TRANSPORTATION RESEARCH RECORD 779

Transportation Information Systems: Applications and Uses

TRANSPORTATION RESEARCH BOARD

COMMISSION ON SOCIOTECHNICAL SYSTEMS
NATIONAL RESEARCH COUNCIL

NATIONAL ACADEMY OF SCIENCES
WASHINGTON, D.C. 1980

Transportation Research Record 779
Price $4.20
Edited for TRB by Brenda J. Vumbaco

modes
1 highway transportation
2 public transit

subject areas
12 planning
16 user needs

Library of Congress Cataloging in Publication Data
National Research Council. Transportation Research Board.
Transportation information systems, applications and uses.

(Transportation research record; 779)
Reports for the Transportation Research Board's 59th annual meeting.
1. Traffic surveys–Addresses, essays, lectures. 2. Traffic surveys–Planning–Addresses, essays, lectures. 3. Traffic surveys–United States–Addresses, essays, lectures. I. National Research Council (U.S.). Transportation Research Board. II. Series. TE7.H5 no. 779 [HE369] 380.5s 81-38319
ISBN 0-309-03124-9 [388.3'14'0723] AACR2 ISSN 0361-1981

Sponsorship of the Papers in This Transportation Research Record

GROUP 1–TRANSPORTATION SYSTEMS PLANNING AND ADMINISTRATION
Leon M. Cole, Library of Congress, chairman

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

Committee on Transportation Information Systems and Data Requirements
Joseph M. Manning, Gordon Fay Associates, Inc., chairman
Robert E. Barraclough, Daniel E. Benson, H.W. Bruck, Kenneth J. Dueker, David J. Dunlap, Robert T. Dunphy, Christopher R. Fleet, Richard L. Forstall, Alexander French, Larry R. Goode, Yehuda Gur, Dayton P. Jorgenson, Roy E. Lave, Jr., William A. Lesansky, Alfred J. Marston, James J. McDonnell, Morton A. Meyer, Philip O. Patterson, Alan E. Pisarski, James E. Sale, John L. Skinner, Jr., Garrison P. Smith, Jr., Arthur B. Sosslau, Robert N. Tap, Ronald W. Tweedie

James A. Scott, Transportation Research Board staff

The organizational units, officers, and members are as of December 31, 1979.

Contents

Authors of the Papers in This Record

Boyd, J. Phillip, Atlanta Regional Commission, 230 Peachtree Street, N.E., Suite 200, Atlanta, GA 30303
Cannon, John, Bureau of the Census, U.S. Department of Commerce, Suitland, MD 29233
Dreisbach, Anthony F., Southwestern Pennsylvania Regional Planning Commission, 564 Forbes Avenue, Pittsburgh, PA 15219
Fisher, R. Fred, Georgia Department of Transportation, No. 2 Capitol Square, Atlanta, GA 30334
Hoang, Lap T., Florida Department of Transportation, 605 Suwannee Street, Tallahassee, FL 32301
Hupp, R. Craig, Southeast Michigan Council of Governments, 800 Park Building, Detroit, MI 48226
Palombo, Carmine, Southeast Michigan Council of Governments, 800 Park Building, Detroit, MI 48226
Poteat, Victor P., Florida Department of Transportation, 605 Suwannee Street, Tallahassee, FL 32301
Stuart, Robert C., College of Architecture and Urban Studies, Virginia Polytechnic Institute and State University, P. O. Box 560, Blacksburg, VA 24060
Torene, Robert, Bureau of the Census, U.S. Department of Commerce, Suitland, MD 20233
Williams, G. Jack, Georgia Department of Transportation, No. 2 Capitol Square, Atlanta, GA 30334

Commercial Data Sources for Urban Transportation Planning

ROBERT C. STUART

Under current pressures to improve file quality, increase multiple use, and reduce cost, a number of metropolitan planning organizations (MPOs) have resorted to the use of commercial data, particularly that from Dun and Bradstreet and R.L. Polk and Company. The purpose of this research, sponsored jointly by the Federal Highway Administration (FHWA) and the Urban Mass Transportation Administration (UMTA), is to assess this trend. MPOs, transit agencies, and other data users in 12 metropolitan areas were interviewed about their experience with the data. Users generally reported satisfaction with the commercial data but also noted some problems and limitations. Nevertheless, the research concluded that a commercial data source, on its merits, will be selected by an increasing number of MPOs. The choice must depend on local circumstances. Under some circumstances, commercial sources may be less expensive, more current, and more quickly available. Such sources provide comparisons with other time points, subareas, or urban areas; offer computer graphics; serve multiple functions; or have fewer confidentiality limitations. Commercial data sources are competitive with other sources such as primary surveys, the U.S. Census, or state employment security files. Recommendations are made for (a) more adequate FHWA and UMTA guidelines for household and employment files, (b) definitive tests of commercial sources with traditional sources, (c) staffing by FHWA and UMTA to implement the first two recommendations, (d) central purchase of Dun and Bradstreet establishment files for all metropolitan areas, and (e) additional commercial services from R.L. Polk and Company.

Transportation agencies are now trying to cope with conflicting pressures. They are being asked to improve the quality of data for a wider set of users at lower costs. Under these pressures, a number of metropolitan planning organizations (MPOs) have resorted to the use of commercial data sources as an aid in transportation planning. These sources are by-products of other data-collection purposes such as business credit ratings or city directories. Because these primary purposes require that data be on computer tape, these sources promise data that are inexpensive, current, readily available for use, already coded to small areas, without confidentiality restrictions, and suitable for multipurpose use. The question addressed by this paper is, "In fact, do the commercial sources live up to this promise?"

A substantial answer to this question is provided by recent contract research sponsored jointly by the Federal Highway Administration (FHWA) and the Urban Mass Transportation Administration (UMTA). The purpose was to investigate the experience of urban transportation planning agencies with the two most widely used commercial data sources--Dun and Bradstreet and R.L. Polk and Company. The research also sought to identify relevant features of these two sources, compare them with other sources to determine their potential, and recommend appropriate action to the MPOs, FHWA, UMTA, and the two companies. The basic methodology consisted of field and telephone interviews with MPOs, transit agencies, city and county planning staffs, and other local users of the two commercial data sources. Selection of interviewees was designed to include both MPOs that had used a commercial source and those that had considered and rejected it and to secure a range of population sizes and geographic locations. The metropolitan areas for which direct interviews were made are Denver; Portland, Oregon; San Antonio; San Diego; Philadelphia; and Saginaw and Battle Creek, Michigan. The first part of this paper compares the strengths and weaknesses of these two commercial sources with competing public sources; the second makes 10 recommendations related to the use of data-collection services.

STRENGTHS AND WEAKNESSES OF COMMERCIAL DATA SOURCES

MPOs maintain three principal files--the household file, the employment file, and the trip file. For each file, potential data sources are identified, and commercial sources are described in some detail (on the assumption that descriptions of public data sources are generally available). Finally, based on criteria that often guide selection of a data source, the strengths and weaknesses of the sources are noted.

Household File

A household file typically contains statistical summaries of households and persons by small areas and by certain socioeconomic characteristics (e.g., household size, income, age, and housing type). It may also have a listing of all individual households from which the summaries were derived (e.g., Polk provides such a listing, the U.S. Census does not). In either case, the household file provides the basis for generation of trip "productions" in the traditional transportation models.

Five alternative sources could be considered for creation of a household file:

1. The 1980 decennial and proposed 1985 mid-decade U.S. censuses;
2. A primary survey;
3. Local permits and approvals;
4. International Data Development, Inc.; and
5. R.L. Polk and Company.

The last two are commercial sources. These are briefly discussed here.

International Data Development, Inc., is new, has been widely used for urban transportation planning, and, therefore, could not be field checked in this study. It synthesizes low-cost summary data at the tract level or zone level. These data represent a set of current estimates for household data items usually required for transportation modeling, including households by structure type (single family and multifamily), by income (five income categories), by automobile ownership (0, 1, 2, 3, or more), and total population.

The estimates are derived by using two current data bases--"automobile households" and "telephone households"--from the files of the Reuben H. Donnelley Corporation, which compiles and prints most telephone books. The products of International Data Development were apparently designed for market-planning purposes. The availability of its data for transportation planning is very new (1978) and has not yet been subject to the kind of critical evaluation that transportation planners have accorded the more widely used R.L. Polk and Dun and Bradstreet data sets.

R.L. Polk and Company prepares city directories for some 3200 U.S. communities covering approximately 70 million persons and ranging from very large places (e.g., Birmingham, Atlanta, San Francisco, Houston, and Memphis) to towns with less than 2500 persons. Annually Polk also compiles Reports of Cars and Trucks on the Road and New Motor Vehicle Registrations based on nationwide data.

In order to produce its city directories, Polk conducts annual door-to-door interviews (canvasses) of households, businesses, and industrial establishments. Polk also takes counts of vacant dwellings and commercial structures. The canvasses achieve typically an 80-95 percent interview rate and are adjusted for noninterviews. Polk uses a field listing method similar to that often employed in land use surveys conducted by transportation planners. One important difference is that the interviewer is equipped with a computer-prepared printout of the previous year's listing. This is used to check the accuracy and completeness of coverage and to record all changes. This listing uses induced errors, which the enumerator must detect and correct, as a means of controlling the quality of field operations.

As a by-product of the city directory survey, a computer tape containing all household and business records is available at relatively low cost. These records can be geocoded to census block, tract, traffic zone, or other locally desired subareas.

Polk provides all of the household data normally required for urban transportation planning. Survey items include household size, number of adults, number of children, owner or renter status, vacancies, number of units in a structure, and number of cars and trucks registered at the address. Derived items include household income and elderly and handicapped population. Polk data are also available as a two-year trend comparison called Profiles of Change, which provides a wide range of neighborhood indicators. Some of these indicators, such as female head of household, promise to be useful in pinpointing transit market segments and uncovering early signs of shifts in travel behavior.

Comparative advantages and disadvantages of each of the five sources for use in a household file include the following.

1. Availability: Polk data are not available for some metropolitan areas. In others they may be available only for jurisdictions and, therefore, would require some supplementation to be suitable for areawide transportation planning. Other sources are available in all areas.

2. Coverage of urban fringe: The developing fringe of the metropolitan area is a place where travel volumes and patterns can change rapidly and drastically. The U.S. census covers the fringe completely but is quickly outdated. The same would be true of a primary field survey.

Tabulation of building permits or occupancy permits can be a strong source in fringe areas. International Data Development covers the fringe with current estimates but concedes that it may be inaccurate in some fringe areas.

Polk data generally cover only the central city and the surrounding built-up urban area. In some regions, however, almost the entire area is covered. A hybrid system, combining the Polk file for the built-up area, local tabulations, and local selective field checks of new development in the fringe, would appear to provide a cost-effective, multipurpose file.

3. Processing time: The time between the decision to secure data and its availability is relatively long for either the census or a complete primary survey but relatively short--a matter of weeks--for the two commercial sources.

4. Currentness: Because of the time interval and the processing time involved, the census can only be rated as fair given these criteria. Comprehensive primary surveys for transportation data have become even more infrequent. The two commercial sources and tabulation of local permits can provide annual reports. Where available, Polk is usually the only current source of household size and vacancy rates--two factors that can change rapidly in subareas.

5. Cost: For a user, the cost of a primary survey would generally be highest and the census would be the lowest (the cost, for the most part, would have already been paid through federal appropriations). The commercial sources generally tend toward the low side.

6. Multiple use: When the cost of a data set can be spread over several users, the cost per use can be substantially reduced. The Polk file appears to have the greatest potential for multiple use. In a number of cities and areas it is already being shared among various departments and agencies with consequent substantial reductions in the cost per use.

7. Statistical facility: Polk and the two local sources list individual households, and their characteristics permit types of statistical analysis that are simply not possible with the small-area summaries provided by the census (due to confidentiality rules) or by International Data Development (due to its techniques).

Employment File

Employment files, required for generating trip "attractions" for the transportation models, are increasingly being used for additional purposes. These include economic development planning, econometric modeling, transportation market segmentation, and civil defense planning. Four data sources for employment files will be discussed here: (a) state employment security files--the most widely used base; (b) a primary field, telephone, and/or mail survey; (c) Dun and Bradstreet data; and (d) R.L. Polk data. The last two are commercial.

In addition, a wide range of other public and commercial sources is used, usually to supplement one of the above sources. These include city and manufacturing directories, commuter computer files, and the 1970 U.S. Census place-of-work data.

Dun and Bradstreet is the most widely used commercial source for creating employment files for transportation planning. The source is a by-product of the credit-rating and the business information compiled on approximately 3.8 million establishments in the United States. Dun and Bradstreet information appropriate for use in transportation planning for each location of a company includes business name; secondary business name; street address; mail address; Standard Industrial Classification (SIC) Code (primary and up to five secondary codes); line of business of each establishment location; annual sales volume (if available); number of employees; number of employees total; identification of the facility as a branch, subsidiary, or headquarters unit; area code and telephone number; name and title of chief executive; and various geographic codes. Dun and Bradstreet can geocode these data items to census tracts inexpensively for 60-75 percent of firms by using the proprietary Reuben H. Donnelley address-coding guide. They will also code the remainder manually--or the MPOs can do this.

Compared with other sources, the most serious weakness of Dun and Bradstreet is underreporting, especially of small firms, particularly in the services classifications. This can be resolved by factoring up tract totals by SIC category to aggregate to known county totals. Public and semipublic (e.g., church) employment is completely omitted, requiring a supplemental survey. A final weakness may be a degree of overreporting of employment by employers. The company's strengths compared with employment security files are in its (a) greater provision of employment data by location in multiple-location businesses, (b) coding by census tract, and (c) reporting of an employment count rather than by classification-of-size categories.

The Polk city directory canvasses list businesses and industries as well as households. Data on tape for each commercial establishment includes the following items of particular interest to transportation planners: firm name, address, telephone number, four-digit SIC classification, change of occupants, vacancies, demolitions, and new construction.

As with the household file, Polk will geocode all of this data to census tract, block, traffic zone, or other small area. A great merit of Polk is that it produces the most complete list of establishments that shows separate locations by address. Compared with other secondary sources, it is strong on small establishments, especially services.

The principal weakness of Polk data as a basis for an employment file is its lack of an establishment-employment figure. There are several ways in which Polk can provide the basis for subarea employment total by SIC classification, the most cost effective of which might be to match geocoded Polk employers with those in the state employment security listing.

Trip File

The Polk data are unique in their ability to provide a source for current journey-to-work data. Because approximately half of all transit passengers are making work trips, Polk provides a useful, unique data source for bus-route planning and for identifying market segments for whom customized service can be planned. The journey-to-work file can be developed by using computer methods for matching employers' names. The Alamo Area Council of Governments (San Antonio), for example, produced a 36 percent sample of work trips from the Polk data. The other two sources of trip tables--the U.S. Census place-of-work and journey-to-work questions and the origin-and-destination (O-D) travel survey cannot always provide current data for two reasons: They are conducted too infrequently and require, typically, a year or two (or more) to process the data for use. Also, as mentioned earlier, the high cost of the O-D survey has reduced the likelihood of new up-to-date data from this source. Substitute methods, such as synthesizing trip generation and distribution patterns, while often acceptable for highway planning, are not promising for transit-route planning.

GENERAL FINDINGS

This section covers several conclusions that do not fit the three-file organization discussed above.

Decision Factors That Affect Data Source Selection

There is no single best data source for all urban transportation-planning situations. The wide variation in local conditions makes generalizations difficult and qualifications necessary.

Use of Commercial Data Sources

Dun and Bradstreet was the source selected for employment files in two major metropolitan areas investigated--Denver and Portland. In Portland, Dun and Bradstreet was considered the least-costly and most-timely single source when compared with the cost of processing the state employment security data. In Denver, it was considered a necessary complement to state employment security data in the creation of a very complete and highly sophisticated file. In both areas, the Dun and Bradstreet file is easily geocoded and could easily be updated.

In two areas--San Diego and Battle Creek--where state employment security data had long been used to create employment files, Dun and Bradstreet was analyzed against the employment security file and was rejected in favor of continued use of the latter. The high rate of discrepancies between the two files was a factor in these decisions.

Southeastern Virginia MPOs and transportation planners for the Alamo Area and Saginaw expressed satisfaction with their continuing use of the R.L. Polk Profiles-of-Change data set. In addition, the planners for VIA Metropolitan Transit in San Antonio were enthusiastic about the unique value of the work-trip file. The study found the following current and indicated uses for Polk data in transportation planning:

1. Work-trip file (San Antonio);
2. Basic source for household file (Saginaw);
3. Current vacancy rates and household size--two factors that greatly affect trip generation and can change sharply in particular subareas (San Antonio, Saginaw, and southeastern Virginia);
4. Monitoring for land use and socioeconomic shifts that indicate changes in trip generation and modal choice (San Antonio, Saginaw, and southeastern Virginia)--as the attention of transportation planners continues to focus on transportation system management and brokerage, current annual data will become increasingly necessary; and
5. A source for the employment file, in conjunction with state employment security data, county business patterns, or Dun and Bradstreet data.

It may be significant that in all three of the metropolitan communities that use Polk data, the cost is borne in part by other users of the data. The other users are agencies concerned with community development block grants and the planning and monitoring strategies to control the quality of neighborhood housing.

The feasibility of using Polk data to create employment files could neither be established nor discounted by the study because no MPO was using it or had considered using it for this purpose. However, there are indications that Polk data may have potential.

Quality of Employment Files

The employment files of MPOs appear to have more problems and inaccuracies than their household files. This conclusion regarding the probable quality of most MPO employment files is based on two study observations: (a) the frequency with which employment file users reported sharp, often inexplicable, variations in time-series data (e.g., southeastern Virginia) and (b) the many discrepancies between two regularly used sources, such as Dun and Bradstreet and state employment security files, whenever these are printed out side by side (e.g., San Diego and Battle Creek).

This conclusion suggests that employment file strategies and procedures could be a fruitful subject for research and training to improve file quality.

Multipurpose Information Systems

In contrast to the single-purpose information systems of transportation studies in the 1960s, all of the agencies interviewed emphasized the multiple uses and users of their data sets. Given the Proposition 13 climate, inflation, tight funding, and other pressures for cost-effectiveness, it can be expected that more MPOs will find ways to justify the cost of data sources through multiple use. Dun and Bradstreet, as well as Polk, have great potential as multipurpose information systems.

Potential of Commercial Data

What is the potential for commercial data--particularly the products of Dun and Bradstreet and R.L. Polk and Company--in urban transportation planning? A considered answer to this central question is a prerequisite to more specific recommendations. Based on interviews with data users (and taking into account the strengths and weaknesses inherent in that methodology), the following conclusions are drawn and then discussed in detail.

1. An assessment of the potential of commercial data must recognize that a great diversity of local metropolitan transportation-planning situations exists and that each commercial source competes with several better-established public sources. For example, an MPO with an investment in geocoding state employment security data may not find the use of Dun and Bradstreet cost effective. (However, see recommendation 5 below.) On the other hand, an MPO that has no automated means of updating its employment security file may find that Dun and Bradstreet is its best buy. In still another MPO, where employment data have uses beyond transportation planning, the additional information in the Dun and Bradstreet file may well warrant its purchase. In short, there is, at least currently, no universal best source for all situations.

2. Commercial data sources were selected. Their users were unanimously satisfied. Some were clearly enthusiastic. All felt that their expectations for the data at the time of selection were borne out in their subsequent use.

3. A variety of reasons was given for selection of commercial data, and these reasons appear plausible and credible. Both Dun and Bradstreet and Polk were selected for use in situations where they were the least expensive, most current, and most readily available sources. They facilitate the frequent updating needed to provide time series, thus revealing trends and early indications of change in trip generation, modal choice, and transportation need. They also facilitate comparisons between subareas and with other urban areas. They have a strong potential for purposes other than transportation planning, which may permit acquisition and processing costs to be shared by other agencies. These private sources are also less burdened with confidentiality limitations than are census and employment security data. Nor are they affected by the federal paperwork-reduction programs required by current Office of Management and Budget guidelines.

4. Both Dun and Bradstreet and Polk have made major efforts and significant progress in adapting their sources to public use. A case in point is Polk's announced decision to provide motor vehicle registration data coded to place of residence (for the use of MPOs and transportation planning only). A few serious problems remain, but these appear to be no greater than the equally serious, but different, problems encountered in the use of public data sources.

5. If a forecast can be risked, it would be that the existing trend toward increased use of commercial data for urban transportation planning will continue. The basis for this forecast is that the advantages listed in item 3 above, which led to the selection of Dun and Bradstreet and Polk by the MPOs interviewed, are likely to be even more pronounced in the future--particularly the continued tightening of funding available for transportation-planning data.

RECOMMENDATIONS

1. Use of commercial data--MPOs and state transportation agencies, when updating their socioeconomic files, should investigate commercial data sources.

The appraisal of the potential of commercial sources presented in the first part of this paper supports this recommendation.

2. Guidelines for integrated, multipurpose MPO household files--FHWA and UMTA, with the help of a mutual-aid network of users, should prepare and maintain guidelines for integrated, multipurpose household files and disseminate them to MPOs and to current and potential local, state, and federal users of MPO data.

The guidelines, as recommended, are both a product and a process. The product, similar to the Urban Transportation Planning System (UTPS) aids, would document the best MPO household-file operating procedures for use by other MPOs and file providers and users. The process would create an informal mutual-aid network of MPO data staffs.

The guidelines would serve a number of purposes. First, they would give impetus to and support the existing trend toward regional, integrated, multipurpose household files, with consequent improvements in file quality and reductions in user cost. Second, the guidelines would provide valuable support to UTPS efforts. The UTPS has been very effective in equipping MPOs, transit planners, and states with a standard versatile software set. Its Achilles' heel is that the quality of the output data is no better than the quality of the input data. The guidelines could help ensure the vitality of the models by encouraging sound selection and processing of data inputs. Third, new junior staff in many states and MPOs could benefit from training aids such as the proposed guidelines. Fourth, the guidelines would help staffs adapt data files to new uses as conditions now frequently require. As more attention is given to such areas as transit, paratransit, transportation system management, and market segmentation, current data will become increasingly important. Finally, the guidelines would serve as a national source of information for users and potential users of the kinds of data that are in MPO household files.

3. Qualification of household data sources using the 1980 Census--FHWA and UMTA should initiate comparisons of Polk and International Data Development household data with the results of the 1980 Census to identify and explain discrepancies and evaluate the source.

Because Polk data are being used more frequently and have definite potential as an intercensal data source, there is a strong case for definitive tests of the quality of Polk data through a careful comparison with the 1980 Census. Since Polk data are used so often, precisely how much variation from census data to expect should be known.

The case for a test takes on added significance when the increasing use of synthesized data is considered. (The entire International Data Development file is synthesized.) Polk suggests such procedures for household income, handicapped persons, and elderly persons. If these prove to be reasonable surrogates, they represent significant cost-saving breakthroughs. Similar tests are also needed of household files based on building permits, electrical connections, and other local sources.

One concern underlies recommendations 4, 5, and 6, which follow. The employment files of many MPOs appear to have more problems and inaccuracies than their household files and, in fact, may in some cases be below acceptable standards. The three recommendations below each provide part of a comprehensive, economical response to the problem.

4. Employment file guidelines--FHWA and UMTA, with the help of a mutual-aid network of users, should prepare and maintain guidelines for MPO employment files and disseminate them to MPOs and to current and potential local, state, and federal users of MPO data.

The surest way to tackle the employment file problem is to use the technical competence within MPOs and states through a mutual-aid network, assisted by FHWA and UMTA. The process would both generate and disseminate procedural aids (guidelines) and provide channels for the exchange of improved and new procedures. This recommendation closely parallels that made concerning guidelines for household files. The reasons cited to justify that recommendation apply here also. Additional justification can be found in the greater difficulty of achieving an adequate employment file. There are more pitfalls, and staffs performing this task are more in need of guidelines, mutual aid, and training.

5. Central purchase of Dun and Bradstreet--FHWA and UMTA should make a central purchase of the Dun and Bradstreet establishment files for the 286 metropolitan areas, which, when combined with the guidelines and MPO mutual-aid network, will provide an inexpensive, independent data source for checking, correcting, and updating employment files.

A major cause of the poor condition of many MPO employment distribution files is the difficulties encountered--both in original file preparation and in updating--in geocoding state employment security files and breaking out multiple locations. The indicated remedy would be an independent, inexpensive source, such as Dun and Bradstreet, that had multiple locations and was coded to street addresses. The Dun and Bradstreet file could be obtained nationwide at a cost of $0.08/establishment. It could then be used by MPOs to check their existing files. Discrepancies between employment security and Dun and Bradstreet files, if greater than a locally determined threshold size, would then be resolved by telephone, mail, or on-site interviews. The guidelines would provide suggested procedures.

6. Qualification of employment data sources--FHWA and UMTA should initiate comparisons of the major sources of employment data to identify and explain discrepancies and to evaluate the sources.

Comparisons of Dun and Bradstreet, employment security, and other sources of employment data show major discrepancies, and there are sharp differences of opinion among the MPOs interviewed as to the accuracy of the various sources. This lack of agreement simply cannot be resolved without an in-depth comparative study of these sources. The study would produce definitive information on the relative accuracy and other relevant characteristics of the principal employment data sources to permit better-informed MPO decisions as to which source or sources to use in particular circumstances. This type of in-depth comparison by MPOs usually cannot be done under normal operating pressures.

7. Polk trip-file processing--R.L. Polk and Company should offer journey-to-work files to transit providers and planners by use of the San Antonio method.

Current work-trip information is important for planning conventional transit as well as paratransit. The San Antonio experience demonstrates that a trip file suitable for this purpose can be generated inexpensively from Polk data. Polk data appear to be the only means for inexpensively producing a work-trip table during the intercensal period. It can empirically reveal recent shifts in transit ridership patterns that are unlikely to be reflected through simulation.

8. Qualification of the Census and Polk trip files--FHWA and UMTA should initiate comparisons of work-trip tables from the 1980 Census, Polk, and, if possible, a recent O-D survey; note and explain discrepancies; and evaluate the sources.

This proposal parallels recommendations 3 and 6. It calls for tests of the 1980 Census journey-to-work results comparable to those conducted after the 1970 Census. New elements, since the 1970 tests, are the use of Polk data to produce a trip file and the scarcity of current O-D studies for comparison. The latter underscores the importance of the former.

9. Polk one-year tapes--R.L. Polk and Company should carefully and objectively explore the market feasibility of providing one-year inventory tapes to meet the needs of transportation and other agencies.

Polk has already acted on this recommendation as a direct result of the study. This will greatly reduce the cost of Polk data for transportation-planning purposes.

10. FHWA and UMTA staffing--Establish a small FHWA and UMTA central staff group to develop the guidelines, facilitate the mutual-aid functions (recommendations 2 and 4), and provide supervision and professional direction to accomplish recommendations 3, 5, 6, and 8.

Recommendations 1 through 6, and 8, all call for FHWA and UMTA leadership to aid the MPOs. This aid could be provided by a small central staff group, including at least one full-time person. This is a very important recommendation. Without its implementation, the advantages of commercial data and the general improvements in transportation-planning data files outlined will be much more difficult to realize. This small investment in staffing could have major payoffs in less expensive and more current data, better-quality files, and multiple cost-shared use.

ACKNOWLEDGMENT

This research was aided by a grant from the Urban Mass Transportation Administration and the Federal Highway Administration. I would like to thank the many persons interviewed during the preparation of this report--especially those at FHWA, along with Merton Meyer of R.L. Polk and Company and Richard Looney of Dun and Bradstreet.

REFERENCE

1. R.C. Stuart. Commercial Data Sources for Transportation Planning. Urban Planning Division, Federal Highway Administration; Urban Mass Transportation Administration, June 1979.

Notice: The Transportation Research Board does not endorse products or manufacturers. The two commercial data sources discussed here were specified in the author's contract with the Federal Highway Administration and the Urban Mass Transportation Administration because they are considered the principal commercial data sources used in urban transportation planning. The use of their names is essential to the nature of this report.

Estimating Vehicle Miles of Travel by Using Random Sampling Techniques

LAP T. HOANG AND VICTOR P. POTEAT

The development of estimates of vehicle miles of travel (VMT) by state and local agencies has been an area of concern for many years. The problem of costs versus accuracy has long been a deciding element in developing VMT estimation procedures. In an effort to lower costs and maintain accuracy, the Federal Highway Administration (FHWA) developed a statistically based procedure for VMT calculation. FHWA's product is presented in the Guide to Urban Traffic Volume Counting (GUTVC). The method proposed in the GUTVC was based on random selection and sampling of highway links that were stratified by volume, facility type, and area type. Based on statistical sampling theory, a desired confidence interval could be assumed and a sample size calculated. The sample was randomly selected from the population of links. The selected links were then matched with a randomly selected day for sampling (i.e., taking counts). In comparison, the current method used by the Florida Department of Transportation (FDOT) employs sample counts and gasoline consumption information to arrive at VMT estimates. The purpose of this FDOT study was to test the GUTVC method for applicability as well as economical and statistical reliability under real-world conditions. FDOT found that VMT calculated by using the procedures presented in the GUTVC were reasonably close to VMT estimates currently developed in-house. In addition, by using traffic counts obtained through the sampling process to create data files, it is possible to entertain thoughts of reduced costs over long periods of time for data collection.

Vehicle miles of travel (VMT) is a measure used by a vast range of professions as an analytical tool. Some examples of VMT uses are found in highway management, traffic engineering, marketing, urban transportation planning, and trend monitoring. Therefore, a reliable inexpensive method of calculating VMT is needed.

In response to a Federal Highway Administration (FHWA) request for participants to test a newly developed procedure for VMT estimation, the Florida Department of Transportation's (FDOT's) Urban Transportation Planning Section agreed to conduct a pilot study. The procedure tested is presented in FHWA's Guide to Urban Traffic Volume Counting (GUTVC) (1). The method tested uses statistical random sampling. FDOT's current method uses gasoline consumption data and selected counts. The study area selected was Brevard County, which is located on the east coast of Florida, south of Daytona Beach.

The approach taken by FDOT was one of practical application to existing data and street networks. The techniques used and results obtained are presented here.

DATA PREPARATION

The organization of base data for preliminary uses was the initial step in preparing for the calculation of sample size. Historical traffic counts were used as a basis for stratification of links by volume. The existing Brevard County Urban Area Transportation Study link-node map was used as the base map for link identification and selection. As suggested in the GUTVC, FDOT's study area was categorized by area types [rural, central area, and central business district (CBD)] and facility types (local, arterial, and freeway-expressway). The GUTVC suggested that volume strata within each area type should be developed. FDOT stratified only the arterials by volume because sufficient historical data were not available for local and freeway links. The range of those volume strata used was 5000 average daily traffic (i.e., 0-5000; 5000-10 000). An additional step taken by FDOT to reduce travel and sampling costs was to separate the study into three geographic sectors--north, central, and south.

To remain consistent with those values suggested in the GUTVC, FDOT used a confidence interval of 68 percent and the following relative errors by functional class: local streets, 15 percent; arterial streets, 5 percent; and freeway-expressway, 10 percent. These values were used in sample-size calculations. Table 1 gives strata, area, and statistical data for arterial links.

SAMPLE-SIZE CALCULATIONS

The formulas and methods used by FDOT to calculate sample size can be found in Appendix C of the GUTVC (1). An outline of FDOT's calculations appears below.

Local Streets

In order to perform the necessary calculations, the following data were compiled: total local street miles (N_h) = 1023.25; temporal variance ($S_2{}^1$) = 300; spatial variance ($S_2{}^2$); mean volume (V_h) = 1000; estimated number of links (L) = 2729;

Table 1. Arterial link data for north, central, and south sectors.

Area Type	Volume Stratum	Street Miles	No. of Links	Link Length (miles) Average	Link Length (miles) Variance	Composite Spatial and Temporal Variance	Composite Stratum Standard Deviation
North sector							
CBD	2	1.00	4	0.25	0.10	1830	879
	3	0.18	1	0.18	0.00	2121	382
Central area	1	30.69	93	0.33	0.25	1581	814
	2	13.76	43	0.32	0.20	1830	1610
	3	4.03	13	0.31	0.20	2121	2585
	4	3.78	9	0.42	0.25	2304	4480
Rural	1	46.02	39	1.18	0.25	1581	1967
	2	16.17	21	0.77	0.25	1830	2345
	3	1.53	3	0.51	0.25	2121	3307
Central sector							
CBD	2	0.39	3	0.13	0.05	1830	444
	4	0.26	1	0.26	0.00	2304	599
Central area	1	39.69	81	0.49	0.25	1581	995
	2	14.85	45	0.33	0.20	1830	1617
	3	7.68	16	0.48	0.25	2121	3287
	4	3.12	12	0.26	0.08	2304	1523
	5	6.37	13	0.49	0.25	2704	5779
	6	0.20	1	0.20	0.00	2889	578
	7	1.74	6	0.29	0.10	2932	3359
Rural	1	8.28	12	0.69	0.25	1581	1257
	2	15.00	12	1.25	0.25	1830	2958
	3	5.70	6	0.95	0.25	2121	3718
South sector							
CBD	3	1.35	5	0.27	0.08	2121	1152
Central area	1	61.41	89	0.69	0.25	1581	1257
	2	46.56	97	0.48	0.25	1830	2070
	3	24.36	58	0.42	0.25	2121	3250
	4	8.10	18	0.45	0.25	2304	4496
	5	4.96	8	0.62	0.25	2704	5870
	6	0.70	2	0.35	0.10	2889	2930
Rural	1	44.03	37	1.19	0.25	1581	1982
	2	12.98	11	1.18	0.25	1830	2860

average link length (l_h) = 0.375 mile; and variance in link length (S^2) = 0.25 mile.

The process is as follows:

1. Calculate the composite spatial and temporal variance.

$$S_h^2 = \sqrt{[(S_1^2)^2 + (S_2^2)^2]}$$
$$S_h^2 = 671 \qquad (1)$$

2. Calculate the composite stratum variance. The composite stratum variance (S_h^2) takes into consideration the variations of volumes per link and lengths of link.

$$S_h{'}^2 = [\bar{\ell}_h^2(S_h^2)^2] + [\bar{V}_h^2(S_\ell^2)^2]$$
$$S_h{'}^2 = 125\,815 \qquad (2)$$

3. Calculate the composite stratum standard deviation.

$$S_h{'} = \sqrt{S_h{'}^2}$$
$$S_h{'} = 355 \qquad (3)$$

4. Calculate the mean vehicle miles per link.

$$\bar{X}_L = (\Sigma \bar{V}_h N_h)/(\Sigma L)$$
$$\bar{X}_L = 375 \qquad (4)$$

Because local streets were not further stratified, the number of local streets to be sampled (N_L) can be calculated by using this formula:

$$N_L = (Z^2 S_h^2)/(e^2 \bar{X}_L^2) \qquad (5)$$

where Z = normal variate = 1.0 and e = relative error = 0.15. Thus, N_L = 40. Good data were obtained for 27 links.

Arterial Streets

A stratified random sampling technique was applied to the arterial streets. The use of volume strata enabled FDOT to reduce the required number of links to be sampled by using strata with less variance. These data were obtained for each volume stratum: number of links, miles of street, mean volume, temporal variance, and spatial variance. After examining the adequacy of the base data, the following calculations were made for each sector of the county per area type and volume stratum.

1. Calculate the composite spatial and temporal variance.

$$S_h^2 = \sqrt{[(S_1^2)^2 + (S_2^2)^2]} \qquad (6)$$

2. Calculate the composite stratum variance.

$$S_h{'}^2 = [\bar{\ell}_h^2(S_h^2)^2] + [\bar{V}_h^2(S_\ell^2)^2] \qquad (7)$$

3. Calculate the composite stratum standard deviation.

$$S_h{'} = \sqrt{S_h{'}^2} \qquad (8)$$

4. Calculate the stratum weight factor.

$$W_h{'} = \text{number of links in stratum} \div \text{total number of links in sector} \qquad (9)$$

5. Calculate: (a) $W_h{'} S_h{'}$ and (b) $W_h{'} S_h{'}^2$.

6. Calculate the mean vehicle miles per link for the total sector of study area.

$$\bar{X}_L = (\Sigma V_h N_h)/(\Sigma L_T) \qquad (10)$$

7. Calculate the total links to be sampled in each sector.

$$N_L = (\Sigma W_h{'} S_h{'})^2/[(e^2\bar{X}_L{}^2/Z^2) + (1/L)\Sigma(W_h S_h{}^2)] \tag{11}$$

8. Calculate the number of links for each stratum from the total links.

$$n_h = (N_L W_h{'} S_h{'})/(\Sigma W_h{'} S_h{'}) \tag{12}$$

The following table gives the results of FDOT's sample-size calculations:

Sector and Facility Type	No. of Links to Be Sampled
North	
Freeway	11
Arterial	88
Local	7
Central	
Freeway	24
Arterial	51
Local	12
South	
Freeway	17
Arterial	71
Local	8
Total	289

Freeway-Expressway

The number of freeway-expressway links to be sampled was found in the following manner.

1. These data were taken from historical counts: $e = 0.10$; $Z = 1.0$; L = number of links = 164; N = number of miles = 189.62; $\bar{\ell}$ = average link length = 1.15 miles; $S_h{}^2$ = composite spatial and temporal variance = 3100; $\bar{V}_h$ = mean volume = 5000; and $S_\ell{}^2$ = variance in link length = 0.50.

2. Calculate the composite stratum variance.

$$S_h{'}^2 = [\bar{\ell}_h{}^2(S_h{}^2)^2] + [\bar{V}_h{}^2(S_\ell{}^2)^2]$$
$$S_h{'}^2 = 18\ 959\ 225 \tag{13}$$

3. Calculate the mean vehicle miles per link.

$$\bar{X}_L = (\Sigma V_h N_h)/(\Sigma L_T)$$
$$\bar{X}_L = 5781 \tag{14}$$

4. Calculate the number of links to be sampled.

$$N_L = (Z^2 S_h{}^2)/(e^2\bar{X}_L{}^2)$$
$$N_L = 57 \tag{15}$$

The 57 links were distributed to the three sectors of the study area based on each sector's proportion of the total number of freeway-expressway links. Good data were obtained for 52 links.

SAMPLE SELECTION

After calculating the required sample size, the next step was to determine which links were to be sampled and on what day each link would be sampled. Local streets and freeway-expressway links were selected by using simple random sampling techniques. With the use of a random number table, the links to be sampled were chosen. The day for sampling was chosen in a similar fashion.

Arterial links to be sampled were selected and sampled by using a stratified random sampling technique. The technique was similar to that used for freeway-expressway and local strata based on historical traffic-count data. Determination of the day for sampling was performed by using a random number table. To select links for sampling, (a) all links in the category under consideration were numbered and (b) a random number table was entered and numbers selected until the desired number of links was picked.

DATA COLLECTION

Sample data were collected by sector of county (north, central, or south). This was done to decrease daily travel by field crews. With the sampling of links over random locations, it was important for field crews to plan their routes to minimize travel. A field crew of two handled all data collection. Due to the random selection of sample days, the number of counts made on a given day varied from 1 to 11.

The counting period was 24 h. Only weekday counts were made due to the high cost of weekend counts. (Of course, weekend counts could be included at an increased cost.) FDOT factored the counts, made during three separate months, in order to obtain counts that represented single-month data.

VMT CALCULATIONS

After the task of collecting data was completed, tabulation of the data by county sector, facility type, area type, and volume stratum (as needed) was accomplished. Data were grouped and VMT calculated for (a) freeways in the north, central, and south sectors; (b) local streets in the north, central, and south sectors; and (c) arterials for the appropriate strata and area types for the north, central, and south sectors.

Formulas and techniques presented in the GUTVC were used for VMT calculations. The formulas used for VMT calculations are described below.

$$X_{hj} = V_{hj} l_{hj} \tag{16}$$

where

X_{hj} = VMT for link j of stratum h,
V_{hj} = traffic for link j of stratum h, and
l_{hj} = length of link j of stratum h.

$$V_h = (N_h X_{hj})/(n_h) \tag{17}$$

where

V_h = total VMT in stratum h,
N_h = total number of links in stratum h, and
n_h = number of links sampled in stratum h.

Freeway-Expressway VMT

Freeway-expressway links were sampled by individual sectors. Because all freeways in the study area were four lane, no stratification was used. An example of freeway-expressway VMT calculations follows.

1. Given--south sector, area type = rural; link length ($l_{1\ 9}$) = 0.40 mile; factored sample volume ($V_{1\ 9}$) = 4563 vehicles; and $X_{1\ 9}$ = (0.40) (4563) = 1825 vehicle miles.

2. X_{hj} is calculated for each freeway-expressway link with this sector and totaled: ΣX_{hj} = 83 554 vehicle miles.

3. Given--n_h = number of links sampled = 16; N_h = population of links in sector = 54; and VMT = (83 554 ÷ 16) (54) = 281 995 vehicle miles.

Table 2 summarizes freeway-expressway VMT calculations.

Table 2. Freeway link VMT.

Sector	Area Type	ΣX_{hj}	n_j	N_h	VMT
North	Rural	105 409	11	33	316 227
Central	Rural	164 900	24	74	508 442
South	Central area	1 786	1	3	5 358
	Rural	83 554	16	54	281 995
Total			52	164	1 112 022

Table 3. Local street VMT.

Sector	ΣX_{hj}	n_j	N_h	VMT
North	1740	7	682	169 526
Central	3934	12	1064	348 814
South	1792	8	983	220 192
Total				738 532

Table 4. Arterial link VMT for north, central, and south sectors.

Area Type	Volume Stratum	ΣX_{hj}	n_j	N_h	VMT
North sector					
CBD	1	10 731	4	4	10 731
	2	2 191	1	1	2 191
Central area	1	14 179	20	117	82 947
	2	49 516	15	43	141 946
	3	31 773	7	13	59 007
	4	36 086	9	9	36 086
Rural	1	59 783	18	69	229 168
	2	69 586	11	21	132 846
	3	19 985	3	3	19 985
Central sector					
CBD	2	5 360	3	3	5 360
	4	4 106	1	1	4 106
Central area	1	17 670	9	187	367 143
	2	27 193	7	47	182 582
	3	35 778	6	17	101 371
	4	13 735	1	13	178 555
	5	111 987	9	14	174 202
	6	6 756	1	2	13 512
	7	36 732	5	6	44 078
Rural	1	3 121	2	30	46 815
	2	46 778	4	13	152 029
	3	49 298	3	6	98 596
South sector					
CBD	3	1 631	1	5	8 155
Central area	1	18 255	10	129	235 490
	2	77 890	20	99	385 556
	3	64 616	17	60	228 056
	4	99 000	6	18	297 000
	5	116 298	6	8	155 064
	6	15 560	2	2	15 560
Rural	1	69 863	6	49	570 548
	2	25 317	3	5	42 195

Local Street VMT

Local streets were not stratified by volume due to insufficient historical data. Thus, the only breakdown of local streets within the study area was by sector. The example below illustrates the technique for local street VMT calculation.

1. Given--north sector; link length ($l_{1\,1}$) = 0.38 mile; factored volume ($V_{1\,1}$) = 1278 vehicles; and $X_{1\,1}$ = (0.38) (1278) = 486 vehicle miles.
2. The total of X_{hj} in the north sector was ΣX_{hj} = 1740 vehicle miles.
3. Given--n_h = number of links sampled = 7; N_h = population of links in sector = 682; and VMT = (1740 ÷ 7) (682) = 169 526 vehicle miles.

Table 3 summarizes local street VMT for all three sectors of the study area.

Table 5. Project-calculated variances and relative errors.

Facility Type	Calculated Variance (000s)	Relative Error: Calculated	Relative Error: Estimated
North arterial	112	0.16	0.05
Central arterial	171	0.12	0.05
South arterial	205	0.11	0.05
Freeways	211	0.19	0.10
North local	38.9	0.23	0.15
Central local	129	0.37	0.15
South local	80.3	0.36	0.15

Arterial Street VMT

The calculation of arterial street VMT required a more detailed process due to the stratified random sampling technique used. Arterial links were sorted by county sector, area type, and volume stratum. The volume stratum used had a range of 5000 vehicles. An example of arterial VMT calculations is given here.

1. Given--central sector, area type = central area; volume stratum = number 7; N_h = 6 and n_h = 5.
2. Total VMT for the sample link day was X_{hj} = 36 732 VMT.
3. Total VMT for volume stratum number 7 of central sector, VMT = (X_{hj} ÷ n_h) (N_h), or VMT = (36 732 ÷ 5) (6) = 44 078 vehicle miles.

These calculations were made for each volume stratum within each area type of each sector. Table 4 summarizes data for each sector of the study area. To get the total VMT for the study area, individual facility-type totals were summed. Total study area VMT is shown in the table below:

Facility Type	VMT
Freeway	1 112 022
Arterial	4 020 880
Local	738 532
Total	5 871 434

VARIANCE CALCULATIONS

To obtain a measurement of the statistical reliability of the VMT estimates and provide a basis for subsequent modifications to sample-size calculations and data-collection processes, the variance in estimated VMT was calculated by using the formula below:

$$S_X{}^2 = \left[\sum_{h=1}^{R} (N_h/n_h{}^2)(1/n_{h-1})\right] \left[n_h \sum_{j=1}^{n} X_{hj}{}^2 - \left(\sum_{j=1}^{n} X_{hj}\right)^2\right] \quad (18)$$

where

$S_X{}^2$ = variance of estimated VMT,
N_h = total number of links in stratum h,
n_h = number of links sampled in stratum h,
X_{hj} = VMT for link j of stratum h, and
R = number of strata.

The calculated variances for the study area are shown in Table 5.

RELATIVE ERROR

The relative error of those VMTs calculated previously was found by using the following equation: relative error = S'_X/VMT where S'_X = variance of group in question. Once the relative errors are calculated for each group, a comparison of results versus initial estimates used for sample-size calculations can be made. For sequential VMT estimates, the actual relative error can be used in sample-size calculations. Table 5 compares initial estimates and final calculated relative error, both at a 68 percent confidence interval.

COMPARISON OF PROJECT VMT WITH CURRENT VMT ESTIMATES

The current method of VMT calculation used by FDOT is summarized here. VMT of the county level is found for state highway (including federal route) systems only. Total VMT--that is, VMT for all facility types--is found for the entire state by use of gasoline consumption data. VMT by county (i.e., state highway system) is found by applying traffic counts to the length of highway section from which they were taken to arrive at VMT for the sample link.

The current estimate of VMT for the study area on state highway streets is 3 634 400 VMT/day. Without a precedent of local and county system VMT, a comparison between FDOT's current method and the GUTVC method was difficult to make for the entire street and highway system. However, a comparison was made by using data collected during the study on state system streets and calculating state system VMT by using the method tested as opposed to FDOT's current figure. The VMT calculated by the GUTVC's method for state system streets was 3 694 856. FDOT's estimate was 3 634 400, as noted above. The difference between the two methods is relatively small.

CONCLUSION

Although the comparison between VMT calculated by methods presented in the GUTVC and FDOT's existing VMT estimation program is good, the high relative errors calculated in this case study point out some deficiencies in the application of the GUTVC's techniques. The use of correct and complete base data for sample-size calculations would do the most to reduce relative error. FDOT feels that a completed local street network would result in more accurate local street VMT estimates. It is obvious that the number of local street links sampled should be increased. The error in the link file on which freeway and arterial links were developed could have caused some of the variance and relative error experienced. The possible error in link lengths was discovered during FDOT's Brevard County Transportation Planning Update.

In order to reduce the costs of data collection, a data base should be developed in which historical data can be kept for future use. By creating a traffic-count data file, data can be used for more than VMT uses and reduce the amount of data collection necessary. This will not work in reverse because special counts do not supply the variety of facility-type data needed. FDOT also found that, by sectoring the study area geographically, travel costs were reduced substantially, accuracy in VMT estimates was not badly affected, and a useful form of VMT data was created.

Although the possibility of error in both sample-size calculation and data collection does exist, the comparison between the GUTVC's estimate and FDOT's estimate is relatively good. The GUTVC method provides a general breakdown of VMT by area and facility type--something not currently available by FDOT's method--and flexibility in the choice of area size from which one may select a sample.

ACKNOWLEDGMENT

Our study was funded by a research grant from the Federal Highway Administration on Testing Urban Vehicle Travel.

REFERENCE

1. Guide to Urban Traffic Volume Counting. Federal Highway Administration, U.S. Department of Transportation, 1975.

Small-Scale, Ongoing Home-Interview Survey in Pennsylvania

ANTHONY F. DREISBACH

This paper reports on the design, administrative aspects, and selected findings of the Southwestern Pennsylvania Regional Planning Commission's (SPRPC's) small-scale, ongoing home-interview survey implemented in the agency's six-county region surrounding Pittsburgh. As initially conceived, the survey was an ambitious effort with multiple objectives. The implemented survey was a modest effort with a single, surveillance-related objective: to measure key travel characteristics and to compare and isolate changes in the characteristics over time. Achievement of the objective would reestablish and serve to retain the credibility of SPRPC's transportation data base. As the objectives were revised over a two-year period, so, too, was the sample size. The final sample design, based on satisfying the surveillance objective still remaining within acceptable funding limits, involved a total of 1600 household interviews during 1978—with an additional 1600 scheduled for interviews during the 1979-1980 period. Guide-Post Research, Inc., a market research consultant employing experienced interviewers, performed the data collection. However, SPRPC maintained overall survey control—managing the effort from data collection through processing. The consultant achieved an interview completion rate of 91.1 percent for the first year of data collection. Selected data tabulations from the 1978 portion of the survey are also presented as an example of the information available and its accuracy. The tabulations indicated that no unusual travel activity occurred in the region during 1978.

On April 10, 1978, 18 households in the six-county area of the Southwestern Pennsylvania Regional Planning Commission (SPRPC) were personally contacted. The household members were interviewed to obtain transportation-related data. The 18 households were the first interviewed as part of SPRPC's ongoing home-interview survey--scheduled through 1980.

This paper reports on the evolution of SPRPC's home-interview survey. The survey was intended to show the transportation data needs of SPRPC, how the survey was shaped to satisfy these needs, and what management aspects were involved in a small-scale travel survey. This paper also presents selected tabulations from the first year (1978) of data collection. The tabulations are not given for a detailed discussion of the findings but rather as an example of the type of information available and the accuracy associated with such information.

EVOLUTION OF SURVEY

As conceived in early 1976, a home-interview survey was viewed as a means of achieving four major objectives: (a) to reestablish the credibility of SPRPC's transportation data base by the surveillance of travel and related characteristics; (b) to obtain data required to retain a long-range transportation planning capability; (c) to obtain data needed in carrying out a planning process oriented toward shorter-term project planning; and (d) to obtain data needed to satisfy (anticipated) federal requirements for reporting transportation indicators.

Surveillance Objective

In 1976, SPRPC was confronted with problems probably experienced by other agencies that had not conducted a major transportation data-collection survey for a decade. (SPRPC's major origin-destination survey was conducted in 1967.) That is, the commission's transportation data base was dated and its credibility, consequently, questioned. A home-interview survey was viewed as the means of reestablishing the credibility of the data base. A current bank of travel information, in turn, would support decisions by the commission members. It was also felt that agency credibility would be enhanced by exhibiting, by means of newsletters and press releases, a constant awareness of travel activity.

Because major use of the data would be by governmental jurisdictions represented on the commission--i.e., the city of Pittsburgh and the six counties--data would be collected for these areas. Monitoring travel activity to isolate changes, a task that would retain credibility, led to an ongoing survey approach.

Planning Objective

Since the transportation data base was dated, the validity of the travel demand forecasts, made by using models developed from the 1967 survey data, was questionable, as was the long-range transportation plan that was developed and evaluated by means of the demand forecasts.

A survey involving personal interviews of household members would obtain data needed to evaluate and update the travel-forecasting models. Ultimately, that task would lead to the reevaluation of the long-range plan.

Because of (a) severe funding restraints, (b) the increasing need for resolving structural deficiencies of the existing transportation system and (c) the uncertainties associated with implementing elements of a long-range plan, the transportation-planning process at SPRPC was becoming short-range oriented in 1976. With emphasis on a process becoming so altered, the need arose to develop transportation models capable of assessing specific project impacts within subareas of travel corridors for shorter time periods than previously considered. Although viewed as a means of retaining a technically sound, long-range planning capability, a home-interview survey was also viewed as a way of acquiring data necessary to develop models that could be used to address some of the issues within the short-range planning framework.

Reporting Requirements Objective

In 1975, the Transportation Research Board established the Advisory Committee on Urban Transportation Data-Reporting Needs and Requirements. The purpose of the committee was "to provide assistance to the Urban Mass Transportation Administration and the Federal Highway Administration in developing a national urban transportation-reporting system" (1).

The exact nature of the data to be included in the national system was unknown when the home-interview survey surfaced at SPRPC. However, what was known at that time was that vehicle miles of travel (VMT) would be an item included in the system. (SPRPC was aware of the evolving reporting system because a staff member served on the advisory committee.)

A home-interview survey was considered a sufficiently accurate means for obtaining data needed to estimate VMT. Specifically, a survey could be designed to collect adequate origin-destination data for zone-to-zone trip table development. Trips could then be assigned to the transportation network, and VMT calculated and summarized by the geographic areas and highway functional classes required for national reporting.

Initial Survey Design

To achieve the objectives for time-series data collection and to maintain financial feasibility, methodology developed by Wickstrom and Pisarski (2) was used in part in designing a continuing home-interview survey. The sample design was based on replacing data from the 20 000 households interviewed during 1967. It called for 5000 household interviews annually (about a 0.6 percent sample), which would have resulted in a completely new survey file and new trip tables after four years.

However, agencies expected to fund the survey (as part of the SPRPC Unified Planning Work Program) questioned the cost-effectiveness of the survey design, as well as its ability to fully satisfy the objectives. In general, the agencies did not consider the survey capable of obtaining data needed for addressing short-range transportation planning issues. (To retain economic feasibility, a less-than-detailed questionnaire was developed.) They also believed that VMT could be estimated with more accuracy by using volume data from a traffic-count program. They also felt a much smaller survey could satisfy the surveillance objective. Finally, because the survey could not be justified on the basis of the other three objectives, it could not be justified on the sole basis of the long-range planning objective--given the de-emphasis of that element of the transportation-planning process. However, the importance of time-series data collection to compare and isolate changes in travel characteristics was recognized--as was the importance of having a system of models that could be used to address issues confronted in short-range planning.

The concept of an ongoing home-interview survey was not categorically rejected. Rather, initial objectives and survey size were questioned. A less ambitious design was needed.

Redefined Objectives, Redesigned Survey

The home-interview survey design was revised in accord with the redefined objectives. It now reflected the data needs of a modeling system

Table 1. Home-interview survey sample design.

Geographic Area	Household Interviews 1978 (Full Sample)	1979 (One-Half Sample)	1980 (One-Half Sample)
City of Pittsburgh	200	100	100
Allegheny County (excluding Pittsburgh)	400	200	200
Armstrong County	200	100	100
Beaver County	200	100	100
Butler County	200	100	100
Washington County	200	100	100
Westmoreland County	200	100	100
Regional total	1600	800	800

responsive to short-range planning issues and of a monitoring program for time-series comparisons of travel characteristics. There was no major long-range planning objective nor a survey objective to satisfy requirements of the national reporting system. [VMT would be obtained from a traffic-count program based on a link-sampling procedure (3).]

Modeling needs for short-range planning hinged on an individual-choice mode-split model. A detailed questionnaire was developed to obtain data for model development (i.e., disaggregate data set).

Based on the interviewing level thought to be financially justifiable to the funding agencies, the sample was reduced to 1000 households per year throughout the six-county region. However, the agencies believed the revised design could not satisfy the monitoring objective (because the regional household distribution was the sampling basis, the monitoring function would suffer in the low-populated counties). The questionnaire developed to collect a disaggregate data set was judged far too lengthy and, therefore, too costly.

Final Objective and Survey Design

The major result of discussions on the redesigned survey was agreement that the county-level monitoring function was the prime survey objective. The collection of a disaggregate data set was deemed a secondary objective and, in a later mutual decision, excluded entirely from consideration during the final design effort.

The survey sample was designed by using standard statistical methods, tempered by a realization of available funding for this type of data-collection effort. Specifically, the ongoing aspect of the survey (a feature based on the policy decision intended to reestablish and then retain the credibility of SPRPC's transportation data) required that the survey be annually budgeted in the agency's Unified Planning Work Program. With other planning activities (such as transportation systems management) competing at increasing rates for limited planning funds, it was recognized that the survey cost should be kept sufficiently low to ensure continued funding through 1980.

Remaining cognizant of the cost issue, the final sample design task was initiated by generating a list of the key travel characteristics to be measured and monitored for each county and Pittsburgh. Of the items on the list, person trips per household was considered the most important to monitor. However, it was recognized that annual detection of changes in this item at the county level would require excess precision and, therefore, too large a sample, too costly a survey. Consequently, 10 percent accuracy (with 90 percent confidence) was accepted for estimating person-trip rates. The resulting sample design was 200 households per county and Pittsburgh.

To further reduce the overall sample size and because it was unlikely that more than 10 percent annual change in person-trip rates would occur in the counties, a decision was made to monitor biennially at the county level after first measuring the travel characteristics during 1978. That is, a full sample of households would be interviewed during 1978, while one-half of the sample would be interviewed during 1979 and the remaining one-half during 1980. Though county comparisons would be made on a biennial basis, sufficient accuracy was expected for annual regional comparisons.

The final sample design and interview schedule appear on Table 1. Note that Allegheny County (excluding Pittsburgh) has double the sample of other areas because an attempt was made to obtain better accuracy for estimating transit-related items (most transit service is available in Allegheny County). The decision to biennially monitor at the county level allowed Allegheny County's sample to be doubled. Based on an annual average, the total three-year sample was close to the 1000 interviews/year generally thought to be financially acceptable.

Although cost was a major issue throughout the survey design period, no budgetary problems materialized. In fact, annual total survey costs have represented less than 4 percent of the agency's total work program budget. Total survey costs have averaged about $66 000/year.

PREFIELD WORK

Questionnaire Design

A questionnaire was developed to obtain data needed to estimate selected travel and related characteristics. It was pretested and, based on the results, slightly revised. The questionnaire was also evaluated following the spring 1978 season of interviews. However, no modifications were made because no questions were reported misunderstood by the respondents.

The data items collected are listed below:

1. Household information--residents in household; out-of-region visitors; value of structure (if owned); monthly rent (if rented); age of structure; vehicles available by type, make, model, and year; household gross income; and persons contributing to household's gross income;
2. Person information (recorded for each person five years of age or older)--relation to household head; sex and age; automobile-driving status; occupation status; physical, mental, or other health conditions that (a) limit kind or amount of work, (b) prevent person from working, (c) limit or prevent use of transit, or (d) limit or prevent driving a car;
3. Work place and occupation information (recorded for all workers)--job status (full or part time), occupation, name and type of business, street address of work place; and
4. Trip data (recorded for each person five years of age or older)--trip purpose, trip-end location, mode of travel, vehicles used (if household vehicle), persons in vehicle, time of trip (start time and duration), captive or choice automobile user (work trips only), type and cost of parking (automobile-driver work trips only), captive or choice transit user, and blocks walked to and from transit stop.

Sample Selection

The total 3200 samples for the home-interview survey were random--systematically selected based on the housing distribution within each county. By using SPRPC's street series maps, home-interview samples were plotted on locational maps for use by interviewers. A list of sample addresses was also prepared for the interviewers and for overall survey control by SPRPC staff.

Consultant Selection

Throughout the survey design period, there was some question concerning who would perform the interviewing of the selected households. SPRPC advocated the use of a consultant with personnel experienced in interviewing; other agencies believed that, since the survey was ongoing, SPRPC should hire and train additional staff. When not interviewing, additional staff members were to be used for processing survey data.

After much discussion, all agencies agreed to the use of a consultant. SPRPC already had sufficient staff members for processing work; hiring additional personnel for the single purpose of interviewing would not justify the cost. The consultant selected to perform the interviewing was Guide-Post Research, Inc.--a Pittsburgh-based market research firm.

Although the selected consultant used experienced interviewers, the unique aspects of a travel survey involving origin-and-destination questions required additional training. The week before interviewing, the consultant's personnel assigned to the survey participated in an intensive three-day training session conducted by SPRPC staff members who managed the 1967 origin-destination travel survey. Additional training sessions were scheduled prior to each season of interviewing.

FIELD WORK

Personnel assigned to the field-work phase of the survey included a survey field supervisor, six interviewers, a quality-control clerk, and the director of survey operations.

The survey field supervisor was responsible for scheduling and assigning all work. The supervisor also assisted in follow-up interviews and carried out the preliminary editing of survey returns.

Interviewers were responsible for collecting data by talking personally, when practical, with each household member 16 years of age or older in accordance with interviewing procedures detailed in the procedures manual (4). Trip data for members younger than 16, but older than 5 years, were obtained from a responsible household member--usually the household head.

Interviewers worked a Tuesday-to-Sunday schedule and made their initial contact with the household either the day before or the day after the designated travel day (only weekdays were considered travel days; no weekend data were obtained). Contact before the travel day was made to explain the survey and distribute trip diaries to household members. The diaries were to be used by participants for recording their trips on the travel day; the interview would be conducted the next day.

The alternative to pre-travel-day contact was the cold-call method. This method simply involved contacting the household the day after the designated travel day and conducting the personal interview then.

Agencies funding the survey suggested that a test be made to determine which of the two contact approaches would yield better results in the form of both household cooperation and accurate responses. Consequently, both approaches were tested during the survey's first week. Results indicated that better cooperation was gained from households in rural areas by using the pre-travel-day approach. In urban areas, the cold-call method resulted in better cooperation. As far as data quality was concerned, however, no difference was detected between the two approaches. Nonetheless, to maximize the number of cooperative households, the cold-call approach was used exclusively in urban areas, while pre-travel-day contact was made only in rural areas for the remainder of the survey.

The quality-control clerk was housed in the survey office and received completed questionnaires from the field supervisor two or three times a week. Duties of this clerk included editing each returned survey questionnaire and conducting respondent callbacks as part of the quality-control procedure. Fifteen percent of the households were randomly selected and phoned for data-verification purposes.

The director of survey operations monitored survey progress and held meetings with all field personnel when necessary to correct procedural difficulties. The director also acted as liaison between the survey office and the SPRPC central office to ensure smooth operation and flow of completed interviews.

CENTRAL OFFICE WORK

The SPRPC central office staff consisted of two editor-coding clerks, a survey supervisor, and the manager of data services.

The editor-coding clerks carried out complete edits of the survey forms. They also reviewed self-coding item and coded selected data fields. Items requiring coding included worker occupations, trip-end and employment-site geographic locations (respondents were provided maps to help locate trip destinations; in most cases, including rural areas, the maps were not used), and land use activity.

The survey supervisor conducted all SPRPC quality-control callbacks; these were in addition to the 15 percent performed by the consultant. SPRPC's quality-control procedure involved callbacks to 25 percent of the households (later reduced to 15 percent as interviewer proficiency increased). Households were randomly selected; none failed the quality-control test. Any that would have failed would have been reinterviewed at the consultant's expense, as the contract with the consultant stipulated.

The manager of data services coordinated all survey functions, developed procedures, and interpreted survey policy. The interface of manual and computer operations was also the manager's responsibility.

SURVEY RESULTS: SELECTED FINDINGS FOR 1978

Although concern was expressed over the use of a consultant for data-collection purposes, this decision proved a wise one. The consultant carried out all duties assigned by SPRPC in a professional manner and satisfactorily completed the field work within the specified budget ($22.30/interview in 1978), quality standards, and time limitations. The overall noninterview rate of 8.9 percent and low refusal rate of 1.9 percent evinced a definite dedication to the survey by the consultant's field personnel.

Information obtained from the 1458 households with completed interviews was processed and formed the basis for analysis. As mentioned, only selected findings are reported and briefly discussed. These are offered to illustrate some data tabulations developed by using the survey information and also to show the accuracy of the findings.

Person Trips per Household

Household trip rates on an average weekday in 1978 are listed on Table 2. The rates were developed by using unlinked trip data.

The design accuracy (relative error) for estimating household trip rates was ±10 percent (90 percent confidence). The achieved accuracy, however, was slightly less--approximately 11 percent for all counties, except Allegheny and the city of Pittsburgh. Accuracy for these areas was calculated at 8 and 12.7 percent, respectively. Larger-than-expected data variability contributed to wider error ranges for rates calculated by county, although the ranges were acceptable. A ±5 percent relative error was calculated for the regional household trip-rate estimate.

Based on the calculated errors, the city of Pittsburgh's household trip rate was found to be significantly lower than rates for the other areas.

The major reasons were lower income (1977 median household income in the city was $9600, compared with a regional median of $13 100) and fewer automobiles available to city households.

With only one exception, there was no significant difference among the trip rates for the six counties. The exception was Armstrong County. This county's rate of 5.91 was significantly lower than the Beaver County rate of 7.52 per household. In this case, however, the major factors influencing household trip productions were similar for the counties. (The trip rates of these counties will be given special attention as the survey continues.)

Finally, the 1978 regional average household trip rate of 5.95 was not significantly less than the trip rate of 6.20 calculated by using SPRPC's 1967 home-interview survey data.

Automobiles per Household

The number of automobiles available for personal use by household members on an average weekday in 1978 also appears in Table 2. Relative errors for this item were 12.7 percent for the city, 5.5 percent for Allegheny County, between 7 and 8 percent for the remaining counties, and 3.4 percent for the region.

As expected, Pittsburgh households had the lowest number of automobiles available. Availability in all other areas was similar, with Washington County the exception. Households in this county had significantly more automobiles than Pittsburgh households, but significantly fewer than the other counties. Slightly lower income levels in Washington County explained the difference.

The 1978 regional average of 1.41 automobiles available was significantly higher than the 1.1 value calculated in 1967. Additional analysis of the survey data revealed that the increase in automobiles per household was not due so much to households purchasing and obtaining a first car as it was to their acquisition of a second and third, as illustrated by Table 3 (5,6). [Table 3 was developed for the Pittsburgh standard metropolitan statistical area (SMSA) for comparison purposes. The Pittsburgh SMSA includes Allegheny, Beaver, Washington, and Westmoreland Counties.]

Mode of Travel

Mode distribution for all trip purposes is reported in Table 4. It indicates that travel for all purposes was overwhelmingly by automobile. Except for the city, mode distribution also was essentially constant among the counties (values in Table 4 are generally subject to a 0.3 percent error, although the error is about 0.2 percent for Allegheny County and about 0.1 percent for the regional values). At the regional level, a slight decrease in transit use was experienced since 1967. The higher use of other modes in 1978 was attributable to a larger share of school trips on school buses since 1967.

Table 5 lists mode-use percentage (subject to a 0.3-0.4 percent error) for work trips by four household income groups. The low-income group contained the greatest percentage of transit users and carpoolers and the smallest percentage of people who drive to work alone. As household income increases, there appears a jump in the drive-alone mode from the low-income group to the $8000-$14 999 group, after which the drive-alone mode percentages leveled off.

As the sudden increase occurred for the drive-alone mode, so, too, did a sudden decrease in transit occur for the low-income group. The percentage of transit use was similar for the $8000-$14 999, $15 000-$24 999, and $25 000 or more income groups, following the sharp decline from almost 19 percent use by individuals in the low-income group.

Although changes in the percentage of carpoolers occurred among the income groups, this mode is equally used on a percentage basis by all income groups.

Trip Purpose

The trip-purpose distribution by county is offered in Table 6. The values are subject to a 0.2 percent error for Allegheny County, 0.1 percent for the regional values, and 0.3 percent for the remaining values.

There was no significant difference among the trip-purpose percentages across geographic areas, nor was a significant difference detected between the 1978 regional distribution and the distribution based on 1967 survey data.

SUMMARY AND CONCLUSIONS

The initial home-interview survey concept was a fairly ambitious effort to obtain data needed for various transportation-planning activities at SPRPC. The survey implemented was a small-scale, ongoing effort with a surveillance-related objective; the initial concept was molded by multiagency input over a two-year period.

From an administrative viewpoint, the survey has proved successful. The small sample survey, requiring only a limited number of personnel, was easily managed. Also, deadlines were met without exception, no cost overruns were experienced, data quality was monitored and maintained, and data processing was performed quickly and efficiently.

From the standpoint of accuracy of results, the survey can also be considered successful because achieved accuracy for the key travel characteristic--household trip rates--was close to that used for designing the sample.

Has the survey achieved its prime objective? Data collected during 1978 have been compiled for the major governmental units of the region, compared with available 1967 data, and readied for comparisons with 1979 and 1980 data. In this respect, the survey has satisfied the surveillance objective. However, the value of surveillance is found in its ability to serve decision-making functions. By itself, surveillance is a wasteful activity (2). Information from the survey is

Table 2. Person trips and automobiles per household in 1978.

Geographic Area	Person Trips per Household	Automobiles per Household
City of Pittsburgh	4.08	0.87
Allegheny County (excluding Pittsburgh)	6.26	1.53
Armstrong County	5.91	1.58
Beaver County	7.52	1.63
Butler County	6.43	1.65
Washington County	6.22	1.40
Westmoreland County	6.45	1.58
Regional average	5.95	1.41
Regional average in 1967	6.20	1.10

Table 3. Household automobile ownership distribution for Pittsburgh SMSA.

	Percentage Distribution		
Year	No Automobile	One Automobile	Two or More Automobiles
1978	19.0	38.8	42.2
1974 (5)	19.5	48.2	32.3
1970 (6)	20.5	51.3	28.2

Table 4. Percentage of 1978 person trips by mode of travel.

Geographic Area	Automobile Driver	Automobile Passenger	Transit	Other[a]
City of Pittsburgh	49.0	26.0	22.3	2.7
Allegheny County (excluding Pittsburgh)	58.2	24.5	4.3	13.0
Armstrong County	61.9	22.6	–	15.5
Beaver County	65.4	21.1	–	13.5
Butler County	58.4	24.1	–	17.5
Washington County	59.2	24.1	–	16.7
Westmoreland County	63.0	22.4	–	14.6
Regional average	58.7	23.9	5.3	12.1
Regional average in 1967	56.5	26.5	7.5	9.6

[a]Includes transit for all areas except Pittsburgh and the balance of Allegheny County.

Table 5. Percentage of 1978 mode distribution for work trips by household income (1977 dollars).

	Household Income ($)			
Mode	0-7999	8000-14 999	15 000-24 999	25 000 or More
Drive alone	53.6	72.9	72.4	69.3
Carpool	27.5	21.4	17.7	22.7
Transit	18.9	5.7	9.6	7.4
Other	–	–	0.3	0.6

Table 6. Percentage of 1978 trip-purpose distribution.

Geographic Area	Home	Work	Shop	School	Other
City of Pittsburgh	43.4	18.0	10.9	5.2	22.5
Allegheny County (excluding Pittsburgh)	43.6	14.9	11.1	8.6	21.8
Armstrong County	43.3	15.8	11.9	8.7	20.3
Beaver County	42.5	13.2	10.7	8.2	25.4
Butler County	44.3	14.1	11.6	10.3	19.7
Washington County	40.5	14.7	8.8	9.6	26.4
Westmoreland County	43.6	17.2	10.0	8.7	20.5
Regional average	43.2	15.5	10.7	8.3	22.3
Regional average in 1967	41.3	17.0	12.3	7.7	21.7

available for use by the regional decision makers and has been used in reports to this body. Because no unusual travel activities were detected during 1978, the degree to which the survey data have affected transportation decisions cannot be assessed--except to say that current data must surely have eased credibility problems associated with using dated information.

Some survey data (coupled with secondary source information) has been used in developing a travel demand-forecasting system responsive to short-range planning issues (while also providing the capability to make a longer-range forecast). To the extent that the survey data have been used in the demand-forecasting process, they have also served the decision-making function.

ACKNOWLEDGMENT

The work reported in this paper was supported by the Federal Highway Administration (FHWA), the Urban Mass Transportation Administration, and the Pennsylvania Department of Transportation. Technical assistance provided by many staff members of these agencies, especially James McDonnell of FHWA, is gratefully acknowledged. I would also like to thank George Ferguson, Robert Kochanowski, Wade Fox, and Charles DiPietro for reviewing the paper and Val Ferretti for editing the manuscript. The views expressed in the paper are not necessarily those of the sponsoring agencies.

REFERENCES

1. Advisory Committee on Urban Transportation Data-Reporting Needs and Requirements. Proposed Urban Transportation Data-Reporting Requirements for States and Metropolitan Planning Organizations: A Report to the U.S. Department of Transportation. National Academy of Sciences, Washington, DC, 1976.
2. G. V. Wickstrom and A. E. Pisarski. The Use of Longitudinal Information Systems for Continuing Urban Transportation Planning. Metropolitan Washington Council of Governments, Washington, DC, Sept. 1969.
3. Wilbur Smith and Associates. Urban Traffic Volume Counting Manual (Preliminary). Federal Highway Administration, U.S. Department of Transportation, Sept. 1974.
4. Home-Interview Survey Procedure Manual. Southwestern Pennsylvania Regional Planning Commission, Pittsburgh, April 1978.
5. U.S. Bureau of the Census. 1974 Annual Housing Survey. U.S. Government Printing Office, 1976.
6. U.S. Bureau of the Census. 1970 Census of Housing: Detailed Housing Characteristics. U.S. Government Printing Office, 1972.

The 1977 Census of Transportation: An Update

ROBERT TORENE AND JOHN CANNON

This paper reports on the status of the 1977 Census of Transportation, which consists of four surveys: National Travel Survey, Commodity Transportation Survey, Truck Inventory and Use Survey, and Survey of Nonregulated Motor Carriers and Public Warehousing. It deals with the methodology used in these surveys compared with that followed in the 1972 census. This report also examines the problems and benefits realized from the 1977 approach. Finally, it suggests additional improvements and alternatives possible for future transportation census taking, particularly for the 1982 census.

The Census of Transportation was initially conducted for the year 1963 and has been repeated for 1967, 1972, and 1977. The census was established to cover the major data gaps that already existed. In 1966, Donald E. Church, chief of the Transportation Division, U.S. Bureau of the Census, observed that "prior to the 1963 census, nearly all of the available transportation statistics were by-products of regulatory and promotional activities. Statistics were adequate for some aspects but poor or completely missing for others...[and] there were critical gaps." The most critical data gaps were identified in the areas of commodity flow, personal travel, trucking resources, and nonregulated motor carriage. The objective was to close these gaps without duplicating data already available from other government or private sources. This led to the adoption of a program consisting of a series of individual surveys: National Travel Survey (NTS), Commodity Transportation Survey (CTS), Truck Inventory and Use Survey (TIUS), and Survey of Nonregulated Motor Carriers (SNMC).

The 1977 Census of Transportation has been substantially modified and improved since the previous census was taken covering the year 1972. The methodology of each survey that is part of the Census of Transportation has undergone a critical review and evaluation between census years and, within limitations imposed by budgetary and employment-ceiling restraints, has been revised with the aim of improving the usefulness of these statistics. Unfortunately, some changes have had an adverse impact on timely release of these data.

Each survey will be discussed here in terms of changes in methodology, the reasons for those changes, and their consequences. Potential modifications for the 1982 Census of Transportation will also be examined.

1977 NATIONAL TRAVEL SURVEY

The purpose of the 1977 NTS was to provide statistical data on the volume and characteristics of nonlocal travel by the civilian noninstitutional population in 1977. For the survey, nonlocal travel was defined as any trip extending 100 miles or more from origin to destination. Specifically excluded from the survey were (a) travel taken as part of an operating crew on a train, plane, bus, truck, or ship; (b) commuting to a place of work; (c) travel by students between home and school; and (d) travel by members of the Armed Forces while on active duty.

The survey was based on a sample of households selected from each of the 50 states and the District of Columbia. Interviews were conducted with the sample households to obtain information on the number of trips taken by members of the household and on certain trip-related characteristics. Selected social and economic characteristics of household members were also collected. The survey was planned and jointly sponsored by the U.S. Department of Transportation, the U.S. Travel Service, and the Bureau of the Census.

Methodology

The 1977 NTS is based on information obtained primarily by personal interview from a national probability sample of approximately 25 000 housing units representing the total civilian noninstitutional population of the United States. Each sample household was designated to be interviewed at approximately three-month intervals over the survey period to develop a record of all nonlocal trips taken by household members during calendar year 1977.

Interviewing was accomplished primarily through a personal visit to the household. Each member 14 years of age or older was asked to report all the trips he or she took during the reference period. Household members not present at the time of the initial visit were later telephoned and asked to report their trips. Trips of persons 13 years of age or younger were reported by a knowledgeable adult member who was present at the time of the visit.

The 1977 NTS was based on a sample of households. Persons whose usual place of residence was located at the sample addresses during the first (April) enumeration formed the basic sample of households interviewed in subsequent visits.

As previously mentioned, several interviews were conducted with the sample household throughout the survey period. In all interviews subsequent to the initial one, households were asked to report trips taken since the last interview. A "bounding" technique was designed to prevent reporting of trips in a current reference period that actually were taken and reported in a previous reference period.

Comparison with Previous Surveys

Previous travel surveys were conducted as parts of the 1963, 1967, and 1972 Censuses of Transportation. The 1963 survey involved approximately 6000 sample households, interviewed by personal visit at quarterly intervals to obtain trips taken during 1963. In 1967 and 1972, questionnaires were mailed to a sample 12 000 and 24 000 households, respectively, at approximately quarterly intervals to obtain data on trips taken during the respective calendar years. Telephone interviews or personal visits were made only when it was necessary to clarify certain information reported, or to follow up nonresponses to the mail inquiry.

Probably the most significant change in the 1977 survey from the previous two travel surveys was the method of data collection. In 1977, the initial contact with the sample household in each interview period was made by a personal visit preceded by a letter. Household members who were not available at the time of the visit were interviewed later by telephone. In the two previous travel surveys, information was obtained primarily by mailing the questionnaire to the sample household, with the request to complete the form and return it by mail to the Bureau of the Census. It is not possible to measure the impact of this methodological change exactly, but tests conducted prior to the 1977 survey clearly indicated that the personal visit interview technique elicited the reporting of more trips and more accurate and complete information relating to the trips reported.

Other methodological aspects of the 1977 survey that differ from previous travel surveys included a change in the technique used to report trips of 100 miles or more and refinement of the adjustment factor used to convert straight-line distance to route miles. The 1963 and 1967 surveys included as a trip "each time a person goes to a place at least 100 miles away from home or is out-of-town one or more nights." For the 1972 survey, the definition of a trip was changed to "each time a person goes to a place at least 100 miles away from home and returns," thus omitting all travel under 100 miles regardless of whether one or more nights were spent away from home.

The 1977 survey used the 1972 trip definition although the technique for collecting information on such trips was modified. Respondents were asked to report all trips of 75 miles or more (one-way), and those calculated through the computerized system to be less than 100 miles in distance were excluded from the estimates of trips derived from the survey. This change in questionnaire wording was instituted in an effort to minimize the failure of respondents to report trips of 100 miles or more due to the possible misconception of actual distances traveled.

Another important revision in the 1977 survey involved the adjustment factors used in determining route miles for a trip. In 1963 and 1967, surveys used the one-way straight-line miles in presenting data on distances traveled. The 1972 and 1977 surveys present distances traveled in terms of route miles. Route miles are round-trip, straight-line miles adjusted upward for the general circuitous nature of the route traveled and are considered a more accurate measurement of actual miles traveled on a trip. The size and type of the adjustment was based on the mode of transportation. In 1972, a standard upward adjustment (circuity factor) was applied to all trips taken by a particular mode of transportation. Research conducted for automobile and truck trips prior to the 1977 survey resulted in separate circuity factors for trips taken by these modes between any two states, as well as for trips taken entirely within a state.

In addition to the circuity factor, automobile and truck trips were adjusted upward by 25 percent of the straight-line mileage in 1972 to account for a general tendency to take side trips and to travel along the most interesting or fastest, rather than shortest, routes. In 1977, after applying the circuity and the 25 percent side-trip factors for automobile and truck trips, the actual route miles for a trip were reduced by 14 percent. This adjustment was based on results from a study conducted in conjunction with the survey that clerically measured the length of trips by using actual route and side-trip information supplied by respondents. The study was based on a representative sample of the trips reported in the first and fourth quarters of the survey. There has also been a substantial increase in the amount of trip-related information obtained in the 1977 survey compared with previous surveys (1).

Difficulties and Problems Encountered

The major problem in the 1977 NTS was the sample methodology used. The 1977 NTS used a sample of households and stayed with the same household throughout the interview period. The main purpose of this methodology was to indicate the dispersion of travel among households and persons. If one or more trips were taken by any members of a household during the year, the household was classified as a traveling household for the year. Likewise, if a person took one or more trips during the year, the person was counted as a traveler for the year. Although the results of the 1977 NTS were not adversely affected by the methodology used, the time required to prepare the file and calculate the inflation or weighting factors was much longer than first anticipated. This was a major factor in the delay in the publication of the final report.

Looking Ahead to 1982

The major concern is one of providing timely data on travel volume and also providing information on the characteristics of traveling and nontraveling households and persons. As was mentioned earlier, providing the traveling and nontraveling characteristics added greatly to the delay in publishing the final 1977 NTS report. Obviously, alternative methodologies will be reviewed to determine which would be best for the NTS. One alternative, however, which will likely be seriously considered, is a split sample with a small national sample set aside and used to provide the traveling and nontraveling characteristics and a much larger sample used to provide information on trip volume and characteristics. It should then be possible to process the larger sample much more quickly because it would not be used for traveling and nontraveling characteristics.

COMMODITY TRANSPORTATION SURVEY

The purpose of the CTS is to provide statistics on the volume and characteristics of commodity shipments by manufacturing establishments in the United States. The survey is based on shipping data obtained from a net sample of 16 000 manufacturing establishments selected from the 1977 Census of Manufacturers' (COM's) universe of manufacturing establishments. Data collected include weight and value of commodities shipped by means of transport, as well as areas of origin and destination.

The CTS was substantially modified in 1977 in order to strengthen the conceptual framework by linking the sampled data back to COM universe totals. To accomplish this, several major steps were taken:

1. The CTS universe was redefined,
2. The value of shipments was determined from CTS respondents, and
3. CTS establishment data were directly verified at the local level with COM establishment data.

CTS Universe

The most important limitation of the earlier surveys was that the defined universe of the survey, namely intercity shipments of the larger manufacturing establishments, made the results difficult to verify or relate to other figures. (The COM excludes private-sector establishments with no paid employees and all government-owned and government-operated establishments.) Therefore, the universe has been broadened to include all shipments from all establishments classified in manufacturing. As a result, the Census Bureau will be able to edit, review, and analyze the CTS data against known universe totals and to verify and link the commodities "shipped" in the CTS to those "produced" in the COM.

The categories added to the 1977 CTS are

1. Local shipments (less than 25 miles) from in-scope establishments;
2. All shipments of establishments that shipped

90 percent or more of their volume less than 25 miles;

3. All shipments from establishments with less than 20 employees;

4. All shipments included in the following Standard Industrial Classifications (SICs): 19, ordinance and accessories; 2026, fluid milk; 2051, bread and other bakery products, except cookies and crackers; 2097, manufactured ice; 241, primary forest products; and 27, printing, publishing, and allied industries;

5. Shipments made by the following means of transport: (a) own power or towed, including motorized vehicles, aircraft, or vessels that are sometimes moved from the manufacturer under their own power, and commodities such as prefabricated buildings that are towed away; (b) pipeline, primarily affecting shipments of petroleum products from refineries; and (c) parcel post shipments through the U.S. Postal Service; and

6. Shipments originating in Alaska and Hawaii.

Value of Shipments

In order to link the CTS and COM data, a common data item was required. It was decided that the most useful and appropriate device was to request the value of the shipment from CTS respondents. Because traffic records are not usually priced out, i.e., do not carry the value of the shipment, many respondents were required to access other documents such as sales invoices and sales memos. However, there is no evidence that this added workload adversely affected the response rate. Although there was a 20 percent overall nonresponse rate, this compares favorably with the response rates experienced in the other economic censuses. Furthermore, of all those who responded, more than 90 percent reported both weight and value.

Benchmark CTS Data to COM

Each 1977 CTS establishment report was directly verified by a computer match to the corresponding COM establishment report. The CTS value-of-shipment data were summarized at a five-digit Standard Transportation Commodity Code (STCC) level and compared with the appropriate code(s) reported on the census record. (A "bridge" between the STCC and SIC classification systems was developed by the Census Bureau in order to link the reported data.) Those reports that do not fall within previously established tolerances were flagged for review and possible correction.

In 1972, commodity coding was clerically verified by comparison with the SIC code of the establishment. Establishments with a majority of their shipments coded differently from the industry code were referred for review. It is probable that the more rigorous commodity verification conducted for 1977 uncovered a significant number of coding errors by respondents that would not have been corrected by using the 1972 procedure. The 1977 procedure also served as a check on respondents' sampling of shipping documents.

As a final step, the individual CTS establishment reports were adjusted at each five-digit Transportation Commodity Classification (TCC) level by the ratio:

$$A_{ec} = V_{ec}/v_{ec} \quad (1)$$

where

A_{ec} = adustment for five-digit TCC commodity c in establishment e,

V_{ec} = COM value of five-digit TCC commodity c in establishment e, and

v_{ec} = CTS estimated value of five-digit TCC commodity c in establishment e.

The estimation procedure will also take advantage of the known 1977 COM universe totals. Estimates of value and weight shipped will be made for four geographic levels of origin (United States, census division, states, and production areas). Adjustment factors were computed that will reconcile the CTS aggregated data to the COM levels. They will take the following form:

$$A_{ig} = V_{ig}/v_{ig} \quad (2)$$

where

A_{ig} = adjustment to the plant expansion factor for all establishments in industry (SIC) i and area g,

V_{ig} = COM aggregated value of industry (SIC) i in area g, and

v_{ig} = CTS aggregated value of industry (SIC) i in area g.

There are certain assumptions that must hold if these adjustments are to provide unbiased results. The most important are as follows:

1. A high correlation between weight and value exists among shipments of a particular five-digit STCC commodity.

2. Value is reported on the same basis in both the CTS and the COM. Indications are that respondents did report value consistently between the two surveys.

Other Modifications to Survey Design

The sample design for the 1977 CTS is a two-stage stratified sampling plan. In the first stage, a sample of approximately 19 500 individual manufacturing establishments was selected from the Standard Statistical Establishment List (SSEL), a computerized list of all employers and their establishments. Large establishments, as determined by the total employment size code, had a greater probability of selection than smaller establishments. Establishments in certain designated geographical areas of high production activity (production areas) had a greater probability of selection than other establishments, to enable the publication of data for these areas. In the second stage of sampling, each establishment selected shipping documents from its own records. Large establishments were requested to select a greater number of documents than smaller ones. For large establishments (≥500 employees), an average of 120 documents was requested; for medium establishments (100-499 employees), an average of 60 documents; and for small establishments, an average of 30 documents. Correspondingly, there were three separate forms for large, medium, and small establishments. Each establishment was asked to report the number of documents in its files and then to sample at a specified interval starting from the first document.

In the 1972 survey, an average of 150 shipping documents were selected per establishment, regardless of establishment size. Studies showed that, for small and medium establishments, the variance between establishments was greater than the variance within establishments. These component contributions to the variance were made more nearly equal by decreasing the number of shipments per establishment and selecting 50 percent more establishments. The figures of 30, 60, and 120 shipments for small, medium, and large establishments, respectively, were derived from an approximate cost-variance formula. This distribution has the added benefit of decreasing the reporting burden on the individual establishment.

Evaluation of Respondents' Self-Sampling

As outlined above, the sample design for the CTS is a two-stage stratified sampling plan, with the second stage requiring each establishment to sample its shipping documents according to a sampling plan provided by the Census Bureau. How well these instructions are followed has never been certain, and, consequently, a study of this process was conducted as part of the evaluation process for the 1977 economic censuses (2). The conclusions noted in that study are as follows:

> The evaluation has shown conclusively that a large percent[age] of manufacturers did not take a sample of their shipping documents properly. Mistakes were made in undercoverage, overcoverage, haphazard sampling, and [use of] an incorrect sampling rate. Some of these mistakes were not preventable due to the system of recordkeeping used by the establishments. However, a large percentage of the mistakes could have been avoided if the respondents were willing to follow the instructions.
>
> The mistakes made in sampling apparently have a serious effect on the estimates in the 1977 CTS. This was shown in the third phase of the analysis and was supported by the results of the sign test in the second phase of the analysis. Because the analysis was [done] prior to the CTS edit and adjustments to the Census of Manufactures, final conclusions will not be made until after the future analysis is completed. However, it is being suggested that the procedure for collecting the data in the CTS be carefully studied in order to modify the existing procedure or create a new procedure.

It should be noted that these conclusions are preliminary and were made prior to the match to the COM records. A future analysis will be performed to determine whether the corrections resulting from the match have an effect in reducing the bias.

Modifications to the Report Form

In addition to the items required for the expanded scope of the survey, other data items have been expanded somewhat, with the additional items available, for the most part, from the shipping document.

In 1972, when the shipping document contained more than one commodity, the respondent was instructed to code the entire shipment to the commodity that contributed the greatest proportion to the total weight. The 1977 survey instructed the respondent to list and code each commodity separately (up to a maximum of seven).

The name and street address of the consignee, along with city and state of destination, were requested. This information will provide a potential for studies involving input-output analysis at both national and regional levels, as well as serve as a unique marketing tool for identifying industries that purchase particular commodities. By matching the consignee's name and address to the SSEL, it will be possible to derive the consignee's characteristics (e.g., industry, employment, and receipt size) and to summarize this information in a disclosure-free way. This is an experimental project not yet funded. However, this approach has been strongly endorsed by the Gross National Product Data Improvement Project Committee (3), Bureau of Economic Analysis, U.S. Department of Transportation, and U.S. Department of Energy.

An additional breakdown was added to distinguish motor carriers regulated by the Interstate Commerce Commission (ICC) from those exempt from ICC regulation.

Publication Plans

The publication plans for 1982 call for an expansion in the amount of data available to the public. The major changes include

1. An increase in the number of production areas from 27 to 49;
2. An increase in data classified by industry (shipper groups) for analysts who wish to study the shipping characteristics of establishments and industries and who wish to link the CTS flow data to regional and national production consumption data;
3. An increase in the number of more analytical summary tables and more cross-classification within tables;
4. The use of actual numbers in all tables rather than percentages; and
5. The expansion of commodity group tables to show data at the two-digit major group level.

Benefits and Trade-Offs of Modifying 1977 CTS

As a result of the foregoing modifications, the CTS has been greatly improved through (a) a stronger conceptual framework that allows cross checking to other data sets; (b) greatly improved editing and estimation procedures through benchmarking to the COM; (c) more usable tabulations, including expanded geographic and industry detail; and (d) fewer suppressions for reasons of confidentiality and reliability of data.

What has been traded away in this process? The answer cannot be complete until users have a chance to scrutinize the publications and public use tape. These are currently scheduled for release in 1980. A partial list would include

1. Loss of comparability with prior CTS data,
2. Fewer individual observations (In 1972, the 13 000 responding establishments provided 1.6 million shipments, an average of 123 shipments/plant. In 1977, the 16 000 responding establishments provided slightly more than 1 million shipments, an average of 64 shipments/plant. The effects of this reduction in the number of observations will not be completely known until final data are tabulated; however, preliminary figures indicate that the sampling variability is comparable to the 1972 experience), and
3. Less timely release of data (In 1972, the individual publications were issued from October 1974 to August 1975. The current time schedule calls for publication to begin about July 1980, or about nine months later than the comparable 1972 data, and end in October 1980).

Things to Do in 1982

The 1977 CTS will make a significant contribution toward solving the jigsaw puzzle that is commodity flow freight data. There is much left to do. Unfortunately, the proposed budget for the 1982 Census of Transportation is unchanged from the 1977 level of activity. Nevertheless, several improvements are possible.

The 1977 CTS was, in many ways, a pioneering effort. In particular, the development of appropriate matching, editing, and estimation procedures required much more effort than initially expected. With the experience gained in 1977, the operational schedule for 1982 can be improved. The number of discrete points of origin may be extended. This is potentially possible in two major ways. In 1977, the number of establishments reporting was extended and the number of shipments from each reporting plant was reduced. The results of this redesign will be evaluated with a view toward further extending the principle in 1982. This may permit an increase in the number of origin points, thereby resulting in a wider geographic distribution and description of the transportation network. Small shippers whose products are primarily for local consumption will also be identified. If these can be successfully identified, they will be asked to report on a summary, short form that would simplify the reporting burden and processing. With the resulting savings, more resources can be directed to further increasing the points of origin.

TRUCK INVENTORY AND USE SURVEY

The TIUS provides data on the physical and operational characteristics of the nation's truck population. It is based on a probability sample of private and commercial trucks registered (or licensed) in the state during 1977.

The 1977 survey was conducted in a similar fashion to the 1972 survey. However, there were important changes in methodology and content.

Methodological Changes

Stratification of the sample was based on small versus large trucks (body type) rather than light versus heavy trucks (registered weight) as in previous surveys. The small-truck stratum included pickups, panels, vans, multisteps, and walk-ins with a gross vehicle weight of 14 000 pounds or less. All other vehicles were selected as large. This revision in stratification was an effort to reduce the amount of crossover from one stratum to the other. Although the number of crossovers in 1977 compared with 1972 has not been quantified yet, preliminary indications are that this has been somewhat reduced in 1977.

As in past surveys, the registration files maintained by R.L. Polk and Company served as the sampling frame from which the sample was drawn except for a few states that were sampled directly by the Census Bureau. (In 1977, only Oklahoma was sampled directly by the Census Bureau.) To develop the publication totals for the 1972 and earlier surveys, the proportions derived from the sample reports were benchmarked to Federal Highway Administration (FHWA) totals of private and commercial truck registrations. These FHWA estimates are based on calendar-year summary reports from the individual states that reflect differences in truck definitions used by the states for vehicle registration. This can be expressed as

$$T_c = (t_c \cdot T_{FHWA})/t_t \quad (3)$$

where

T_c = estimated number of trucks with characteristic c,
t_c = number of trucks with characteristic c in sample,
t_t = total number of in-scope trucks in sample, and
T_{FHWA} = total number of private and commercial trucks reported by FHWA.

For 1977, this procedure was revised so that the individual state universe estimates represent the base from which the sample was drawn, adjusted to a common date (July 1, 1977). These estimates serve as the benchmark to which the survey results were adjusted to produce the more detailed estimates shown in the publications. They were developed through a review of the characteristics of each vehicle registered in each state.

The estimates were developed independently for small and large trucks and can be expressed in the following manner:

$$T_c = [(ts_c \cdot TS)/ts_t] + [(tl_c \cdot TL)/tl_t] \quad (4)$$

where

T_c = estimated number of trucks with characteristic c,
ts_c = number of small trucks in sample with characteristic c,
ts_t = total number of small trucks in sample,
TS = Polk count (July 1, 1977) of small trucks adjusted by census-determined out-of-scope trucks,
tl_c = number of large trucks in sample with characteristic c,
tl_t = total number of large trucks in sample, and
TL = Polk count (July 1, 1977) of large trucks adjusted by census-determined out-of-scope trucks.

The item-by-item computer edit performed more consistency checks between data entries to identify and correct major errors and/or contradictions in reporting. At this time, missing data were imputed for annual and lifetime miles, type of engine, and number of trucks and truck-tractors operated from the base of operation.

The procedure for imputing annual and/or lifetime miles was revised from the technique used in 1972. Rather than using a standard estimating technique of imputing missing data with cell averages, a mean ratio for each model year was computed and an algorithm was developed that used a combination of reported variables (e.g., annual miles, lifetime miles, and age of vehicles) and the computed mean ratio to determine values for missing data.

Content of the Questionnaire

The questions asked of respondents were substantially increased to provide greater detail on the physical and operational characteristics of the sampled trucks. The full questionnaire is reproduced in the TIUS published reports.

Improvements for 1982

We believe a substantial improvement has been made in the usefulness of the TIUS by providing greatly expanded detail and making the user's job easier by showing actual numbers in the tables rather than percentage figures. However, there is clearly one major improvement due for 1982--more timely release of data. Due to a combination of circumstances, some beyond our control and some not, the 1977 reports were about 18 months behind the 1972 release dates. Our goal for 1982 is to return to and perhaps beat the 1972 release dates.

NONREGULATED MOTOR CARRIERS AND PUBLIC WAREHOUSING

This survey is the only phase of the Census of Transportation that deals with privately owned business enterprises that were engaged primarily in providing for-hire transportation to the public. It has had a rather checkered career in that it began in 1963 as part of the Census of Transportation, was transferred to the Census of Selected Services in 1967 and 1972, and returned to the fold in 1977.

The survey presents data only on that portion of the motor carrier industry not subject to economic regulation by the ICC and the public warehousing industry. As a result, it has a rather limited audience, primarily intended to fill a void of information needed for the national accounts.

It is being conducted under substantially the same methodology as the previous surveys except that certain questions on expenses and fringe benefits were added. Also, the categories of revenue freight equipment were made comparable to the definitions used by ICC.

There are currently no plans to expand this survey in 1982. However, the Advisory Committee on Gross National Product Data Improvement recommended that there should be a "census of transportation industries with a full complement of questions on inputs and outputs by relevant subindustries." The major obstacle in implementing this recommendation is the restriction in Title 13, U.S. Code, under which the Census Bureau operates, that prohibits the Census Bureau from collecting data for means of transportation for which statistics are required by law to be filed with a designated body (3).

CONCLUSION

The Census of Transportation was originally designed to fill in major data gaps without unnecessary duplication. Therefore, the general structure and objectives of the transportation program differ from those of the orthodox economic censuses, largely because of the availability of various types of transportation data from other government and private sources (4).

While the individual surveys have been expanded and improved on over the years, the goal of the Census of Transportation has not changed since it was conceived nearly 20 years ago. It would appear that the original objective has been substantially achieved. Policymakers, planners, and other users have much more information than existed 20 years ago and are able to make decisions more effectively.

However, there is still much left to do. In this day when questions about energy and transportation are critical, perhaps a new goal should be espoused--the goal of providing an integrated, comprehensive body of statistics that serve rather than confuse the user community. The current state of transportation statistics is characterized by a whole series of conflicting pieces of information, a veritable jigsaw puzzle. New approaches must be conceived if we are to be able to put the pieces together and develop a body of statistics that accurately describe the real world.

REFERENCES

1. Travel During 1977, Census of Transportation. Bureau of the Census, U.S. Department of Commerce, Final Rept., TC77N, 1979.
2. A. R. Tupek and A. M. Perez. Can U.S. Businesses Take a Sample of Their Own Records? Presented at the American Statistical Association Meeting, Washington, DC, Aug. 13-16, 1979.
3. Gross National Product Data Improvement Project Report. Office of Federal Statistical Policy and Standards, U.S. Department of Commerce, Oct. 1977.
4. D. E. Church. Plans for the 1963 Census of Transportation. Transportation Division, Bureau of the Census, U.S. Department of Commerce, Dec. 10, 1962.

Evaluation of the FHWA Vehicle Classification and Automobile-Occupancy Sampling Manual

R. CRAIG HUPP AND CARMINE PALOMBO

The Federal Highway Administration recently sponsored the development of a how-to manual for conducting automobile-occupancy and/or vehicle classification studies. A draft version of the manual, Guide for Estimating Urban Vehicle Classification and Occupancy, was field tested by the Southeast Michigan Council of Governments in cooperation with the Michigan Department of Transportation. This paper describes how the guide was used to set up a regional vehicle classification and automobile-occupancy study. Study procedures and results are discussed. Overall, the guide was found to be an excellent how-to manual for vehicle classification and automobile-occupancy studies. Perhaps the largest contribution of the guide is the short-count sampling approach it recommends. This approach results in a significant survey cost saving with no loss of accuracy.

The U.S. Department of Transportation continuously reviews and evaluates the kinds of data required to support national and local decisions on transportation. In support of this process the Transportation Research Board established the Advisory Committee on Urban Transportation Data-Reporting Needs and Requirements in 1975. The committee identified vehicle type and automobile occupancy as two important traffic variables that describe the highway system and recommended that the Federal Highway Administration (FHWA) develop and test sampling methods to obtain these data (1). The FHWA implemented this recommendation by contracting with Peat, Marwick, Mitchell, and Company to develop a sampling manual entitled Guide for Estimating Urban Vehicle Classification and Occupancy (2). In addition, the FHWA worked with several metropolitan planning organizations (MPOs) to have them test the procedures described in the guide.

The Southeast Michigan Council of Governments (SEMCOG), the MPO for the southeast Michigan region, in conjunction with the Michigan Department of Transportation (MDOT), was one of two MPOs that tested the guide. This paper describes how the guide was used in setting up a vehicle-classification and automobile-occupancy survey and what the results were of the survey that SEMCOG conducted. This paper places emphasis on the practical aspects of developing the survey-sampling plan and analyzing the results. The statistical approach on which the guide is based and that is the key to this economical approach to data collection is not discussed in detail here. The statistical approach incorporated in the guide is summarized in Field Data and Sampling Procedures for Measuring Regional Vehicle Classification and Occupancy (3).

The guide presents procedures to collect automobile-occupancy and vehicle-classification data and to estimate vehicle miles of travel and person miles of travel. The travel estimates are developed by combining vehicle-classification and automobile-occupancy data with traffic-count data from a regional traffic-counting program. These data collected on an areawide basis can be used to support five technical analyses:

1. Evaluate the effectiveness of short-range transportation programs,
2. Assess transportation-related air quality,
3. Assess the energy efficiency of travel,
4. Validate urban transportation planning models, and
5. Monitor general trends in traffic and travel characteristics.

The procedures in the guide can also be used on a corridor or project basis as part of a before-and-after study or as part of project planning.

DEVELOPING THE SAMPLING PROGRAM

The guide stresses the need to clearly state the survey objectives before a sampling plan is developed. The principal objective of SEMCOG's study was to obtain up-to-date regional data for automobile occupancy and vehicular classification. At the time the study was undertaken the most recent data for these items were from a 1965 origin-destination survey. In addition, these data will be used to track trends in ridesharing. Finally, these data will be used to improve SEMCOG's vehicular-emissions and fuel-consumption models and to check the results of SEMCOG's model recalibration process. Obviously, another important objective of the study was to evaluate the guide for its clarity, content, and usefulness in performing vehicle-classification and automobile-occupancy surveys. These considerations are discussed more fully below.

The sampling program consists of the specification of the desired tolerance and confidence level of the estimate, the location-stratification scheme, the selection of sampling locations, and the selection of categories under which these data will be collected. The key to developing a good sampling plan is to use prior knowledge of the characteristics of the data that are to be collected. Prior data can be used to design a program that samples the minimum number of sites required to estimate desired population parameters with a given accuracy and precision.

Definitions

The following definitions were used in the vehicle-occupancy portion of the study:

1. One-person passenger vehicles--includes noncommercial pickup trucks, vans, and private automobiles and taxis;
2. Two-person passenger vehicles;
3. Three-person passenger vehicles; and
4. Passenger vehicles carrying four or more passengers.

The following definitions were used in the vehicle-classification portion of the study:

1. Panel and pickup trucks that have less than 1-ton cargo-carrying capacity, defined as pickup trucks and vans that were obviously commercial and that either contained side markings or construction equipment;
2. Other single-unit trucks, defined as having the cab and cargo area within the same frame served by two or three axles;
3. All truck combinations not otherwise classified (this classification consisted mainly of tractor-trailer trucks);
4. Motorcycles and/or motor scooters, with or without sidecars;
5. Buses, not counted;
6. Motor homes, classified as other single-unit trucks; and
7. Small pickup trucks, classified in the same way as larger pickup trucks (presence of commercial equipment or markings meant inclusion in the panel-and-pickup under 1-ton category; otherwise, they were considered as passenger vehicles subject to the occupancy criteria).

Stratification of Sampling Sites

When a sample population group is known or suspected to include subgroups that have more homogeneous characteristics than the population as a whole, a stratified sampling plan designed to sample each subgroup can provide both a more efficient estimate (in terms of sample size) of average characteristics of the population as a whole and more information about each subgroup within the population. Sample size is a direct function of the variation of the population measured in terms of the variance or standard deviation of the population parameter being sampled. Sampling by subgroup reduces the variation and hence the sample size needed to obtain a given level of accuracy.

For example, existing MDOT and SEMCOG data indicated that the proportion of trucks varies significantly by functional class. It was suspected that it also varied by area type (urban or rural), although sufficient data were not available to verify this assumption. Because of this, SEMCOG decided to use a stratified sampling approach for vehicle-classification sampling. This dictated that a stratified approach also be used for automobile occupancy because automobile-occupancy data would be collected at the same time.

In addition, other planning efforts at SEMCOG had indicated a need to know whether the proportion of trucks was the same on high-volume roads as on low-volume roads. It was decided to stratify sites based on traffic volumes as well as functional class and area type in order to determine whether this were the case. Arterial links were stratified by volume to determine whether the proportion of trucks varied by traffic volume. An average daily volume of 35 000 vehicles/day was selected as a stratification cutoff point because a review of traffic volumes in the region indicated that most of the roads of regional significance were carrying volumes in excess of 30 000 to 35 000 vehicles/day. This stratification applied only to urban arterials because all freeways in the region carry more than 35 000 vehicles/day and rural arterials do not.

It should be noted that, of the three stratifications (functional class, area type, and traffic volume), the first was selected because of known variations in the population while the other two were selected to test for suspected variations. Note also that, even though these stratifications were developed to improve sampling for truck travel, they are also relevant to automobile occupancy. Intuitively, urban freeways would be expected to have a greater proportion of work-trip travel and, hence, lower occupancies during peak periods than other combinations of functional class and area type. Conversely, rural freeways would be expected to have higher occupancies because of recreational travel and because longer trip lengths promote ridesharing. Thus the stratification for trucks was expected to provide useful information about automobile occupancy within a given tolerance and confidence level.

Calculation of Sample Size

Sample size is a function of the tolerance and level of confidence desired in the sample estimate and the variation in the parameter being estimated. For this study two parameters were estimated, average automobile occupancy and proportion of trucks. Whichever parameter requires the larger sample size for a given tolerance and level of confidence will determine the total sample required. In this study, the unit being sampled was the one-way link day, where the link is a unit distance on the highway network.

The guide provides the following formula to calculate the minimum sample size required to estimate population parameters within a given tolerance and confidence level.

$$N = (Z^2 \cdot S^2)/D^2 \tag{1}$$

where

N = minimum sample size in one-way link days;
Z = normal variate for the (1-α) level of confidence, two-tailed test where α is the level of confidence;
S = composite standard deviation of the sample; and
D = tolerance or acceptable difference between the estimated value of the population parameter and the true value.

The composite standard deviation of the sample S is a function of the manner in which the population parameter varies as a function of several other variables, including the following:

1. The variation across link days within a season,
2. The variation from season to season,
3. The variation across time periods during a day as a result of short counts,
4. The variation between lanes where the short-count procedure includes sampling between lanes,
5. The variation introduced by human error in the survey process, and
6. Other sources.

The composite standard deviation is equal to the square root of the sum of the squares of the standard deviations attributed to each of these sources of variation.

Depending on the sampling program, some of these sources of variance will not apply. For example, the seasonal variation factor was not relevant because the SEMCOG survey was conducted during only one season and an annually adjusted average automobile occupancy or proportion of trucks was not desired. Variation introduced by human error was ignored, although it does exist. Variation across traffic lanes was also omitted on SEMCOG's sample-size estimate because it was not significant. Hence, the only two sources of variation that SEMCOG considered in calculating the minimum sample size were variation across link days and variation across time periods during the day.

The guide provides default values for these standard deviations to be used where there are insufficient local data to permit their calculation. SEMCOG used the default data values. As discussed in the section on results, the composite standard deviation (SD) observed in the data SEMCOG collected was typically smaller than the composite standard deviation obtained from the guide's default values. The default values used are noted below:

Measure	Variation Source	Default Values Used	Composite SD
Proportion of trucks	Location and day	0.040	0.041
	Within day	0.009	
Average automobile occupancy	Location and day	0.063	0.065
	Within day	0.017	

The desired tolerance and confidence levels represent a trade-off between the desired accuracy of the estimate and data-collection costs. SEMCOG specified a 95 percent confidence level (α). This means a sample size was chosen that would result in the estimated average automobile occupancy or proportion of trucks falling within the desired levels of tolerance about the true value of the estimated parameter 19 times in 20.

SEMCOG established maximum tolerance levels D for both average automobile occupancy and proportion of trucks at ±0.03 for freeways locations and ±0.04 for arterial locations. That is, the average automobile-occupancy estimate will be within ±0.03 persons/automobile for freeways or ±0.04 persons/automobile for arterials of the true value that would be found if all vehicles passing all locations were counted. For vehicle classification, total proportion of trucks would be within ±0.03 trucks/total vehicles for freeways and ±0.04 trucks/total vehicles for arterials of the true value. In relative terms, assuming the true average automobile occupancy is 1.30 persons/automobile, specified maximum tolerance of the estimate will be approximately ±3 percent of the average occupancy. For the proportion of trucks the relative tolerance would be greater. Assuming the proportion of trucks on arterials and freeways is 0.05 and 0.15, respectively (i.e., 5 percent and 15 percent of total traffic), the relative tolerance would be ±80 percent and ±20 percent for arterials and freeways, respectively.

However, because these parameters (i.e., average automobile occupancy and proportion of trucks) do not have the same level of variance, different sample sizes would be required to achieve the same tolerance levels. Because the two parameters were being sampled at the same time, the sample that had the most variance determined total sample size. Sample size for both automobile occupancy and proportion of trucks was calculated for each survey cell. In each case the automobile-occupancy parameter required the larger sample size. In terms of the total sample required, the minimum sample required to estimate automobile occupancy was 69 one-way link days and, for proportion of trucks, 28 one-way link days.

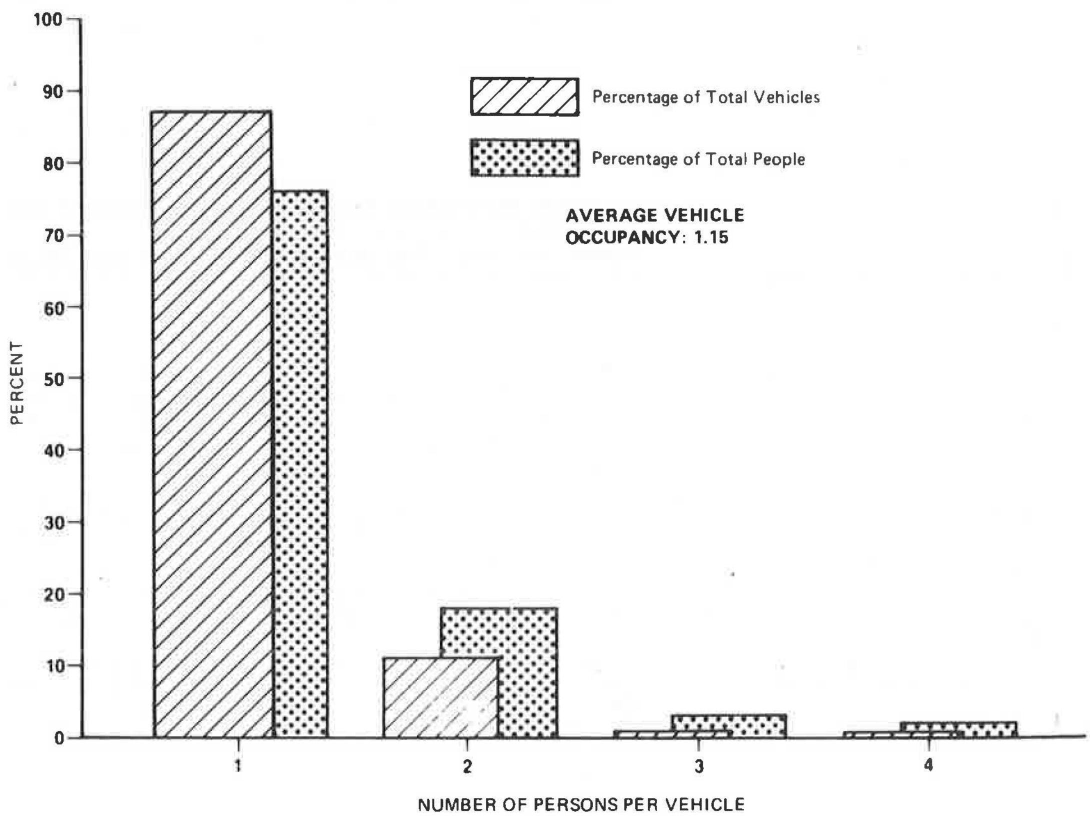
Because more than the minimum sample size was collected for proportion of trucks, the final

Table 1. Study results.

Time Period	Stratification Cell	Automobile Occupancy: Average Occupancy	Composite SD	Tolerance[a]	Vehicle Classification: Proportion of Trucks	Composite SD	Tolerance[a]
7:00-9:00 a.m.	Urban freeway	1.13	0.02	0.01	0.16	0.03	0.02
	Rural freeway	1.24	0.04	0.02	0.26	0.04	0.02
	Urban arterial[b]	1.21	0.08	0.04	0.05	0.02	0.01
	Urban arterial[c]	1.14	0.03	0.01	0.08	0.03	0.01
	Rural arterial[b]	1.25	0.12	0.07	0.13	0.04	0.02
	Total	1.17		0.015	0.11		0.0086
11:00 a.m.-1:00 p.m.	Urban freeway	1.28	0.04	0.02	0.24	0.04	0.02
	Rural freeway	1.45	0.05	0.03	0.32	0.05	0.03
	Urban arterial[b]	1.41	0.07	0.03	0.09	0.03	0.01
	Urban arterial[c]	1.34	0.05	0.02	0.09	0.03	0.01
	Rural arterial[b]	1.43	0.07	0.04	0.15	0.05	0.04
	Total	1.36		0.016	0.14		0.0096
2:00-6:00 p.m.	Urban freeway	1.24	0.02	0.01	0.18	0.02	0.01
	Rural freeway	1.35	0.04	0.03	0.25	0.03	0.02
	Urban arterial[b]	1.39	0.04	0.02	0.06	0.02	0.01
	Urban arterial[c]	1.31	0.04	0.02	0.06	0.02	0.01
	Rural arterial[b]	1.43	0.07	0.04	0.10	0.03	0.02
	Total	1.33		0.012	0.10		0.0061

[a]At 95 percent confidence level. [b]Less than 35 000 vehicles/day. [c]More than 35 000 vehicles/day.

Figure 1. Distribution of automobile occupancy, urban freeway, morning peak hours.



tolerances for this parameter were smaller than those for average automobile occupancy. Final tolerances for proportion of trucks were ±0.019 for freeway locations and ±0.025 for arterials. The final sampling matrix is shown below:

Area Type	Functional Class: Freeway	Arterial (vehicles/day): ≤35 000	≥35 000
Urban	18	11	11
Rural	18	11	X

Sampling Methods

The unit sampled was the one-way link day. Hence, a two-way street represents two possible sampling sites. The guide (2) suggests using short-count data-collection techniques to reduce data-collection costs. This technique involves the periodic sampling of traffic characteristics. The most practical way to accomplish this is to sample lanes one at a time on high-volume multilane facilities. For example, on a six-lane urban freeway, a single surveyor would cover one travel direction and would collect data from the three lanes in that direction for 15 min each. By comparison, an alternative approach frequently used by MDOT involves six surveyors (one for each lane), a supervisor, and, possibly, a backup person to continuously count all lanes in both directions. The cost advantages of the one-way, short-count approach are obvious.

Selecting the Sample Links

The next step was to select a random sample of highway links for each cell of the sampling matrix. The highway network used for sampling was SEMCOG's computer-coded regional highway network, which included all highway links classed as minor arterials or above. Approximately 5000 miles of streets and highways are included in this network. SEMCOG's computerized file for this network represents each road in the system as a series of links of varying length that have as their termini intersections with other roads in the regional system. For each link, a variety of jurisdictional, physical, functional, and operational data are maintained, including functional class, area type, and volume.

To ensure the validity of the random sampling process, it was necessary to consider that the highway links were made up of a series of sublinks. Each sublink was 0.16 km (0.1 mile) long, and had an equal probability of being selected. In practical terms, this meant weighting the link selection process so that the probability of selecting a link was in direct proportion to its length. That is, a link 0.8 km (0.5 mile) long should have a probability of being selected that is 5 times greater than a link only 0.16 km (0.1 mile) long. To accomplish the weighted link-selection process a computer program was written that analyzed each link in the computerized highway network, assigned it to the proper survey cell, and then randomly sampled the sublinks in the survey cell.

PERFORMING THE SURVEY

The survey was performed as a joint effort by SEMCOG and MDOT. MDOT provided survey crews to cover all freeway locations. SEMCOG retained a consultant to survey all arterial locations. Some 69 locations were surveyed.

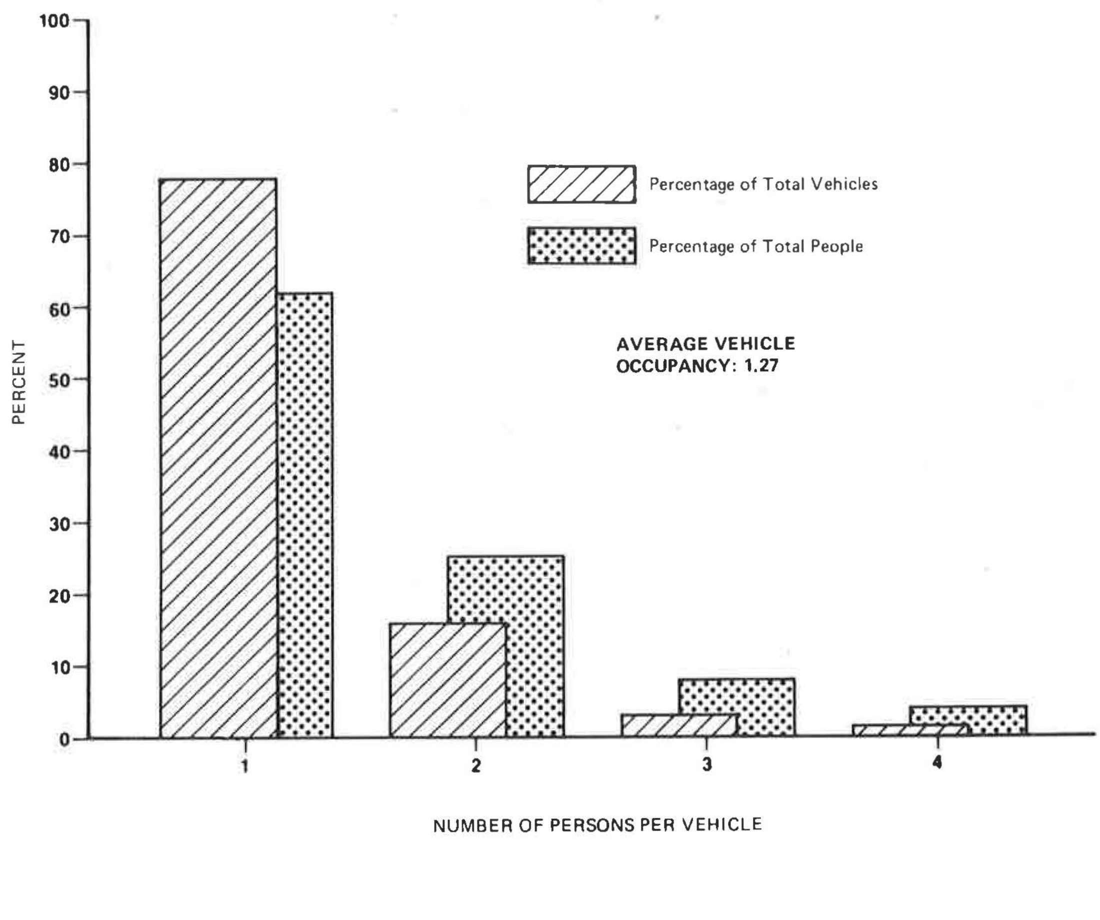
Data were collected for the periods 7:00-9:00 a.m., 11:00 a.m.-1:00 p.m., and 2:00-6:00 p.m. Each lane was counted for 15 min, followed by a 5-min interval to record the counts and reset the counters to zero.

The table below summarizes the survey costs, including SEMCOG staff time to administer the consultant contract. By using part-time labor, the consultant was able to survey at one-half of MDOT's cost per site because MDOT used full-time department survey crews.

Item	Cost ($)
Consultant costs (data collection on arterials, 33 link days)	2 750
Estimated MDOT costs (data collection on freeways, 36 link days)	6 000
SEMCOG staff costs (including staff time for administration of contract, coding, and computer programming, estimated at approximately 600 h)	7 500
Total	16 250

Had MDOT's usual practices for recording vehicle classification and occupancy data been used, it is estimated that the costs of the study would have been substantially higher. In conducting

Figure 2. Distribution of automobile occupancy, rural arterials, morning peak hours.



conventional studies, MDOT uses three-person crews working 10-h days at an estimated $325 per crew per day, including vehicles, meals, and overnight accommodations. Each person is responsible for monitoring one lane; hence, on a six-lane freeway, two crews would be needed. By calculating costs in this manner, estimated MDOT costs for surveying the freeway locations would have been $11 700. Actual costs to MDOT were approximately $6000. Consultant costs, in comparison, would have been approximately $3000 for the freeway locations. This represents an added saving available due to the use of part-time survey crews.

RESULTS

The average passenger automobile-occupancy and proportion-of-truck data including the tolerance and observed composite standard deviation are given in Table 1. In general, the tolerances and composite standard deviation are smaller than default data used by SEMCOG in applying the guide's formula.

In general, the average passenger-vehicle occupancy figures are inversely proportional to the traffic volumes on the particular roadway. That is, the higher occupancy figures were recorded on the lower-volume facilities and the lower occupancy figures were recorded on the higher-volume facilities. Higher occupancy figures were recorded in the more rural counties than in the other counties.

For all facility types monitored, the recorded average passenger-automobile occupancy was lowest during the morning peak period (7:00-9:00 a.m.) and highest during the midday period (11:00 a.m.-1:00 p.m.). For each of the three time periods monitored, the recorded average automobile occupancy was lowest on urban freeways and highest on rural arterials. In the more urban counties, the recorded average occupancy on arterials was higher than on freeways for all three time periods monitored. In the more rural counties, the recorded average automobile occupancy was higher on freeways than on arterials.

For all facility types monitored, the recorded proportion of trucks was highest during the midday period (11:00 a.m.-1:00 p.m.). For each of the three time periods monitored, the recorded proportion of trucks was highest on rural freeways and lowest on urban arterials.

The number of vehicles is shown in Figures 1 and 2 as a percentage of the total number of passenger vehicles monitored during the morning peak hours for both urban freeways and rural arterials with 24-h two-way traffic volumes of less than 35 000 vehicles per day. The percentage of total people carried during these periods by passenger vehicles is also shown. These facility types were chosen because the study indicated that the lowest average vehicle occupancy was recorded on urban freeways and the highest average vehicle occupancy was recorded on rural arterials.

The overwhelming majority of vehicles monitored during the three time periods on these facility types were one-passenger vehicles. However, it should be noted that the average occupancy figure or the percentage of one-passenger vehicles is a deceptive measure of the extent of ridesharing. As shown in Figure 1, although approximately only 12 percent of all vehicles were multioccupant vehicles, they carried over 25 percent of the people. In other words, even at an average automobile occupancy of 1.15 persons/vehicle, 25 percent of all people were sharing rides. In the case of rural arterials, Figure 2 shows that approximately 38 percent of all people were sharing rides during the morning peak periods. In effect, there was 50 percent more ridesharing on rural arterials during morning peak periods than on urban freeways at the same time. In the off-peak period on rural arterials (not shown), more than 55 percent of all people were sharing rides.

EVALUATION OF THE GUIDE

SEMCOG found the guide to be a straightforward how-to manual for collecting automobile-occupancy and vehicle-classification data. The overall sampling approach is clearly explained. The default values for standard deviation appear to be excellent based on SEMCOG's results. Any planner or traffic engineer should be able to use the guide with little or no reference to other materials.

Perhaps the largest contribution of the guide is the short-count sampling approach. This affords a dramatic saving in the survey cost at little loss in accuracy. It reduces survey costs to such an extent that SEMCOG expects to continue its automobile-occupancy and vehicle-classification survey on an annual basis. The tolerances obtained for the average automobile-occupancy estimates at the 95 percent confidence level are so small for most sampling cells that changes over time in average automobile occupancies on the order of 3-5 percent can be detected.

REFERENCES

1. Advisory Committee on Urban Transportation Data-Reporting Needs and Requirements. Proposed Urban Transportation Data-Reporting Requirements for States and Metropolitan Planning Organizations: A Report to the U.S. Department of Transportation. National Academy of Sciences, Washington, DC, 1976.
2. Peat, Marwick, Mitchell, and Co. Guide for Estimating Urban Vehicle Classification and Occupancy. Federal Highway Administration, Sept. 1978.
3. R. A. Ferlis. Field Data-Collection and Sampling Procedures for Measuring Regional Vehicle Classification and Occupancy. TRB, Transportation Research Record 701, 1979, pp. 1-6.

Atlanta Vehicle Occupancy Monitoring

R. FRED FISHER, G. JACK WILLIAMS, AND J. PHILLIP BOYD

This report describes the implementation of a statistically defined survey technique for collecting vehicle classification and occupancy data in the Atlanta region. The paper describes the results of a stratified, areawide survey for collecting passenger occupancy rates. The potential movement of people provided by the capacity of our roadway system is virtually an untapped resource, according to the data collected for this study. Sponsored by the U.S. Department of Transportation, efforts are being made to improve the usefulness of passenger vehicles through current programs that include vanpooling, ride-sharing programs, and park-and-ride lots. The success of these ventures, which are likely to become more significant in the future, can be measured by a dependable vehicle-occupancy monitoring program. This research has proved that the Guide for Estimating Urban Vehicle Classification and Occupancy provides a statistically acceptable method to measure vehicle occupancy rates. The minimum sample requirement for determining occupancy rates by area and facility type is desirable for an annual program of this nature.

In recent years the need for monitoring the movement of people by vehicle has touched all aspects of society. Not only is the information valuable to transportation engineers and planners, but it has also become vital to those evaluating energy consumption, environmental quality, and trends in the nation's economy.

In 1979, the Georgia Department of Transportation (GDOT) and the Atlanta Regional Commission (ARC) participated in testing the Guide for Estimating Urban Vehicle Classification and Occupancy (1), which was prepared for the Federal Highway Administration. The overall objective in testing the procedures set forth in the guide was to establish a methodology for estimating vehicle classification and occupancy in the Atlanta region on an annual basis, during peak periods, and for daylight hours. The Atlanta region is the more than 2000-mile^2 area composed of Clayton, Cobb, DeKalb, Douglas, Fulton, Gwinnett, and Rockdale Counties in north central Georgia.

A joint effort was made by local planning agencies in order to define purposes for monitoring vehicle occupancy and classification so that potential use of this data could be maximized. These purposes are as follows:

1. Provide monitoring data for the Atlanta Regional Transportation Plan;
2. Provide a basis for assessing the impact of implementing the Metropolitan Atlanta Rapid Transit Authority's rapid rail lines;
3. Evaluate the effectiveness of such programs as carpooling, park-and-ride facilities, and high-occupancy vehicle lanes;
4. Validate transportation planning models;
5. Assess energy efficiency of travel;
6. Assess air quality related to transportation;
7. Provide data for project planning and design; and
8. Provide data for person miles traveled (PMT) when estimating procedures for vehicle miles of travel (VMT) are implemented in the Atlanta region.

SURVEY DESIGN

The survey was structured to test procedures in the guide and provide results compatible to existing data-collection and modeling techniques. The primary criterion was to develop a survey that would gather statistically reliable data at the lowest possible cost. To meet these objectives, the effort was divided into two surveys.

Stratified Areawide Study

A stratified areawide survey was designed to provide areawide data with eight stratifications. There were two highway functional classifications (freeway and nonfreeway) and four geographic areas, as illustrated in Figure 1. The methodology in the guide (1) was used in determining sample size for the areawide survey. This was considered a critical element in testing the validity of procedures set forth in the guide. The sample size of link days was estimated separately for vehicle classification and automobile occupancy to determine which travel measure controlled the design. A link day was defined as the sampling unit--i.e., a particular highway segment on a particular day.

Vehicle Classification

The first approach considered for determining sample size was by regional vehicle classification. This measure is defined as the ratio of VMT of a particular vehicle type to the total regional VMT. Before a minimum sample size can be computed, the composite standard deviation of the proportion of vehicles must be estimated. After assuming the standard deviation of the proportion of truck traffic, S_t, was the same as that of passenger vehicles, the composite standard deviation was computed for each stratification as follows:

$$S_{th} = (S_{tlh}{}^2 + S_{tsh} + S_{tdh}{}^2)^{1/2} \quad (1)$$

where

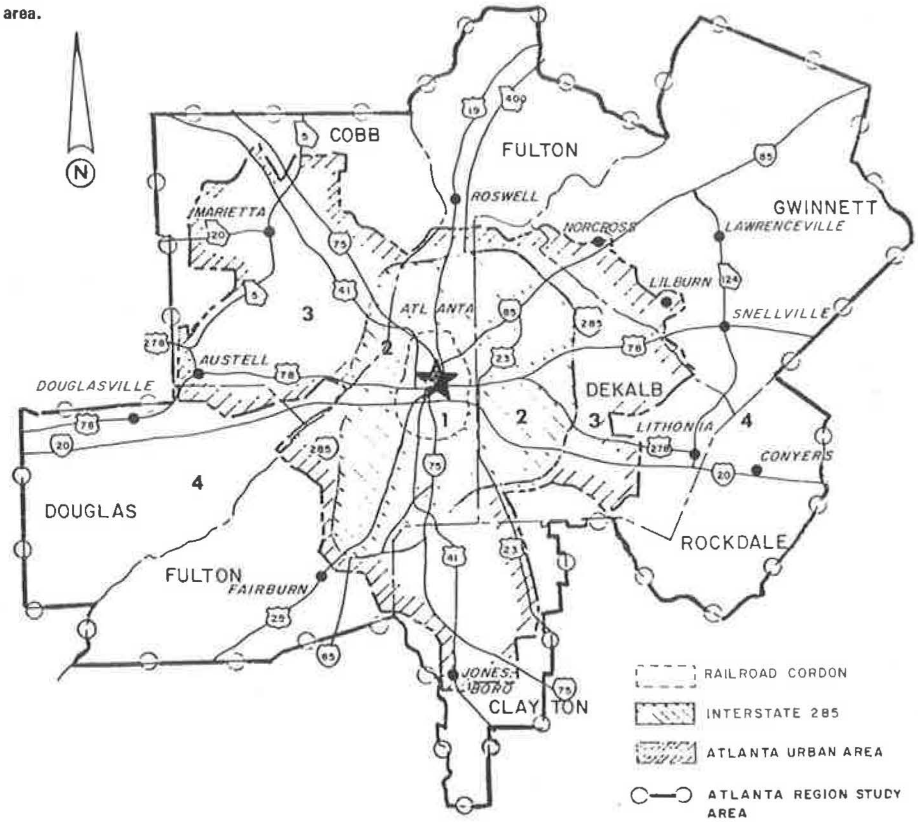
- S_{th} = composite standard deviation of the proportion of trucks in stratum h,
- S_{tlh} = standard deviation of the proportion of trucks across link days within a season in stratum h,
- S_{tsh} = standard deviation of the proportion of trucks across seasons in stratum h, and
- S_{tdh} = standard deviation of the proportion of trucks across time periods during the day (as a result of hourly counts) in stratum h.

The following assumptions were made in calculating the composite standard deviations for each stratum:

1. S_{tlh} = (maximum % trucks - minimum % trucks)/3.5 (2)
2. S_{tsh} = 0 (the survey would not be conducted across seasons), and
3. S_{tdh} = 0.009 (recommended for all strata, based on default value).

Factor	Computation
Composite standard deviation	$S_{oh} = (S_{olh}{}^2 + S_{osh}{}^2 + S_{odh}{}^2)^{1/2}$
	$= (0.63^2 + .000^2 + .017^2)^{1/2} = 0.065$
Sample size per stratum	$N_h = (Z^2 \cdot S_{oh}{}^2)/(D_{ot}{}^2)$
90 percent confidence level	$N_h = (1.645^2 \cdot 0.065^2)/(0.03^2) = 12.7 \approx 13$
80 percent confidence level	$N_h = (1.28^2 \cdot 0.065^2)/(0.03^2) = 7.7 \approx 8$
Areawide tolerance	$D_o = Z(\sum_h W_h{}^2 \cdot S_{oh}{}^2/N_h)^{1/2}$
90 percent confidence level	$D_o = 1.645(\sum_h W_h{}^2 \cdot 0.065^2/13)^{1/2} = 0.011\ 88$
80 percent confidence level	$D_o = 1.280(\sum_h W_h{}^2 \cdot 0.065^2/8)^{1/2} = 0.011\ 78$

Figure 1. Atlanta vehicle-occupancy study area.



Thus, the lower confidence level could be selected with little loss of precision and would require a sample of only eight link days per stratum as opposed to 13 link days per strataum at the higher confidence level.

In analyzing the two methods for computing sample size, it was apparent that the average passenger-vehicle-occupancy approach should be used for determining the sample size. Eight link days were selected from each stratum for a total of 64 samples required for the areawide survey.

A sample selection technique on random highway links and random days was used to allocate the 64 required samples. A computer listing and map of all the links for the current Atlanta Regional Traffic Assignment Network was used to define the survey population. The network was delineated by direction and by the eight stratifications.

A computer program was developed to generate a series of random numbers within the range of the link-number assignment network. The series was of sufficient quantity that only one draw would be necessary for the selection of the samples in the various stratifications.

The samples were selected by using listings of all the links and random numbers. Because the link-node listing was in numerical order areawide, each link number selected had to be verified as to its stratification. If the link did not meet the required stratification, the selection was rejected, and the next random number was selected. This process was repeated until all 64 required samples were selected for the areawide survey.

Once the standard deviations were computed, the total sample size that would allow for a reliable estimate of the proportion of trucks within desired levels of precision was computed as follows:

$$N = \left[Z^2\left(\sum_h W_h S_{th}\right)^2\right] / D_t^2 \qquad (3)$$

where

- N = number of link days required for monitoring vehicle classification and occupancy for the entire region,
- W_h = proportion of VMT occurring in stratum h to total regional VMT,
- Z = normal variate at specific confidence level,
- S_{th} = composite standard deviation for stratum h computed in Equation 1, and
- DT_t = desired tolerance (acceptable difference between estimated and true proportion of trucks).

The following assumptions were made in making this computation:

1. Z = 1.96 (value for a two-tailed statistical test at a 90 percent level of confidence) and
2. DT_t = ±0.02 (desired objective for the tolerance level).

The required minimum sample of link days within each stratum was then computed by the following equation:

$$N_h = N\left[(W_h S_{th}) / \left(\sum_h W_h S_{th}\right)\right] \qquad (4)$$

The results of these computations for computing composite standard deviation and sample sizes are shown in Table 1. As illustrated, these computations resulted in an areawide sample size of 22 link days. Once the sample sizes were determined by strata, the precision of the areawide estimate was computed as follows:

$$DT_t = Z\left[\left(\sum_h W_h^2\right)(S_{th}^2/N_h)^{1/2}\right] \qquad (5)$$

Table 1. Results of computation of sample size by vehicle classification when S_{tsh} = 0, S_{tdh} = 0.009, Z = 1.96, and DT_t = 0.02.

Facility	Factor	Inside Railroad Cordon	Railroad Cordon to I-285	I-285 to Urban Arterial Boundary	Urban Arterial Boundary to Region Boundary
Freeway	W_h (%VMT)	0.05	0.25	0.14	0.13
	S_{tlh}	0.014	0.042	0.057	0.071
	S_{tsh}	0.000	0.000	0.000	0.000
	S_{tdh}	0.009	0.009	0.009	0.009
	S_{th}	0.017	0.043	0.058	0.072
	N_h	1	5	4	4
Nonfreeway	W_h (%VMT)	0.03	0.19	0.10	0.11
	S_{tlh}	0.029	0.042	0.057	0.029
	S_{tsh}	0.000	0.000	0.000	0.000
	S_{tdh}	0.009	0.009	0.009	0.009
	S_{th}	0.030	0.043	0.058	0.030
	N_h	1	4	2	1

The areawide tolerance equaled the desired value of 0.02.

Average Passenger-Vehicle Occupancy

The second method used for determining sample size considered the measure of average passenger-vehicle occupancy. The composite standard deviation was computed as follows:

$$S_{oh} = (S_{olh}^2 + S_{osh}^2 + S_{odh}^2)^{1/2} \quad (6)$$

where

S_{oh} = standard deviation of average occupancy across link days within a season in stratum h,
S_{olh} = standard deviation of average occupancy across link days within a season for stratum h,
S_{osh} = standard deviation of average occupancy across seasons in stratum h, and
S_{odh} = standard deviation of average occupancy across hours during the day (as a result of hourly counting) in stratum h.

The following assumptions were made for this computation:

1. S_{ol} = 0.063 (recommended value based on previous studies),
2. S_{os} = 0 (the survey would not be conducted across seasons), and
3. S_{od} = 0.017 (recommended value based on previous studies).

Once the composite standard deviation was computed, the desired sample size of link days was then determined by using the following equation:

$$N_h = (Z^2 S_{oh}^2)/(DT_{oh})^2 \quad (7)$$

where

N_h = sample size in link days in stratum h,
Z = normal variate,
S_{oh} = composite standard deviation of average occupancy in stratum h, and
DT_{oh} = desired tolerance for stratum h (acceptable difference between the estimated average occupancy and the true value).

Once the sample sizes were determined by strata, the tolerance achieved at the regional level was computed as follows:

$$DT_o = Z\left[\left(\sum_h W_h^2\right)\left(S_{oh}^2/N_h\right)\right]^{1/2} \quad (8)$$

Tests were made of the sample-size computation by using confidence levels of 90 percent and of 80 percent, with a tolerance of ±0.03 for both tests. Results indicated, as shown below, a very small difference in the areawide tolerance (see Figure 2 for W_h) between the two confidence levels:

A random number, ranging in value equivalent to the days of the survey, was selected for each sample. This number was assigned to the sample to designate the date of observation.

Focused Multiple-Location Survey

A focused multiple-location survey was designed to fulfill needs for travel information for specific screenlines and corridors within the region. Some 22 locations were selected to monitor freeways and arterials crossing the railroad cordon, which included the central business district (CBD), and the Interstate perimeter highway (I-285). It was determined that the data from these locations coupled with the ongoing GDOT traffic-counting program would provide a good measure of person travel. In addition, five locations were selected to measure "before" occupancy rates on highways adjacent to proposed park-and-ride lots within the region.

For calculating sample-size requirements, the composite standard deviation was assumed equal to 0.018 for both vehicle classification and occupancy. To achieve a 90 percent confidence and ±0.03 tolerance, one day of data collection was required at each location. Thus, a higher accuracy for total traffic entering each cordon (railroad and perimeter) would be achieved.

Due to the characteristics of this survey, observations were required in both directions at each location (areawide survey observations were one direction only). The focused observations were randomly selected on days by using the same process for random days in the areawide survey.

DATA COLLECTION

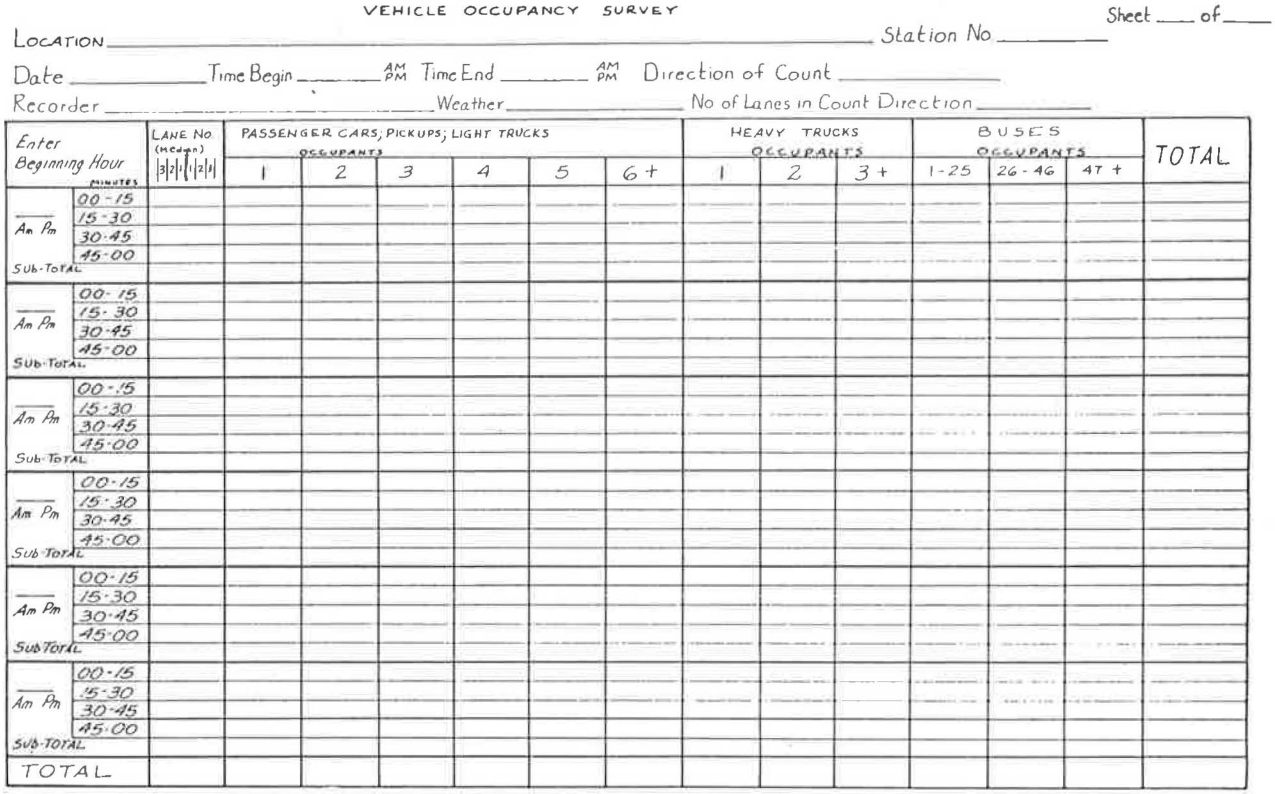
Once the monitoring stations were selected, a work plan was prepared for collecting the vehicle classification and occupancy data. A listing of the monitoring stations that described the assigned station number, station type, geographic area, facility type, number of lanes, and location description was prepared.

Scheduling

The technique used to monitor the traffic was to manually observe each lane for a 15-min duration and a total of 45 min/h. The remaining 15-min period was used to record the observed information and provide a break for the monitor. Observations were made by one person parked on the shoulder of the road in such a manner as to maximize observation of passing vehicles and minimize disruption to normal traffic flow due to the presence of the surveyor's vehicle.

A work plan for the monitoring program was then established, and the required work crew was scheduled. Field data collection began on March 31, 1979, and was completed on May 30, 1979. Observations were made from 7:00 a.m. to 7:00 p.m. The individuals selected to make the manual observations were assembled and given their work schedule for the entire study period with the appropriate classifying instructions and counting forms. By scheduling the observations for the duration of the study, the field personnel knew

Figure 2. Form used during Atlanta-region vehicle-occupancy survey.



VEHICLE OCCUPANCY SURVEY

Sheet ___ of ___

LOCATION ___ Station No ___

Date ___ Time Begin ___ AM PM Time End ___ AM PM Direction of Count ___

Recorder ___ Weather ___ No of Lanes in Count Direction ___

Enter Beginning Hour	Minutes	Lane No (Median) 3 2 1 | 1 2 3	Passenger Cars; Pickups; Light Trucks Occupants 1	2	3	4	5	6 +	Heavy Trucks Occupants 1	2	3 +	Buses Occupants 1-25	26-46	47 +	TOTAL
___ Am Pm	00-15														
	15-30														
	30-45														
	45-00														
Sub-Total															
___ Am Pm	00-15														
	15-30														
	30-45														
	45-00														
Sub-Total															
___ Am Pm	00-15														
	15-30														
	30-45														
	45-00														
Sub-Total															
___ Am Pm	00-15														
	15-30														
	30-45														
	45-00														
Sub-Total															
___ Am Pm	00-15														
	15-30														
	30-45														
	45-00														
Sub-Total															
___ Am Pm	00-15														
	15-30														
	30-45														
	45-00														
Sub-Total															
TOTAL															

their days off in advance. Their knowledge of the complete work schedule served as a performance incentive.

Field Data-Entry Form

The primary objectives for designing the data-entry form were to properly identify the monitoring stations and provide necessary inputs for computing accurate vehicle classification and occupancy rates. The necessary information included the station location description, station identification number, beginning and ending hour of observation, time interval of observation by lane, vehicle-count data, and direction of observation.

The vehicle-count data were recorded in 15-min intervals by lane number in the following classification groups:

1. Passenger cars, pickups, and light trucks;
2. Heavy trucks; and
3. Buses.

Each of these three categories was further classified by the number of occupants per vehicle.

The data form was also designed to include information that could influence the resulting classification and occupancy rates. Such items included the weather conditions, lane configuration, date of observation, and field recorder's name. The field data-entry form is illustrated in Figure 2.

Field Data-Collection Supervision

The field supervisor visited each site during the period of observation to ensure adherence to the schedule and proper conformance to the various procedures previously outlined. The survey forms from the previous day's observation were collected and verified. The field supervisor was also responsible for tabulating the field data into hourly summaries as the data sheets were collected. This enabled the supervisor to assure the accuracy and legibility of the data as they were recorded. The count summaries were later used in compiling classification and occupancy rates. If counts appeared to be incorrectly recorded, based on the supervisor's observation of the traffic flow, questionable entries were resolved. The field supervisor maintained a file of the completed forms. At the conclusion of the field data-collection phase, the data were transmitted to the clerical unit for processing.

DATA ANALYSIS

Data Preparation

The data from each station were input into a computer file. The input items included facility type, geographic area, station identification number, direction, number of lanes, date, beginning hour, and hourly classification subtotals.

The field data were expanded to a common 12-h base according to the number of lanes. The factors were used to expand each vehicle classification for every hour of count data. The following ratios were used to compute the expansion factors: 1-lane factor = total possible count period/actual total period counts = 48/36 = 1.33; 2-lane factor = 48/18 = 2.67; 3-lane factor = 48/12 = 4.00; and 4-lane factor = 48/9 = 5.33.

Station Tabulations

A computer program was written to summarize the

Figure 3. Sample printout of Atlanta automobile-occupancy summary data.

ATLANTA AUTO OCCUPANCY SUMMARY

STATION TYPE 2
AREA 3
FACILITY TYPE 2
STATION NO 780 Memorial Drive between Collinwood Drive and Rockbridge Road
DIRECTION 4

TIME	PASSENGER CAR	HEAVY TRUCK	BUS	ALL VEHICLES
700- 800	1.17	1.29	29.63	1.40
800- 900	1.33	1.38	26.78	1.57
900-1000	1.37	1.30	22.32	1.56
1000-1100	1.37	1.25	22.17	1.50
1100-1200	1.42	1.33	18.30	1.58
1200-1300	1.44	1.36	18.02	1.57
1300-1400	1.44	1.32	21.42	1.53
1400-1500	1.37	1.24	36.67	1.55
1500-1600	1.33	1.14	28.04	1.49
1600-1700	1.39	1.17	30.03	1.60
1700-1800	1.35	1.10	21.42	1.42
1800-1900	1.41	1.21	30.03	1.56
MEAN	1.37	1.21	24.90	1.53
STD DEV	.123	.147	5.465	.140
PEAK HOURS (700-900)	1.25	1.35	28.08	1.49

Table 2. Resulting standard deviation for each stratification after computation of average passenger-vehicle occupancy rates.

	Freeway			Nonfreeway		
Functional Classification Area	Samples	Occupancy Rate	Standard Deviation	Samples	Occupancy Rate	Standard Deviation
Inside railroad cordon	8	1.35	0.033	8	1.40	0.108
Railroad cordon to outside I-285	9	1.35	0.046	6	1.41	0.074
Outside I-285 to urban arterial boundary	7	1.36	0.058	9	1.35	0.105
Urban arterial boundary to region boundary	9	1.57	0.122	8	1.44	0.029

vehicle occupancy data by direction, station number, facility type, geographic area, and station type. These tabulations provided hourly occupancy rates in four categories: passenger cars, heavy trucks, buses, and all vehicles. The mean and standard deviation of the occupancy rates were computed for each category of each station. An occupancy rate for the period between 7:00 and 9:00 a.m. was also computed by category to represent the morning peak-hour rates. Figure 3 shows a typical station summary.

Sample Precision

Due to publication constraints, only the data analysis of the stratified, areawide passenger-vehicle occupancy is presented. Before the precision of the sample data could be assessed, the composite standard deviation had to be computed.

The composite standard deviation of the average passenger-vehicle occupancy rates were computed for each stratification by the following formula:

$$S_{oh} = N_h \left[\sum_i (P_{ih} - OCC_h VPV_{ih})^2 \right] / \left(\sum_i VPV_{ih} \right)^2 \qquad (9)$$

where

S_{oh} = composite standard deviation of average occupancy rates in stratum h,
N_h = total number of sample locations in stratum h,
P_{ih} = factored number of persons counted at sample location i in stratum h,
OCC_h = estimated average passenger-vehicle occupancy in stratum h, and
VPV_{ih} = factored number of passenger vehicles counted at sample location i in stratum h.

Table 2 illustrates the resulting standard deviation for each stratification. The objective standard deviation (0.065) was exceeded in four of eight strata. The areawide occupancy rate equaled 1.39 persons/vehicle. The actual precision of the estimates was then determined by using Equation 8 and substituting the S_{oh} values calculated above and the W_h values from Table 1. The resulting areawide precision level was 0.013, which was well within the desired areawide precision level of 0.020.

SUMMARY

The major objective of this study was to test the procedures set forth in the guide (1) and to investigate the possible implementation of an annual monitoring program in the Atlanta region. The statistical tests applied to the results indicated a satisfactory degree of confidence. Only a minimal effort ($23 000) was required to collect this data when compared with similar activities in other urban areas of a comparable size. The 1979 survey results will be used to improve precision of future surveys while (it is hoped) reducing costs. Realignment and aggregation of geographic stratifications will allow a reduction in sample-monitoring sites without compromising precision.

Due to the success of this research, an on-going monitoring program is being established to generate vehicle occupancy and classification rates for the metropolitan Atlanta area. Data will be collected on a quarterly basis at selected sites in order to measure seasonal variations in occupancy rates.

These data will be used extensively by local planning units to evaluate the effectiveness of transportation management programs, validate transportation planning models, and monitor general trends in travel characteristics. Local environmental units will use the data to assess transportation-related air quality measures. The measure of vehicle classification and occupancy rates is an important statistic in today's environment, and all indications are that it will become more important in the future.

REFERENCE

1. Peat, Marwick, Mitchell, and Co. Guide for Estimating Urban Vehicle Classification and Occupancy. Federal Highway Administration, Sept. 1978.