	16	:	38	:	54	-9999
T	N	3024.204	W	9110.636	24-Aug-99	16	:	39	:	4	-9999
T	N	3024.216	W	9110.691	24-Aug-99	16	:	39	:	38	-9999

The data in table 18 have been read into an Excel® file, so that the column settings split up the main contents of the file. A blank line and headers can be seen between the first and second tracks in table 18. An example of the output from the first visual basic program is shown in table 19. This example shows the addition of the track point identification number, the track number, the transformed latitude and longitude reordered for TransCAD, the local time and date, and calculated time, distance, and speed. The calculated time and distance are used to check for possible continuation of the same trip and to identify intermediate stops. In table 19, the large value for the first elapsed time indicates that the initial track point of the first track shown is a real new trip. The time for the last track point of the first track of 403 seconds makes this a case where an automatic joining of the two tracks would not take place (the value has to be less than 400 seconds) and the time and distance for the first point of the second track are not indicative of a continuation track. There are no indications in table 19 of intermediate stops in the segments of track shown.

Table 19
Example of output of the first program

ID	Longitude	Latitude	Track	Local Date	Local H	Local M	Local S	Time(Hr)	Time(Sec)	Distance	Speed	DupD	Addres	Purpos	Nearest
1	-91175052	30400900	1	8/24/99	11	29	14	3.27	11760	0.27	0.08	1			
2	-91174017	30401887	1	8/24/99	11	31	9	0.03	115	0.09	2.87	2			
3	-91172737	30403088	1	8/24/99	11	31	18	0.00	9	0.11	45.02	3			
4	-91171734	30403115	1	8/24/99	11	31	41	0.01	23	0.06	9.35	4			
5	-91172617	30403751	1	8/24/99	11	38	24	0.11	403	0.07	0.61	5			
ID	Longitude	Latitude	Track	Local Date	Local H	Local M	Local S	Time(Hr)	Time(Sec)	Distance	Speed	DupD	Addres	Purpos	Nearest
1	-91174025	30403692	2	8/24/99	11	38	39	0.00	15	0.08	20.13	1			
2	-91175508	30403134	2	8/24/99	11	38	54	0.00	15	0.10	23.11	2			
3	-91177260	30403391	2	8/24/99	11	39	4	0.00	10	0.11	38.07	3			
4	-91178185	30403606	2	8/24/99	11	39	38	0.01	34	0.06	6.04	4			

The empty columns at the right hand side of the table are provided for information to be added when the track is viewed in TransCAD. This is explained in more detail in the next subsection of this report. In addition to providing the entire track file in the format shown in table 19, the first program also provides individual track files for importing into TransCAD. These are dBase® files, an example of which is shown in table 20.

Table 20
Example of a single track file created by the first program

ID	LONGITUDE	LATITUDE	TRACI	LOCAL_	LOCAL_H	LOCAL_M	LOCAL_S	TIME(HR)	TIME(SEC	DISTA	SPEED	DUPI	ADDRE	PURP	NEAR
1	-91174025	30403692	2	8/24/99	11	38	39	0.00	15	0.08	20.13	1			
2	-91175508	30403134	2	8/24/99	11	38	54	0.00	15	0.10	23.11	2			
3	-91177260	30403391	2	8/24/99	11	39	4	0.00	10	0.11	38.07	3			
4	-91178185	30403606	2	8/24/99	11	39	38	0.01	34	0.06	6.04	4			

This information is the same as the bottom half of table 19, but is output in one file for each track in order to import into TransCAD. At this point, each track file is imported into TransCAD, and the ends of the track are identified. As noted in the subsequent description of the protocol, information has already been obtained about the most frequent destinations of the household members, and these, together with the home location, have been positioned on a TransCAD layer. The ends of the track are now either identified as belonging to one of these destinations, to home, or to some other unidentified location. The address column is used to enter information about the beginning and end of the track (no intermediate locations are identified), and the purpose is deduced from the destination, if known. For example, if a track begins at home and ends at a reported workplace for the driver of the vehicle, the address is identified as that workplace, and the purpose is assumed to be work. If the location is not clearly identified, then the nearest street intersection is also indicated in the third column. Figure 10 shows an example of a track file displayed for this part of the analysis.

Figure 10
Example of track file displayed in a GIS

In figure 10, the track goes from near LSU to Ace Hardware. Other locations for this individual are shown on the map, such as "Home" and "Albertsons." This clearly shows the nature of the displayed track file as a point layer in the GIS. However, the point layer is not displayed in a way that necessarily makes it easy to communicate the information back to the respondent. The above map does allow the track to be displayed showing the origin and destination, although it does not indicate which is which. Another problem is that the dataview for the track file does not remain in the chronological order of the data points although it can be sorted to that order.

Using a map such as that of figure 10, the origin and destination locations are identified, the purposes are surmised when possible, and the results are entered into the dataview file in the appropriate columns. This file is then exported as a database file and input into a second set of visual basic programs.

The purpose of the subsequent visual basic programs is to remove the intermediate track points between each pair of origins and destinations and concatenate the tracks into a file that provides a summary of the information about the travel recorded by each GPS device for each 24-hour period. The data are also rearranged to group data pertaining to the origin together as the first entries on each row of this final file and to group data pertaining to the destination as the last entries on each row of the final file. An example of the output of this program is shown in table 21.

Table 21
Example of the final concatenated file

Track Sta	Date Start	Time Start	Elapsed	Place Star	Track E	Date End	Time End	Track Tim	Cum Dista	Place End	Purpose	Near Node
2	23-Aug-99	17:00:20	5566	WORK (LT	2	23-Aug-99	17:15:01	0:14:41	3.24	HOME	HOME	
3	23-Aug-99	17:54:39	2378	HOME	3	23-Aug-99	18:04:08	0:09:29	3.44	120 ARQUETTE AVE		
5	23-Aug-99	20:41:15	7499	stgerters	5	23-Aug-99	20:45:33	0:04:18	2.00	gffdgds	dfgdf	
5	23-Aug-99	20:45:42	9	dgfdds	5	23-Aug-99	20:50:44	0:05:02	1.95	asdfefwe		

In this file, the starting time and location are shown, together with the elapsed time in seconds, since the previous track ended. This is followed by the information on the end of the track, including the time of arrival at the destinations, the total distance traversed, and the place where the trip ended. The final items of information are the assumed trip purpose and the nearest node when appropriate. These data can then be used to describe the trip to the respondent. As is described in the next section of this report, the output data file is designed to be used directly in the telephone script that forms the final retrieval portion of the protocol.

The final procedure developed in this stage of the project was to represent the tracks recorded by the GPS in such a way that a map could be constructed that respondents or interviewers could use to assist in attaching the remaining recall data to each track. The method used for this was to create each track as a route in a route system layer, such as would be used to represent a transit system. The result of this process is depicted in figure 11. In this map, each track is shown in a different color, and it is possible to number the tracks to show the order in which the trips were performed. In order to achieve this, the following steps are necessary:

- Each of the address locations on the GIS is connected to the street network by creating a dummy street link if necessary.
- The street layer is then built as a network that includes these dummy street links or driveways.
- Each track is displayed as a point layer, and the beginning point, ending point, and intermediate turns are clicked with the mouse while on an editable route system layer.
- Each route, corresponding to a trip, is saved in the route system layer with the track number as an identifier.

Figure 11
Representation of one day of travel from track files

This activity completes the analysis of the track-point data collected from the GPS. A map, such as the one shown in figure 11, can now be produced for each vehicle in which a GPS device was located and for each day that the GPS device was used. The map can be sent to the respondent and/or can be used by an interviewer. The use of the map and the final file from processing are discussed in the next section.

Development of an Application Protocol

A prototype protocol for deploying the GPS units has been developed. In this protocol, a household is contacted by telephone from some prior listing or random-digit dialing and is recruited for the study. In the protocol that has been developed to date, households are selected from households that participated in an earlier time-use diary survey, which provided home telephone number, home address, workplace and school addresses, number of workers in the household, number of school-age children, household size, and number of vehicles. The recruitment call also records address information for the home, workplaces, schools, and supermarkets used by household members, or it confirms that the data previously obtained on these locations is correct. An appointment is made to take the GPS devices to the household, locate them in the household's vehicles (they simply require to be plugged into the cigarette lighter and placed in a position where there is no more than a partially-obstructed view of the sky), and request the driver of the vehicle to turn the device on each time the car is started. The devices are to be used for a period of two or three days, and one is provided for each of the household's operating vehicles.

After the prescribed time has elapsed, the devices are collected from the household, and the track point data are retrieved from each device. As described in the previous section, the researchers have developed visual basic programs that manipulate the data to provide separate database files of each definable track (representing a trip from ignition on to ignition off), which can then be read into GIS software and transferred into a map layer. The addresses recorded in the recruitment call are also geocoded into a separate map layer, allowing identification of when the trip end occurred at home, work, school, or one of the recorded supermarkets. Other trip ends can be provided with an address range from the street map layer by querying that layer in the GIS at the location of ignition on and ignition off. The tracks from the GPS device are then concatenated for each day that each GPS device was in use. Data are entered on the elapsed time since the preceding trip, the length of the current trip, its starting point and time, and its ending point and time.

As currently developed, the protocol is completed by a second telephone call to the household. In this call, household members are given the details of each trip and asked to respond to questions to identify who was the driver, what was the driver's purpose, if there were any passengers in the vehicle, and who they were and what their purposes were. Additional data are also collected on whether any payment was made for parking and whether any stops were made between on and off locations for purposes where the engine was not turned off. At this point, the data collected would be equivalent to what was collected in the Battelle experiment. However, the process is more user friendly, uses a much cheaper device, and can readily capture data from more than one vehicle in a household. Both procedures require the driver to start the device for recording the trip, but the Battelle device required the driver to delay the start of his or her trip in order to enter various pieces of data about the persons participating in the trip while the newly-proposed process requires no further action or delay on the part of the driver.

Tests of Transferability

To test the transferability of data to metropolitan areas in Louisiana, the similarity of travel characteristics among different data bases to data recently collected in Baton Rouge was tested. Several travel characteristics were considered, including trip rates, mode share, and trip length frequency distributions. Only households in urban areas and those surveyed on a weekday were included in the data analyzed. The results are reported in tables 22-25.

In table 22, the trip rates from NPTS 95 data for Metropolitan Statistical Areas (MSAs) of population from 500,000 to 1,000,000 are listed with those from Baton Rouge Personal Transportation Survey and the North Central Texas Council of Governments (NCTCOG) Survey. NPTS 95 data for MSAs of population from 500,000 to 1,000,000 were chosen to match the metropolitan size of Baton Rouge. The NCTCOG data covers the Dallas-Ft.Worth area, which has a population in excess of 3,000,000 and, therefore, can be expected to display different travel behavior on those data items that are sensitive to metropolitan size such as trip length and mode share. The values in table 22 include weighted and unweighted trip rates. The weighted values adjust the sample data for non-response, under-reporting, or other inconsistencies of the sample with observed totals of the population on a variety of factors as described earlier.

The transferability of the trip rates between data sources can be measured by the similarity of their values from the two sources. A suitable statistical test of similarity is the z-test of the difference between two means as shown earlier in equation 11. Those that were significantly different at the 0.05 level of significance from the trip rates of the BRPTS data are shown with an asterisk in the table.

Table 22

Comparison of trip rates by purpose

Trip purpose	Trips/household/day					
	NPTS 95$_{MSA500K\text{-}1M}$		BRPTS		NCTCOG	
	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted
All	10.5	9.7	10.9	9.7	9.7*	9.5
HBW	1.5*	1.5	1.9	1.6	1.6*	1.6
HBO	5.5	5.0	5.6	4.9	4.8*	4.7
NHB	3.5	3.0	3.5	3.2	3.2	3.2

The results show that NCTCOG weighted trip rates are significantly different from the BRPTS values for all purposes except NHB, whereas the NCTCOG and NPTS unweighted trip rates for metropolitan areas of similar size to Baton Rouge are not significantly different. Part of the reason is that the weighting factor used in the NCTCOG data is solely an expansion factor to figure the sample to population size while the weighting factor used in the NPTS 95 and BRPTS 97 data is both an expansion and adjustment factor. Thus, the difference in weighted trip rates between the NCTCOG data and BRPTS 97 data is more a function of the difference in types of weighting factors than a difference in the underlying value of the trip rates.

As argued in the introduction of this report, it is expected that the transferability of data will be improved when it is transferred between segments. To test this premise, trip rates for all-purpose trips were derived for the nine household size/vehicle ownership segments used in this study from the NPTS 95, BRPTS 97, and NCTCOG data. The results are shown in table 23. Again, the trip rates that are significantly different from BRPTS trip rates at the 0.05 level of significance are shown with an asterisk.

Table 23
Comparison of all-purpose trip rates by segment

Segment	All-purpose trips/household/day					
	NPTS $95_{500K-1M}$		BRPTS		NCTCOG	
	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted
1	4.6*	4.3	8.2	5.9	6.9	6.8
2	4.8	4.5	4.7	4.2	4.0	4.2
3	8.5	7.5	8.0	7.0	7.3	7.4
4	9.2	8.4	10.3	8.7	7.4*	7.7*
5	14.3	13.2	13.7	11.4	12.2	12.5
6	11.8	11.0	12.5	11.2	11.5	11.8
7	14.2	12.9	15.0	12.8	11.6*	11.4
8	18.5	16.2	18.8	16.0	18.4	18.4*
9	20.6	18.8	21.4	18.0	17.5*	17.4

Trip rates from the NCTCOG data are significantly different from those from BRPTS for five of the 18 segment comparisons (i.e. weighted and unweighted segments). In contrast, segment trip rates from the NPTS 95 data for MSAs of similar size to Baton Rouge were significantly different in only one of the 18 comparisons. The case where a significant difference between the NPTS and BRPTS trip rates occurs (i.e. segment 1, weighted) is where the weighted trip rate for the BRPTS data is unusually high, even though the unweighted trip rate was within expected limits. However, in multiple comparisons such as this, the chance that at least one comparison will be found to be significantly different when they are all equal increases with the number of comparisons *[54, p. 273]*. The Bonferroni test for multiple comparisons adjusts for these situations by modifying the significance level depending on the number of comparisons. Specifically, the chosen significance level is divided by the number of comparisons when determining the significance level of the test statistic. For example, in a situation where a significance level in excess of 0.05 would normally be considered sufficient evidence of a significant difference, with 18 simultaneous comparisons, one or more comparisons must be significant at the 0.05/18 or 0.00277 significance level in order to draw the same conclusion.

Mode share by trip purpose among the three data sets is shown in table 24. Mode shares that are statistically different from those observed in the Baton Rouge Personal Transportation Survey are shown with an asterisk in the table. The 0.05 level of significance was used.

Trip length distributions by purpose are shown in table 25. The trip lengths are expressed in terms of reported travel time and are categorized by 15-minute intervals. Normally, travel times derived from the network are used to determine the trip length frequency distribution, but only reported travel times were available in this study. Large intervals of travel time were chosen in this analysis to compensate for the tendency of travelers to report their travel time in five and 15 minute increments. The intervals include observations that equal the lower limit of each interval and exclude those that equal the upper limit. The values in table 25 are the percentages of all trips of the specific purpose that fall within each of the travel time intervals. As before, the data involves only weekday trips from households in urban areas and the NPTS 95 data is only for households in MSAs with populations between 500,000 and 1,000,000. Proportions of trip lengths that are different from those BRPTS values at the 0.05 level of significance are shown with asterisks in the table.

The trip length distributions from the NCTCOG data display longer trips on average than those observed in the BRPTS data. However, this is not unexpected since the NCTCOG metropolitan area is considerably larger than that of Baton Rouge and trip length distributions are influenced by the size of the metropolitan area. The NPTS 95 data for MSAs of population between 500,000 and 1,000,000 represents MSAs of similar size to Baton Rouge. The observed trip length distribution for all trip purposes are indeed more similar between the NPTS and Baton Rouge data, although a significant difference still exists for the shorter trips. The limited analysis of trip length distributions conducted in this study suggests that trip length distributions may indeed be stable among metropolitan areas of similar size as observed by other researchers *[55]*.

Table 24

Comparison of mode share

Trip purpose	Mode	Mode share					
		NPTS 95 $_{500K-1M}$		BRPTS		NCTCOG	
		Wtd.	Unwtd.	Wtd.	Unwtd.	Wtd.	Unwtd.
HBW	Auto driver	0.87	0.89	0.89	0.90	0.91	0.92*
	Auto pass.	0.09	0.08	0.09	0.09	0.06	0.06
	Bus	0.03	0.02	0.01	0.01	0.00	0.00
	Walk	0.01	0.02	0.01	0.01	0.03	0.02
	Taxi	0.00	0.00	0.00	0.00	0.00	0.00
HBO	Auto driver	0.62	0.61	0.62	0.63	0.53*	0.56*
	Auto pass.	0.27	0.28	0.27	0.27	0.35*	0.32*
	Bus	0.05	0.05	0.05	0.05	0.00	0.00
	Walk	0.07	0.06	0.06	0.04	0.11	0.11*
	Taxi	0.00	0.00	0.00	0.00	0.00	0.00
NHB	Auto driver	0.69*	0.70*	0.73	0.73	0.62*	0.58*
	Auto pass.	0.24	0.23	0.23	0.22	0.21	0.17*
	Bus	0.02	0.02	0.01	0.02	0.09	0.12*
	Walk	0.05	0.05	0.03	0.03	0.09	0.12*
	Taxi	0.00	0.00	0.00	0.00	0.00	0.00

Table 25

Comparison of trip length distributions

Trip purpose	Travel time interval (mins.)	Percentage of trips in time interval					
		NPTS $95_{500K-1M}$		BRPTS		NCTCOG	
		Wtd.	Unwtd.	Wtd.	Unwtd.	Wtd.	Unwtd.
HBW	0-15	42	45*	36	33	21*	21*
	15-30	41	39	41	44	34*	33*
	30-45	11	10	14	14	29*	29*
	45-60	4	3	5	5	9	9
	>60	3	3	4	4	8	8
HBO	0-15	63*	65*	60	61	54*	54*
	15-30	28	26	28	28	31	31
	30-45	6	6	6	6	9	10
	45-60	2	2	2	2	3	3
	>60	2	2	4	3	2	3
NHB	0-15	64*	65*	59	60	54*	53*
	15-30	27	26	29	28	29	28
	30-45	6	5	6	7	12	12
	45-60	2	2	2	2	3	3
	>60	3	2	4	4	3	3

The comparison of household structures in different areas is shown in table 26. Household structure is shown in terms of households size and vehicle ownership according to the segmentation scheme identified earlier. The comparison shows a reasonably stable household structure between the three data sets analyzed with greater similarity existing among the weighted sample values than among the unweighted values.

Table 26

Comparison of household structure

Segment number	Segment description	Percentage of households					
		NPTS $95_{500K\text{-}1M}$		BRPTS		NCTCOG	
		wtd.	unw.	wtd.	unw.	wtd.	unw.
1	zero vehicles., all hh.sizes	7	7	9	8	8	5
2	1+ vehicles., household size 1	24	19	25	18	21	23
3	1 vehicle, household size 2	9	12	8	10	7	7
4	2+ vehicles, household size 2	21	24	20	24	22	27
5	1 vehicle , household size 3+	9	8	6	5	10	7
6	2 vehicles, household size 3	6	7	9	10	8	7
7	3+ vehicles, household size 3	6	5	4	5	5	5
8	2 vehicles, household size 4+	11	12	12	12	12	11
9	3+ vehicles, household size 4+	8	7	8	8	8	7

Updating transferred data

As reported earlier, a survey was conducted as part of this study among residents of the Baton Rouge metropolitan area to test the concept of establishing a local panel and using it to adjust imported data to local conditions. The panel that was established consisted of only 108 households rather than the approximately 500 households considered necessary to provide accurate estimates of local conditions. However, the purpose of the panel in this study was to establish the process by which a panel could be constructed, rather than to establish a functional panel. The small panel established in this study achieved the purpose of testing procedures of efficiently gathering local data but did not provide accurate estimates of local conditions. At the same time, in order to demonstrate how panel data can be used to update transferred data, panel data was used to update NPTS 95 data and was then compared with the values from the BRPTS survey.

The Bayesian updating process described earlier was used to update transferred data with values from the panel. Data from NCTCOG and data from NPTS 95 for MSAs with populations between 500,000 and 1,000,000 were used as transfer data. Since the Baton Rouge panel survey provided only unweighted trip rates, it was used to update the unweighted trip rates from the NCTCOG and NPTS 95 data. To obtain trip rates from the panel survey, observed trip rates in each category were weighted by the proportion of the population who belong to each category. Data from NPTS 95 data for households in MSAs with populations between 500,000 and 1,000,000 were used to estimate these proportions.

As a demonstration of the updating process using multiple data sets, trip rates from both the NCTCOG and NPTS 95 for metropolitan areas with populations between 500,000 and 1,000,000 were updated with trip rates from the Baton Rouge panel. The results are shown in table 27 along with the trip rates from the BRPTS data for comparison. The results show that the resulting trip rates are very close to those observed in the BRPTS. The panel has a relatively small impact on the updated values because of its small sample size. The updated values are not significantly different from the BRPTS values at the 0.05 level of significance.

Table 27

Comparison of updated and observed unweighted trip rates

Trip purpose	Trips/household/day				
	NPTS $95_{500K\text{-}1M}$	NCTCOG	BR Panel	Updated	BRPTS
All	9.7	9.5	11.1	9.6	9.7
HBW	1.5	1.6	1.6	1.6	1.6
HBO	5.0	4.7	6.1	4.8	4.9
NHB	3.0	3.2	3.3	3.1	3.2

The Bayesian updating process can be used to update mode shares as well, because share estimates are normally distributed for any share containing five or more observations *(56, p. 358)*. The standard error of the share needed in the updating process can be obtained from sample data using the following expression *(56, p. 359)*:

$$s_i = \sqrt{\frac{x_i}{n}(1-\frac{x_i}{n})} \qquad (13)$$

where,

s_i	=	standard error of the estimate of the share i.
x_i	=	number of observations of mode i in the sample.
n	=	number of observations in the sample.

Using NPTS 95 data for metropolitan areas with populations between 500,000 and 1,000,000 as the transfer source, mode shares were transferred to Baton Rouge and updated with the Baton Rouge panel. The results are shown in table 28 together with the values from the BRPTS. The updated values that are statistically different from the BRPTS data values at the 0.05 level of significance using equation 12 are shown with an asterisk in table 28.

The results show that the updated mode shares are, for the most part, very similar to those observed in the BRPTS. Of the fifteen mode share comparisons, four are significantly different from the BRPTS values at the 0.05 level of significance. The updated values evidently are very similar to those of the NPTS 95 for metropolitan areas with populations of 500,000 to 1,000,000. The reason is that the updating process weights the estimates from the individual sources in relation to their variance, the variance is inversely proportional to the sample size on which the estimate is based. Thus, the small sample size for the Baton Rouge panel results in a large variance of the estimate, leading to a low weight in the updating process. The weights could be manually manipulated, although the need did not appear in the estimate of mode share. However, in the estimation of trip length frequencies, the impact of giving the Baton Rouge panel estimates the same weight as those from the NPTS 95 data was investigated.

Table 28

Updated mode shares

Trip purpose	Mode	Mode share			
		NPTS 95	BRPS	Updated	BRPTS
HBW	Auto driver	0.88	0.65	0.88	0.90
	Auto pass.	0.08	0.34	0.09	0.09
	Bus	0.02	0.00	0.02	0.01
	Walk	0.02	0.01	0.02*	0.01
	Taxi	0.00	0.00	0.00	0.00
HBO	Auto driver	0.61	0.43	0.61	0.63
	Auto pass.	0.28	0.49	0.28	0.27
	Bus	0.05	0.05	0.05	0.05
	Walk	0.06	0.03	0.06*	0.04
	Taxi	0.00	0.00	0.00	0.00
NHB	Auto driver	0.70	0.55	0.70*	0.73
	Auto pass.	0.23	0.41	0.23	0.22
	Bus	0.02	0.01	0.02	0.02
	Walk	0.05	0.04	0.05*	0.03
	Taxi	0.00	0.00	0.00	0.00

Trip length frequencies from NPTS 95 data for metropolitan areas with populations of 500,000 to 1,000,000 were updated with data from the Baton Rouge panel. The results are shown in table 29. As before, values that differ significantly from the BRPTS values at the 0.05 level of significance are shown with asterisks in the table. The updated values are generally different to those observed in the BRPTS survey. Twelve of the fifteen percentages compared are significantly different at the 0.05 level of significance.

Table 29
Updated trip length distributions

Trip purpose	Travel time interval (mins.)	Percentage of trips in time interval			
		NPTS 95	BRPS	Updated	BRPTS
HBW	0-15	45	15	45*	33
	15-30	39	35	39*	44
	30-45	10	32	10*	14
	45-60	3	10	3*	5
	>60	3	8	3*	4
HBO	0-15	65	29	65*	61
	15-30	26	39	26*	28
	30-45	5	15	6	6
	45-60	2	6	2	2
	>60	2	11	2*	3
NHB	0-15	65	39	65*	60
	15-30	26	0.41	26*	28
	30-45	5	0.01	5*	7
	45-60	2	0.04	2	2
	>60	2	0.00	2*	3

The weight that values from the Baton Rouge panel receive in the updating process can, as previously mentioned, be manually manipulated. For example, if it is assumed that the panel receives one-half the weight given to data from NPTS 95, then the results in table 30 are obtained. As can be seen, the values are now considerably closer to those of the BRPTS survey although 10 of the 15 values are still significantly different from the BRPTS values.

Table 30
Updated trip length distribution with manual adjustment of weights

Trip purpose	Travel time interval (mins.)	Percentage of trips in time interval			
		NPTS 95	BRPS	Updated	BRPTS
HBW	0-15	45	15	35	33
	15-30	39	35	38*	44
	30-45	10	32	18*	14
	45-60	3	10	5	5
	>60	3	8	5	5
HBO	0-15	65	29	53*	61
	15-30	26	39	30*	28
	30-45	5	15	9*	6
	45-60	2	6	3*	2
	>60	2	11	5*	3
NHB	0-15	65	39	56*	60
	15-30	26	0.41	29	28
	30-45	5	0.01	8*	7
	45-60	2	0.04	3*	2
	>60	2	0.00	3	3

Manual adjustment of the weights in the updating process must be based on sound reasoning in order to be justified. For example, it may be justified to give greater weight to local data than is implied by the variance of estimates from the data because they reflect local conditions. It also may be the case that one data set is considered more appropriate to a particular area due to similarity of conditions or vintage of the data between the two areas. However, no guidance is available on how weights may be appropriately adjusted; it is a matter of judgement on the part of the analyst.

CONCLUSIONS

This study set out to investigate the transferability of metropolitan transportation planning data, the potential of a smal,l local panel to provide updating capabilities to transferred data, and the use of GPS in data collection for transportation planning. The conclusions that can be drawn from conducting this investigation are as follows:

1. Considerable existing data are available that can be usefully applied to metropolitan transportation planning. More data are available today than ever before, since all public data sets providing transportation information have either been maintained since their inception (e.g. Census, NPTS), replaced with comparable data sets (e.g. American Travel Survey replacing the previous National Travel Survey), or initiated as new data sets (e.g. the Commodity Flow Survey and the American Community Survey). In addition, while relatively few transportation surveys were conducted in the 1980s, numerous metropolitan travel surveys were conducted during the 1990s. These transportation surveys comprise a valuable source of transfer data.

2. Most existing data are available free of charge and are readily accessible. The Bureau of Transportation Statistics of the Federal Highway Administration maintains a directory of transportation data sources and administers many of the existing data sets available. Other important sources of publicly-administered data are the Federal Transit Administration, the Bureau of Economic Analysis of the Department of Commerce, the Bureau of Labor, and pipeline and waterway inventory data, collected by the Oak Ridge National Laboratory.

3. Data considered for transfer include simple relationships, distributions, proportions, and patterns in the data. These aggregate expressions of data can be used to estimate some traditional travel demand models and update others. The expressions can be transferred from one or more external sources and can be updated with local data.

4. The use of a panel to provide current, local updating of transfer data appears feasible. A panel of approximately 500 households with each repeat survey being conducted no less frequently than every year would be necessary for this purpose. Based on the cost of other panels, maintaining a panel of this size and conducting annual surveys would cost approximately $25,000 per year. Further testing is required to determine whether

a panel of this nature could be drawn from urban households across the state and function adequately or whether a panel will be required in each metropolitan area.

5. The use of GPS to identify automobile travel appears feasible, although further testing is required to test its performance relative to traditional travel survey methods. Initial testing indicates that a more complete and accurate record of travel may be obtained with GPS, than by conventional methods. However, some aspects of the operation of the instrument and the process of retrieving trip attribute data in a subsequent contact with the traveler require further study before a proper comparative assessment can be made. Some trips are lost with GPS when the traveler forgets to turn on the instrument or information on portions of the trip is forfeited when the instrument loses contact with the satellites due to tree canopies or building canyons. However, initial observation suggests that these problems may be small and the benefit of obtaining accurate, geocoded travel information may far outweigh the current limitations of the process. A greater question is how successfully supplemental information on trips identified by GPS, may be obtained in follow-up contact with the traveler. The traveler must identify who was the driver, the number of occupants of the vehicle, and the trip purpose of the driver and each occupant. Further research is required to test this aspect of the GPS travel data collection process.

RECOMMENDATIONS

The research conducted in this study was an exploratory investigation into methods of cost-effective data collection for metropolitan transportation planning. The initial investigation has determined that several procedures do provide cost and time savings while others require further testing before their benefit can be determined. Subsequently, it is recommended that the following subject areas receive further research:

1. The establishment of a panel to provide up-to-date, local information should be tested further. Specifically, it must be determined whether a statewide panel will suffice or whether local panels within each individual metropolitan area are required. Panel size should be more definitively determined and the issue of survey frequency must be investigated further, taking into account cost, panel attrition, and the sense of continued involvement required of participants. The possibility that the panel may be used for other purposes in the metropolitan area also needs to be investigated.

2. The use of GPS in the collection of travel data should be investigated further. In particular, the issue of retrieving trip attribute data from respondents following return of the GPS instrument, the matter of recording all trips, the interpretation of the track point data into the correct number of trips, and the potential for motorists to feel that a GPS that monitors every move they make in their vehicle is too great an invasion of their private life all need to be studied further.

3. The methods investigated in this study should be applied to a metropolitan area on a trial basis to test the efficiency of the procedures in an actual application. The application should be conducted in parallel to the transportation planning activity normally conducted using conventional procedures. The results from the two procedures should be compared together with a review of the effort and time involved in each case.

REFERENCES

1. Richardson, A.J.; Ampt, E.S.; and Meyburg, A.H., *Survey Methods for Transport Planning*. Eucalyptus Press, Parkville, Australia, 1995.

2. Karash, K.H. and Schweiger, C., "Identification of Transportation Planning Data Requirements in Federal Legislation", Final Report to the John A. Volpe National Transportation Systems Center, U.S. DOT, Report DOT-T-94-21, July 1994.

3. Transportation Research Board, *Transportation Data Needs: Programs for a New Era*, Transportation Research Circular No. 407, April 1993.

4. TMIP, *TMIP Newsletter*, Issue 10, December 1999, Texas Transportation Institute, Arlington, TX, pp. 4-5.

5. Bradley, K. "Directory of Transportation Data Sources," Report Number DOT-VNTSC-BTS-95-1, Bureau of Transportation Statistics, U.S. Department of Transportation, March, 1995.

6. Institute of Transportation Engineers (ITE), "Use of Census Data in Transportation Planning," ITE Publication No. IR-011B, 1987.

7. Christopher, E.J. "Census Data Use in Illinois by a Large Metropolitan Planning Organization," *Transportation Research Board Conference Proceedings No. 13,* Transportation Research Board, Washington D.C., 1996, pp. 21-32.

8. Stopher, P.R. and Metcalf, H.M.A. "Methods for Household Travel Surveys." *NCHRP Synthesis of Highway Practice No. 236*, Transportation Research Board, Washington, D.C., 1996.

9. Stopher, P.R. "Household Travel Surveys: Cutting-Edge Concepts for the Next Century." *Transportation Research Board Conference Proceedings No. 10,* Transportation Research Board, Washington, D.C., 1996, pp. 12-14.

10. Stecher, C.C.; Bricka, S.; and Goldenberg, L. "Travel Behavior Survey Data Collection Instruments." *Transportation Research Board Conference Proceedings No. 10,* Transportation Research Board, Washington, D.C., 1996, pp. 154-169.

11. Stopher, P.R. "Use of an Activity-Based Diary to Collect Household Travel Data." *Transportation*, Vo. 19, 1992, pp.159-176.

12. Axhausen, K.W. "Travel Diary Content: A Brief Review of Its Development," Presented to the 74th Annual Meeting of the Transportation Research Board, Washington, D.C., 1995.

13. Kish, L. *Survey Sampling*, John Wiley & Sons, New York, 1967, pp. 469-477.

14. Murakami, E., and Watterson, W.T. "Developing a Household Travel Panel Survey for the Puget Sound Region." *Transportation Research Record No. 1285*, 1990, pp.40-46.

15. Van Wissen, L.G.J. and Meurs, H.J. "The Dutch Mobility Panel: Experience and Evaluation." *Transportation*, Vol. 16, 1989, pp. 99-119.

16. Raimond, T., and Hensher, D.A. "A Review of Empirical Studies and Applications," in *Panels for Transportation Planning: Methods and Applications*, edited by T.F. Golob, R. Kitamura, and L. Long, Kluwer Academic Publishers, Boston, 1997, pp. 15-74.

17. Kish, L. "Timing of Surveys for Public Policy." *Australian Journal of Statistics*, Vol . 28, No. 1, 1985, pp. 1-12.

18. Kitamura, R. "Panel Analysis in Transportation Planning: An Overview," *Transportation Research*, 1990, Vol. 21A, No. 6, pp. 401-415.

19. Cambridge Systematics, Inc., *Travel Survey Manual*, Prepared for U.S. DOT and U.S. EPA, TMIP, Washington, D.C., July 1996.

20. Transportation Research Board, "Conference on Household Travel Surveys: New Concepts and Research Needs," *Conference Proceedings 10*, Transportation Research Board, Washington, D.C., 1996.

21. Battelle Transportation Division, *Global Positioning Systems for Personal Travel Surveys: Lexington Area Travel Data Collection Test*, Report to the Federal Highway Administration, U.S. DOT, Washington, D.C., September 1997.

22. Jack Faucett Associates with COMSIS Corporation and Mid-Ohio Regional Planning Commission, "Guidance Manual for Managing Transportation Planning Data," National Cooperative Highway Research Program Report 401, Transportation Research Board, Washington, D.C., 1997.

23. Jack Faucett Associates with COMSIS Corporation and Mid-Ohio Regional Planning Commission, "Multimodal Transportation Planning Data," Final Report prepared for National Cooperative Research Program, Transportation Research Board, Washington, D.C., 1997.

24. Jack Faucett Associates with COMSIS Corporation and Mid-Ohio Regional Planning Commission, "Compendium of Data Collection Practices and Sources," Project 8-32(5), prepared for National Cooperative Research Program, Transportation Research Board, Washington, D.C., 1997.

25. Hill, D.M., and von Cube, H.G. "Development of a Model for Forecasting Travel Mode Choice in Urban Areas," Highway Research Record 38, Washington D.C., 1963.

26. Hill, D.M., and Dodd, N. "Studies of Travel Between 1954 and 1964 in a Large Metropolitan Area," Highway Research Record 141, Washington D.C., 1966.

27. Koppelman, F. S.; Kuah, G; and Wilmot, C.G. "Transfer Model Updating with Disaggregate Data," *Transportation Research Record 1037*, Transportation Research Board, Washington D.C., 1985, pp. 102-107.

28. Walker, W. T.; Brady, S.H.; and Taylor, C. "Updating Existing Travel Simulation Models with Small-sample Survey Data Using Parameter Scaling Methods," *Transportation Research Record 1607*, Transportation Research Board, Washington D.C., 1997, pp. 55-61.

29. Institute of Transportation Engineers (ITE), "Urban Travel Characteristics Database," An

Informational Report of the Institute of Transportation Engineers, April, 1995.

30. Federal Transit Administration (FTA), "Characteristics of Urban Transportation Systems," Revised edition, September, 1992.

31. Institute of Transportation Engineers (ITE), "Trip Generation," Third printing, 1991.

32. Oi, W. Y., and Shuldiner, P.W. "An Analysis of Urban Travel Demands." Northwestern University Press, Evanston, Illinois, 1962.

33. Tardiff, T.J. "Specification Analysis for Quantal Choice Models." *Transportation Science*, Vol. 13, No. 3, 1979.

34. Koppelman, F.S., and Wilmot, C.G. "The Effect of Omission of Variables on Choice Model Transferability," Transportation Research, Part B: Methodological, Vol. 20 B, No. 3, June, 1986, pp. 205-213.

35. Ben-Akiva, M.E., and Lerman, S.R. *Discrete Choice Analysis: Theory and Application to Travel Demand.* MIT Press, Cambridge, Massachusetts, 1985.

36. McFadden, D. "Properties of the Multinomial Logit (MNL) Model," Working Paper No. 7617, Urban Travel Demand Forecasting Project, Institute of Transportation Studies, University of California, Berkeley, September, 1976, pp. 26-28.

37. Atherton, T.J., and Ben-Akiva, M.E. "Transferability and Updating of Disaggregate Travel Demand Models," *Transportation Research Record 610*, Transportation Research Board, Washington D.C., 1976. pp. 12-18.

38. Maddala, G.S. *Econometrics*, McGraw-Hill, New York, 1977, p.412.

39. Hensher, D.A. "Longitudinal Surveys in Transport: An Assessment," edited by E.S. Ampt, A.J. Richardson, and W. Brög, *New Survey Methods in Transport*, VNU Science Press, The Netherlands, 1985, pp. 77-97.

40. Murakami, E. and Ulberg, C. "The Puget Sound Transportation Panel," in *Panels for*

Transportation Planning, edited by T.F. Golob, R. Kitamura, and L. Long, Kluwer Academic Publishers, Boston, 1997, pp. 159-192.

41. Horowitz, J.L. "Accounting for Response Bias," in *Panels for Transportation Planning,* edited by T.F. Golob, R. Kitamura, and L. Long, Kluwer Academic Publishers, Boston, 1997., pp. 207-210.

42. Brownstone, D. and Chu, X. "Multiply-Imputed Sampling Weights for Consistent Inference with Panel Attrition," in *Panels for Transportation Planning,* edited by T.F. Golob, R. Kitamura, and L. Long, Kluwer Academic Publishers, Boston, 1997, pp. 259-276.

43. Meurs, H. and Ridder, G. "Attrition and Response Effects in the Dutch National Mobility Panel," in *Panels for Transportation Planning,* edited by T.F. Golob, R. Kitamura, and L. Long, Kluwer Academic Publishers, Boston, 1997, pp. 211-232.

44. Pendyala, R.M., and Kitamura, R. "Weighting Methods for Attrition in Choice-Based Panels," in *Panels for Transportation Planning,* edited by T.F. Golob, R. Kitamura, and L. Long, Kluwer Academic Publishers, Boston, 1997, pp. 233-258.

45. AMPG, *Report on the Pretests of Survey Instruments and Procedures for the North Central Texas Council of Governments Household Travel Survey,* Unpublished report to the North Central Texas Council of Governments, Arlington, TX, 1995.

46. Tooley, M.S. "Incentives and Rates of Return for Travel Surveys." *Transportation Research Record No.1551,* Transportation Research Board, Washington, D.C., 1996, pp. 67-73.

47. Richardson, A.J.; Ampt, E.S.; and Meyburg, A.H. "Nonresponse Issues in Household Travel Surveys," *Transportation Research Board Conference Proceedings 10,* Transportation Research Board, Washington, D.C., 1996, pp. 79-114.

48. Paaswell, R.E. "Why Panels for Transportation Planning?" in *Panels for Transportation Planning,* edited by T.F. Golob, R. Kitamura, and L. Long, Kluwer Academic Publishers, Boston, 1997, pp. 3-14.

49. Siaurusaitis, V. J., and Saben, L.J. "Transportation Planner's Handbook on Conversion Factors for the Use of Census Data", Paper 970179, Transportation Research Board's 76th Annual Meeting, Washington D.C., January, 1997.

50. Breiman, L.; Friedman, J.H.; Olshen, R.A.; and Stone, C.J. *Classification and Regression Trees.* Wadsworth, Belmont, California, 1984.

51. Miller, I., and Freund, J.H. "Probability and Statistics for Engineers." Second Edition, Prentice-Hall, Englewood Cliffs, New Jersey, 1977.

52. Robinson, J.P. *How Americans Use Time: A Social-Psychological Analysis of Everyday Behavior.* Praeger Publishers, New York, NY, 1977, p.7.

53. Robinson, J.P., and Godbey, G. *Time for Life: The Surprising Ways Americans Use Their Time.* Pennsylvania State University Press, 1997.

54. Norusis, M.J. "SPSS for Windows Base System User's Guide Release 6.0," SPSS Inc., 1993.

55. Levinson, D.M., and Kumar, A. "Multimodal Trip Distribution: Structure and Application." *Transportation Research Record 1466*, Transportation Research Board, Washington D.C., 1994, pp. 124-131.

56. Freund, J.E., and Williams, F.J. *"Elementary Business Statistics: The Modern Approach,"* Prentice-Hall, Englewood Cliffs, New Jersey, 1982.

APPENDIX A: LIST OF CENSUS DATA FILES

U.S. Census Bureau

Agriculture:

1992 Census of Agriculture, Final County File

1992 Census of Agriculture, Final State File

1987 Census of Agriculture -- Volume 1. State (Final)

1987 Census of Agriculture -- Volume 1. County (Final)

1982 Census of Agriculture -- Volume 1. State (Final)

1982 Census of Agriculture -- Volume 1. County (Final)

1974 and 1969 Census of Agriculture -- Combined Final

Housing Surveys:

1980 Census Detail Data File for the American Housing Survey, National Sample File

American Housing Survey, National Core File

American Housing Survey, National Core and Supplemental File

American Housing Survey, Metropolitan Statistical Area (MSA) Core and Supplement Files

American Housing Survey, Standard Metropolitan Statistical Area Travel-To-Work Microdata Files

American Housing Survey, Standard Metropolitan Statistical Area Travel-To-Work Summary Files

American Housing Survey, National Travel-To-Work File, 1975

New York City Housing and Vacancy Survey: 1996, 1993, and 1991

New York City Housing and Vacancy Survey Longitudinal

New York City Housing and Vacancy Survey: 1987, 1984, 1981 and 1978 Longitudinal

New York City Housing and Vacancy Survey: Single Room Occupancy, 1987

Property Owners and Managers Survey

Compendia Data:

County and City Data Book

County Statistics, CO-STAT 4

State and Metropolitan Area Data Book, 1982

Current Construction Surveys:

Building Permits Survey Description

Building Permits Current Month File

Building Permits Monthly Cumulative File

Building Permits Annual Cumulative File

Building Permits Annual Summary File

County Business Patterns:

County Business Patterns

Current Population Survey, CPS:

Current Population Survey Description

Current Population Survey, 1998

Current Population Survey, 1997

Current Population Survey, 1996

Current Population Survey, 1995

Current Population Survey, 1994

Current Population Survey, 1993

Current Population Survey, 1992

Current Population Survey, 1991

Current Population Survey, 1980 through 1990

Current Population Survey, 1968 through 1979

1980 Decennial Censuses:

1980 Alphabetical Index of Occupations and Industries

1980 American Indian Supplemental Questionnaire, Summary Tape File

1980 American Indians, Eskimos and Aleuts for Historic Areas of Oklahoma (PC80-2-D)

1980 Census Tract Relationship File, 1970-1980

1980 Congressional District Boundary Files

1980 Congressional District Equivalency Files

1980 County-To-County Migration Flows

1980 Equal Employment Opportunity -- United States

1980 Equal Employment Opportunity -- Puerto Rico

1980 Equal Employment Opportunity -- Supplements I & II

1980 Geographic Identification Code Scheme -- GICS

1980 Group Quarters Population by Age, Sex, Race, and Spanish Origin

1980 Industry and Occupation Classification Systems 1970 - 1980

1980 Inter-County Migrant File

1980 Journey To Work (Subject Reports, PC80-2-6C, 6D, 6E)

1980 Map Index File

1980 Master Area Reference Files -- MARF

1980 Neighborhood Equivalency File

1980 Number of Workers by County of Resisence by County of Work

1980 Occupation by Industry (Subject Reports, PC80-2-7C)

1980 Characteristics (Subject Reports, PC80-2-7A)

1980 Public Law 94-171 All States -- PL 94-171

1980 Public Use Microdata Samples -- County Group Equivalency File

1981 Residential Finance Survey

124

1980 Spanish Surname List

1980 Summary Tape File 1 -- STF 1

1980 Summary Tape File 2 -- STF 2

1980 Summary Tape File 3 -- STF 3

1980 Summary Tape File 4 -- STF 4

1980 Summary Tape File 5 -- STF 5

1980 Census of Population & Housing Special Tabulations:

1980 Age, Occupation, Industry by Residency, Class of Worker for States and SMSA's

1980 Age by Race by Sex by Nativity by Spanish Origin

1980 Age, Sex and Race Groups

1980 Age, Sex and Race Groups, Part II

1980 Age and Sex for Race Groups Modified to OMB Definitions for all Census Tracts in Illinois

1980 Age, Sex, Tribal Affiliation and Spanish Surname for New Mexico and Arizona ED's and Counties

1980 Age, Sex and Ethnicity by Census Tracts in Los Angeles, California

1980 Characteristics of Teachers

1980 Detailed Occupation by Industry, Earnings and Education

1980 Department of Defense Mobilization Analysis -- STF D1

1980 Department of Defense Recruitment Analysis -- STF D2

1980 Department of Defense EEO by Age -- STF D3

1980 Detailed Characteristics of Single Female Parents and Displaced Homemakers

1980 Detailed Occupation by Industry by Class of Worker for Service Delivery Areas in Minnesota

1980 Detailed Race by Census Tracts in New York City -- STF 2D

1980 Detailed Race by Census Tracts in New York City -- STF 4D

1980 Households by Income and Age of Householder; and Owner-occupied units by Value and Age of

Householder

1980 Migration Data for Selected Counties in California -- STF 5

1980 Poverty Level and Children -- STF S-1

1980 Public Use Microdata Sample F, PUMS-F

1980 Occupations by Earnings by Education

1980 Recode to Scool Districts -- STF 1F

1980 Recode to Scool Districts -- STF 3F

1980 Selected Household Statistics by ZIP Code for California

1980 Selected Housing Tallies for ZIP Code in Tennessee Valley Authority States

1980 Sex and Spanish Origin by Citizenship and Occupation for Selected Areas in the United States

1980 STF 2, 3 & 4 for Labor Market Areas in Maine

1990 Decennial Census of Population and Housing:

Public Law 94-171

1990 Census of Population and Housing Redistricting File -- Puerto Rico

1990 Summary Tape file 1 and Related Products

1990 Summary Tape File 1A -- STF1A

1990 Summary Tape File 1B -- STF1B

1990 STF1B Geographic Header File Special Auxiliary File

1990 Summary Tape File 1C -- STF 1C

1990 Summary Tape File 1D -- STF 1D

1990 Summary Tape File 2 -- STF 2 and Related Products

1990 Summary Tape File 2A -- STF 2A

1990 Summary Tape File 2B -- STF 2B

1990 Summary Tape File 2C -- STF 2C

1990 Summary Tape File 3 -- STF 3 and Related Products

1990 Summary Tape File 3A -- STF 3A

1990 Summary Tape File 3B -- STF 3B

1990 Summary Tape File 3B ZIP Codes/Block Equivalency

1990 Summary Tape File 3C -- STF 3C

1990 Summary Tape File 3D -- STF 3D

1990 Summary Tape File 3X -- STF 3X

1990 Summary Tape File 3A Record Sequence Example File

1990 Summary Tape File 4 -- STF 4 and Related Products

1990 Summary Tape File 4A -- STF 4A

1990 Summary Tape File 4B -- STF 4B

1990 Summary Tape File 4C -- STF 4C

1990 Summary Tape File 420 Journey to Work -- STF 420

1990 Summary Tape File 420 Place of Work 20 Destination File -- Puerto Rico

1990 Census Equal Employment Opportunity (EEO) Files

1990 Census Equal Employment Opportunity (EEO) Supplemental Tabulation File -- Part 1 Private

1990 Census Equal Employment Opportunity (EEO) Supplemental Tabulation File -- Part 2 Federal

1990 Summary Tape File User-Defined Areas Program -- UDAP

1990 Census of Population and Housing: Population and Housing Counts -- STF-S-1

1990 Census of Population: Number of Workers by County of Residence by County of Work, STF-S- 5

1990 Census of Population and Housing: Modified Age/Race, Sex and Hispanic Origin Files -- MARS

1990 Census Transportation Planning Package

1990 Census Special Tabulation Program -- Veterans Administration STP 5

1990 Census Special Tabulation Program -- Aging STP 14

1990 Census Special Tabulation Program -- Housing and Urban Development Block Group File STP 19

127

1990 SSTF-1 Foreign Born Population of the United States

1990 SSTF-2 Ancestry of the Population in the United States

1990 SSTF-3 Persons of Hispanic Origins in the United States

1990 SSTF-4 Characteristics of Adults with Work Disabilities, Mobility Limitations, or Self-Care Limitations

1990 SSTF-5 Asian and Pacific Islander Population in the United States

1990 SSTF-6 Education in the United States

1990 SSTF-7 Metropolitan Housing Characteristics

1990 SSTF-8 Housing the Elderly

1990 SSTF-9 Housing Characteristics of New Units

1990 SSTF-10 Mobile Homes

1990 SSTF-11 Language Use In the United States

1990 SSTF-12 Employment Status, Work Experience and Veteran Status

1990 SSTF-13 Characteristics of American Indians by Tribe and Language

1990 SSTF-14 Occupation and Industry

1990 SSTF-15 Geographic Mobility for Metropolitan Areas

1990 SSTF-16 Fertility

1990 SSTF-17 Poverty Areas in the United States

1990 SSTF-18 Condominium Housing

1990 SSTF-19 Older Population of the United States

1990 SSTF-20 Journey to Work in the United States

1990 SSTF-21 Characteristics of the Black Population

1990 SSTF-22 Earnings by Occupation and Education

1990 Automated Industry and Occupation Coder

1991 Residential Finance Survey

128

1988 Dress Rehearsal Census of St. Louis City, East Central, MO and Eastern Washington STF 1A

1986 PL 94-171, 1986 Test Census of Central Los Angeles County, CA

Economic Censuses & Surveys:

City Reference File 1992

Economic Censuses, 1987

Economic Censuses, 1982

Economic Censuses, 1977

Economic Censuses, 1972

Enterprise Statistics, IRS Link

Fuels and Electrical Energy Consumption, 1967 - 1982

Manufactures' Shipments, Inventories, and Orders -- M3 Published & Unpublished January 1958 - March 1993

Annual Survey of Manufactures - Fuels and Electrical Energy Consumed

Annual Survey of Manufactures - Fuels and Electrical Energy Nonpurchased Fuels Consumed, 1978

Annual Survey of Manufactures - Hydrocarbon, Coal and Coke 1980

Research and Development in Industry -- Correlation Matrices 1957 - 1977

Research and Development in Industry -- 1957 - 1977

Research and Development in Industry -- 1958 - 1983

Geographic ReferenceFiles:

Ce nsus Tract Comparability Files (Pre- 1990)

Ci ty Reference Files - 1977, 1982, 1983, 1987, 1992

Co ngressional District Boundary Files (Pre- 1990)

Co ngressional District Equivalency Files (Pre- 1990)

Co ntiguous County File

19 79 County and MCD by ZIP Code

Governments:

Annual Survey of Governments, Employment

Annual Survey of Governments, Finance

Annual Survey of Governments, Education Finance (File C)

Annual Survey of Governments, Employment Retirement Systems (File D), 1988

Consolidated Federal Funds Report

Census of Governments, 1987

Census of Governments, 1982

Census of Governments, 1977

Directory of Governments, 1988 (Names and Addresses)

Federal Assistance Award Data System

International Statistics

International Data Base World Population

Women in Development IV

Miscellaneous:

English Language Proficiency Study 1982(EPLS)

Survey of Fishing, Hunting and Wildlife - Associated Recreation, 1980 and 1996

Survey of Natural and Social Scientists and Engineers: 1982, 1984, 1986, and 1989

National Survey of College Graduates: 1993 and 1995

Population Estimates and Projections:

National Population Projections by Age, Sex, Race, and Hispanic Origin: 1995 to 2050 (PE-37)

County Population Estimates (Revised) by Age, Sex and Race: 1980-1989

County Population Estimates by Age, Sex and Race: 1991

Population Projections for State by Age, Sex and Race: 1993 to 2020 -- Detailed

County Population Estimates (Experimental) by Age, Sex and Race: 1980-1985

County Population Estimates (Experimental) by Age, Sex and Race: 1980, 1982 & 1984

County Population Estimates (Provisional) by Age, Sex and Race: 1980-1982

Federal-State Cooperative Program, 1982 Population Estimates

Intercensal Estimates of the Population of Counties by Age, Sex and Race: 1970-1980

Intercensal State & County Population Estimates: 1970-1980

Population & Per Capita Income Estimates for Governmental Units

Population Estimates by County with Components of Change: 1981-1987 (Provisional)

Projections of National Population by Age, Race and Sex

Projections of State Populations by Age, Race and Sex

Projections of the Spanish Population of the U.S.: 1983-2080

Public Use Microdata Samples -- PUMS:

1990 Public Use Microdata Samples (PUMS) -- Description

1990 Public Use Microdata Samples (PUMS) -- 1 Percent

1990 Public Use Microdata Samples (PUMS) -- 3 Percent (Elderly)

1990 Public Use Microdata Samples (PUMS) -- 5 Percent

1980 Public Use Microdata Samples (PUMS) -- Description

1980 Public Use Microdata Samples (PUMS A) -- 0.1 Percent National

1980 Public Use Microdata Samples (PUMS A) -- 5 Percent

1980 Public Use Microdata Samples (PUMS B) -- 1 Percent

1980 Public Use Microdata Samples (PUMS B) -- 0.1 Percent National

1980 Public Use Microdata Samples (PUMS C) --1 Percent

1980 Public Use Microdata Samples (PUMS C) --1 Percent National

1980 Public Use Microdata Samples (PUMS A, B and C) -- Puerto Rico

1980 American Indian Supplemental Questionnaire Public Use Microdata Samples

1980 Public Use Microdata Samples County Group Equivalency File

Pre-1980 Public Use Microdata Samples

Survey of Income & Program Participation --SIPP

Survey of Income & Program Participation -- Description

1996 Wave 1 and 2 Person Month

1993 Longitudinal Panel File

1993 SIPP Panel List

1992 Panel, SIPP, Wave 10 Longitudinal File

1992 Panel, SIPP, Waves 1-7 Longitudinal File

1992 SIPP Panel List

1991 SIPP Panel List

1991 Longitudinal Panel File

1990 SIPP Panel List

1990 Panel: Waves 1 Through 5 Longitudinal Panel File

1990 Panel: Longitudinal Full Panel File (All Waves)

1989 SIPP Panel List

1988 SIPP Panel List

1988 Full Panel Research File (Longitudinal)

1987 SIPP Panel List

1987 Full Panel Research File (Longitudinal)

1986 SIPP Panel List

1986 Full Panel Research File (Longitudinal)

1985 SIPP Panel List

1985 Full Panel Research File (Longitudinal)

1984 SIPP Panel List

1984 SIPP Annual Weights

1984 Full Panel Research File (Longitudinal)

1984 SIPP Wave 7 & 1983 Wave 3 Rectangular Core & Topical Module

SIPP on Call

Topologically Intergrated Geographic and Encoding Referencing System -- TIGER

1992 TIGER/Census Tract Street Index -- Version 2 (CSTI Version 2)

1992 TIGER/Line Files

1992 TIGER/Geographic Names File

1990 TIGER/Line Files

1990 TIGER/American Indian/Alaska Native Areas FIle

1990 TIGER/Census Tract/Block Numbering Area File

1990 TIGER/Line County Boundary Files

1990 TIGER/Line County Subdivision File

1990 TIGER/103rd Congressional District Boundary Files

1990 TIGER/103rd Congressional District Block Equivalency File

1990 TIGER/103rd Congressional District Geographic File

1990 TIGER/Census Tract Comparability File

1990 TIGER/GICS -- Geographic Identification Code Scheme

1990 TIGER/Geographic Reference File -- Names

1990 TIGER/Line Mapping Software

1990 TIGER/Map Sheet Corner Point Coordinate File (Census Version)

1990 TIGER/Urbanized Areas Limit File

Last update: 08/04/98 kfm

APPENDIX B: CENSUS STATE DATA CENTERS, LOUISIANA

Census State Data Centers: Louisiana

Louisiana State Census Data Center
Office of the DataBase Commission (ODBC)
P.O. Box 94095
Baton Rouge, LA 70804
150 N. 3rd Street
Baton Rouge, LA 70802
Ms. Karen Paterson
(225) 219-4025
FAX (225) 219-4027
webmaster@doa.state.la.us
kpaters@doa.state.la.us

Division of Business and Economic Research
University of New Orleans
Lake Front
New Orleans, La 70148
Ms. Pat Conner
(504) 280-7326
v1mdb@uno.edu

Division of Business Research
Louisiana Tech University
P.O. Box 10318
Ruston, LA 71272
Ms. Vickie Blue
(318) 257-3701
blue@cab.latech.edu

Reference Department
Louisiana State Library
P.O. Box 131
Baton Rouge, LA 70821
Ms. Virginia Smith
(504) 342-4914
vsmith@pelican.state.lib.la.us

The Louisiana Population Data Center
Department of Sociology
Room 126, Stubbs Hall
Louisiana State University
Baton Rouge, LA 70803-5411
Mr. F. Andrew Deseran
(504) 388-1113

136

adeseran@lapop.lsu.edu

Center for Business and Economic Research
University of Louisiana at Monroe
Monroe, LA 71209-0101
Dr. Jerry Wall
(318) 342-1215
brwall@ulm.edu
Louisiana Electronic Assistance Program

Print outs, extracts, tape copies, and software for accessing and downloading Census data.

CD-ROM products and services

Online data services

News letter, technical journal

Foreign Trade

Mapping capabilities

Ability to copy Census maps

TIGER

APPENDIX C: AMERICAN COMMUNITY SURVEY

About the American Community Survey

What is the American Community Survey?

The American Community Survey is a new approach for collecting accurate, timely information needed for critical government functions. This new approach provides accurate, up-to-date profiles of America's communities every year. Community leaders and other data users will have timely information for planning and evaluating public programs for everyone from newborns to the elderly.

The decennial census has two parts: 1) it counts the population; and 2) for the administration of federal programs and the distribution of billions of federal dollars, it obtains demographic, housing, social, and economic information by asking a 1-in-6 sample of households to fill out a "long form."

Since this is done only once every 10 years, long-form information becomes out of date. Planners and other data users are reluctant to rely on it for decisions that are expensive and affect the quality of life of thousands of people. The American Community Survey is a way to provide the data communities need every year instead of once in ten years. It is an on-going survey that the Census Bureau plans will replace the long form in the 2010 Census.

Full implementation of the survey would begin in 2003 in every county of the United States. The survey would include three million households. Data are collected by mail and Census Bureau staff follow up those who do not respond.

The American Community Survey will provide estimates of demographic, housing, social, and economic characteristics every year for all states, as well as for all cities, counties, metropolitan areas, and population groups of 65,000 people or more.

For smaller areas, it will take three to five years to accumulate sufficient sample to produce data for areas as small as census tracts. For example, areas of 20,000 to 65,000 can use data averaged over three years. For rural areas and city neighborhoods or population groups of less than 20,000 people, it will take five years to accumulate a sample that is similar to that of the decennial census. These averages can be updated every year, so that eventually, we will be able to measure changes over time for small areas and population groups.

Goals of the Program

The goals of the American Community Survey are to:

- Provide federal, state, and local governments an information base for the administration and evaluation of government programs.
- Improve the 2010 Census.
- Provide data users with timely demographic, housing, social, and economic data updated every year that can be compared across states, communities, and population groups.

Why Are We Doing the American Community Survey?

Data users have asked for timely data that provide consistent measures for all areas. Decennial sample data are out-of-date soon after they are published, about two years after the census is taken. Their usefulness declines every year thereafter. Yet billions of government and business dollars are divided among jurisdictions and population groups each year based on their social and economic profiles in the decennial census.

The American Community Survey can identify changes in an area's population and give an up-to-date statistical picture when data users need it, every year, not just once in ten years. Communities can use the data, to track the well-being of children, families, and the elderly; determine where to locate new highways, schools, and hospitals; show a large corporation that a town has the workforce the company needs; evaluate programs such as welfare and workforce diversification; and monitor and publicize the results of their programs.

The American Community Survey is conducted using the best mail self-response techniques of the decennial census combined with follow-up techniques that produce high-quality data. For households that do not respond by mail, the quality of data is improved by using well-trained, permanent interviewer staff using computerized interviewing, which incorporates edits into the collection process. Using a permanent coding staff provides additional improvements in data quality.

As an on-going survey, the American Community Survey is a flexible vehicle, capable of adapting to changing customer needs. Once it is fully implemented, the potential is there to add questions of national policy interest or specialized supplements to help identify the situations of special population groups.

How does the American Community Survey work?

The American Community Survey:

- Uses the Master Address File (MAF), a complete listing of all residential addresses and group quarters in the country, for sample selection.
- Mails or delivers American Community Survey questionnaires each month to sample addresses.
- Uses commercial vendor lists to obtain telephone numbers for addresses that did not mail back their American Community Survey questionnaires and conducts telephone interviews.
- Selects a one-in-three sample of the addresses still not interviewed and conducts personal interviews.
- Improves the infrastructure for the federal statistical system by providing customized samples for subpopulations of interest, by providing the ability to increase sample sizes in the American Community Survey, and by providing a vehicle for collecting data on supplemental topics for population groups or specific geographic areas.

Master Address File

The Census Bureau maintains a national Master Address File (MAF). The MAF was constructed by a computer match of the U.S. Postal Service (USPS) Delivery Sequence File (DSF), the 1990 Census Address Control File (ACF), and the Topologically Integrated Geographic Encoding and Referencing (TIGER) files. Thereafter, periodic updates from the USPS DSF, census surveys, and field listing activities keep the MAF current.

The MAF can be created automatically for all areas that have city-style address systems where the mail is delivered using these addresses. For areas that do not have a city-style address system, the Census Bureau creates a MAF by conducting an address listing operation.

The MAF will be used as a sampling frame for the American Community Survey, as well as all of the Census Bureau's demographic surveys.

A critical element in the overall success of the ACS is the ability to keep the Census Bureau's MAF up-to-date and accurate from year to year, especially in rural areas. The MAF serves as the main source of the housing unit sample for the ACS. In addition, the housing unit counts contained in the MAF play an important part in the editing, weighting, and data tabulation process. Thus, the overall accuracy of the MAF is a paramount concern.

The need for an up-to-date MAF spawned the development of a new program called the American Community Survey - Coverage Program (formerly called the Community Address Updating System). This program, which is currently under development, has two major objectives:

- To obtain address information about new housing units and add those units to the MAF; and
- To correct and update the existing addresses in the MAF.

Sample Selection

Each month, we will select a systematic sample of addresses from the most current MAF for the American Community Survey. The sample will represent the entire United States. Each month, a sample will be randomly selected. No address will receive the American Community Survey questionnaire more than once in any five-year period.

A larger proportion of addresses will be sampled for small governmental units (American Indian reservations, counties, and towns). The monthly sample size is designed to approximate the sampling ratio of Census 2000, including the oversampling of small governmental units.

Data Collection

The American Community Survey will be conducted using three methods of data collection to contact households: Self-enumeration through mail-out/mail-back; Computer Assisted Telephone Interviewing (CATI); and Computer Assisted Personal Interviewing (CAPI).

Self-enumeration through mail-out/mail-back methodology--The self-enumeration procedure uses several mailing pieces: a prenotice letter, the American Community Survey questionnaire, and a reminder card. A replacement questionnaire will be mailed to addresses in the sample if the original questionnaire is not completed and returned to the processing office within the prescribed amount of time. Sample addresses that do not respond by mail will be contacted using the follow-up procedures CATI, CAPI, or both.

Computer Assisted Telephone Interviewing (CATI)--The CATI operation is conducted approximately six weeks after the American Community Survey questionnaire is mailed. We will attempt to obtain telephone numbers and conduct telephone interviews for all households that do not respond by mail. Census Bureau telephone interviewing staff will conduct these interviews.

Computer Assisted Personal Interviewing (CAPI)--Following the CATI operation, a sample will be taken from the addresses which remain uninterviewed. These addresses will be visited by Census Bureau field representatives, who will conduct personal interviews to obtain the information on the American Community Survey.

Implementation

The American Community Survey is being implemented in three parts:

- Demonstration period 1996-1998
- Comparison sites 1999-2002
- Full implementation nationwide starting in 2003 and continuing

The American Community Survey demonstration period began in 1996 in four sites. In 1997, the survey was conducted in eight sites to evaluate costs, procedures, and new ways to use the information. In 1998, the American Community Survey expanded to include two counties in South Carolina that overlapped with counties in the Census 2000 Dress Rehearsal. This approach allowed the Census Bureau to investigate the effects on both the American Community Survey and the census due to having the two activities going on in the same place at the same time.

In 1999, the number of sites in the sample increased to 31 comparison sites. The comparison with Census 2000 is designed to collect several kinds of information necessary to understand the differences between 1999-2001 American Community Survey and the 2000 long form.

The comparison sites include various situations in which these differences are expected to be prominent. They were selected to have at least one site in each of 24 strata representing combinations of county population size, difficulty of enumeration, and 1990-1995 population growth. The selection also attempts to balance areas by region of the country, and seeks to include several sites representing different characteristics of interest, such as racial or ethnic groups, highly seasonal populations, migrant workers, American Indian reservations, improving or worsening economic conditions, and predominant occupation or industry types.

The purpose of the comparison sites is to give a good tract-by-tract comparison between the 1999-2001 American Community Survey cumulated estimates and the Census 2000 long-form estimates, and to use these comparisons to identify both the causes of differences and diagnostic variables that tend to predict a certain kind of difference.

In 2002, we will continue to collect data in the 31 comparison sites to maintain the continuity of the survey.

In 2003, plans are to implement the American Community Survey in every county of the United States with an annual sample of three million housing units. Once the survey is in full operation, American Community Survey data will be available every year for areas and population groups of 65,000 or more beginning in 2004.

For small areas and population groups of 20,000 or less, it will take five years to accumulate a large enough sample to provide estimates with accuracy similar to the decennial census. That means updated information for areas such as neighborhoods will be available starting in 2008 and every year thereafter. A43

Data Dissemination

An American Community Survey goal is to provide data to the users within six months of the end of a collection or calendar year. For states, populous counties, and other governmental units or population groups with a population of 65,000 or more, the American Community Survey can provide direct estimates for each year. For smaller governmental units or population groups (those with a population of less than 65,000), estimates can be provided each year through refreshed multi-year accumulations of data.

Plans include the release of a microdata file each year patterned after the five percent Public Use Microdata Sample (PUMS) file of the 1990 decennial census records. The microdata file allows for two different units of analysis: housing unit and person. The microdata file includes as many records as possible and shows the lowest level of geography possible within confidentiality constraints. Users of the American Community Survey data can customize tabulations to examine the information in the way that best serves their needs.

In addition, the American Community Survey will provide summarized data for population and housing estimates, cross tabulated by various characteristics, down to the block-group level. The summarized data will be similar to the Summary Tape Files (STF) of the 1990 decennial census records, and are designed to provide statistics with greater subject and geographic detail than is feasible or desirable to provide in printed reports.

The microdata files, tabulated files, and associated documentation will be available on CD-ROM, as well as on this web site.

The American Community Survey and the Federal Statistical System

The American Community Survey offers a number of features that can improve the federal statistical system. They are:

- Increased sampling options;
- Flexibility in design and content; and
- More frequent data for evaluation.

Because the current federal statistical system is decentralized, surveys are conducted independently of one another. Each one must collect the same core data: number of occupied units, number of people, and the general characteristics of people. After these core data are collected, each survey focuses on its specific needs. The American Community Survey can provide better estimates of the core data as well as provide a vehicle for collecting some specific survey data, thereby reducing this duplication.

The American Community Survey can screen for households with specific characteristics. These households could be identified through the basic survey, or through the use of supplemental questions. Targeted households can then be candidates for follow-up interviews, thus providing a more robust sampling frame for other surveys. Moreover, the prohibitively expensive screening interviews now required would no longer be necessary.

State and local governments are becoming more involved in administering and evaluating programs traditionally controlled by the federal government. This devolution of responsibility is often accompanied by federal funding through block grants. The data collected via the American Community Survey will be useful not only to the federal agencies, but also to state, local, and tribal governments in planning, administering, and evaluating programs.

Finally, the American Community Survey will provide more timely data for use in area estimation models that provide estimates of various concepts for small geographic areas. In essence, detailed data from national household surveys (whose samples are too small to provide reliable estimates for states or localities) can be combined with data from the American Community Survey to create reliable estimates for small geographic areas.

Source: U.S. Census Bureau
Last revised: Friday, 20-Apr-01 16:28:45

APPENDIX D: 1990 PUBLIC USE MICRODATA SAMPLES

1990 PUBLIC USE MICRODATA SAMPLES (PUMS) Description

◆ *Availability:* Order from Customer Services, 301-457-4100; FAX 301-457-3842

◆ *Census contact:* Microdata Access Branch, 301- 457-1139.

◆ *Description:* The 1990 Public Use Microdata Samples (PUMS) contain individual records of responses to questionnaires with unique identifiers (names, addresses, etc.) removed so that the confidentiality of respondents is protected. These files, on computer tape, enable users to produce their own tabulations within the limits of the data provided. For 1990, the Bureau will produce PUMS for the United States and those outlying areas which meet a 100,000 minimum population size threshold. Currently, the standard PUMS products are the 5 percent and 1 percent samples for the United States and Puerto Rico, and a special 3 percent sample dealing specifically with the elderly population. The 5 percent and 1 percent samples are similar in content to the "A" and "B" files made available in 1980.

Besides the obvious difference in file size, the 5 percent and 1 percent files differ in the geography around which the files are constructed. For example, the Public Use Microdata Area (PUMA) is the lowest level of geography identified on any PUMS file. The 5 percent sample is basically a county level file; that is the PUMA can be a single county (or county equivalent), a group of counties, a place, or county/place parts if that county had more than 200,000 persons. On the other hand, the 1 percent sample is basically a metropolitan area file. For this, the PUMA will be an MSA, groups of MSAs, parts of MSAs when the MA is larger than 200,000 persons, and groups of nonmetropolitan areas.

The file structure of all PUMS files is hierarchical, with special features to aid in their use with commercial/existing software packages. While no "value-added" user software will be provided by the Census Bureau, the technical documentation will include a section with suggestions and pointers on how to use these files. All PUMS files have two record types; one for housing units, and one for persons. The number of records per file are determined by the sample size.

◆ These files also will be available on CD-ROM.

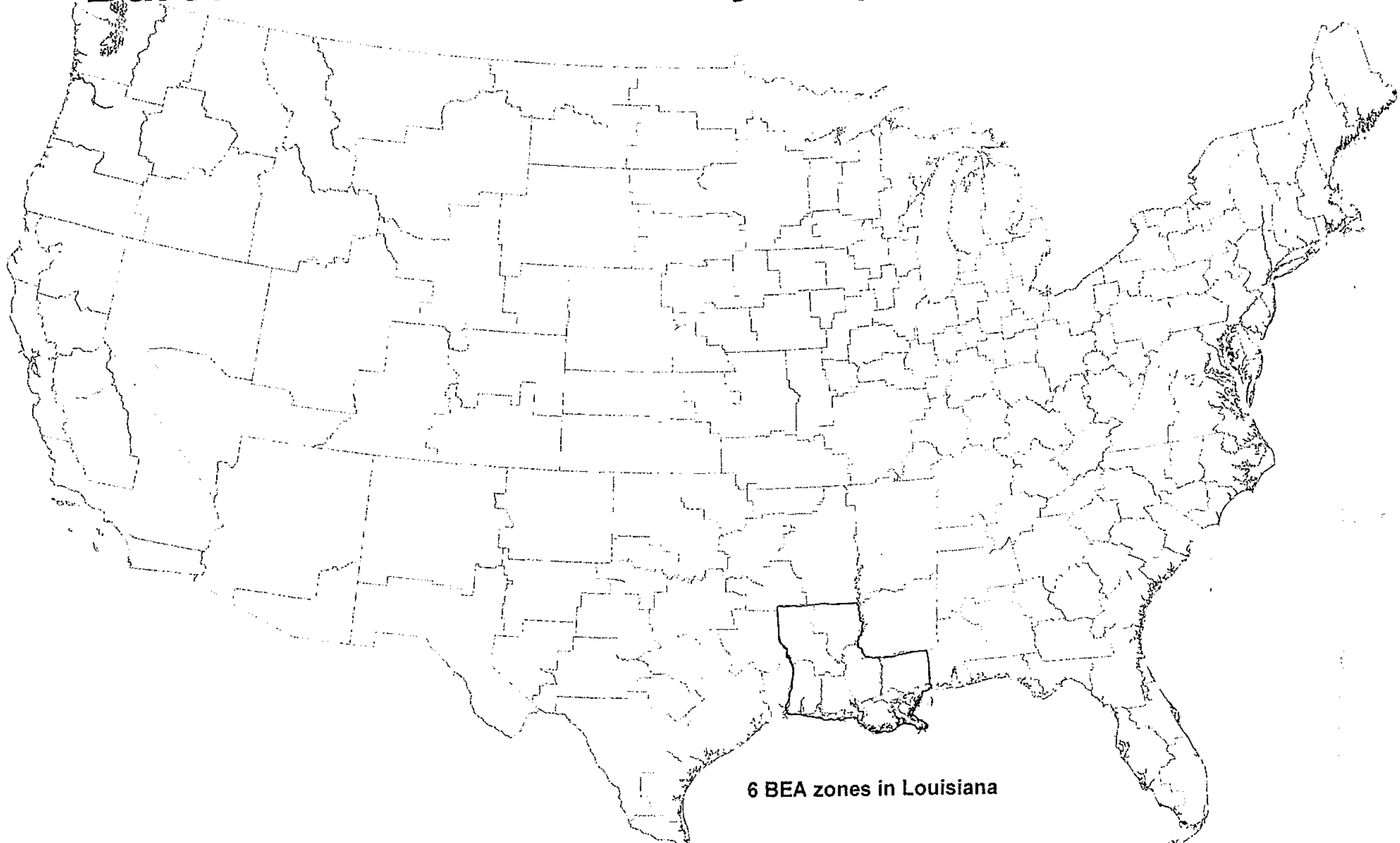
◆ Data Development XVI-61; Puerto Rico Data Development XVI-60

◆ Technical documentation included free with product - $15 when purchased separately.

Last update: *Monday, 12-Jul-99 13:54:59*

APPENDIX E: BUREAU OF ECONOMIC ANALYSIS ZONES

Bureau of Economic Analysis (BEA) zones (183 off)



APPENDIX F: REGIONAL ECONOMIC INFORMATION SYSTEM 1969-1995

REGIONAL ECONOMIC INFORMATION SYSTEM 1969-95

U.S Department of Commerce, Economic and Statistics Administration, Bureau of Economic Analysis

SETUP:

To run the SETUP program, put the REIS CD in your CD drive, select
File/Run from Program Manager, and type

    E:\SETUP

(where E is your CD drive letter) and press Enter.

The SETUP program tries to install these files to your WINDOWS\SYSTEM subdirectory:

    VER.DLL
    SETUPKIT.DLL
    VBRUN300.DLL

At the end of the SETUP program, a program group is created and saved in your WINDOWS directory. The size of the program group is 2.27 KB. The main program is REIS. All programs are run directly from the REIS CD. The programs include

- REIS
- QuickREIS
- BEA-EA Projections
- County Summary
- FIPS
- MSA Projections
- Gross State Product
- State Projections
- Journey to work

CD-ROM CONTENTS

The following document describes the contents of the CD-ROM organized by directory. The REIS database covers the U.S., States, metropolitan areas, BEA Economic Areas, and counties for the years 1969-1995

CA05 - Personal Income by major source and Earnings by industry

Personal income (Table CA05) is a measure of income received; therefore, estimates of State and local area personal income reflect the residence of the income recipients. The adjustment for residence is made to wages and salaries, other labor income, and personal contributions for social insurance, with minor exceptions, to place them on a place-of-residence (where-received) basis. The adjustment is necessary because these components of personal income are estimated from data that are reported by place of work (where earned). The estimates of proprietors' income, although presented on the table as part of place-of-work earnings, are largely by place of residence; no residence adjustment is made for this component. Net earnings by place of residence is calculated by subtracting personal contributions for social insurance from earnings by place of work and then adding the adjustment for residence, which is an estimate of the net inflow of the earnings of interarea commuters. The estimates of dividends, interest, and rent, and of transfer payments are prepared by place of residence only.

The files in the \CA05 subdirectory of the REIS CD-ROM contain the Table CA05 estimates. The data files are in a flat ASCII format. The files contain annual estimates from 1969 to 1995. Each State has its own data file, CA05nn.DFX, where nn = the State post office abbreviation. It is suggested that the **REIS** program be used to access CA05.

The file for the metropolitan areas is CA05MSA.DFX.
The file for the State metro/nonmetro portions is CA05PORT.DFX.
The file for the BEA economic areas is CA05ECON.DFX.

<u>CA25</u> - Full and Part-time Employment by major industry

Total Full-time and Part-time Employment by Industry (Table CA25) contains estimates of employment in Standard Industrial Classification (SIC) Division ("one-digit") detail. Employment is measured as the average annual number of jobs, full-time plus part-time; each job that a person holds is counted at full weight. The estimates are on a place-of-work basis. The estimates are organized both by type (wage and salary employment and self-employment) and by industry.

The files in the \CA25 subdirectory of the REIS CD-ROM contain the Table CA25 estimates. The files contain annual estimates from 1969 to 1995.Each State has its own data file, CA25nn.DFX, where nn = the State post office abbreviation. It is suggested that the **REIS** program be used to access CA25.

The file for the metropolitan areas is CA25MSA.DFX.
The file for the State metro/nonmetro portions is CA25PORT.DFX.
The file for the BEA economic areas is CA25ECON.DFX.

<u>CA30</u> - Regional Economic Profile

The Regional Economic Profile (Table CA30) provides general economic data that are derived from REMD's more detailed table (CA05, CA25, and CA35). Data are organized by both place of residence and place of work. The place of residence profile includes estimates of total personal income, population, and per capita personal income. The place of work profile includes estimates of total earnings, total employment, and average earnings per job.

The files in the \CA30 subdirectory of the REIS CD-ROM contain the Table CA30 data. The data files are in a flat ASCII format. The files contain annual data from 1969 to 1995. Each State has its own data file, CA30nn.DFX, where nn = the State post office abbreviation. It is suggested that **REIS** program be used to access CA30.

The file for the metropolitan areas is CA30MSA.DFX.
The file for the State metro/nonmetro portions is CA30PORT.DFX.
The file for the BEA economic areas is CA30ECON.DFX.

CA35 - Transfer Payments

Transfer Payments (Table CA35) are payments to persons for which they do not render current services. As a component of personal income, they are payments by government and business to individuals and nonprofit institutions. At the county level, approximately 75 percent of total transfer payments are estimated on the basis of directly reported data. The remaining 25 percent are estimated on the basis of indirect, but generally reliable, data. Payments to individuals are divided into several groups with program detail for some of the groups. The groups are retirement and disability insurance benefit payments, medical payments, income maintenance benefit payments, unemployment insurance benefit payments, veterans benefit payments, federal education and training assistance payments (other than for veterans), other government payments, and payments by business. Payments to nonprofit institutions are divided into federal government payments, State and local government payments, and business payments (corporate gifts).

The files in the \CA35 subdirectory of the REIS CD-ROM contain the Table CA35 data. The data files are in a flat ASCII format. The files contain annual data from 1969 to 1995. Each State has its own data file, CA35nn.DFX, where nn = the State post office abbreviation. It is suggested that **REIS** program be used to access CA35.

The file for the metropolitan areas is CA35MSA.DFX.
The file for the State metro/nonmetro portions is CA35PORT.DFX.
The file for the BEA economic areas is CA35ECON.DFX.

CA45 - Farm Income and Expenses

Farm Income and Expenses (Table CA45) provides detailed estimates of gross farm income and production expenses. Gross farm income consists of estimates for the following items: cash receipts from marketing of crops and livestock; income from other farm-related activities, including recreational services and the sale of forest products; government payments to farmers; value of food and fuel produced and consumed on farms; gross rental value of farm dwellings; and the value of the net change in the physical volume of farm inventories of crops and livestock. Production expenses consist of: purchases of feed, livestock, seed, fertilizer and lime, and petroleum products; hired farm labor expenses (including contract labor); and all other production expenses (e.g. depreciation, interest, rent and taxes, and repair and operation of machinery). Production expenses and gross farm income excluding inventory change are used to calculate realized net income of all farms (gross farm income, excluding inventory change, minus production expenses equals realized net income). Realized net income is then modified to reflect current production through the change-in-inventory adjustment and to exclude the income of corporate farms and salaries paid to corporate officers. These modifications yield BEA's estimate of farm proprietors' income.

The files in the \CA45 subdirectory of the REIS CD-ROM contain the Table CA45 data. The data files are in a flat ASCII format. The files contain annual data from 1969 to 1995. Each State has its own data file, CA45nn.DFX, where nn = the State post office abbreviation. It is suggested that **REIS** program be used to access CA45.

The file for the metropolitan areas is CA45MSA.DFX.
The file for the State metro/nonmetro portions is CA45PORT.DFX.
The file for the BEA economic areas is CA45ECON.DFX.

<u>SUMMARY</u> - Summary county information

The files in the \SUMMARY subdirectory of the REIS CD-ROM contain the Tables CA1-3 and CA34. The data files are in a flat ASCII format. The files contain annual data from 1969 to 1995. Each table has its own data file, CA13.DFX and CA34.DFX. These files contain estimates for the United States, all States and counties, the metropolitan areas within each State, and the metro/nonmetro portions within each State and the United States. The files for the metropolitan areas only are CA13MSA.DFX and CA34MSA.DFX.

These files may be easily displayed using the **SUMMARY.EXE** program located in root directory of the REIS CD-ROM. This program is also accessible through the **REIS.EXE** program.

CA34 Total Wage & Salary Disbursements, Total Wage & Salary Employment, and Average Wage per Job

CA1-3 Personal Income, Per Capita Personal Income, and Population

CA91 Gross Inflows and Outflows of Earnings, and Net Residence Adjustment

<u>SQ</u> - State quarterly personal income and earnings, IQ1990-IQ1997, as released in July 1997

<u>BF8595 - BEARFACTS:</u> one page summary of economic conditions, 1985-95, for every area.

This directory contains the BEARFACTS files for the 1985-95 time period. Each State has its own file: BF8595nn.TXT where nn = State PO abbreviation. The metropolitan areas are in the file, BF8595MR.TXT. All files in this directory are accessed by the REIS program. It is recommended that the **REIS** program be used to display and print the BEARFACTS tables.

BF9495 - BEARFACTS: one page summary of economic conditions, 1994-95, for every area.

This directory contains the BEARFACTS files for the 1994-95 time period. Each State has its own file: BF9495nn.TXT where nn = State PO abbreviation. The metropolitan areas are in the file, BF9495MR.TXT. All files in this directory are accessed by the REIS program. It is recommended that the **REIS** program be used to display and print the BEARFACTS tables.

GSP - Gross State Product (1977-1994) estimates, documentation files, and program files

GSP for a State is derived as the sum of gross state product originating in all industries in the State. In concept, an industry's GSP, referred to as its "value added," is equivalent to its gross output (sales or receipts and other operating income, commodity taxes, and inventory change) minus its intermediate inputs (consumption of goods and services purchased from other industries or imported). As such, it is often referred to as the State counterpart of the Nation's gross domestic product (GDP) . In practice, GSP estimates are measured as the sum of distributions by industry of the components of gross domestic income --that is, the sum of the costs incurred (such as compensation of employees, net interest, and indirect business taxes) and the profits earned in production. The difference between GDP and gross domestic income is the statistical discrepancy.

BEA prepares GSP estimates for 63 industries. For each industry, GSP is composed of three components: Compensation of employees, Indirect business tax and nontax liability (IBT). Other GSP. All three of these component estimates are, in current dollars, along with total current-dollar GSP. It also contain estimates of real GSP, expressed as chain-weighted quantity indexes. chain-weighted 1992 dollars (1982-1994), and fixed-weighted 1992 constant dollars.
It is recommended to use the **Gross State product** program to display the tables

PROJECT - Projections of income and employment to the year 2045. Each subdirectory contains data files and program files (DOS and Windows).

Subdirectories:

STATE - all States

Double click on the program **State Projections** to go the Data Viewer mode. When you enter the program, you are in the Data Viewer mode. You are able to select any of 5 data series (Employment, Earnings, GSP, Population, or Personal Income), and to view the data for all years, with the following options:

- Choose one State or Region, and view all variables under the selected data series.
- Choose one variable, and view that variable for all States and Regions.
- Print the displayed data.
- Write the displayed data to a file, in space / quote delimited format.
- Alternatively, you may bypass viewing the data by going to the File menu and choosing to Select Multiple Variables for Export.

The published State projections volume presents historical data for 1983, 1988, and 1993 (1992 for gross state product), and projected data for 1998, 2000, 2005, 2010, 2015, 2025, and 2045. These data files contain projected data for the same years as are contained in the published volume, but they contain additional years of historical data. For employment and earnings by industry, population by three broad age groups, and for personal income by major component, historical data are presented for 1969-93. For gross state product (GSP) by industry, historical data are presented for 1977-92. The inclusion of historical and projected GSP is a new feature of the current edition of the BEA Regional Projections. GSP is not included in the forthcoming projections for metropolitan areas or BEA economic areas, however, because no historical GSP estimates exist for those levels of geography.

MSA - metropolitan areas defined as of 1995

Double click on the program **MSA Projections** to go the Data Viewer mode. When you enter the program, you are in the Data Viewer mode. You are able to view the data for all years, with the following options:

- Choose one geographic area, and view all variables.
- Choose one variable, and view that variable for all areas.
- Print the displayed data.
- Write the displayed data to a file, in space / quote delimited format.
- Alternatively, you may bypass viewing the data by going to the File menu and choosing to Select Multiple Variables for Export.

For each of 313 regions (United States plus 312 metropolitan areas) 40 lines of historical and projected data are presented. For each of the 40 lines, historical data are presented for 1969-1993 and projections are presented for 2000, 2005, 2010, 2015, 2025, and 2045, a total of 31 data values for each of the 40 lines per region.

ECONAREA - BEA economic areas defined as of 1995

Double click on the program **BEA-EA Projections** to go the Data Viewer mode. When you enter the program, you are in the Data Viewer mode. You are able to view the data for all years, with the following options:

- Choose one geographic area, and view all variables.
- Choose one variable, and view that variable for all areas.
- Print the displayed data.
- Write the displayed data to a file, in space / quote delimited format.
- Alternatively, you may bypass viewing the data by going to the File menu and choosing to Select Multiple Variables for Export

For each of 173 regions (United States plus 172 economic areas) 40 lines of historical and projected data are presented. For each of the 40 lines, historical data are presented for 1969-1993 and projections are presented for 2000, 2005, 2010, 2015, 2025, and 2045, a total of 31 data values for each of the 40 lines per region.

JTW - Census county-to-county commuting flows for 1960, 1970, 1980, and 1990.

The Journey to Work Database Viewer is a Windows application that will allow you to view and export the Journey to Work data. The program performs the same functions as the DOS-based PULLCNY.EXE program but has a few enhancements. A program group and icon for the Journey to Work Database Viewer will appear. Double-click on the **JWVIEWER (Journey to Work)** icon to run the program. Selecting Open Database from the File menu brings up a sub-menu with two choices:

1) JWFREQ (1960-90 Commuting)
This database contains workers flows between a Place of Residence and a Place of Work for the years 1960, 1970, 1980, and 1990.

2) JWIDGIND (1980-90 Commuting and Wages)
This database contains workers flows between a Place of Residence and a Place of Work showing the average wages earned at the one digit level for the years 1980 and 1990.

MISC - miscellaneous data and information files.

The miscellaneous data and information include

- Changes to the area definitions on this REIS CD-ROM (69-95) compared to last year's REIS CD-ROM (69-94) for counties, BEA-economic areas, and MSA.
- Phone directory of offices in BEA.
- BEA Economic Area Component County List.
- BEA Metropolitan Statistical Area Component County List.
- BEA user group as of July 9, 1997, listed statewise.
- User's Guide to BEA Information. The Bureau of Economic Analysis (BEA) provides basic information on such key issues as economic growth, regional development, and the Nation's role in the world economy. The guide contains program descriptions and entries for specific products. The first section, entitled "General," describes the products that cut across the range of BEA's work. The following sections describe the products related to BEA's national, regional, and international economics programs.

ROOT - Location of all .VBX and .DLL files needed for the computer, and some of the Windows-based programs available on this CD-ROM. See \GSP, \PROJECT, and \JTW subdirectories for Gross State Product, Regional Projections, and Journey to Work Windows programs.

METHODS - Income methodology files.

This section describe the sources of the data and the methods that were used to prepare the annual estimates of personal income for counties for 1990-95.

CODE - miscellaneous code and title files including area name and **FIPS** code files.

This is the CODE directory of the REIS CD-ROM. It contains geographic code files. All are in ASCII format and most are in a delimited format which may be imported into spreadsheet programs. nnCODE.PRN files, where nn = State PO abbreviation list of counties with corresponding FIPS codes and MSA codes

MSACODE.PRN list of MSA's, ECONCODE.PRN list of BEA economic areas, MSALIST.PRN list of MSA's with component counties, ECONLIST.PRN list of BEA economic areas with component counties. *.TTL files have more lists of FIPS codes and area names. Use the **FIPS program** to view FIPS codes and area names easily.

STANNUAL - contains the State annual series.

This directory contains the State annual personal income and employment estimates. Estimates for 1969-95 were released in September, 1996. The State estimates (presented in the SA series) differ from the State sum of the county estimates released on August 27, 1997 because of the estimating schedule for the major regional series described below. The State annual estimates are for the years 1969-95; they are identical to the files provided by BEA on diskette. Some of the tables' files were compressed using PKZIP by PKWARE, Inc. These files were then converted to an auto-executing format. All the data for tables SA05, SA07, SA25, SA27, SA35, SA45, and SA50 have been put into these archives. Each table is in its own subdirectory. A documentation file, README.DOC, is available within each subdirectory.

Subdirectories:

- *SA05* - Total Personal Income by major source and Earnings by industry
- *SA07* - Wage & Salary Disbursements by Industry
- *SA25* - Full & Part-time Employment by major industry
- *SA27* - Full & Part-time Wage & Salary Employment by major industry
- *SA35* - Transfer Payments
- *SA45* - Farm Income & Expenses
- *SA50* - Tax & Non-tax Payments

SUMMARY - Total Personal Income, Per Capita Personal Income, Disposable Personal Income, Disposable Per Capita Personal Income, and Total Population for 1969-1995.

STANPREL - Preliminary Total Personal Income, Per Capita Personal Income, Disposable Personal Income, Disposable Per Capita Personal Income, and Total Population for 1969-1996. The files are in comma-delimited format.

APPENDIX G: FRAMEWORK SUGGESTED IN NCHRP 401

Suggested Data-Organization Framework

(F) - Denotes freight data

Data Components

Supply Attributes (S)
Demand Attributes (D)
System Performance (P)
System Impacts (I)

Supply Attributes (S)

S.H. Highway

S.H.1 Systems Data

S.H.1.1 Mileage and lanes (total lane miles and number of lanes, lane miles of HOV, intercity highway miles)
S.H.1.2 Capacity (including highway link capacities)
S.H.1.3 Functional road class
S.H.1.4 Nodes and segments (GIS or highway route)
S.H.1.5 Land use data for system expansion
S.H.1.6 (F) Intraurban truck routes (by route number)
S.H.1.7 Other

S.H.2 Service Data

S.H.2.1 Access (connections to other modes, highways, and roadways)
S.H.2.2 (F) Interurban access (GIS or highway route numbers; principal routes for trucks entering and exiting urban areas carrying interurban freight)
S.H.2.3 Intermodal access (rail, water, air, by highway route mile)
S.H.2.4 Data on service providers
S.H.2.5 Fare or fee structure data (tolls, parking)
S.H.2.6 (F) Drayage services
S.H.2.7 Other

S.H.3 Facilities Data

S.H.3.1 Inventory of facilities (bus terminal and stops, rest areas, park and ride lots, truck terminals, intermodal facilities, cargo transfer equipment, etc.)
S.H.3.2 Land use data for use in planning for route modifications, terminal and warehouse locations

S.H.3.3 (F) Delivery and Pickup (On-street, off-street parking by principal intraurban routes).
S.H.3.4 Other

S.H.4 Condition Data

- S.H.4.1 Pavement data by highway route (pavement serviceability rating, long-term pavement performance counts)
- S.H.4.2 Any data pertinent to condition of routes, bridges, ramps, etc. that affect the efficiency of interurban truck access to the urban area or truck pick-up and delivery activities
- S.H.4.3 Age of various road classes
- S.H.4.4 Other

S.H.5 Project Data

- S.H.5.1 List of all state projects proposed for the next 3 years (minimum) - including funding data
- S.H.5.2 List of all metropolitan (MPO) projects proposed for the next 3 years (minimum) - including funding data
- S.H.5.3 Major investment data (planned supply augmentation projects)
- S.H.5.4 Project history data (past capacity expansion and maintenance, project information such as project dates, type of construction, rehabilitation, etc.)
- S.H.5.5 Project evaluation data
- S.H.5.6 Planned expansions and modifications
- S.H.5.7 Project maintenance data
- S.H.5.8 Other

S.R. Rail

S.R.1 Systems Data

- S.R.1.1 Miles of passenger and freight track
- S.R.1.2 Nodes and segments by rail line (GIS)
- S.R.1.3 Capacity and current utilization by principal routes
- S.R.1.4 Land use data for system expansion
- S.R.1.5 Other

S.R.2 Service Data (Terminals)

- S.R.2.1 Access (intermodal access)
- S.R.2.2 Cities serviced
- S.R.2.3 Percent trains on time by principal routes
- S.R.2.4 Passenger service frequency
- S.R.2.5 (F) Freight service frequency
- S.R.2.6 Data on service providers
- S.R.2.7 Fare and fee structure data
- S.R.2.8 Other

S.R.3 Facilities Data

- S.R.3.1 Number of passenger and freight cars
- S.R.3.2 Passenger track miles

S.R.3.3 Inventory of facilities at each stop
S.R.3.4 Inventory of road crossing equipment
S.R.3.5 Intermodal terminals by location
S.R.3.6 (F)Cargo transfer equipment
S.R.3.7 (F)Cargo storage facilities
S.R.3.8 Inventory of infrastructure (e.g., guideways)
S.R.3.9 Other

S.R.4 Condition Data (Systems)

S.R.4.1 Age (cars, tracks, tunnels, bridges, crossing equipment, etc.)
S.R.4.2 Road crossing condition data
S.R.4.3 Bridge condition (by route)
S.R.4.4 Tunnel clearances (by route)
S.R.4.5 Service record (tracks, facilities, cars, road crossing)

S.R.5 Project Data (Systems and Terminals)

S.R.5.1 List of all state projects proposed for the next 3 years (minimum) - including funding data
S.R.5.2 List of all metropolitan (MPO) projects proposed for the next 3 years (minimum) - including funding data
S.R.5.3 Major investment data (planned supply augmentation projects)
S.R.5.4 Project history data (past capacity expansion and maintenance, project information such as project dates, type of construction, rehabilitation, etc.)
S.R.5.5 Project evaluation data
S.R.5.6 Planned expansions and modifications
S.R.5.7 Project maintenance data
S.R.5.8 Other

S.T. Transit Systems

S.T.1 Systems Data

S.T.1.1 Inventory of all routes
S.T.1.2 Capacity and current utilization
S.T.1.3 Route or track miles
S.T.1.4 Vehicle miles
S.T.1.5 Inventory of intermodal connections
S.T.1.6 Other

S.T.2 Service Data

S.T.2.1 Access data
S.T.2.2 Percent on time by transit mode and route
S.T.2.3 Fare and fee structure data
S.T.2.4 Other

S.T.3 Facilities Data

S.T.3.1 Inventory of facilities (garages, park & ride lots, stations, stops)
S.T.3.2 Inventory of transit vehicles (light rail cars, buses, subway cars, etc.)
S.T.3.3 Other

S.T.4 Condition Data

S.T.4.1 Statistics on services performed on all transit mode vehicles (maintenance schedule, service records)
S.T.4.2 Statistics on services performed on all transit mode facilities (maintenance schedule, service records)
S.T.4.3 Historical statistics on services performed on all transit mode infrastucture (e.g light rail guideway)
S.T.4.4 Age of vehicles, facilities, and infrastructure

S.T.5 Project Data

S.T.5.1 List of all state projects proposed for the next 3 years (minimum) - including funding data
S.T.5.2 List of all metropolitan (MPO) projects proposed for the next 3 years (minimum) including funding data
S.T.5.3 Major investment data (planned supply augmentation projects)
S.T.5.4 Project history data (past capacity expansion and maintenance, project information such as project dates, type of construction, rehabilitation, etc.)
S.T.5.5 Project evaluation data
S.T.5.6 Planned expansions and modifications
S.T.5.7 Project Maintenance data
S.T.5.8 Other

S.P. Ports and Inland Waterways

S.P.1 Systems Data

S.P.1.1 Ports (GIS)
S.P.1.2 Inland waterway segments (GIS)
S.P.1.3 Locks and capacity
S.P.1.4 Capacity and current utilization by principal routes
S.P.1.5 Land use data for port expansion
S.P.1.6 Other

S.P.2 Service Data

S.P.2.1 Access by all modes
S.P.2.2 Shiplines/Ferry Service lines serving each port
S.P.2.3 Sailing frequencies by destination
S.P.2.4 (F) Barge lines serving each port

S.P.2.5 Multimodal connections
S.P.2.6 Months river is open
S.P.2.7 Cities/Regions serviced
S.P.2.8 Fare and fee structure data
S.P.2.9 Other

S.P.3 Facilities Data

S.P.3.1 Number of providers (boats/ferries)
S.P.3.2 Number of passenger docking facilities
S.P.3.3 Inventory of passenger facilities at port
S.P.3.4 (F) Cargo transfer facilities by port (including handling equipment)
S.P.3.5 (F) Cargo storage facilities
S.P.3.6 Berth Capacity
S.P.3.7 Other

S.P.4 Condition Data

S.P.4.1 Dredging schedules
S.P.4.2 Docks and berths (age, service records, maintenance schedules)
S.P.4.3 Navigation aids (age, service records, maintenance schedules)
S.P.4.4 Boats and ferries (age, service records, maintenance schedules, U.S. Coast Guard certificates)
S.P.4.5 Channel depth and width
S.P.4.6 Locks (age and maintenance schedule)
S.P.4.7 Other

S.P.5 Project Data

S.P.5.1 List of all state projects proposed for the next 3 years (minimum) - including funding data
S.P.5.2 List of all metropolitan (MPO) projects proposed for the next 3 years (minimum) - including funding data
S.P.5.3 Major investment data (planned supply augmentation projects)
S.P.5.4 Project history data (past capacity expansion and maintenance, project information such as project dates, type of construction, rehabilitation, etc.)
S.P.5.5 Project evaluation data
S.P.5.6 Planned expansions and modifications
S.P.5.7 Project maintenance data
S.P.5.8 Other

S.A. Air (Airports)

S.A.1 Systems Data

S.A.1.1 Runways (number and lengths)
S.A.1.2 Land use data for airport expansion

S.A.1.3 Number of airports
S.A.1.4 Capacity and current utilization by principal routes
S.A.1.5 Other

S.A.2 Service Data

S.A.2.1 Number of providers (airlines serving city)
S.A.2.2 Cities served
S.A.2.3 (F) Freight service frequency
S.A.2.4 Intermodal access and connections
S.A.2.5 Percent on time by airline and route
S.A.2.6 Fare or fee structure (range of prices, prices per passenger mile)
S.A.2.7 Other

S.A.3 Facilities Data

S.A.3.1 Passenger transfer facilities (bus stops, train stations, parking)
S.A.3.2 (F) Cargo transfer equipment
S.A.3.3 (F) Cargo storage facilities
S.A.3.4 Inventory of airport facilities (gates, walkways,etc.)
S.A.3.5 Other

S.A.4 Condition Data

S.A.4.1 Terminal condition data (age, service records, maintenance schedule)
S.A.4.2 Runway data (age, service record, maintenance schedule)
S.A.4.3 Airplane data by airline (age, service record, maintenance schedule)
S.A.4.4 Cargo transfer equipment
S.A.4.5 Other

S.A.5 Project Data

S.A.5.1 List of all state projects proposed for the next 3 years (minimum) - including funding data
S.A.5.2 List of all metropolitan (MPO) projects proposed for the next 3 years (minimum) - including funding data
S.A.5.3 Major investment data (planned supply augmentation projects)
S.A.5.4 Project history data (past capacity expansion and maintenance, project information such as project dates, type of construction, rehabilitation, etc.)
S.A.5.5 Project evaluation data
S.A.5.6 Planned expansions and modifications
S.A.5.7 Project Maintenance data
S.A.5.8 Other

Demand Attributes (D)

D.1 Economic Data

- D.1.1 Income data by household and region -- historical, current and projected
- D.1.2 Employment data by SIC code and region -- historical, current and projected
- D.1.3 Vehicle ownership data by household and region
- D.1.4 Travel cost data (e.g., auto operating costs, parking costs, transit fares, tolls, etc.)
- D.1.5 Proxy data for projecting income and employment by household, SIC code, and region
- D.1.6 (F) Industrial operations (Location, SIC code and employment)
- D.1.7 (F) Wholesalers and distributors (Location, SIC code and employment)
- D.1.8 (F) Commodity production data by SIC and geographic detail -- historical, current and projected
- D.1.9 (F) Commodity consumption data by SIC and geographic detail -- historical, current and projected
- D.1.10 (F) Export/import data by point of exit/entry (seaports, airports and highway and rail border points)
- D.1.11 (F) Proxy data for projecting commodity production and consumption data (projections of employment, income, etc., by geographic area)
- D.1.12 Other

D.2 Demographic Data

- D.2.1 Population and labor force data (e.g., population size, density, geographic distribution) -- historical, current and projected
- D.2.2 Household characteristics (e.g., household size, number of children, number of licensed members) -- historical, current and projected
- D.2.3 Other

D.3 Land Use Data

- D.3.1 Acreage data (e.g., acres of land by major use, square footage by major use) -- historical, current and projected
- D.3.2 Housing data (e.g., occupancy densities, type distributions, location distributions)
- D.3.3 Employment data (e.g., employment densities, type distributions, location distributions)
- D.3.4 Access data (e.g., accessibility to services, mix and intensity of services)
- D.3.5 Zoning data (e.g., information on zoning restrictions, planned land uses)
- D.3.6 Other

D.4 (F) Commodity Flow Data -- (historical, current and projected)

- D.4.1 Commodity flow data by O-D
- D.4.2 Modal split on commodity flow data by O-D
- D.4.3 Factors affecting modal split
  - D.4.3.1 Relative modal rates
  - D.4.3.2 Delivery time by O-D
  - D.4.3.3 Other

D.4.4 Other

D.5 Travel Data

D.5.1 Trip generation data (e.g., person trips by purpose, vehicle trips by purpose, transit trips by purpose, non-motorized trips, etc.) -- historical, current and projected

D.5.2 Trip distribution data (e.g., trip length distributions, trips by time-of-day, etc.) -- historical, current and projected

D.5.3 Special generator data (e.g., tourism, conventions, special events, etc.)

D.5.4 Traffic volume data (e.g., annual average daily traffic, design hourly volume, peak hour traffic percentage, directional split, peak period volume, turning movements, zone to zone modal split, external station traffic counts, etc.) -- historical, current and projected

D.5.5 VMT data (e.g., VMT mix, VMT by functional road class, VMT by time-of-day, etc.) -- historical, current and projected

D.5.6 (F) Shipper modal selection factors

D.5.6.1 Delivery times by O-D

D.5.6.2 Relative modal costs

D.5.7 Other

D.6 Travel Behavior Data

D.6.1 Mode choice data (e.g., air, rail, highway, port, transit fare matrices, parking costs, mode availability variables such as vehicle ownership and percent of houses and jobs within walking distance to transit, etc.) -- historical, current and projected

D.6.2 Route choice data (e.g., network assignment, pretrip planning, out-of-pocket and time costs, etc.)

D.6.3 User preference data (e.g., willingness to pay, rider preferences, carpooling, ridesharing, etc.) -- historical, current and projected

D.6.4 (F)Time-of-day for pickup and deliveries

D.6.5 (F)Carrier behavior data

D.6.5.1 Discriminatory pricing

D.6.5.2 Intermodal agreements

D.6.6 Other

System Performance Attributes (P)

P.1 Safety Data

P.1.1 Incident data (e.g., number, type, location, and duration of traffic incidents, etc.)

P.1.2 Accident data (e.g., number of accidents, deaths, injuries by mode)

P.1.3 Security data (number and type of security incidents by mode and service populations, etc.)

P.1.4 Medical services data (e.g., response time, number of providers, etc.)

P.1.5 Other

P.2 Performance Measures

P.2.1 Highway performance data (e.g., recurrent and non-recurrent congestion, person and vehicle miles/hours of delay, lane and vehicle miles of roadway operating at substandard level-of-service, average system speed, incident location and response, average delay per person/vehicle, etc.) -- historical, current and projected
P.2.2 Transit performance data (e.g., average system speed, on-time performance, vehicle hours per trip, etc.) -- historical, current and projected
P.2.3 Intermodal system performance data (e.g., transfer time between modes, delay along terminal access routes, etc.) -- historical, current and projected
P.2.4 Efficiency data by mode (e.g., load factors per unit of capacity, percent on-time performance, average delay time, percent service interruptions, etc.) -- historical, current and projected
P.2.5 User cost data (e.g., cost per trip and unit of travel, travel time, etc.) — historical, current and projected
P.2.6 (F) Delivery times by O-D and mode/intermodal
P.2.7 (F) Cargo damage by mode/intermodal
P.2.8 (F) Congestion at terminals
P.2.9 (F) Shipment costs
P.2.10 Other

System Impact Attributes (I)

I.1 Air Quality Data

I.1.1 Vehicle registration data (e.g., vehicle populations by class, fuel type, vintage distributions, etc.) -- historical, current and projected
I.1.2 VMT data (e.g., VMT by functional road class and time-of-day, VMT mix by vehicle class, mileage accumulation rates by vintage, traffic counts by time-of-day, VMT forecasts, etc.) -- historical, current and projected
I.1.3 Speed data (e.g., average speed by functional road class, vehicle class, time-of-day, geographic area, etc.) -- historical, current and projected
I.1.4 Trip data (e.g., number of trips by purpose, cold versus hot starts, etc.) -- historical, current and projected
I.1.5 Impact assessment data (e.g., emissions contributions by vehicle class and pollutant, TCM effectiveness estimates--VMT reduction, etc.) -- historical, current and projected
I.1.6 Interurban emission contribution by mode (highway, rail, air, port) -- historical, current and projected
I.1.7 Other

I.2 Other Environmental Data

I.2.1 Visual and aesthetic impacts (e.g., information on interrupted views, neighborhood view of facility, simulations or drawings that scale facility to neighborhood, etc.)
I.2.2 Noise and vibration impacts (e.g., local noise criteria, etc.)
I.2.3 Ecosystems (e.g., existing wildlife or vegetation resource data, acres of wetlands affected by construction of transportation facilities, etc.)

1.2.4 Archeological and cultural impacts (e.g., inventory of national register sites, aerial photographs and periodicals, etc.)

1.2.5 Parklands (e.g., information on size, owner, and location, accessibility, function, etc.)

I.3 Land Use Data

I.3.1 Socio-economic impact (e.g., information on displacement of residents, dwellings, and businesses, conceptual design drawings, land market value data, parcel mapping, etc.)

I.3.2 Neighborhood impacts (e.g., neighborhood plans, demographic composition, socio-economic composition, etc.)

I.3.3 Other

I.4 Energy Data

I.4.1 Energy consumption impacts by mode -- historical, current and projected

I.4.2 Energy efficiency impacts by mode -- historical, current and projected

I.4.3 Energy price impacts -- historical, current and projected

I.5 Economic Growth Data

I.5.1 Local employment impacts (e.g., expected job creation of transportation projects in local area, etc.)

I.5.2 Regional employment impacts (e.g., expected job creation of transportation projects in local area, etc.)

I.5.3 (F) Access to natural resources

I.5.4 (F) Access to domestic markets

I.5.5 (F) Access to ports and foreign markets

APPENDIX H: WEIGHTING OF NPTS 1995 DATA

3-F. CONFIDENTIALITY ASSURANCE

CONFIDEN-TIALITY MEASURES

The following measures were taken in producing this public use data set to assure respondent confidentiality:

- All direct identifiers, such as telephone numbers, zip codes, county codes, names of individuals, and addresses were removed from the files.
- Metropolitan Statistical Areas (MSAs) of less than 1 million population and states with less than 2 million population are not specifically identified on the datafile.
- Other geographic variables were examined to prevent identification of geographic areas with less than 50,000 population (1990 Census). These variables included the MSA size code, Census division, and the specifically-identified MSAs and states.
- The data files contain a number of population and workforce variable estimates at Census Tract and Block Group levels. These variables will help describe the area of the sample members' household and work locations. The values published for these variables were rounded and/or placed into intervals to lessen the likelihood of users identifying specific areas from these variables.
- The specific dates of travel day and travel period trips were removed from the file.
- Data values for certain other variables were coded into intervals or suppressed, and some data distributions were capped. For example, detailed year/make/model information for antique and classic autos could compromise respondent confidentiality if fully revealed. In the public use files, rare make and model codes were re-coded as "other" makes and models. The year data for 1919 to 1969 model vehicles was re-coded into intervals.

3-G. WEIGHT CALCULATIONS

WEIGHTS

The purpose of weighting in NPTS is to expand the sample data to estimates for the U.S. population. There are four different NPTS weights that are used to compute different kinds of population estimates. The methods used to calculate each of the four weights are discussed in the sections which follow.

HOUSEHOLD WEIGHTS

With the NPTS list-assisted sample design, all in-sample households have a known, nonzero probability of selection. The unadjusted household weight is simply the reciprocal of the household's selection probability.

Since household telephone numbers were selected with equal probabilities within each sample stratum, the initial household sampling weights are computed simply as the ratio of the number of sampling units (telephone numbers) in the sampling frame for a stratum to the number of sample telephone numbers released for calling.

The initial sampling weights were adjusted for multiplicities arising from households that had more than one residential telephone number in the sampling frame, i.e., more than one chance of being in the sample.

Then the household weights were adjusted to sum to 98,990,000, an estimate of the number of U.S. households in 1995, to correct for non-responding households. Note that the estimated number of households includes those with and without telephones.

The household weights were then adjusted to equal marginal totals for the important variables listed below, to correct for non-response and non-coverage, and to reduce non-response bias. The basic concept is to adjust the sampling weights of the survey respondents so that they sum to known external totals, e.g., Census totals. A method of iterative proportional fitting was used to adjust the household weights simultaneously so the sums agreed closely with the following marginal controls:

- Equal weight totals for each of the 12 months of the year.
- Geographic areas - estimated total households in the four Census regions plus sub-regions associated with the add-on areas (39 total areas).
- U.S. level Current Population Estimates of the numbers of Black and non-Black households.
- U.S. level Current Population Estimates of the numbers of Hispanic and non-Hispanic households.
- Five categories of MSA population sizes.
- Four household size categories (1, 2, 3, 4 or more persons).

The adjusted household weights are appropriate for use in weighting all NPTS household variable data and vehicle variable data, since information on vehicles was collected at the household level. This variable is WTHHFIN.

NOTE: It is NOT appropriate to summarize travel day or travel period travel at the household level and then weight the estimate by the household weight. Travel data was collected at the person level, and a derivation of the person weight, such as the trip weight, must be used to obtain accurate estimates of travel day and travel period data. This is primarily because the person weight and the trip weights have been adjusted to account for non-interviewed persons within an interviewed household.

PERSON WEIGHTS

Since there was no sub-sampling of age-eligible persons within NPTS sample households, the household weights would also be appropriate for weighting the person data if data for 100 percent of the eligible persons within sample households had been obtained. Since that was not the case, the person weights were adjusted to compensate for person-level non-response in the 1995 NPTS. The sum of all person weights was adjusted to equal 241,675,000, an estimate of the number of U.S. residents in 1995 five years and older. Post-stratification weight adjustments were also made to adjust the person weights to the following external known totals:

- Equal weight totals for each of the 12 months of the year.
- Geographic areas - estimated total persons in the four Census regions plus sub-regions associated with the add-on areas (39 total areas).
- U.S. level Current Population Estimates of the numbers of Black and non-Black persons.
- U.S. level Current Population Estimates of the numbers of Hispanic and non-Hispanic persons.
- Ten categories of U.S. level age by gender populations (males and females each by the following ages: 5 - 17 years; 18 - 34; 35 - 44, 45 - 64, and 65 years and older).

The adjusted person weight , variable WTPERFIN, should be used to weight all person-level data from the 1995 NPTS survey. Person weights form the basis of the travel day and travel period weights, since person weights are adjusted to account for non-interviewed persons within an interviewed household.

TRAVEL FILE WEIGHTS

The two trip-level weights are simple functions of the adjusted person weights. There is no adjustment to be made for trip-level non-response, since the trip data had to be obtained in order for the person to be treated as a responding person. Each person's travel-day trip weight, variable WTTRDFIN, was calculated by multiplying the final person weight, WTPERFIN, times 365 to expand the person's travel day to an annual total. This weight is appropriate for weighting data from the travel day trip file and the segmented travel day trip file. The travel period weight, variable WTTRPFIN, for a person was calculated by dividing their travel day weight by 14, to reflect the 14-day travel period.

3-H. SURVEY METHOD AND PROCEDURE CHANGES

1995 NPTS CHANGES

In many ways the 1995 NPTS represents a significant change in survey methods and procedures from earlier NPTSs. These survey changes, which are listed in Exhibit 3.1, have had a significant impact on the results of the survey. The greatest impacts are most likely from:

1. Use of a written diary to help remember travel on a specific day. In the pretest conducted in 1994 for the 1995 NPTS, a written diary was compared to the retrospective, or recall, method. The diary method averaged 0.5 trips more per person per day than the retrospective method. (Reference: PlanTrans, Draft report on NPTS Pretest Methods, Spring 1997)

2. The household roster of trips, that maintained a list of trips that household members already interviewed had been on with, or accompanied by, this respondent.

3. The $2.00 incentive that was sent with each travel diary. This may have made the respondents feel obligated to record and report all of their travel.

4. Use of an advance letter to notify potential respondents that they would be recruited for the survey. We believe that the advance letter added legitimacy to the telephone recruitment, which contributed to higher quality data. The effect of the advance letter cannot be measured quantitatively.

5. Confirmation of "no travel" to distinguish from "soft refusals." The proportion of persons who said they made no trips on the assigned travel day was approximately 12

APPENDIX I: CASRO METHOD OF COMPUTING RESPONSE RATES

CASRO METHOD OF COMPUTING RESPONSE RATES

The Council of American Survey Research Organizations (CASRO) has developed a method for computing response rates on surveys that it recommends all survey research organizations should use. The method is applied as follows.

From a telephone (or other) initial contact, there are three results that can be obtained – ineligible contacts (e.g., businesses, out-of-service telephone numbers, etc.), eligible contacts (i.e., households with a qualified person residing there), and contacts with eligibility unknown (e.g., telephones that are not answered and have no recorded message, lines that are always busy, etc.). In many survey reports, this last category of contacts is often added to the ineligible category and ignored thereafter, since ineligible contacts do not affect the response rate. In the CASRO method, they are included. By using the information on the proportion of eligible to ineligible contacts, the unknown eligibility contacts are subdivided into those estimated to be eligible and those estimated to be ineligible. These are then added to the total number of eligible and ineligible contacts. The response rate is then estimated as the fraction of the eligible contacts plus assumed eligible contacts from which a response is obtained.

Suppose a survey involves a total of 5,000 contacts. Suppose that 3,000 of these contacts are determined to be ineligible, 1,000 are determined to be eligible, and 1,000 have eligibility unknown after the required number of attempts to contact have been made. Of the 1,000 that were found to be eligible, suppose that 500 are successfully recruited for the survey. Suppose also that 250 of these actually complete the survey task. If the eligibility unknown contacts are ignored, the response rate may be reported as either 50 percent (based on the number of successful recruitments made) or 25 percent, based more correctly on the number of eligible contacts. However, CASRO defines the response rate as 250/1,250, or 20 percent. In addition to the 1,000 eligible contacts, one-quarter of the contacts for which eligibility was known were eligible (1,000/4,000). Applying this eligibility rate to the 1,000 contacts with unknown eligibility suggests that there would have been 250 eligible contacts in this group, bringing the total to 1,250 assumed eligible contacts.

The CASRO method is stringent and sets a high standard for survey responses. While it has not yet been adopted by the transportation profession as the standard method for calculating response rates, it has been discussed widely as a potential standard.