TRANSPORTATION RESEARCH RECORD 639

Transportation System Evaluation Techniques

TRANSPORTATION RESEARCH BOARD

COMMISSION ON SOCIOTECHNICAL SYSTEMS
NATIONAL RESEARCH COUNCIL

NATIONAL ACADEMY OF SCIENCES
WASHINGTON, D.C. 1977

Transportation Research Record 639
Price $3.00
Edited for TRB by Mary McLaughlin

subject area
84 urban transportation systems

Transportation Research Board publications are available by ordering directly from the board. They may also be obtained on a regular basis through organizational or individual supporting membership in the board; members or library subscribers are eligible for substantial discounts. For further information, write to the Transportation Research Board, National Academy of Sciences, 2101 Constitution Avenue, N.W., Washington, D.C. 20418.

Notice
The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competence and with regard for appropriate balance.

This report has been reviewed by a group other than the authors according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

The views expressed in this report are those of the authors and do not necessarily reflect the view of the committee, the Transportation Research Board, the National Academy of Sciences, or the sponsors of the project.

Library of Congress Cataloging in Publication Data
National Research Council. Transportation Research Board.
Transportation system evaluation techniques.

(Transportation research record; 639)
Reports for the 56th annual meeting of the Transportation Research Board.
Includes bibliographical references.
1. Urban transportation–United States–Congresses. 2. Transportation planning–United States–Congresses. I. Title. II. Series.
TE7.H5 no. 639 [HE308] 380.5'08s [388.4'0973] 78-9158
ISBN 0-309-02668-7

Sponsorship of the Papers in This Transportation Research Record

GROUP 1–TRANSPORTATION SYSTEMS PLANNING AND ADMINISTRATION
E. Wilson Campbell, New York State Department of Transportation, chairman

Transportation Forecasting Section
John R. Hamburg, John R. Hamburg and Associates, chairman

Committee on Transportation Systems Design
Joseph L. Schofer, Northwestern University, chairman
Surendra K. Agarwal, Edward A. Beimborn, Melvyn Cheslow, James E. Clark III, Robert T. Dunphy, Raymond H. Ellis, Gary R. Erenrich, Gorman Gilbert, Hollis R. Goff, Harvey Haack, Lonnie E. Haefner, Thomas N. Harvey, Kevin E. Heanue, Harold Kassoff, Joe Lee, Stephen C. Lockwood, Marvin L. Manheim, Charles William Ockert, Richard H. Pratt, James E. Sale, Ali F. Sevin, William L. Smith, Darwin G. Stuart, Martin Wachs, Richard D. Worrall

Committee on Transportation Information Systems and Data Requirements
George V. Wickstrom, Metropolitan Washington Council of Governments, chairman
Mark M. Akins, Robert E. Barraclough, Daniel E. Benson, H. W. Bruck, Jeffrey M. Bruggeman, Kenneth J. Dueker, David J. Dunlap, Robert T. Dunphy, Christopher R. Fleet, Wade G. Fox, Alexander French, William L. Grecco, Yehuda Gur, Dayton P. Jorgenson, Thomas F. Keane, Joseph M. Manning, Alfred J. Marston, Robert E. Paaswell, Peter S. Parsonson, Alan E. Pisarski, Sydney R. Robertson, Arthur B. Sosslau, Ronald W. Tweedie

Transportation Systems and Planning Innovations Section
Charles V. Wootan, Texas A&M University, chairman

Committee on New Transportation Systems and Technology
H. William Merritt, Transportation Consultant, Arlington, Virginia, chairman
C. Michael Walton, University of Texas at Austin, secretary
J. Edward Anderson, William H. Avery, Robert U. Ayres, George J. Bacalis, Robert A. Burco, Eugene T. Canty, Michael G. Ferreri, William Hamilton, M. D. Harmelink, Clark Henderson, Thomas E. Lisco, Roger L. Merrill, Theodore F. Morf, Robert A. Olmsted, Daniel Roos, Richard Shackson, Colin G. Simpson, Albert J. Sobey, Charles E. Taylor, J. William Vigrass, M. Lucius Walker, Jr.

James A. Scott, Transportation Research Board staff

Sponsorship is indicated by a footnote at the end of each report. The organizational units and officers and members are as of December 31, 1976.

Contents

Design-Synthesis Approach to Transit Planning

Mark H. Scheibe,* Puget Sound Council of Governments, Seattle
Gordon W. Schultz, R. H. Pratt Associates, Inc., Kensington, Maryland

Transportation system design should be oriented toward meeting specific local and regional objectives. In most current planning, objectives are used only to evaluate selected alternatives. This paper presents a design-synthesis approach to transit planning, which allows objectives to be input directly to the process and generates a transit system incorporating characteristics that are selected to optimize the attainment of specific service and cost objectives. The design-synthesis technique specifies transit service in the abstract so that characteristics of service such as frequency, headway, travel speed, and fare can be examined individually without being constrained to a specific system alternative. The paper reports successful applications of the approach in (a) identifying short-range transit improvements for San Diego and Denver and (b) designing long-range transit alternatives for Denver.

In design-synthesis planning, the community objectives that transportation service is provided to meet explicitly propel the transportation system design process. Design-synthesis planning consists of a three-step process (Figure 1).

First, a set of regional and local transportation objectives are determined that typically fall into two categories: service and cost. Service objectives define a minimum or (less likely) maximum desired level of transportation service for a particular geographic or socioeconomic area of the region. Criteria used to measure whether the objective is achieved could include, for example, the percentage of regional employment locations that can be reached in a certain travel time and the level of transit service provided to the area, defined by areawide average excess time (in this case the average time spent walking to a transit stop plus the average time spent waiting for a transit vehicle).

Cost objectives, on the other hand, typically take the form of constraints defining the maximum resources available for providing transportation service. The criteria used to measure the attainment of the objective generally are defined either by the total operating or amortized annual cost of providing transportation or by a productivity criterion. This criterion could be defined, for example, by transit trips per bus kilometer supplied to a given area or by a rate of transit subsidy available to offset operating losses.

In addition to defining transportation objectives and criteria, this first step in the design-synthesis process should also determine the relative weighting of the objectives. Because a transportation service standard cannot usually be met without exceeding a transportation cost standard, it is necessary to define an equilibrium position between supplying service and expending resources by relatively weighting the different objectives.

The second step in design-synthesis planning is to determine the system characteristics necessary to meet the objectives. By use of a mathematical programming approach, a wide range of system characteristics can be examined simultaneously and the best combination selected. This approach, which can be costly depending on the number of feasible options, is best oriented toward minimizing cost in the planning process (for instance, to determine which links should be added to a highway network system). A more heuristic approach would involve examining the characteristics of individual components of a transportation system separately while maintaining a constant level for the other system components, e.g., holding system characteristics such as speed and fare constant while examining the impact of different levels of service coverage and frequency or holding service and speed constant while considering different fare levels. Such an approach is more cost-effective and allows the planner an active role in balancing different system components.

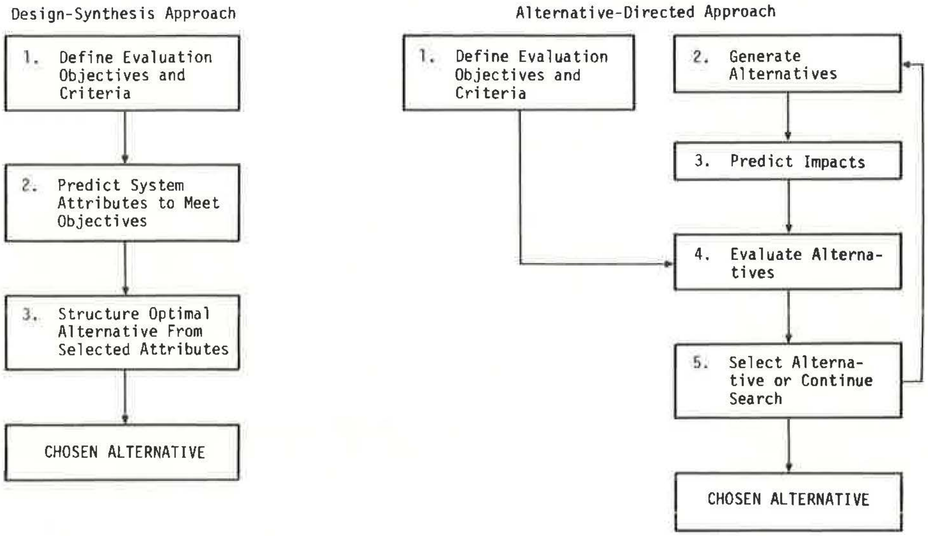
The final step in design-synthesis planning is to design a system incorporating the optimal characteristics obtained in the previous step. If the planning project involves deciding which links should be added to the highway network, this step is straightforward. For the heuristic approach, however, where the intermediate-phase output might be definitions of the level of transit service to be supplied to different subareas of a region, this step might involve locating transit routes to provide service efficiently while maintaining route and system continuity.

A design-synthesis planning approach has several advantages over the most commonly used approach, alternative-directed planning. Alternative-directed planning consists of a five-step iterative process (Figure 1), which begins, like the design-synthesis process, with a definition of transportation objectives. The objectives, however, are ignored in the next two steps, in which the planner generates a number of alternative system configurations and forecasts their impacts. These impacts are then compared to the criteria used to measure attainment of the objectives, and the best alternative is finally chosen. If no alternative is acceptable, additional alternatives are devised and the process is repeated. In spite of its disadvantages, alternative-directed planning is most often used in current transportation planning because design-synthesis models are generally more difficult to construct, particularly those models using mathematical programming.

In contrast to alternative-directed planning, design-synthesis planning requires only a single pass through the process. A single, optimal alternative is generated that incorporates the system attributes devised to meet the criteria defined by the transportation objectives. This single process allows more efficient use of time and cost resources. Another advantage of a design-synthesis approach is that objectives are made active rather than passive by being input explicitly into the procedure. This is particularly important as planning becomes more oriented toward serving a wide variety of regional, social, and environmental objectives. Finally, by designing a system to meet certain demands rather than designing a system before the demands are known, it is possible to provide the appropriate combination of modes, submodes, and services for the specific situation.

This paper describes an efficient and rational approach to design-synthesis planning that has been successfully applied to transit planning in the San Diego and Denver regions. It is presently applicable only to transit planning but is being expanded to consider other high-occupancy modes. The technique predicts and evaluates the impacts of different levels of transit service characteristics without having to consider the characteristics

Figure 1. Design-synthesis and alternative-directed approaches to transit planning.



in the context of a specific transit system alternative. The approach produces information that can determine feasible transit areas and corridors and associate them with the level of transit service needed to produce a required level of patronage or to meet regional transit criteria.

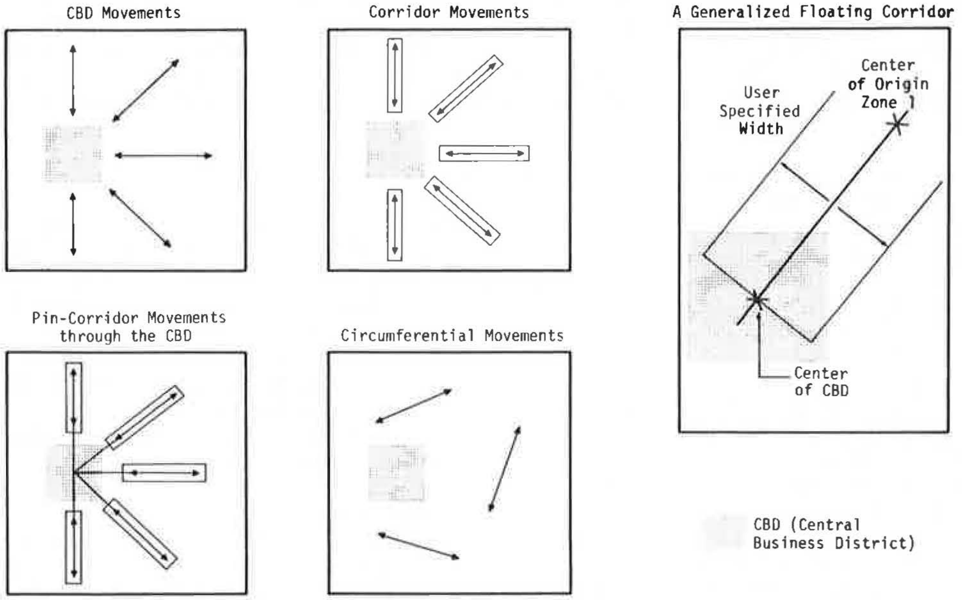
The design process is based on three premises: (a) The entire travel market should be considered in planning a transit system; (b) the planning process should be as free as possible, at least in the initial phases, from prejudicial routing assumptions; and (c) the specification of transit service should be based on policy service levels rather than on specific transit route spacing and headways. The second and third premises are made operational by using the concept of a ubiquitous bus system capable of directly serving each potential transit trip with a single ride from trip origin to trip destination along the shortest available highway route. Obviously such transit service cannot normally be provided, but the assumption of ubiquity aids in systematic analysis by defining the system abstractly instead of specifying alternative routes. Only the concept of ubiquity contradicts the characteristics of regular transit service. All other standard transit trip characteristics are considered in the analysis, including walk to and from the bus, wait for the bus, transit speed, and transit fare. Transit service time defines the walk to and from the bus and the wait for the bus. Transit fare and speed are included in the analysis as exogenous variables. The first premise of design-synthesis planning, that of considering the entire travel market, is handled by using a travel-demand chain of models that forecast trip generation, trip distribution, and mode choice based on socioeconomic and transportation system data.

In order to compare the amount of transit ridership at a given service level with the associated cost, the methodology must estimate cost. Although many elements affect operating cost, one of the more important is the number of bus kilometers operated. The minimum number of bus kilometers needed to provide a specific level of service is considered a surrogate for transit service cost.

The design process is used in the following fashion. The impacts of several levels of transit system characteristics such as service time or transit speed are forecast and defined in terms of the study criteria. For instance, if the objective is a minimum level of transit accessibility at the lowest cost in bus kilometers, the evaluation data produced for each travel-analysis zone would be the transit accessibility at each of several service times and the cost in bus kilometers to provide each level of service. For each individual zone then, the appropriate service time would be that which exceeds the minimum accessibility standard at the least cost in bus kilometers. The output on a regional scale would be a designation for each zone of the level of transit service that should be provided to meet regional transportation objectives. This information could then be used as a foundation for the route-specific design of a transit system.

This approach has been used successfully in several cases. An earlier version of the procedure was used to identify feasible service areas and establish a basic system operating pattern for north suburban Chicago (1, 2). The version of the process discussed here was first used to identify transit service and routing improvements to be included in the 5-year transit development plan for the San Diego region (3, 4). The improvements were primarily designed to meet regional objectives of transit accessibility. The process was next used as part of the Denver Long-Range Transit Analysis sponsored by the Denver Regional Transportation District (5, 6, 7, 8). In that analysis three "year 2000" bus networks were designed. One network, designed to provide the existing level of transit service, served as a base case in the analysis of alternative transit modes. A second network, designed for a substantially better level of transit service but still relying on on-street bus operations, served as a background bus system for the fixed-guideway alternatives being considered and also as a base for the third network. The third network used buses only but included fixed-busway facilities and was designed to provide service comparable to that of the other fixed-guideway systems. The design process also assisted in route selection for the other fixed-guideway systems. The most recent use of the design-synthesis procedure was in the generation of a transit development program for Denver under the auspices of the Regional Transportation District (9). The method was used to define transit service improvements to be included in the program and to identify appropriate areas for internal circulator transit service.

Figure 2. Four types of transit movements and the floating-corridor concept.



DETAILS OF THE METHODOLOGY

A more detailed examination of the design-synthesis methodology is needed to understand how it can be applied. The crucial concept of ubiquity has already been discussed. To complete the specification of transit service, travel orientation of potential transit trips, the definition of walk and wait times with respect to service times, and the specification of other service characteristics, i.e., transit speed and transit fare, must also be discussed.

Travel Orientation

To efficiently design a transit system, the analyst should have information for at least the following four types of transit movements (Figure 2):

1. CBD movements—movements to or from the CBD,
2. Corridor movements—movements occurring within a CBD-focused corridor,
3. Pin-corridor movements—movements that have the potential for transferring in the CBD, and
4. Circumferential movements—all other movements.

The need to investigate the potential of travel to the CBD is obvious, and most conventional transit planning is focused on this trip movement. Corridor trips and trips that transfer in the CBD also play an important role in transit planning because they (a) are normally large in number, (b) can make use of CBD-focused transit routes, and (c) tend to reinforce the CBD routes. The first three trip patterns can all be considered radial; the only difference among them is destination location. The fourth category, circumferential movements, is normally given little attention in transit planning because such movements are difficult to serve. A preliminary investigation, however, will show a high proportion of trips in this category; the ability of transit to serve these movements should thus not be neglected.

Part of the methodology of system design is to define these four movements and use them as categories in reporting results. CBD movements are easily defined: The analyst need only define the CBD (normally as a range of traffic analysis zones) and all trips to and from this area as CBD movements. The definition of corridor movements is more difficult. Instead of a concept of specific corridors based on the geography of the region and the existing transportation system, the concept of floating corridors is used. In this concept, each origin zone has its own unique corridor defined as a rectangular area extending from the origin zone to the CBD (Figure 2). The mathematical definition of this corridor consists of

1. The center line of the rectangle as defined by coordinates of the origin zone centroid and the centroid of the CBD,
2. The slope of the sides of the corridor as defined by the slope of the center line, and
3. The width of the corridor, which is user-defined but has typically been assumed as 1.6 km (1 mile).

The floating-corridor concept defines a corridor movement as any interchange movement from an origin zone that ends within the corridor. For computer analysis, the beginning and end points of an interchange are described by the X and Y coordinates of the zone centroids (i.e., the geographic center of the zones).

Any movement that is not a CBD or corridor movement can be either a pin-corridor or circumferential movement. The distinction between these two movements must be based on transit travel times. A trip is a pin-corridor movement if it can be made more quickly by going to the CBD and then transferring to another radial line; otherwise, the trip is a circumferential movement. Obviously the circumferential movement is always quicker than the pin-corridor movement if both have the same level of service. It is therefore necessary to define two types of service times: one for radial movements (CBD, corridor, and pin-corridor) and one for circumferential movements.

Walk and Wait Times

In the estimation of transit trips by mode-choice models, the time spent walking to and waiting for a transit vehicle (excess time) is an extremely significant variable. It is important, therefore, to specify transit excess time realistically. The assumptions made in the methodology are that (a) walking time is related to the average transit line spacing and (b) waiting time is related to transit line headways. These assumptions are logical and realistic, although they may not apply in certain unique situations.

Values for the walk to the transit line can be derived from transit line spacing. The walk distance is one-quarter the transit spacing. For example, in an area with 1.6-km (1-mile) spacing, the average resident would walk 0.4 km (0.25 mile) to reach a bus line. Normally the average wait time is a simple function of headway. In this paper it is assumed that the walk time for a given travel-analysis zone can be calculated by using the average distance and a walking speed of 4.8 km/h (3 mph). It is also assumed that the average wait time is equal to half the headway. Other assumptions could also be used with this design approach, including ones that assign nonlinear relationships between average walk time and walk distance and between average wait time and headway.

As noted earlier, this design approach uses the concept of transit service time, which is equated to the sum of the average walk time and the average wait time for a given zone. This service time, rather than the individual values of walk and wait time, is the transit characteristic to be optimized. This concept requires the assumption that walk and wait times are equally weighted in travel decisions and also that the service level can be broken down into its spacing and headway components in the design phase.

Once average walk and wait times, stratified by radial and circumferential movements, have been calculated for all travel-analysis zones, the total excess time for any zone-to-zone transit trip can easily be obtained given the trip-orientation assumptions discussed earlier. For CBD and corridor movements, the excess time for a trip from zone X to zone Y equals the sum of the radial walk times for zones X and Y plus the larger of the two radial wait times. For pin-corridor movements, which assume a transfer in the CBD, the total excess time equals the sum of the radial walk times and radial wait times for both zones. The excess time for circumferential trips equals the sum of the two circumferential walk times plus the larger of the two circumferential wait times.

Other Service Characteristics

Two other important characteristics of transit service are running time and fare. Transit running time is calculated by dividing the trip distance by the speed. As mentioned previously, trip distances by transit are considered equal to the minimum-path highway distances for the same movements. Transit speed can be input as a systemwide value as in the Denver studies, stratified by trip type as in the San Diego study, or further stratified by trip location within the region, depending on operating conditions in the study region. Because this procedure deals with transit trips on a zone-to-zone interchange basis, the technique for dealing with transit fares is the same in this as in any other methodology. In all applications of the approach, transit fares have been set and held constant by public policy; thus, impacts of fare changes have not been examined.

Travel Demand

The design-synthesis approach is not dependent on specific travel-demand models. In both San Diego and Denver local travel-demand models were used. Of course, demand estimation is a necessity. At a minimum, a matrix of person trips, a modal-split model, and the socioeconomic and highway system data required for the modal-split model are needed in addition to the transit service specification data produced by the design methodology.

Evaluation Data

The evaluation data produced by the methodology are a function of the criteria used in the particular study. In the Denver and San Diego studies, three main criteria were considered: a transit cost criterion, expressed in terms of bus kilometers; a transit productivity criterion, expressed as the number of transit trips produced per bus kilometer of service provided; and a transit accessibility criterion, expressed as the percentage of regional employment attractions that could be reached in a given travel time by transit. These criteria seem to be applicable in most transit design circumstances although their relative weightings may vary.

Given zone-to-zone transit times and zonal employment attractions, transit accessibility can easily be calculated. Similarly, the transit trips portion of the criterion for trips per bus kilometer is directly output from the travel-demand calculation. But the calculation of bus kilometers, which is crucial because bus kilometers are used as a surrogate for transit cost in the design approach, is not so obvious.

For a given area, the number of bus kilometers is calculated by multiplying the number of bus lines per kilometer—the average transit spacing—by the number of bus lines per hour—the transit line frequency—times a unit distance of 1.6 km (1 mile). Multiplying this by the area of a particular travel-analysis zone gives the number of bus kilometers for the peak hour (assuming peak-hour spacing and headway) in one direction provided to the zone. For a total cost, this calculation must be applied to both radial and circumferential movements. Peak-hour data are used for compatibility with most other modeling processes, but off-peak data could also be used. For comparison purposes, peak-hour, one-direction bus kilometers can be used, but a simple factor from local transit system operating data can be applied to yield daily or annual bus kilometers.

Using the assumptions described earlier that relate walk time and spacing and wait time and headway, bus kilometers can be directly calculated from zonal walk and wait times and zonal service times. One-way hourly bus kilometers per square kilometer equal the reciprocal of the spacing times frequency times a unit distance. Average walk distance is one-quarter the transit spacing, and walk time equals distance divided by average walk speed; the spacing is therefore 4.0 times the product of average walk time and walk speed. Assuming a walk speed of 4.83 km/h (3 mph) [0.8 km/min (0.05 miles/min)],

$$S = 4.0 \times 0.08 \times WK = 0.32 \times WK \qquad (1)$$

where

S = transit spacing (km) and
WK = average walk time (min).

Assuming that the average wait time equals half the headway and frequency (in buses per hour) equals 60 divided by the headway (in minutes),

$$F = 60.0/2.0 \times WT = 30.0/WT \qquad (2)$$

where

F = transit frequency in buses per hour and
WT = average wait time in minutes.

Thus,

$$BM = 1/S \times 1.0 \times F = (3.11/WK) \times (30.0/WT) = 93.17/(WK \times WT) \quad (3)$$

where BM = one-way hourly bus kilometers per square kilometer.

Therefore, bus kilometers provided to a zone can be directly calculated from the walk and wait times for the zone. Further, Equation 3 shows that bus kilometers, and thus transit cost, are inversely proportional to the product of the walk and wait times. The maximum product of walk and wait times would thus be the least cost combination for a given transit service time. This maximum product occurs when walk equals wait time; therefore, the minimum cost for a given service time to a zone is achieved when the average walk time equals the average wait time equals half the service time. (In this case the previously mentioned relationships of walk time to spacing and wait time to headway are assumed. Other least cost solutions are found for other assumptions.)

In the design-synthesis approach it is assumed that, given several alternatives that provide identical service times, the least cost alternative will always be selected. Thus, each unique service time for a particular zone has associated with it a unique number of bus kilometers. Walk and wait times need not be considered individually.

Other evaluation data can also be produced by this methodology. For the Denver and San Diego studies, the other data included transit trip density measures, i.e., trips produced per household and per square kilometer and trips attracted per square kilometer and per employee.

Transit Design Process

We have discussed how the design methodology can be used to specify transit service in the abstract and forecast impacts as measured by various criteria. The precise way in which localized system characteristics that provide information for route-specific design are then determined varies as a function of the criteria used. Following is an example of a design exercise intended to meet the service objective of maximizing patronage within a regional budget constraint. This objective can be measured by using a transit productivity criterion of trips produced per bus kilometer.

Transit operators have observed that the highest transit productivity often occurs with low patronage and few bus kilometers (such a phenomenon was forecast for Denver and San Diego). Logically, as service increases, bus kilometers rise more rapidly than patronage. From Equation 3 it can be seen that they increase inversely to the product of the walk and wait times. As almost any modal-split model will show, a 1 percent decrease in excess time will result in a less than 1 percent increase in patronage, except perhaps for conditions of high initial excess times. Thus, for several alternatives that at least meet a selected productivity standard, the option that produces the largest patronage would have the lowest productivity above the standard.

Determining the optimal level of service to be provided to each zone is a multistep process. First, productivity is predicted by zone and obtained for each of several service times. Next, an interim productivity standard is set. Then, for each zone, the service time is chosen that has the lowest productivity but still exceeds the standard. Total regional bus kilometers are then calculated by using the selected zonal service times and are compared to the regional bus-kilometer constraint. If total bus kilometers do not approximate the constraint, the selection process is repeated. If the allocated bus kilometers are too few, the productivity standard is lowered; if too many, the productivity standard is raised. Our experience has shown that, within two or three iterations, a set of zonal service times can be selected that utilize the available regional bus kilometers and that, because of the process, produce the maximum possible patronage. These iterations involve only the selection step and do not require additional forecasts of impacts. For a complete regional system, this selection process would be done separately for radial and circumferential service.

The output of this process is the designation of the optimal transit service time to be provided to each travel-analysis zone in the region. This information can be used to design a route-specific transit network that provides those service times by translating service time to its walk and wait components and then to transit spacing and headway. This process usually results in some areas receiving better service than was originally selected for them, to maintain route and system continuity. These unproductive kilometers, which must be considered in setting a regional constraint, should constitute between 10 and 30 percent of the regional total, depending on local geography and topography.

Service-time data are helpful in setting short-range policies for a transit system as well as in designing a complete future-year system. Such data show the type of service various areas can support within the selected transportation objectives. This can be compared to the service currently being provided to locate candidate areas for service improvement.

APPLICATIONS OF THE METHODOLOGY

San Diego Study

The San Diego Short-Range Transit Study was the first application of the design-synthesis approach to transit planning. The first step was to attempt to validate the transit-service specification and demand-forecasting capabilities of the methodology. The existing San Diego transit system was simulated by coding zonal radial and circumferential walk and wait times based on the frequency of service and the coverage provided to each zone by the various transit lines in the system. Zone-to-zone transit travel times were then estimated by the design-synthesis procedure and input into the San Diego model (10) to forecast expected regional transit trips. Regional bus kilometers were also calculated from the walk and wait times. The results were that regional transit riders were underestimated by 8 percent, regional bus kilometers by 6 percent, and regional average trips per bus kilometer by 2 percent. The range of error was felt to be extremely small considering (a) the amount of detail needed to specify a transit system and (b) that the design-synthesis approach is intended as a tool to aid in systematic design and not as a replacement for other, more sophisticated network simulation programs.

In this study a set of transit objectives and criteria were developed for San Diego that defined the amount and quality of service that should be provided to various parts of the region. The most significant criteria defined levels of minimum accessibility to be provided by transit. The accessibility standards were stratified by location and by a socioeconomic indicator. The design-synthesis approach was used to identify service improvements that would enable the accessibility standards to be met. Five route extensions, four new routes, and nine headway reductions on existing routes were proposed to enable all of the standards to be exceeded or at least nearly met. In two instances, only 98 percent and 97 percent of standards were achieved.

The San Diego study demonstrated the feasibility of design-synthesis planning. The methodology for transit-service specification produced a reasonable replicate of reality. Most important, the output of the process identified specific service improvements that would bring the transit system closer to its objectives and thus was found to be applicable to the needs of transit operators.

Denver Long-Range Transit Analysis

The purpose of the Denver Long-Range Transit Analysis was to design a complete future-year transit system rather than, as in San Diego, to identify specific improvements to existing service. As mentioned earlier, three transit system alternatives were designed: a local service alternative at the current level of service, a local system providing slightly more than twice as much service as the first alternative, and a local and express service using exclusive busway facilities. All three systems were designed to maximize patronage.

An interesting aspect of this study was the identification of appropriate corridors for exclusive busway facilities. Corridors were also identified for fixed-guideway facilities proposed as part of other system alternatives. Corridors were identified by examining the impacts on system utilization of changes in transit operating speed. Forecasts were made for a particular set of service times and transit speeds of 19.3 and 38.6 km/h (12 and 24 mph). The difference between the two output transit trip matrixes yielded a matrix of trips attracted to transit by the speed increase. When this matrix was assigned to a highway network, portions of the region were located that had the largest transit trip increases resulting from bus-priority treatments. Potential corridors for exclusive transit facilities were identified as those highway links with the largest numbers of assigned trips.

This study demonstrated that the transit design-synthesis approach can assist planners in specifying optimum future transit system alternatives. It also expanded the scope of the approach beyond application to local bus operations and design.

Denver Short-Range Transit Study

The application of the design-synthesis approach in the Denver Short-Range Transit Study was similar in scope to its application in the San Diego study. First, the existing system was simulated. The design approach underestimated patronage by 4 percent, regional bus kilometers by 1 percent, and regional average trips per bus kilometer by 2 percent. The results were even better than those obtained in San Diego.

More complicated criteria were used in Denver than in San Diego. A productivity criterion was used to identify areas where additional service could be provided to increase patronage and still maintain a standard level of trips per bus kilometer. An accessibility criterion identified areas that required additional service to meet standards, and an additional criterion identified the minimum transit travel times necessary to reach a major shopping center and a general hospital.

The service improvements identified by this approach included 21 frequency improvements, 2 route extensions, 1 additional radial route, and 9 areas for intra-area circulator systems. Identification of areas that warrant circulator systems is generally difficult and was made possible in this case by a separate analysis of radial and circumferential trips. Various areas in the region showed a potential for producing significant numbers of circumferential transit trips, and it was found that the majority of these trips were destined for attractions near their points of origin. The most promising of these areas were selected to receive circulator transit service.

This most recent application once again demonstrated the utility and flexibility of the design-synthesis approach. After the impacts were forecast, most of the design effort was undertaken by staff of the Denver Regional Transportation District.

RECENT DEVELOPMENTS

The design-synthesis approach to transit planning is currently being incorporated, with improvements, into the urban transportation planning system (UTPS) of the Urban Mass Transportation Administration (UMTA) computer program battery as a part of the UMTA short-range planning software systems development program. The number of travel orientations is being expanded to handle express trips in designated corridors and various combinations of local and express trips such as those that occur with feeder service. The approach is also being expanded for use in car-pool priority system design and will be capable of specifying exclusive car-pool lanes having differential speeds and explicit entry requirements as well as differential parking costs and terminal times for car-pool vehicles. To analyze the impact of these specifications on mode choice, a default modal-choice model is being incorporated into the program. This model, which is a five-mode, work-purpose, multinomial logit model, can handle separately categories of one, two, three, and four or more persons per vehicle mode, as well as a transit-passenger mode, thus permitting different definitions as to how many riders constitute a car pool. The five-mode model does not need a model of automobile occupancy to predict the number of vehicle trips as does a two- or three-mode model.

The UMTA program, which is compatible with other UTPS programs, was scheduled to be released during 1977 and to be accompanied by a user's manual to serve as a guide for the use of the design-synthesis approach.

CONCLUSIONS

A design-synthesis approach in which transportation service is specified explicitly to meet transportation objectives is the most efficient one for the design and evaluation of transportation plans. It allows planners and policy makers to concentrate on the service desired for an urban area rather than on the specifics of transportation networks. Such an approach requires an abstract specification of transit service so that service characteristics such as frequency, headway, travel speed, and fare can be examined individually without being constrained to a specific system alternative.

The approach can be used to identify the levels of transit service that should be provided to various parts of a region to meet specific criteria. Accessibility and productivity criteria have been used, but other criteria can also be employed. (This process will be mechanized in the new UTPS program.) The concept is equally applicable to short- and long-range design applications and meshes easily with other existing transportation planning tools. It is not dependent on specific forecasting models but can use any available local models. Incorporation of the approach procedure into the UTPS program battery will allow easy access to the methodology by planners. The approach can then be used either through the UTPS program or individual user-coded computer programs.

The design-synthesis approach to transit planning is a useful tool in a structured framework for transit system planning and design, is applicable to a wide variety of planning situations, and is a step toward the development of more effective multimodal design-synthesis planning.

ACKNOWLEDGMENTS

The San Diego Short-Range Transit Study was performed by R. H. Pratt Associates, Inc., as subcontractor to De Leuw, Cather and Company under contract to the Comprehensive Planning Organization of the San Diego region. The Denver Long-Range Transit Development Analysis and Short-Range Transit Study were performed by R. H. Pratt Associates, Inc., as subcontractor to De Leuw, Cather and Company as a part of the Regional Transportation District system management contract. The UTPS program development work is being performed by Peat, Marwick, Mitchell and Company and their subcontractor, R. H. Pratt Associates, Inc., under contract to the Urban Mass Transportation Administration.

REFERENCES

1. R. H. Pratt and G. W. Schultz. A Systems Approach to Subarea Transit Service Design. HRB, Highway Research Record 419, 1972, pp. 37-47.
2. R. H. Pratt and H. W. Bevis. An Initial Chicago North Suburban Transit Improvement Program. North Suburban Transportation Council, Skokie, Ill., May 1971.
3. Systems Analysis Approach to Transit Route and Service Design. R. H. Pratt Associates, Inc.; San Diego Comprehensive Planning Organization, Aug. 1974.
4. Transit Development Plan and Program San Diego Metropolitan Area 1975-1980. De Leuw, Cather and Co., R. H. Pratt Associates, Inc., and D. H. James; San Diego Comprehensive Planning Organization, Feb. 1975.
5. Bus Design Systems Analysis Methodology. R. H. Pratt Associates, Inc.; Denver Regional Transportation District, Nov. 13, 1974.
6. Service Level Investigation. R. H. Pratt Associates, Inc.; Denver Regional Transportation District, Jan. 9, 1975.
7. Bus Network Descriptions. R. H. Pratt Associates, Inc.; Denver Regional Transportation District, Jan. 27, 1975.
8. Transit Concept Comparison, Long Range Transit Development Analysis. Denver Regional Transportation District, April 1975.
9. Transit Development Program 1975-1980. Denver Regional Transportation District, March 1975.
10. Implementation of the N-Dimensional Logit Model. Peat, Marwick, Mitchell and Co., Washington, D.C.; San Diego Comprehensive Planning Organization, May 1972.

Publication of this paper sponsored by Committee on Transportation Systems Design.

**Mr. Scheibe was with R. H. Pratt Associates, Inc., when this research was performed.*

Accommodating Multiple Alternatives in Transportation Planning

Darwin G. Stuart, Barton-Aschman Associates, Inc., Evanston, Illinois
Warren D. Weber, California Department of Transportation, Sacramento

This paper, which is based on procedures used in the San Diego-Los Angeles Corridor Study, examines several methodological improvements that enable a wider range of multimodal alternatives to be included in the transportation planning process. Staging of the planning and evaluation processes is identified as a basic organizing strategy. The design of significantly different alternatives, in terms of primary service characteristics, is described, and alternative multimodal service combinations are emphasized. The paper discusses travel-demand analyses conducted at relatively low cost at a sketch-planning level of detail with multiple computer model runs and efficient model application. A goal-achievement-oriented evaluation framework is specified that permits the quantitative evaluation of a wide range of local and regional performance objectives. The role of judgmental assessment as well as several areas for additional methodological improvement is also discussed.

One of the more frequently expressed concerns in urban transportation planning involves the need for a wider range of alternatives (4, 8, 11). More alternatives are needed, for example, to explore greater variation in levels of transit service or to investigate additional right-of-way location opportunities. Incorporating a larger number of alternatives in the planning process will expand the level of effort involved. Improved methodologies must therefore be developed that better organize the sequence of planning and evaluation activities and accommodate a wider range of transportation planning alternatives.

Although multiple alternatives are important at each major planning level—corridor, subarea, regional system, interregional, state—the interregional planning level is used here for illustration. The general approach used to deal with the major methodological questions can be applied at other levels of planning. The San Diego-Los Angeles Corridor Study, sponsored cooperatively by the California Department of Transportation (CALTRANS), the Southern California Association of Governments (SCAG), and the Comprehensive Planning Organization of the San Diego Region (CPO), is used as a case study. The methodological topics addressed are (a) staging of the planning and evaluation process, (b) broad-brush design of alternatives, (c) travel-demand analysis (at a sketch-planning level of detail), and (d) goal-oriented evaluation of alternatives.

STAGING THE PLANNING PROCESS

When the number of alternatives to be analyzed in the planning process is significantly increased, some strategy must be devised for sequencing the work. Given the increasingly comprehensive depth of analysis (in environmental impact statements for example), it is unlikely that all alternatives can be examined at once. One way to deal with this problem is to stage the planning and evaluation process into two or more phases. Each successive phase can reduce the number of alternatives to be evaluated and, as the number of alternatives is reduced, the level of detail with which each is investigated can be increased. Such a process can affect the design and redesign of alternatives, the level of detail achieved in analyses of travel demand and indirect impact, and the level of detail pursued in the evaluation of alternatives.

The two-stage planning process used in the San Diego-Los Angeles Corridor Study is shown in Figure 1 (1). In the first stage, a series of 21 multimodal alternatives (or packages) were analyzed and evaluated. Only 14 objectives and 27 evaluation criteria were applied. Modal-improvement options were reduced at this stage from 21 to 7, and those 7 were then subjected to a second round of analysis. This second, more thorough stage, which used 25 objectives and 41 evaluation criteria, emphasized the analysis of indirect effects and resulted in the identification of a single, preferred multimodal improvement plan.

DESIGNING ALTERNATIVES

According to the basic staging strategy, the design of modal improvement alternatives can also follow a pattern of increasing detail for a smaller number of alternatives. Nevertheless, much of the basic work in defining alternatives should probably be done in the initial stage of the planning process (14). The primary service characteristics—route location, line-haul speed, number of access points, and frequency of service—must all be established. This was essentially the strategy followed in the case study: The second stage only refined operating and cost characteristics for the set of seven alternatives, to more carefully match supply with forecast demand. Regardless of how the design of alternatives is staged, developing significantly different levels of service among alternatives is essential, both among modes and within the alternative levels of improvement hypothesized for any single mode.

There are two general ways in which the number of transportation alternatives under consideration can be increased.

1. Expand the number of modes investigated (4). In urban area transportation planning this generally calls for a broader consideration of transit alternatives in which different technologies are treated as alternative modes (e.g., bus rapid transit, heavy-rail mass rapid transit, light rail transit, small-group rapid transit, personal rapid transit). Metropolitan transit planning studies are only beginning to give comprehensive consideration to the many technology options. At the intercity or statewide planning level a number of modes, some with further technology options, already exist: automobile-highway, intercity bus, intercity rail, air, and in some cases water. These five modes were included in the San Diego-Los Angeles corridor planning project.
2. Devise a strategy to span the range of reasonable improvement alternatives within a given mode by examining several alternative levels of improvement (13, 16). Two to five alternative levels, from a minimum-improvement base through increasingly ambitious service and facility expansions, may be appropriate. Initially, such alternate improvement levels can be devised to reflect a broad understanding of current urban area or interregional travel patterns, short-range improvement plans, and various technological options reported in the literature. Such improvement levels should be designed for basic service characteristics at a sketch-planning level of detail (by including only generalized route alignments or station locations, for example).

As given in Table 1, 21 different modal-improvement alternatives were defined in the San Diego-Los Angeles case study for four service characteristics: number of routes, number of access points, maximum line-haul speed, and one-way frequency of service (1). Introducing a larger number of alternatives means dealing with intermodal relationships (15), which are crucial in demand analyses. Multimodal demand models currently available for projecting modal market shares hinge on the relative level of improvement in each mode. The different levels of improvement in a mode must be combined with varying levels of improvement in other modes to form multimodal packages for demand-analysis testing. When the number of alternatives is significantly increased, the number of possible multimodal packages quickly becomes unmanageable. A simplifying process that incorporates the staging strategy discussed above is necessary. In the case study the first-stage analysis defined only 21 multimodal packages by holding all modes except the subject mode at a base level of improvement (level 0). This allowed travel demand analysis and other analyses to focus on the relative effects of service improvements, one mode at a time.

Level 0 for each modal-improvement alternative should generally reflect current short-range regional and local transportation plans. This baseline should not only include existing facilities or services but also all relevant projects and programs contained in the 5-year implementation program of the local governments and transit operators concerned. Level 0 might thus be regarded as a no-build or low-capital-intensive alternative. Additional levels of improvement within a given mode can then be devised in an incremental manner, each built on the last. Questions of supply and demand and cost versus revenue can be made a part of the overall evaluation as the alternatives are narrowed down.

ANALYZING TRAVEL DEMAND

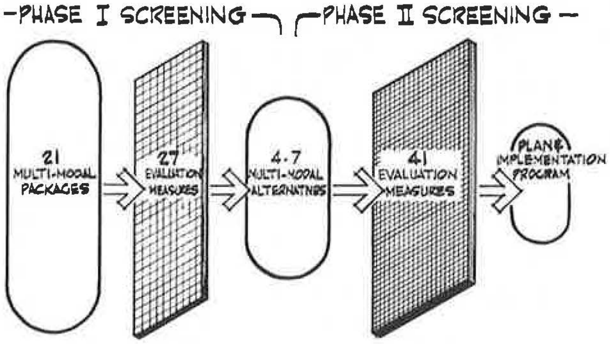
The progress made in recent years in improving urban travel demand models has been aided particularly by the urban transportation planning system (UTPS) package and its component models as well as by various add-on models, subroutines, or modifications that can be incorporated in UTPS, including logit-type mode-choice models calibrated on the basis of disaggregate, individual trip records and direct-demand models combining trip generation, distribution, and mode choice within a single-decision forecasting step. These modeling advances, which promise to improve substantially the overall transportation planning process, are well-documented in the literature (6).

Important progress has been made in developing or adapting models that can be applied at a sketch-planning (large-zone) level of detail; the number of alternatives that can be considered has thus been greatly increased. For example, a recent transit planning case study in the Milwaukee area involved adapting large-zone modeling techniques within the UTPS framework and testing a wide range of regional dual-mode-guideway network configurations (7). The analysis involved three stages: manual sketch-planning (and simplified modeling) analysis of 15 initial baseline systems; computer-based analysis for three refined baseline systems, including mode split and transit network assignment for a 100-zone system; and a series of 150 modeling runs for a variety of parametric analyses to systematically test variations in different service characteristics. Considerable flexibility and range were achieved in the number of alternatives accommodated.

A recent Los Angeles study of regional mode-choice incentives and disincentives achieved similar flexibility and multiple-run capability (18). In this case, a large number of transportation control strategies had to be evaluated relatively quickly and related to various improvements in level of transit service (routes and schedules). These control strategies included restrictions on parking cost and supply, preferential freeway ramp and lane treatment for multiple-occupancy vehicles, constraints on gasoline price and availability, and carpooling incentives. A modified DODOTRANS modeling package at a 107-zone level of analysis was developed and applied to permit the essential quick turnaround time in travel-demand model application (12). Fifty-five combinations of transportation control strategies were then tested.

Other approaches to travel-demand modeling at a sketch-planning level of detail are being developed and applied but cannot be adequately treated here (6). Instead, the intercity demand forecasting performed as part of the San Diego-Los Angeles Corridor Study is briefly reviewed to demonstrate how sketch-planning models are applied in a multiple-alternative context (2).

Figure 1. Basic plan-evaluation process of San Diego-Los Angeles Corridor Study.



The multimodal, direct-demand model used in this case for forecasting total travel demand and mode split among five different modes represents a modification of the DODOTRANS package undertaken by CALTRANS (12). The 1995 multimodal demand forecasts were made for 20 of the initial modal-improvement alternatives (packages) and four refined improvement alternatives. A total of about 40 full modeling runs were made over a time span of 6 months after model calibration was completed (these included additional runs made to account for adjustments of input data).

Both the direct-demand and multimodal features of the demand model are particularly significant for intercity analysis. In the direct-demand approach, as noted above, the three fundamental steps in demand estimation—trip generation, trip distribution, and mode split—are performed simultaneously rather than sequentially, which ensures that both total amount of travel and amount and geographic distribution of travel attracted by each mode are directly related to the supply of transportation provided by each mode. Thus the concept of induced travel—that increase in travel demand that can be related to an increase in the level of service provided by any particular mode—can be represented. The multimodal nature of the model permits the competitive effect of varying levels of service among modes to be tested in each modeling run.

For modeling purposes the San Diego-Los Angeles Corridor Study area was divided into 141 zones: 107 in the Los Angeles area, 31 in the San Diego region, and 3 in Tijuana, Mexico. Relatively coarse transportation networks were then developed for highway, bus, rail, and air routes in relation to this zonal system, both for current conditions, as input to model calibration, and for the 1995 forecast of modal-improvement alternatives as part of the first-stage evaluation. For the nonhighway public modes, terminal-to-terminal matrices were developed for scheduled travel times, fares, and service frequencies. Business and nonbusiness trip purposes were considered. In the second-stage evaluation, 1985 networks for selected alternatives were also developed.

Table 1. Alternative modal service characteristics.

Mode	1995 Improvement Level	Number of Routes	Route Length[a] (km)	Access Points or Terminals[b]	Maximum Speed (km/h)	Daily One-Way Frequency
Automobile-highway	0	6	126	394	88.5	Unlimited
	1	5	266 to 282	473 to 482	88.5	Unlimited
	2	3	292	482	88.5	Unlimited
	3[c]	3	240 to 490	490 to 519	88.5	Unlimited
Intercity bus	0	3	177 to 202	15	88.5	2 to 34
	1	3	177 to 226	16	88.5	6 to 34
	2	4	177 to 250	21	88.5	18 to 31
	3	7	177 to 258	53	88.5	15 to 45
Intercity rail	0	1	207	7	145	5
	1	1	207	7	145	6
	2	1	207	7	177	8
	3	1	202 to 207	7	177	10
	4	1	198	3	323	21
Air	0	5	123 to 195	10	645	5
	1	5	123 to 195	10	387 to 645	6 to 46
	2	3	139 to 171	6	387	5
	3	6	137 to 205	12	290 to 387	5 to 10
Water	1	1	145 to 161	2	84	3
	2	2	145 to 181	3	84	3
	3	3	145 to 194	4	84	2 to 3

Note: 1 km = 0.62 mile.

[a] Total route kilometers are given for the automobile mode; individual route kilometers are given for all other modes.

[b] Double counting is used for bus, air, and water modes if a terminal may be served by more than one route.

[c] Comprises two alternatives for analysis purposes.

The large-zone system developed for the study was quite coarse in nature. Conventional travel-demand analyses in the Los Angeles area are conducted at a 1285-zone level. The sketch-planning framework of the corridor study obviously greatly reduced the work of network preparation as well as computer processing time for model application. The much quicker turnaround time for individual modeling runs and the ability to test several modal improvement alternatives in rapid succession made it possible to examine a large number of alternative multimodal transportation systems spanning the entire corridor study area [250 km (155 miles) in length]. Although a large number of shorter intrazonal trips were eliminated, the large-zone site used in the analysis primarily eliminated only the many shorter trips that were not likely to compete for capacity on interregional transportation facilities (with the exception of some automobile-highway routes). The large-zone system was thus compatible with the longer trips typical of interregional travel demand.

EVALUATING ALTERNATIVES

Because a significant increase in the number of alternatives to be evaluated can greatly increase the amount of information to be processed during the evaluation phase, using some systematic cost-effectiveness framework for plan evaluation is essential (5). Such a framework should connect the evaluation of alternatives to transportation goals and objectives established for the corridor, region, or multiregion study area (17). The objectives should in turn be expressed in some measurable way, wherever possible, by specific quantitative or qualitative criteria. In any case, procedures must be designed for the subjective comparison of alternatives at a significant level, either by assigning weights among goals or by assessing trade-offs among the impacts of various alternatives (9).

Staging is especially critical in the plan-evaluation phase. Because large amounts of information are generated not only for travel demand impacts but for social, economic, and environmental impacts as well, it may be best to examine only a selected set of most significant impacts during the first stage of plan evaluation, when the largest number of alternatives must be reviewed. After this initial range of alternatives has been screened and a smaller number of most promising options have been identified, a more complete set of objectives and criteria is available for subsequent evaluations. The alternatives themselves can also be further refined with regard to such details as cost characteristics, operating scenarios, and route and station locations.

Many methodological options can be applied in the plan-evaluation process (17). The structures of cost-effectiveness evaluation matrices and the range and description of goals and objectives can vary; wide variation can also be expected in the types of criteria applied and in the extent to which the community and the decision makers get involved in plan evaluation (including the extent to which goal-weighting exercises are conducted). Selection of an appropriate sequence of plan-evaluation methodologies thus depends on the unique circumstances of the study area as well as the agency, decision-maker, and community participants involved. Because what works in one area may not work in another, we emphasize the illustrative nature of the evaluation procedures used in the San Diego-Los Angeles Corridor Study, a brief description of which is given below (2,3).

The five broad goals and 25 objectives that guided the case study corridor project are summarized in Table 2. These goals and objectives were synthesized from local, regional, and statewide goal and policy statements based on current plans and on interviews with transportation and land-use planning agencies. The five goals cover overall transportation problems; multimodal balance; social, economic, and environmental consequences; interregional transportation demands; and local and neighborhood impacts. The 25 objectives reflect both regional and local concerns and involve both direct and indirect consequences of transportation improvements.

The goals and objectives given in Table 2 are also grouped under three basic issues: economic feasibility, nonuser impacts, and user benefits. These three issues formed the backbone for plan-evaluation trade-off analyses. At least one evaluation measure or criterion was defined for each objective, as indicated in Table 2. Twenty-seven of the more significant measures (for 14 objectives) were applied during the first phase of evaluation; all were applied during the second phase. Most of these criteria are quantitative, e.g., costs and revenues, air pollutant emissions, service frequencies, and ridership levels. Qualitative measures reflecting judgmental assessments by the study team in such areas as aesthetics, tax-base impacts, and support of the California coast environmental plan were used in a few cases. In addition to the travel-demand and associated cost and revenue analyses, the study team estimated the impacts of many of the objectives in Table 2 by using a variety of environmental and land-use impact analyses.

During each phase of the evaluation, goal achievement was assessed by using a three-step process. First, an impact-analysis matrix was completed for each set of alternatives to compare the levels of modal improvement. Table 3 gives the results of the various feasibility, impact, and benefit analyses conducted by the study team for the intercity bus alternatives tested in phase 1. These impact measures were then converted to relative rankings, according to least negative impact or most positive impact, within each mode. This made comparing the alternatives easier and represents a crude form of normalizing—converting all measures to a common percentage score over a high-low range of impact values for a particular criterion (Table 3). A subjective comparison and a trade-off analysis were then made among the three basic categories of impact: economic feasibility, nonuser impacts, and user benefits. (In Table 3 it would be possible to add information to assign relative weights to objectives and to calculate a single weighted summary score for each alternative, but this was avoided because it was felt that a single summary score would tend to oversimplify the evaluation process and obscure some important differences among the alternatives.) Tables 4 and 5 give a summary of the results of this subjective procedure for the first and second evaluation phases respectively. A judgmental ranking of alternatives in the three basic issue areas is given based on a comparative assessment of goal-achievement evaluations such as that given in Table 3. Although such a subjective procedure may be criticized, it does force each evaluator to reflect carefully on the results of impact analyses and to compare the relative performance of alternatives.

Judgmental rankings permitted a reduction from 21 to 7 basic multimodal alternatives in the first evaluation phase and then a reduction from 7 alternatives to one recommended multimodal combination, as indicated in Table 5. During the first phase of the evaluation only within-mode comparisons were made (Table 3). For example, only the different bus alternatives were compared to identify the most promising initial set of bus alternatives. In the second-stage evaluation, however, comparisons were made between modes, and the 7 final alternatives, as well as an expanded list of evaluation criteria, were listed within single impact-analysis and goal-achievement

Table 2. Criteria for goal-achievement evaluation.

Issue	Goal	Objective[a]	Evaluation Measure	Application
Economic feasibility	Improve multimodal balance	Ridership levels	Number of weekday person trips	Phases 1 and 2
			Weekday mode-split percentage	Phases 1 and 2
		Revenue-cost viability	Annual revenue to operating cost ratio	Phases 1 and 2
		Investment efficiency	Annual operating cost/passenger·km	Phases 1 and 2
			Annual capital cost/passenger·km	Phases 1 and 2
		Implementation feasibility	1985 revenue to operating cost ratio	Phase 2
			1985 revenue to total cost ratio	Phase 2
		Geographic balance	Modal improvement costs by county	Phase 2
		Modal coordination	Number of multimodal terminals	Phase 2
			Judgmental rating of improvement staging	Phase 2
	Effectively meet interregional travel demands	Multimodal rights-of-way	Bimodal route distance	Phase 2
			Trimodal route distance	Phase 2
		Collection-distribution interfaces	Judgmental ratings by mode	Phase 2
		Capacity-demand balance	Volume-capacity ratios on peak links (public modes)	Phase 2
Nonuser impacts	Minimize undesired social, economic, and environmental impacts	Coastal environment	Judgmental ratings by mode	Phases 1 and 2
		Open space resources	Designated open space and parks consumed (hm^2)	Phases 1 and 2
		Ecological and historical resources	Number of intrusions on historical or archaeological sites	Phases 1 and 2
		Agricultural resources	Agricultural land consumed (hm^2)	Phases 1 and 2
			Vacant land consumed (hm^2)	Phases 1 and 2
		Transportation noise	Noise level at 15-m (dBA)	Phases 1 and 2
			Maximum frequency of service	Phases 1 and 2

Note: 1 hm^2 = 2.5 acres; 1 m = 3.3 ft.

[a] Only the basic factor involved is given. The appropriate verb should be supplied; e.g., improve, increase, preserve, reduce, minimize.

tables. Various multimodal combinations of service levels could also be examined, e.g., automobile 1, bus 2, rail 2, air 1 (a possible total of eight combinations). This was partly accomplished in the demand analysis.

CONCLUSIONS

The San Diego-Los Angeles Corridor Study has made some first steps in accommodating a larger number of alternatives in interregional transportation planning. Improvement is needed, however, in the following areas in developing planning methods that will make multiple alternatives more meaningful in the planning process. Needed improvements in travel-demand modeling have been adequately addressed elsewhere (6) and are not included here.

Design of Alternatives

1. A more varied mix of modes is needed. Alternatives tended to be developed one mode at a time; for example, a combined interregional bus-rail alternative was not examined. At the regional system planning level mixed-technology transit alternatives may be especially relevent.
2. Increased short-range emphasis is needed. While minimum-level improvement alternatives (level 0) could be interpreted as short range in nature, additional low-capital-intensive options in transportation system management should be defined. These will tend to become more detailed and local in nature, but must somehow still be contained within a sketch-planning framework.
3. More emphasis should be given to the staging of alternatives. In the case study, recommendations were developed for the single preferred multimodal alternative, in terms of a three-stage series of improvements, and the first 5-year stage was emphasized. Thus it appeared that the staging options themselves could be made a part of the basic alternatives, particularly if more emphasis were given to short-range alternatives (10).
4. More careful attention should be given to the identification of key decision points. In blending short-range with long-range alternatives, decisions that foreclose future options, especially technology choices, should be clearly identified, perhaps in the form of a decision tree indicating those options that remain open at each successive stage of decision making.
5. More direct participation is needed by community groups and individuals as well as decision makers. Generating significant levels of community or decision-maker participation was difficult in the San Diego-Los Angeles Corridor Study mainly because interregional transportation needs were only a small proportion of overall regional travel needs. Increased participation focused on the design of alternatives should be vigorously pursued at smaller scale regional and corridor transportation planning levels.

Evaluation of Alternatives

1. Methods for trade-off comparisons and judgmental matching of alternatives should be more systematic and explicit, especially when they concern impact or issue conflicts. Judgment cannot be eliminated, but the kinds of subjective trade-offs illustrated in Tables 5 and 6 should be more clearly explained, e.g., by more detailed tabular or graphic summaries.
2. Some form of goal weighting, although not essential, may be desirable. Goal weighting, especially in support of more systematic trade-off comparisons, could simplify the comparison process by permitting the calculation of performance indexes for alternative plans (20). A variety of techniques exist for goal weighting.
3. Better procedures must be developed for incorporating the results of parametric analyses in the evaluation process. Parametric analyses can greatly increase

Table 3. Impact estimates and improvement rankings for intercity bus alternatives.

			Improvement Level							
			0		1		2		3	
Issue	Objective[a]	Criterion	Estimate	Ranking	Estimate	Ranking	Estimate	Ranking	Estimate	Ranking
Economic feasibility	Ridership levels	No. of person trips (000)	4	4	5.6	3	6.8	2	8.6	1
		Mode split (%)	3.2	4	4.4	3	5.3	2	6.6	1
	Cost-revenue viability	Revenue to operating cost ratio	1.8	1	1.7	1	1.4	2	1.2	3
		Revenue to total cost ratio	1.8	1	1.5	1	1.2	2	1	3
	Investment efficiency	Operating cost/passenger·km ($)	0.028	1	0.029	2	0.033	3	0.037	4
		Annual capital cost/passenger·km ($)	0	1	0.002	2	0.006	3	0.007	4
Social, economic, and environmental nonuser impacts	Coastal environment	Judgmental rating (support of California coastal plan)	2	2	2	2	2	2	3	1
	Open space resources	Open space, parks, ecological preserves, wildlife habitats consumed (hm^2)	0	1	0	1	0	1	0	1
	Ecological and historical resources	No. of intrusions on historical or archaeological sites	0	1	0	1	0	1	0	1
	Agricultural resources	Agricultural land consumed (hm^2)	0	1	0	1	0	1	0	1
		Vacant land consumed (hm^2)	0	1	1.2	2	2.4	3	5.2	4
	Transportation noise	Noise level at 15-m (dBA)	75 to 85	1	75 to 85	1	75 to 85	1	75 to 85	1
		Maximum frequency of service	88	1	116	2	134	3	284	4
	Neighborhood disruption	No. of community areas severed	0	1	0	1	0	1	0	1
		No. of residential units displaced	0	1	0	1	0	1	7	2
		Residential land consumed (hm^2)	0	1	0	1	0	1	1.6	2
		No. of businesses displaced	0	1	0	1	11	2	18	3
		Commercial and industrial land consumed (hm^2)	0	1	0	1	8	2	5.2	3
	Air quality	CO/passenger·km (g)	0.33	1	0.33	1	0.33	1	0.39	2
		HC/passenger·km (g)	0.06	1	0.06	1	0.06	1	0.07	2
		NO_x/passenger·km (g)	0.12	1	0.12	1	0.12	1	0.12	1
	Energy consumption	No. of automobile trips (000)	105.4	2	104.8	1	104.8	1	106.1	3
		Kilojoules/passenger·km	696	1	430	1	430	1	430	1
User benefits	Modal availability	No. of access points	56	2	56	2	61	1	62	1
		Daily one-way frequency of service	44	4	58	3	67	2	142	1
		Los Angeles to San Diego line-haul travel time (min)	140	1	140	1	140	1	140	1

Note: 1 km = 0.62 mile; 1 hm^2 = 2.5 acres; 1 m = 3.3 ft; 1 g = 0.035 oz; and 1 J = 0.000 94 Btu.
Alternatives are ranked according to least negative impact (9 objectives) or most positive impact (4 objectives), and best performance rating receives a ranking of 1.

[a] Only the basic factor involved is given. The appropriate verb should be supplied, e.g., improve, increase, preserve, reduce, minimize.

Table 4. Phase 1 evaluation of modal alternatives.

	Ranking by Goal									
	Economic Feasibility			Nonuser Impacts			User Benefits			
Alternative	First	Second	Third	First	Second	Third	First	Second	Third	Preferred Alternatives
Automobile-highway	—	—	—	0	1	2	3A	3B	2	1, 2
Intercity bus	1	2	0	1	0	2	3	2	1	1, 2
Intercity rail	1	0	2	1	0	2	4	3	2	1, 2
Rail extension[a]	TJ 3	—	—	TJ 3	—	—	TJ 3	—	—	TJ 3
Air	1	0	2	0	1	2	3	2	1	1

[a] Tijuana only.

Table 5. Phase 2 evaluation of modal-improvement alternatives.

	Ranking by Goal		
Alternative	Economic Feasibility	Nonuser Impacts	User Benefits
Automobile-highway			
1B	—	5	2
2	—	6	2
Bus			
1	1	1	4
2	2	1	4
Rail			
1	2	2	4
2	3	3	3
Air 1	—	4	1

Note: The preferred combination is bus 1, rail 1, air 1, automobile-highway 1B.

the amount of information available on the performance of alternatives. Ranges of performance or impact might be consistently associated with each alternative, or some means might be used to attach probabilities to different consequences.

4. Participation by affected groups and by responsible decision makers is crucial. At regional, subarea, and corridor levels of planning, it is even more important to ensure that all significant needs and impacts are addressed.

5. More effective communication devices are needed, including graphs, charts, pictograms, and color-coded maps and tables that effectively display the differences among alternatives and present increasingly large amounts of information in a form that has meaning for most decision makers. This may be one of the most important research areas in plan evaluation.

REFERENCES

1. San Diego/Los Angeles Corridor Study: Phase IIA, Improvement Alternatives. Barton-Aschman Associates, Inc.; California Department of Transportation; Southern California Association of Governments; and San Diego Region Comprehensive Planning Organization, Aug. 1975.
2. San Diego/Los Angeles Corridor Study: Phase IIB, Travel Demand Analysis. Barton-Aschman Associates, Inc.; California Department of Transportation; Southern California Association of Governments; and San Diego Region Comprehensive Planning Organization, Feb. 1976.
3. San Diego/Los Angeles Corridor Study: Phase III, Final Report. Barton-Aschman Associates, Inc.; California Department of Transportation; Southern California Association of Governments; and San Diego Region Comprehensive Planning Organization, draft, April 1976.
4. P. Benjamin, J. Barber, C. Heaton, G. Paules, and D. Ward. Urban Transportation Alternatives: A Macro Analysis. Transportation Systems Center, U.S. Department of Transportation, Dec. 1974.
5. R. J. Bouchard, E. L. Lehr, M. J. Redding, and G. R. Thomas. Techniques for Considering Social, Economic, and Environmental Factors in Planning Transportation Systems. HRB, Highway Research Record 410, 1972, pp. 1-7.
6. D. Brand and M. L. Manheim, eds. Urban Travel Demand Forecasting. HRB, Special Rept. 143, 1973.
7. Dual-Mode Transit Planning Case Study. Cambridge Systematics, Inc., and Barton-Aschman Associates, Inc.; Urban Mass Transportation Administration, U.S. Department of Transportation, April 1976.
8. R. E. Engelen and D. G. Stuart. New Directions in Urban Transportation Planning. American Society of Planning Officials, Planning Advisory Service Rept. 303, 1974.
9. G. Hall and R. Breuer. User and Community Benefits in Intercity Freeway Corridor Evaluation. HRB, Highway Research Record 399, 1972, pp. 27-39.
10. R. J. Hocking. Time-Stage Strategy in the Transportation Planning Process. TRB, Transportation Research Record 491, 1974, pp. 24-39.
11. W. G. Hansen and S. Lockwood. Metropolitan Transportation Planning: Reforming the Process. TRB, Transportation Research Record 582, 1976, pp. 1-13.
12. M. L. Manheim and E. R. Ruiter. DODOTRANS I: A Decision-Oriented Computer Language for Analysis of Multimode Transportation Systems. HRB, Highway Research Record 314, 1970, pp. 135-163.
13. M. H. Rapp and C. D. Gehner. Criteria for Bus Rapid Transit Systems in Urban Corridors: Some Experiments With an Interactive Graphic Design System. HRB, Highway Research Record 455, 1973, pp. 36-48.
14. J. C. Rea. Designing Urban Transit Systems: An Approach to the Route-Technology Selection Problem. HRB, Highway Research Record 417, 1972, pp. 48-59.
15. J. W. Schmidt, R. K. Arnold, and S. Levy. Specification and Evaluation of Alternative Feeder and Local Transit Systems in a Suburban Area. HRB, Highway Research Record 417, 1972, pp. 37-47.
16. J. B. Schneider and J. W. Clark. Designing Bus Rapid Transit Systems: Exploratory Investigation. TRB, Transportation Research Record 606, 1976, pp. 30-36.
17. J. L. Schofer and D. G. Stuart. Evaluating Regional Plans and Community Impacts. Journal of the Urban Planning and Development Division, ASCE, March 1974.
18. An Analysis of Incentives and Disincentives for Automobile Management in the SCAG Region. Southern California Association of Governments, draft, Oct. 1975.
19. M. Wachs, B. M. Hudson, and J. L. Schofer. Integrating Localized and Systemwide Objectives in Transportation Planning. Traffic Quarterly, April 1974.
20. J. C. Yu and R. C. Hawthorne. Assessing Urban Transit Systems: A Goal Programming Approach. TRB, Transportation Research Record 574, 1976, pp. 35-47.

Publication of this paper sponsored by Committee on Transportation Systems Design.

Development and Application of a Model to Evaluate Transportation Improvements in Urban Corridors

Ronald W. Eash and Edward K. Morlok, Civil and Urban Engineering
Department, University of Pennsylvania, Philadelphia

This paper introduces the linear program model developed at the University of Pennsylvania for evaluating transportation improvements in a high-travel-demand urban corridor. Variables included in the linear program are discussed, and the linear objective function and the constraint equations of the model are outlined. Application to a radial travel corridor in Chicago, Illinois, illustrates the capability of the model; an analysis is made of existing corridor bus service and several corridor capital investments to improve that service. In the analysis of existing bus service, several alternatives to the existing price structure of bus transportation in the corridor were studied; the major result was an evaluation of the shift in mode choice caused by the different pricing schemes and the effects of a change in patronage on bus operating and capital costs. For the study of alternative capital investments, the corridor model computes the patronage attracted by the improvements and adjusts operating and bus capital costs of bus lines serving the corridor.

The overall purpose of the work described in this paper was to develop and test an analytical model for planning transportation improvements in a high-travel-demand urban corridor. The objective was to produce a technique that (a) incorporates anticipated travel demands and establishes air quality and noise standards and financial and energy limitations, (b) searches for transportation alternatives to satisfy those constraints, and (c) identifies alternatives generating the most benefits. This technique was to be an alternative to the sequential transportation models whose primary application is long-range regional urban transportation planning and to the very detailed analyses used in corridor location studies. The goal of the research was a corridor-level approach that could assess a large number of potential corridor transportation improvements and thus enable the determination of the trade-offs between alternatives.

The corridor model developed at the University of Pennsylvania (1) was applied in this research to a radial travel corridor in Chicago, Illinois. Application of the model to an actual planning problem was an important aspect of the corridor model project and was a joint effort of the University of Pennsylvania and the Chicago Area Transportation Study (CATS). The model was applied to the Chicago southwest corridor, the only major radial corridor in the Chicago region in which rapid transit facilities do not operate on exclusive right-of-way.

CORRIDOR PLANNING MODEL

The general form of the corridor transportation planning model used in the analysis is a linear program. The solution of a linear program is a set of variable values that satisfy a series of linear equations or inequalities and also optimize the value of a separate linear equation that is termed the objective function. Mathematically, in a linear program the objective function can be summarized as follows:

$$\text{Maximize (or minimize)} \sum_i c_i x_i \qquad (1)$$

subject to

$$\sum_i a_{ij} x_i \geqslant b_j \qquad j = 1, \ldots, m$$

$$\sum_i a_{ij} x_i = b_j \qquad j = m + 1, \ldots, n$$

$$x_i \geqslant 0$$

where

x_i = choice variables to be evaluated,
c_i = objective function coefficient for x_i, and
a_{ij}, b_j = parameters of the constraint relations.

The model can thus be described by the variables, the objective function, and the constraints.

Choice Variables

In this model choice variables for the highway mode include

1. Volume of traffic on individual links in the highway network,
2. Capacity of individual links,
3. Corridor highway travel times during different periods of the day, and
4. Cost to the user of driving an automobile.

Public transportation choice variables are similar and include

1. Amount of patronage on bus-line segments,
2. Bus travel times, and
3. Bus user costs or fares charged.

But, for public transportation, frequency of service on the lines replaces capacity as a choice variable because frequency of buses on a line determines the number of individual units of capacity or line capacity.

Objective Function

The objective function is to minimize the weighted sum of the costs of providing corridor transportation service plus the vehicle emissions and fuel consumption of any alternative. In more detail, the objective function is composed of the following elements:

1. Capital costs—For highways capital costs are treated as a function of the capacity added to the highway or the improvement in travel time on the route. Capital costs for public transportation are based on the frequency of service during the peak period and any investments made to change line-haul travel times.
2. Public transportation operating costs—These costs depend on frequency of service, which controls vehicle kilometers operated. Travel time also enters this calculation because the number of times a vehicle can be put into service is determined by the time required

for one run of its route.

3. Highway user costs—This cost is calculated from vehicle kilometers of highway travel.

4. Travel time—Highway travel time is the sum of travel times on individual highway links; for public transportation, travel time is a function of link travel times and service frequency.

5. Vehicle emissions—Vehicle emissions are computed from total highway vehicle kilometers.

6. Fuel consumption—The amount of fuel consumed is again calculated from highway vehicle-kilometers.

Constraint Equations

The constraints are as follows:

1. Mode-choice relations—These equations calculate the number of trips for each corridor movement that will use public transportation and allocate the remaining trips to the highway mode. Mode-split fractions are calculated as a function of highway travel times, public transportation travel times and frequencies, and the cost to the user of traveling by either of these modes.
2. Minimum public transportation service levels—These equations relate the minimum frequency of bus service on a line in the corridor to the maximum volume on that line.
3. Summation of flows using a link—Modal flows determined in the mode-split equations are origin-destination movements within the corridor. This set of equations assigns those movements to public transportation and highway links.
4. Bus and highway link capacity—Maximum volumes on a highway link are constrained by these equations to the link capacity. In the case of public transportation, the maximum volume on a link is limited to the capacity of a bus times the number of buses traveling on the link.
5. Highway reverse peak and off-peak travel times—These constraints limit highway travel times in the off-peak period and, in the peak period, reverse direction to values that are consistent with peak-period performance.
6. Noise restrictions—By using these constraints, travel on a highway link can be limited to ensure that standards for maximum traffic noise will be met.
7. Budget restrictions—These constraints ensure that capital and operating costs can be held to specific levels.

The structure of the linear program model applied to the Chicago southwest corridor is shown in Table 1 in the form of a matrix in which an X indicates the use of a choice variable in either the objective function or the constraints.

APPLICATION OF THE MODEL

In the application of the model to the Chicago southwest corridor, travel demand in the corridor was first estimated. The estimate relied on data from a 1970 home interview survey of regional travel undertaken by CATS and the Northwestern Indiana Regional Planning Commission. Data from the home interview survey were supplemented by a mass transit usage survey of regional travel undertaken by CATS in 1974 and miscellaneous traffic counts taken on major corridor arterial highways.

The second phase of the application was to link this travel demand estimate to the linear program formulation of the corridor planning model. An operational test of the resulting linear program was made by using it to analyze the existing bus service in the southwest corridor. Several alternatives to the existing price structure of corridor transportation were studied. The major emphasis of the study was on the impact on choice of mode of these different pricing schemes. Of secondary interest was the question of how the costs of providing corridor bus service change with an increase in patronage.

The final portion of the Chicago application dealt with an evaluation of several capital investments to improve bus service. The alternatives studied included

1. An exclusive-bus-lane facility operating only within the Chicago central business district (CBD),
2. A connection between the CBD bus lanes and a short busway having a high level of service and extending 3.2 km (2 miles) into the southwest corridor,
3. CBD bus lanes and a longer nonstop busway running from the CBD facility to a point 10.5 km (6.5 miles) into the corridor, and
4. A 10.5-km (6.5-mile) extension from the CBD of an exclusive bus facility that serves intermediate corridor destinations.

The corridor model computed the amount of patronage attracted by the time savings of each of the above improvements and adjusted operating and bus capital cost requirements of the bus lines serving the corridor according to the savings accruing from each alternative. The corridor model was also used to investigate other impacts such as user time savings and environmental impacts of each of the four investments.

Existing Corridor Travel

Existing CBD-oriented bus service in the corridor is provided by six bus lines that are used, in a combination of express, limited-stop, and local service, to transport large volumes of passengers. Figure 1 shows the route structure as well as 1974 passenger volumes past selected points (summed over all routes passing that point) by direction during the 6:00 a.m. to 9:00 a.m. morning peak period (2). The total combined two-way daily flow at the maximum load point near Halsted Street is in excess of 27 000 persons/d. The Archer Avenue operation is unusual in its use of three different types of operations along a single route, in the amount of service offered, and in the patronage attracted.

Pricing Options

The corridor model was first applied to the question of how ridership would increase in the southwest corridor if fares were decreased. Using the linear program formulation of the model, this calculation was first done by assuming that frequencies remained at existing peak and off-peak levels. A $0.10 reduction in fare (from $0.45) resulted in a 1.9 percent increase in patronage in the peak and a 6.3 percent increase in the off-peak period. The effect was more dramatic in the off-peak period than in the peak period because of the high ratio of nonriders to Archer Avenue bus users during the off-peak period; even a small percentage change in the mode split between automobile and public transit in the off-peak period added a large number of public transit users.

During the peak period, however, the Archer Avenue bus lines operate nearly at capacity. Average occupancy per bus on the Archer Avenue limited and local service is in the range of 70 to 75 passengers/bus. Expressway lines average only slightly less, approximately 60 passengers/bus (3). Off-peak ridership is considerably less than capacity and averages around 30 percent of total line capacity, including standees, or slightly more than 40 percent of seating capacity. This peak-period capacity constraint means that any reduction in fares in the peak

Table 1. Structure of the linear program corridor model.

	Highway Choice Variables				Transit Choice Variables			
Objective Function	Volume	Capacity	Time	Price	Volume	Frequency	Time	Price
Minimize the sum of								
Capital costs		X	X			X	X	
Operating costs								
Bus						X	X	
Automobile	X							
Travel times			X			X	X	
Vehicle emissions	X							
Fuel consumption	X					X		
Constraints								
Mode-split relation			X	X		X	X	X
Minimum bus service					X	X		
Summation for link flows	X				X			
Bus and highway capacity	X	X			X	X		
Highway nonpeak times	X	X	X					
Noise restrictions	X		X					
Budget restrictions		X	X			X	X	

period must be accompanied by an increase in the peak-period capacity of the Archer Avenue lines. The model indicates that, for the buses in use, existing routes incur an approximate daily capital cost of \$2200 and a daily operating cost of \$21 000. These figures were calculated by using an operating cost of \$0.98/km (\$1.57/mile) for a bus (4) and a bus capital cost of \$23.50/d (5).

Figure 2 is a composite of four plots developed from the corridor model showing the impact on operating and capital costs of diverting travelers from the automobile to the Archer Avenue bus lines. The axes for these plots are as follows:

1. Difference in user cost between bus transit and the automobile,
2. One-way patronage on the Archer Avenue bus lines,
3. Bus capital costs for the Archer Avenue service, and
4. Bus operating costs for the Archer Avenue service.

The intersection of the axes defines existing costs and patronage of the Archer Avenue bus service.

An example of the use of Figure 2 would be tracing the impact of a \$1.00 increase in automobile user costs relative to bus fares. In the upper right quadrant of the figure, the impact of the cost change on patronage can be seen. Peak daily ridership would increase to around 9500 riders, and daily off-peak ridership would climb to about 11 500. At this level of peak-period ridership, additional buses would have to be obtained and bus capital costs would increase to approximately \$2650/d (found by tracing costs across Figure 2 from the upper right quadrant to the upper left quadrant). New operating costs of \$22 500/d can then be located in the lower left quadrant.

Capital costs of the Archer Avenue bus service vary directly with peak-period patronage because, at the time of the study, there was no excess peak capacity. The cost of additional garages or other related capital facilities for buses is not included in the calculation; if the number of buses added were small, there would probably be little effect on total garage requirements. This explains the linear capital-cost relation to patronage in the peak period. But operating costs are not linear, and the relation between operating costs and patronage shown in Figure 2 is kinked. This behavior is explained by the excess capacity available on the Archer Avenue bus lines in the off-peak period. As travelers are diverted to the bus mode, peak-period operating costs rise but new off-peak bus users are absorbed in the excess capacity. This continues until bus transit attains a cost advantage of approximately \$1.00 over the automobile. Beyond this cost advantage, operating costs for the Archer Avenue bus lines increase at a higher rate because of added off-peak service.

Figure 3 shows how the costs of the Archer Avenue service vary with the cost advantage achieved by the service. The capital costs of the required additional buses are shown to be relatively unimportant. Because of the operating characteristics these total costs are not linear but kinked. In plotting the revenue that would be obtained from the existing \$0.45 fare, Figure 3 also implies that the entire bus cost advantage is attributable to automobile cost penalties. Revenue climbs faster than total costs to the left of the point where capacity is exhausted in the off-peak period. As soon as added off-peak service must be provided, marginal cost exceeds an added fare. Revenue does not exceed costs at any point, and the service must always be subsidized. A \$0.75 fare is the minimum fare that would allow revenue to exceed costs, but this would occur only when an additional cost penalty of \$1.00/trip is applied to the automobile user.

Investment Options

The four capital investments proposed to improve the existing Archer Avenue bus service were evaluated with the model. The initial capital investment considered was an upgrading of bus service in the Chicago CBD that would separate bus operations from automobile and commercial traffic. For the running of the linear program, exclusive bus lanes were evaluated from 12th Street (Roosevelt Road) to Wacker Drive on the northern end of the CBD. This 2.17-km (1.35-mile) section within the Chicago CBD is congested and bus travel times are quite high. Scheduled bus travel times through this section are greater than 11 min during the morning peak period and more than 15 min in the peak direction during the more congested evening peak period (6).

The second investment alternative evaluated was an extension of the separate bus right-of-way from the end of the CBD bus lanes at Roosevelt Road to the vicinity of Halsted Street and Archer Avenue. One possible alignment for this extension would be along existing railroad right-of-way. This alternative would add around 3.2 km (2 miles) of exclusive bus right-of-way to the CBD bus lanes. It is assumed that the extension would offer a high level of service and that nonstop buses using it would travel at a top speed of around 80.5 to 88.5 km/h (50 to 55 mph).

The next alternative investigated concerned extending the busway of the second alternative to Pulaski Road. This segment, which could be constructed on available right-of-way in the median of the Stevenson Expressway, would lengthen the busway facility another 7.1 km (4.4

miles) west from Halsted Street. To maintain high speeds on this section of the busway no stops would be made. This would allow high-speed operation and improved service over the existing bus lines operated on the Stevenson Expressway.

The inability of the busway facility proposed in the two previous alternatives to serve trips to intermediate corridor destinations led to the development of the fourth and final alternative capital investment: an exclusive bus facility with stations or access points along its length. In coding this alternative into the linear program, it was assumed that stations would have to be located at approximately 0.8-km (0.5-mile) intervals to serve intermediate movements adequately. This alternative could be realized by (a) exclusive bus lanes on an existing street, (b) a low-design busway facility on which buses would pick up and discharge passengers while stopped, or (c) a higher design busway with stations separated from the through busway lanes to permit overtaking and passing. The third design would permit the facility to be used by the Archer Avenue expressway lines as well as the Archer Avenue limited-stop service. The bus facility would start at the end of the CBD bus lanes at Roosevelt Road and end at Pulaski Road. Given the range of design options, the speed of the buses on this alternative could vary substantially depending on the selected design. As a compromise, it is assumed in the linear program that average speeds of around 32.2 km/h

Figure 1. Peak-period bus routes, frequencies, and patronage in the southwest corridor.

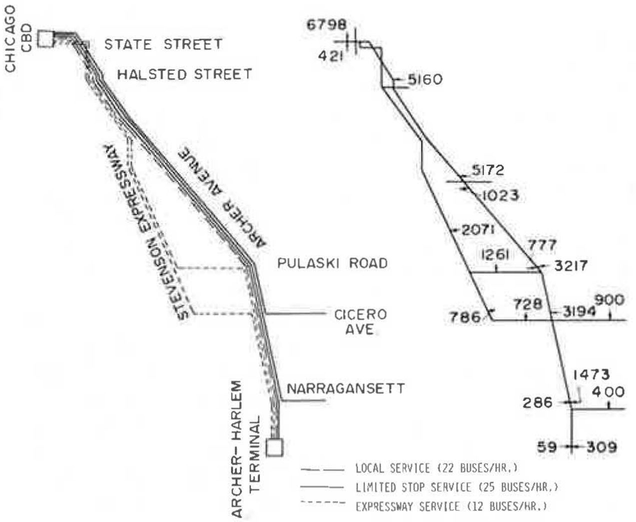


Figure 2. Bus patronage, operating costs, and capital costs versus automobile-cost disadvantage.

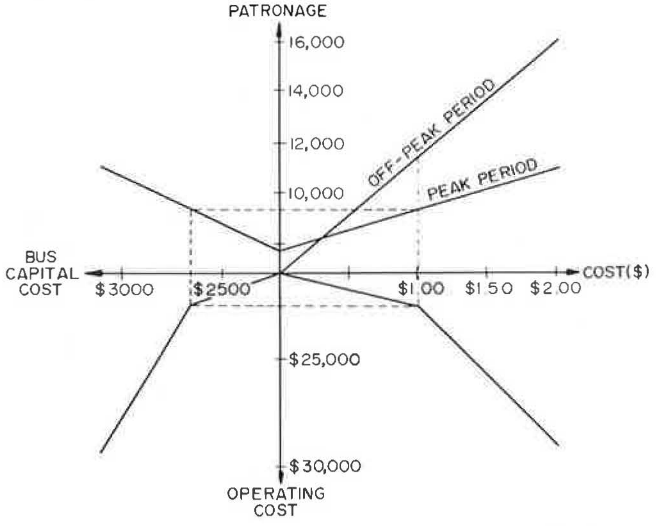


Figure 3. Daily costs and revenue of Archer Avenue bus lines.

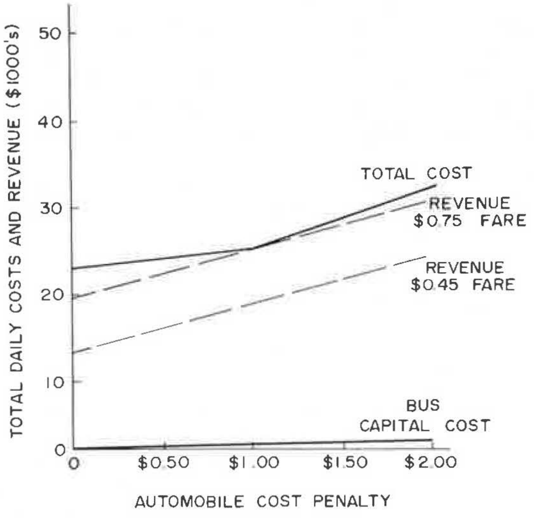


Table 2. Time savings for the four investment alternatives.

	Time Saved per Trip (min)							
	CBD Bus Lanes		Halsted Busway		Pulaski Busway		Pulaski Bus Lanes	
Trip Origin and Destination	Peak	Off Peak	Peak	Off Peak	Peak	Off Peak	Peak	Off Peak
Roosevelt Road to CBD	2.5	3.0	2.5	3.0	2.5	3.0	2.5	3.0
Halsted Street to CBD	2.5	3.0	7.5	7.5	7.5	7.5	5.0	4.0
Pulaski Road to CBD	2.5	3.0	8.0	7.5	12.5	7.5	2.5	3.0
Halsted Street to Roosevelt Road	0	0	5.0	4.5	5.0	4.5	2.5	1.0
Pulaski Road to Roosevelt Road	0	0	5.5	4.5	10.0	4.5	0	0
Pulaski Road to Halsted Street	0	0	0	0	4.0	0	3.5	2.0

Table 3. Patronage and costs of the investment alternatives for southwest corridor bus lines.

		Alternative			
Item	Existing	CBD Bus Lanes	Halsted Busway	Pulaski Busway	Pulaski Bus Lanes
Peak load-point patronage					
Peak	7900	7950	8050	8090	8005
Off Peak	6900	7050	7310	—	7110
Daily bus capital costs ($)	2200	2050	1910	1890	1965
Daily bus operating costs ($)					
Vehicle-kilometers					
Peak	2330	2340	2370	2375	2365
Off Peak	4670	4670	4670	—	4670
Labor					
Peak	4670	4350	4060	4035	4135
Off Peak	9330	8425	7510	—	8010

(20 mph) would be maintained over the length of the route.

The estimated travel-time savings of the four alternatives for selected movements in the southwest corridor are shown in Table 2. All users of the Archer Avenue bus lines do not benefit equally from the different improvements. Only trips to the CBD are able to use the CBD bus lanes. Trips with intermediate corridor destinations are not well served by the nonstop busway alternatives between Roosevelt Road and Halsted Street and Halsted Street and Pulaski Road. An exclusive bus facility with a number of intermediate stops does not improve travel times to the CBD for users at the western end of the corridor who use the expressway lines. However, a busway (or bus lanes) with stops along its length could benefit through bus travelers as well as users having trip ends in the middle section of the corridor.

The following table shows the total daily person hours of user time savings that would result from implementing each of the investment alternatives.

Alternative	Person Hours Saved	Alternative	Person Hours Saved
CBD bus lanes	1020	Pulaski busway	4050
Halsted busway	3400	Pulaski bus lanes	2105

The alternatives are ranked by their approximate investment cost, and the travel time saved rises regularly with the increased cost of the alternative. In terms of incremental travel time saved, the busway extension that runs from Roosevelt Road to Halsted Street is most beneficial, although the nonstop busway extension from Halsted Street to Pulaski Road saves less user travel time marginally than the lower cost investments.

Patronage, Capital Costs, and Operating Costs

The construction of any of the alternatives would increase patronage by reducing travel times. Table 3 gives the calculated increase in peak and off-peak patronage for each of the alternatives. The mode-split equations in the linear program that predict this patronage change are based on the CATS mode-split model (7), which indicates that the fraction of trips made by transit is not very sensitive to transit travel time in the existing range of times in the corridor or in the range of times considered as alternatives in this analysis. This insensitivity accounts for the small ridership increases computed for the alternatives by the corridor planning model.

However, increased patronage would affect the capital and operating costs of the Archer Avenue bus lines in two ways: (a) These costs for the Archer Avenue service would tend to increase because of the increased peak-period use because these lines presently operate near capacity in the peak period, and the impact of increased peak patronage would be additional runs with added equipment; and (b) the exclusive bus right-of-way alternatives would also tend to cause economies in the operating and capital costs of the bus lines because a reduced peak-period cycle time would permit buses to make additional runs in the peak period and reduce the total number of buses required, thereby reducing bus capital costs. Fewer buses in operation also means that fewer operators would be needed and the labor element of bus operating costs would decline.

Table 3 summarizes the operating and bus capital costs for the Archer Avenue service with each of the four alternatives in place. Patronage is given for the peak load point near Halsted Street. The shortened cycle time tends to decrease daily bus capital costs, but these costs do not decrease in direct proportion to the cycle time because of increased peak-period patronage, which creates a need for added bus trips. Vehicle-kilometers refers to those operating costs of a bus that vary with the kilometers operated and account for about one-third of total operating costs; these costs increase in direct proportion to patronage increases in the peak period. The labor component of operating costs varies with the number of buses required for service and changes at the same rate as bus capital costs.

Bus operating and capital cost savings shown in these tables reveal that, as the capital investment increases, incremental cost savings generally decrease. The highest cost alternative, the western extension of the nonstop busway from Halsted Street to Pulaski Road, achieves little bus operating and capital cost savings over a nonstop busway terminating at Halsted Street because the western extension is only suitable for use by the peak-period Archer Avenue expressway bus lines. The three lower cost investments that can be used by all Archer Avenue bus lines show more regular benefits with increased investment.

CONCLUSIONS

The corridor model developed at the University of Pennsylvania was designed to aid transportation planners in the short-range or implementation planning stage of the planning process for improvements in major transportation corridors. The model therefore considers as alternatives not only the construction of new facilities but also pricing options for public and private transportation. The linear program treats these alternatives parametrically, by using time and cost variables, and can consider a wide range of design and operation policies. The technique helps to overcome one of the major weaknesses of traditional transportation planning—the limited number of plans that can be considered by the traditional model system because of time and money requirements.

The application also indicates that the corridor model is substantially operational. Although the mathematical formulation used in this paper is considerably simplified, it is clear that the general mathematical programming approach has advantages in transportation planning that is subregional in character and less detailed than route location planning. The present linear program form is only a starting point, however, for the development of suitable methods to fill that gap.

ACKNOWLEDGMENTS

The work presented in this paper was supported by the Program of University Research, Office of the Secretary of Transportation, U.S. Department of Transportation, and by the Illinois Department of Transportation. The support of these sponsors is gratefully acknowledged and implies no endorsement of the findings.

REFERENCES

1. E. K. Morlok and others. Development and Testing of a Transportation Planning Model for High Accessibility Urban Corridors. Univ. of Pennsylvania, Philadelphia, Final Rept., Vols. 1 and 2 and Appendixes, 1975.
2. Current Mass Transit Usage. Chicago Area Transportation Study, 1975.
3. CTA Bus and Rapid Transit Operating Facts, Winter 1973-74. Chicago Transit Authority, Document RP-Y74508, 1974.
4. Bus and Rail Operating Cost Per Vehicle-Mile, Fiscal Year 1974. Chicago Transit Authority, Document OP-X75091, 1974.

5. Capital Development Program—Phase II. Chicago Transit Authority, 1974.
6. Archer Avenue Local, Limited, and Expressway Employee Timetables. Chicago Transit Authority, 1974.
7. E. V. Ryan. Calibration of the Network Sensitive Mode Split Model. Chicago Area Transportation Study Research News, Vol. 16, No. 3, 1974.

Publication of this paper sponsored by Committee on Transportation Systems Design.

Comparing Modes in Urban Transportation

Slobodan Mitric, Department of Civil Engineering, Ohio State University

Modal comparisons are defined as those studies in which an analyst compares urban transport modes with each other in a generalized framework, attempting to assess relative advantages and disadvantages of modes under a variety of conditions. This paper establishes a link between comparative analyses of transport modes and urban planning processes and generates a basis both for a normative theory of modal comparisons and for a critique of existing works in this field.

An ongoing debate in the field of urban transportation planning revolves around comparative advantages and disadvantages of various transportation modes. It often takes the form of polemics, such as bus versus rail. Considering the variety of conflicting positions and the potential impact of the conflict on policies and investment decisions, a methodological study of this debate is long overdue. The main purpose of this paper, which is a summary of a larger report (1), is to establish a link between comparative analyses of transportation modes and urban planning processes and thus generate a basis both for a normative theory of modal comparisons and for a critique of existing works in this field.

A transportation mode is initially defined as a particular combination of transportation-related structures, vehicles, and strategies of operation. Within the broad setting of urban transportation planning, modes are usually compared in the following specific contexts:

1. When a planner deliberates which modes to include as components of alternative plans for a given urban area (this context will be called a site-specific design of alternatives);
2. When a decision-making body evaluates a set of alternative transportation plans for a given urban area (this activity is expected to end up with a decision or at least a recommendation and will be called site-specific evaluation); and
3. When an analyst compares modes with each other in a general fashion, attempting to determine conditions under which a particular mode is in some sense better than others or to arrive at rankings of several modes under a variety of conditions (studies of this type, which will be called modal comparisons, are usually not site specific although they sometimes make use of data from a single site).

It is customary to analyze decision processes by breaking them down into activities such as clarification of goals, design of alternatives, evaluation, and action. Such activities take place both in the site-specific, urban planning context and in the context of modal comparison; in fact, alternatives considered in these two contexts are similar. Both exercises involve evaluations using similar criteria, and both end with expressions of preference. Nevertheless, they differ in scale and in depth and should not be confused with each other. They also serve different purposes, by answering similar questions from different questioners. Perhaps the most significant difference between them is that modal comparisons arrive at expressions of preference for transportation modes through a technical process and site-specific evaluations arrive at these preferences through a political process.

Modes are what transport plans are made of. The site-specific planner faces numerous possible combinations of transportation structures, vehicles, and operational strategies and, because of time and money limitations, can consider only a few of these combinations in depth. The task will be made easier if he or she is provided with modal descriptions that enable the planner to screen many alternatives quickly and select the few that are promising in a specific context. Comparisons, or descriptions that bring out similarities and differences between the things compared, are well suited for this purpose.

Although the site-specific planner cannot evaluate all modes, somebody must. The design of alternatives for a modal comparison should therefore be based on a structured, exhaustive classification of modes. No single exercise can be expected to compare all, or even many, modes, but it should sample the set of modes in a systematic manner.

Alternatives in transportation planning are evaluated on the basis of their service characteristics, costs, and external (nontransport) effects. There are many ways to select and organize this type of information, including making judgments about which parameters to include, exclude, stress, or deemphasize; choosing between aggregate measures or distributions; and exercising a preference for quantitative or qualitative information. If evaluation criteria used in modal comparisons are to be useful, they should broadly correspond to criteria used in the site-specific decision process.

Evaluation criteria in urban transportation planning have changed substantially in the past 20 years in both theory and practice, reflecting changes in planners' perceptions of what constitutes the transportation problem. Until the mid-1960s, the prevalent view of urban transportation was that of a closed, functional system designed to achieve narrow but precise objectives. Then, as a result of the revolt against freeways and the general increased awareness of environmental ills, urban transportation was recognized as an open system with far-reaching economic and political consequences. Transportation planners became involved with external effects of their alternatives in the areas of distribution and environment. The role of technicians, who (by virtue of recommending the "best" plan) were once the sole planners and decision makers, was diminished by legal requirements for citizen participation in all activities of the planning process. In the emerging planning process, technicians would be producers of facts and alternatives and spokespersons for the unrepresented public interest, but they "must not be the focus of making recommendations" (2). Having come far from its early concern with benefit-cost ratios, evaluation theory now recognized the importance not only of decision outcomes but also of decision processes (3).

The new planning process, here called the open process, differs from the old engineering-economic model of decision making in the following major ways:

1. Functions of the planner and the decision maker are vested in different people.
2. The decision-making body consists of a number of groups that may espouse different value sets and thus different evaluation criteria.
3. Fragmentation of the role of the decision maker rules out any attempt by the planner to propose an optimal solution. The evaluation process is political, and the decision that is eventually reached is a political compromise.
4. Engineering-economic decisions were based on aggregate impacts; open evaluation is based on trade-offs.
5. There is no unified set of goals to guide the planner in the design of alternatives. The final goal set is the product of the evaluation process.
6. To inform and broaden the political debate, the planner deliberately designs alternatives to suit competing goal sets and presents them in a fashion that makes diverse trade-offs explicit.
7. It is understood that transportation projects can be used to achieve nontransportation—i.e., developmental, economic, political, environmental—ends.
8. Engineering-economic decision models required that evaluation criteria be quantifiable and commensurate. These conditions are dropped in the open processes, and the result is fuller descriptions of alternatives (and the risk of overloading decision makers with information).

These characteristics of open decision-making processes are significant for evaluation criteria used in modal comparisons. When site-specific criteria are complex in number and kind, modal-comparison criteria must also be complex. If there is no best alternative in a site-specific context, there is even less chance of one in a generalized comparison. Therefore, the goal of modal comparisons is describing modes in a manner that illuminates the functional, economic, environmental, and aesthetic trade-offs they offer.

EXISTING MODAL COMPARISONS

The archetype of all modal comparisons is that of Meyer, Kain, and Wohl in their study of urban transport (4). The main features of this generalized comparison of automobile with bus and rail rapid transit are as follows:

1. The environment studied is a single suburb-to-downtown corridor during peak hours. Passenger volumes are given and uniformly distributed along the corridors, and almost all traffic is assumed to be downtown oriented.
2. Service standards are developed based on walking distance, waiting time, in-vehicle travel speed, and seating area in the vehicle (substituted for comfort). Because the aim of the study was to examine the case for transit versus the automobile, values for each element are based on what suburban drivers are presumed to expect of transportation services.
3. An automobile system and two classes of transit alternatives—bus transit, including local and other express buses, and rail rapid transit—are designed for the corridor to satisfy the adopted service standards.
4. Conclusions about the relative worth of modes are based on average origin-to-destination cost per seat trip, agency costs for transit modes, and agency plus automobile ownership and operating costs for the automobile alternative.
5. It is concluded that the economic case for transit can be made at one-way, peak-hour design volumes greater than 5000 passengers (at the maximum load point).
6. The comparison between bus-based and rail-based transit alternatives is favorable to the bus. At medium residential densities, bus transit is significantly cheaper than rail for all design passenger volumes. At high densities, the lowest cost curves for bus and rail coincide for all practical purposes.

Several events in the early 1970s made modal comparisons an important field of study. The first new regional rail system in the United States since World War II—the San Francisco Bay Area Rapid Transit System (BART)—was completed after a decade of troubled efforts. A similar system in Washington, D.C. (Metro) was experiencing similar difficulties. During the same period significant federal funds in the form of capital grants became available to urban public transportation projects. Cities such as Atlanta, Buffalo, and Baltimore, which proposed to build rail rapid transit systems, applied for the largest capital grants. Even sprawling Los Angeles had a brief, unsuccessful encounter with rail rapid transit. Significantly, many new modal comparisons coincided with or followed these events. Three large, recent studies are discussed here.

A study by Boyd, Asher, and Wetzler (5) compares three alternatives: (a) rail rapid transit, (b) express buses operating on arterial streets only, and (c) express buses operating on arterial streets during collection and distribution and using an exclusive busway for line-haul. The third alternative is referred to as integrated bus.

The Boyd, Asher, and Wetzler study follows the method of Meyer, Kain, and Wohl but with the following important differences.

1. The restrictive assumption of equal service is dropped. Alternatives are designed to provide different types of service, and these differences are reflected in door-to-door travel times.
2. Alternatives are compared in terms of generalized costs, which consist of agency costs plus time costs of travelers.

The overall conclusion is that, under the study conditions, bus systems have lower generalized costs than rail rapid transit systems.

A study by Bhatt (6) compares 16 modal alternatives, 14 of which are bus and rail systems. The alternatives differ mainly in their method of collection and distribution. The general method of Meyer, Kain, and Wohl is followed but the assumption of equal service is dropped. Bhatt does not follow the Boyd, Asher, and Wetzler method of converting time into equivalent doller costs; instead, results are presented in both cost and time dimensions. The study findings favor bus-based alternatives.

The most technically ambitious effort to date to develop intermodal cost comparisons and draw policy implications was made by Keeler and others (7). [Pozdena's study (8) was done as part of the same project.] Most of their data are site specific and are taken from BART, bus properties, and the highway system in the San Francisco Bay Area. Marginal cost pricing is explicitly introduced (e.g., by charging drivers the marginal congestion costs). The work will be of special significance in transit cost modeling.

In their modal comparison Keeler and others follow in the footsteps of Boyd, Asher, and Wetzler. But they differ in the greater econometric sophistication and the local origin of their data, as well as in the close scrutiny they give to the automobile mode. Their results show that the bus-based system has lower generalized costs than rail rapid transit for all study conditions and lower costs than the private automobile for all but the lowest design volumes. The study concludes, among other things, that BART should never have been built.

All four studies, but especially the last three, appear to prove that urban rail transit has no future. These findings contradict those of a number of site-specific studies in which consultants or local planners recommend rail transit. The four studies reviewed here and similar studies will be referred to in the following methodological analysis as economic modal comparisons because the majority of their authors are economists who emphasize cost analyses while assuming a given demand.

METHODOLOGICAL ANALYSIS

As stated above, the nature of site-specific decision processes in urban transportation determines the nature of the modal comparisons that attempt to inform these processes. If modal comparisons are to inform open transport planning, they should underscore the many differences and similarities (trade-offs) among modes. This analysis of economic modal comparisons focuses on (a) perception of the problem (or goal clarification), (b) approach to evaluation of modal comparisons, and (c) design of alternatives.

Perception of the Problem

The conclusions of modal-comparison studies reveal a tendency to structure comparisons toward picking a winner among modes. For example, Bhatt (6) states: "High performance exclusive busways require substantial investment but are less costly and faster than rail rapid transit in almost all environments and volume levels." This implies that the role of modal comparisons is to help the site-specific planner by eliminating some alternatives. The site-specific design of alternatives would be greatly simplified if such conclusions were considered to be true. The planner could eliminate rail rapid transit and turn all attention to the various bus-based alternatives. Uncertainties facing vehicle manufacturers would disappear, bus producers would enjoy sizable economies of scale, and so on.

What the planner actually needs is information on trade-offs; the above findings offer none (none, at least, between bus and rail rapid transit). It appears that economic modal comparisons attempt to preempt the roles of site-specific planners and decision makers by decreasing rather than enriching their decision agendas. In other words, they solve the wrong problem.

Approach to Evaluation

The belief that such large issues as the elimination of rail transit can be resolved in a generalized study is apparently based on authors' certainty about the correctness of their decision model (9): "The techniques used in this report can be applied in different communities to evaluate economic cost of proposed transportation alternatives, thus providing a basic economic foundation for recommendations." In effect, this is a variant of the site-specific decision-making process, but the site is a flat, featureless plain peopled by commuters who make modal choices with textbook rationality. The community is so homogeneous that there exists a well-defined welfare function. The transport system is an exact replica of a single suburb-to-downtown corridor. The costs are cross-sectional averages, or very particular site-specific cases. Of course the result of these efforts is the best solution, determined under monolithic conditions. The features of the engineering-economic model of decision making are easily recognizable here.

Meyer, Kain, and Wohl wrote their book at the time when these concerns were yet to be strongly articulated among transport professionals. Their decision model is very simple. The client body communicates its uniform service standards to the planner, who then designs to meet the specified standards and selects the design that minimizes total cost. That is, the planner is also the decision maker. The environment of the system enters the model through the description of the corridor (e.g., length, population density) and through the assumption of given demand (presumably derived from a land-use forecast). All cost estimates are cast in a deterministic form requiring literally dozens of assumptions. Transport alternatives are unchanging, as are values. Obviously, this model, which is a prime example of an early engineering-economic model, imposes iron restraints on the design of alternatives. By adopting a service standard for speed, for example, it biases the outcome against a mode that, all other things being equal, could offer a higher speed.

The new wave of economic modal comparisons recognized this difficulty and achieved an improvement by designing alternatives for different service characteristics. But those characteristics must still be translated into travel-time scores. Although the client body may now trade time for money in the model, all travelers must value time at a uniform rate of X dollars per hour where X is taken from studies of current modal choices. By using this rate the aggregate time score of an alternative is converted into dollars and added to capital and operating costs to obtain a generalized cost figure—"a single, comparable datum" (8).

The concept of generalized costs suffers from a number of problems, among them the problem of completeness. Generalized costs are supposed to measure both costs and service characteristics of an alternative but, as long as the average components of travel time are the only aspects of service represented, generalized costs would systematically underrate those alternatives whose advantages lie in other service areas. In other words, the technique is blind to such service measures as safety, reliability, and comfort. This bias would be particularly strong against alternatives that operate on an exclusive right-of-way (rapid transit) or that have elaborate, costly safety devices and practices (rail rapid transit). All service characteristics important to individual modal choice and to the needs of society should be incorporated into generalized costs if this technique is to be a useful tool for modal comparisons. Of course, there are difficulties in measurement and interpretation of measurements, e.g., whether to measure characteristics or perceptions of characteristics. Some concepts are too complex to be captured by a single quantitative measure; this includes even valuation of time, the area in which measurement has progressed the farthest. There is and always ought to be a place and a need for qualitative statements.

Generalized Cost and Individual Modal Choice

In some of the later economic modal comparisons the assumption of equal service (4) was replaced by "equal shadow price of travel time." Both assumptions imply, the second one more weakly, that there is a direct correspondence between group standards, which are represented by the shadow price, and individual preferences for transportation services. Thus, if the client body communicated to the planner that a shadow price of X dollars per hour of travel time should be assumed, then when a system was actually built travelers would be observed using that travel-time value as if they indeed valued time at that rate. It is known from many studies of travelers' preferences that a value put on a service characteristic is not a unique number but a distribution that depends on such things as taste, income, and trip purpose. The assumption of equal valuation is convenient in that it seems to circumvent the need for an explicit model of modal choice; that is, passenger attraction need not be estimated in a modal comparison. Thus, the concept of given demand is implicitly endorsed.

Relation Between Individual and Social Choice

In some aspects of decision making a group may purposely choose a standard different from that of many (or any) of its individual members. For example, empirical research may show that safety plays no role in travelers' choice of mode, that it is implicitly valued at zero, and yet a decision-making body may choose to place a high value on safety and invest accordingly. On the other hand, a group may value some characteristic less than individuals do. Standard practice in economic modal comparison has been to value walking and waiting time three times higher than in-vehicle time (this 3:1 ratio has frequently been observed in actual modal choices). This implies that the opportunity cost to society of waiting time is three times the cost of in-vehicle time. Some modal comparisons are particularly sensitive to this assumption, especially comparisons between systems requiring feeders, transfers, or integrated lines.

Generalized cost, as it is used in economic modal comparisons, is therefore an incomplete and limited measure of service. It neither adequately replaces an analysis of passenger attraction nor reflects group valuation when that valuation differs from individual valuations. Individual transit users, interest groups, local government, and transit operators all have their distinct points of view, and modes cannot be meaningfully compared unless the point of view is specified. Unfortunately, that is not possible in a generalized modal comparison, at least not by means of an analytical approach.

Design of Alternatives

Modal comparisons have paid surprisingly little attention to what constitutes a mode and have not attempted to disentangle relations between the input and output (cost and service) characteristics of alternatives. Indeed, because modal comparisons identify a mode with a particular vehicle technology, as represented by some typical design arrangements, it appears that vehicle technologies, and not modes, are being compared (e.g., rail and bus).

There is some diversity in bus-technology alternatives, particularly in Meyer, Kain, and Wohl, but the typical system based on rail technology is almost always rail rapid transit. Costs for this typical system are usually borrowed from BART or Metro. The impression is given, in fact, that rail transit equals BART or Metro and vice versa. If a study shows that BART is more expensive than a number of bus-based alternatives, a subtle cost generalization is made over all rail-based designs.

Important issues are implicit in the way in which economic modal comparisons select and characterize alternatives and in the conclusions they draw. These issues are discussed below. (BART is frequently used as an example only because it is the best-known, new, large-scale transit system in this country.)

Mode Concept

There is substantial agreement among transportation engineers and planners that a mode should not be defined according to its vehicle technology. A morphological concept of mode connects the portions of service (output) space with pertinent characteristics of inputs such as way, vehicles, and rules of operation (10). Whether these connections (mode classifications) are made coarse or fine grained depends on the purpose of the exercise. For example, by using the following three-way classification a mode could be conceived as a large subsystem of an urban transport system:

1. Degree of exclusivity of right-of-way (e.g., entirely exclusive, partially shared, fully shared);
2. Technology class (type of guidance, vehicle size, dynamic properties, fuel consumptions); and
3. Operational strategy (express, local, or skid-stop; single-unit or train operation; strategy for fare collection; safety procedures).

The virtues of a morphological approach for the systematic exploration of all alternatives in a specific context are well known (11). Detailed accounts of its application to transportation modes are also available (10, 12). What is important here is that the degree of exclusivity of right-of-way, and not the technology class, is the most important determinant of service output. An exclusive right-of-way offers designers and managers the potential to maximize the overall efficiency of transit while emphasizing reliability and safety. This characteristic largely determines the cost of a mode. As discussed above, economic modal comparisons note the high costs but not the corresponding benefits.

In the morphological view of urban transport modes, BART is a regional transit system that operates on a fully exclusive right-of-way and uses rail technology. This same technology can be used for a whole range of modes, some considerably cheaper (light rail, for example) and others conceivably more expensive. A BART-type system could also be less expensive depending on site-specific conditions. It is possible, with numerous advantages and disadvantages, to use bus technology for such a system. Unfortunately, economic modal comparisons do not note these trade-offs. In drawing conclusions, modal comparisons emphasize vehicle technologies, and yet technological aspects are almost totally absent from the analyses. The absence of a clear concept of mode prevents them from making a systematic selection of alternatives. An example is the mismatch that results when BART is compared with freeway flyers.

Technology and Cost

BART is expensive only partly because it uses rail technology. The major share of BART capital expenses can be attributed to such factors as exclusive right-of-way, extensive tunneling, elevated structures, the underwater tube, and lengthy delays in construction. Problems with rolling stock have partly resulted from trying to introduce too many innovations simultaneously. Labor agreements have also had a complicated impact on BART operating costs.

The historic correlation between rail and underground operation is strong. But, in modal comparisons, the expense of the so-called rail alternatives is in great part due to an erroneous identification of rail technology with tunnels. When Keeler and others (7) say it makes no economic sense to build another BART, they are actually saying that it makes no economic sense to dig tunnels and construct underwater tubes. Authors of all recent modal comparisons follow this practice, in spite of a clarifying study by Deen and James (13).

Dealing with costs and other historical data is a complicated matter, especially when cross-sectional data are used to derive averages. One reason is that designs for systems with similar functional characteristics can run from the spartan to the luxurious (e.g., the cost of stations). Another reason is the potential for a learning-curve effect in constructing successive versions of a system or a vehicle.

It is one of the purposes of modal comparisons to inform the site-specific planner about the consequences of choices. To achieve this purpose, historical correlations between right-of-way and costs, technologies and costs, or operational strategies and costs should be examined for causal chains. The subject is a sensitive one requiring substantial research (14).

Another controversial topic is that of the propriety of assigning all costs of a given system to its functional purpose. During the construction of BART, citizens of Berkeley went to court and forced a section of the system that was to pass through the city to be located underground (14). As a result, BART registered a cost increase attributed to environmental considerations. Keeler and others included this item and many like it in their total generalized costs (7). This type of expense ultimately became a part of the cost per passenger.

Some BART stations appear, at least to some people, to be quite lavish. Many rapid transit systems around the world share this characteristic. Although costs may sharply increase because of these embellishments, there is no corresponding increase in performance, especially none measured by generalized costs.

Keeler and others compare freeway-flyer buses with BART on the basis of generalized costs. Historically, the former alternative could have been gradually introduced in the San Francisco Bay Area without any new construction, new route by new route, by purchasing new buses as the patronage increased or by attempting to modestly stimulate patronage. An engineering-economic model could have been used to design and evaluate the additions. If the whole project or some part of it did not work well, one would at worst have some buses to sell. BART, however, is an alternative of a different nature. It is primarily supposed to carry people within the region, but it is supposed to do much more. Correctly or not, it is expected to stimulate a change of activity patterns, even life-styles, in the whole region. The former alternative is an incremental one designed to follow land-use development and observed user preferences. BART is a "big leap," designed to shape activity patterns and change user preferences for transport, and more.

It is clear from these examples that the major tool of economic modal comparisons, generalized costs, is particularly inadequate in the presence of externalities and multiple purposes, especially when such impacts are so large that they overshadow the functional impact of an alternative.

Comparison of Alternatives

On the basis of the previous discussion, comparisons of dissimilar alternatives appear to be full of pitfalls when a single, limited criterion of evaluation is used. The troubles start, however, when an analyst is not deductively aware of the difference in the alternatives. This creates distortions both in the selection and characterization of alternatives and in the choice of evaluation criteria.

The morphological approach to the concept of mode would help in such cases by systematically organizing all alternatives on the basis of a selected set of parameters and the associated scales. Nearness in morphological space (or planes) indicates similarity between alternatives and suggests the proper evaluation criteria. For example, given a fully exclusive way and bus technology, an effective comparison could be made among all operating strategies. Given the current debate about technologies, it might be a good strategy to make many comparisons within the same vehicle-technology groups. Lehner (15) gives an example of this strategy in his comparison of light rail and rail rapid transit.

Comparisons based on the morphological approach would not be global but partial and significantly deeper. Drawing samples from the entire population of modes would aid in the design of coherent research programs that avoid excessive overlapping.

Other Issues

It was pointed out earlier that modal comparisons tend to (a) concentrate on a single downtown-oriented corridor having an insignificant amount of local travel, and (b) treat only peak journeys to work and charge most (or all) capital costs of alternatives to peak use. A few brief comments are warranted.

1. Downtown-oriented corridors that serve no local travel are not the only situation encountered in our cities; neither do they have a special claim on the future. Of course, in comparing modes in an environment of heavy local travel, technological details such as the width of doors, prepayment of tickets, and overloading potential become especially important.
2. To the best of my knowledge, not one economic modal comparison has examined the integration of corridors into a system, especially in the context of transit [although Pozdena (8) did make a start]. This has resulted, among other things, in transfers being treated as an inconvenience rather than an efficient way of connecting zones that lack a direct corridor.
3. The concentration on peak travel to work, including charging capital costs mostly to peak use, and the almost complete absence of references to off-peak travel reveal an underlying assumption that transit exists only for peak journeys to work. Meyer, Kain, and Wohl (4) argue that off-peak volumes are so low that the automobile's advantages multiply "because the avoidance of discomfort, inconvenience, and other travel conditions seem to be more important to off-peak than peak travelers." This is a common way to look at transit, especially if one assumes multiple-car families, fixed values, and short-range planning. Meyer, Kain, and Wohl are quick to point out that new ways of operating old technologies

should be considered; yet they stop short of applying the same wisdom to new values and differently organized cities. Their concepts of equal service and the bedroom corridor are unquestioningly projected into the future. Nevertheless, values change and cities change. Alternative philosophies of transit, assuming different values and transformed cities, should find their way into modal comparisons.

SUMMARY AND RECOMMENDATIONS

The main problems with economic modal comparisons can be summarized as follows:

1. Modal comparisons are not based on a clear understanding of the decision processes in urban transportation but, by implication, on an outdated, engineering-economic model of decision making having well-defined, commensurate goals and guided by criteria of aggregate efficiency.
2. They are oriented toward single-answer, global comparisons that emulate site-specific planning based on the engineering-economic model.
3. When comparisons are global but the site is not specific, there is a need to make numerous assumptions (including idealized environments) and to use average data. These factors frequently affect findings more than do the characteristics of the alternatives studied.
4. Modal comparisons make static, undifferentiated assumptions about individual and social travel preferences and future urban patterns. Their assumption of given demand eliminates the essential aspect of transportation alternatives, which is the comparative ability to attract passengers.
5. The criterion of evaluation, generalized cost, cannot account for some important service characteristics of alternatives nor reflect important externalities. It does not allow for the fact that site-specific evaluation deals with multipurpose projects involving the client, the planner, and the decision maker and that these entities frequently differ about goals and evaluation criteria.
6. Because the selection of alternatives is not based on a clear concept of mode, it is not systematic. Alternatives are often dissimilar in a manner that cannot be measured by the adopted technique; some differences are purely functional, others stem from external effects. Some alternatives represent incremental changes to the current transportation system, and others represent drastic departures.
7. Conclusions of economic modal comparisons are stated in terms of vehicle-technology groups, such as bus and rail. Such conclusions stir unproductive controversy and divert the attention of site-specific planners from more important questions.

It is recommended that economic modal comparisons of the type discussed in this paper be abandoned. To be useful, modal comparisons should

1. Recognize the multiplicity of interests and values among urban residents and that planning requires both projection and vision,
2. Recognize the difference between incremental and large-scale changes in urban transportation and the corresponding difference in evaluation criteria,
3. Recognize that modal comparisons serve to inform site-specific planners about service and cost tradeoffs related to alternative transport designs,
4. Use the morphological concept of mode as a basis for selecting the alternatives to be compared and the evaluation criteria,
5. Allow greater depth of analysis by comparing alternatives with incremental differences in characteristics and costs and thus require the analyst to scrutinize the data and build detailed causal chains between input characteristics, costs, services, and travelers' reactions to these factors, and
6. Leave global, single-answer studies behind and instead perform many partial comparisons (if there is no best plan under site-specific conditions, there can hardly be one under generalized conditions).

ACKNOWLEDGMENTS

The research on which this paper is based was supported in part by a grant from the graduate school of the Ohio State University. I am indebted to David Hulefeld and Joan Orosz for helping with the project and to Wolfgang Homburger for his valuable comments on an earlier draft of the paper. Responsibility for statements made here is mine.

REFERENCES

1. S. Mitric. Comparisons of Modes in Urban Transport: A Methodological Analysis. Department of Civil Engineering, Ohio State Univ., Columbus, Technical Rept., 1976.
2. W. Hansen and S. Lockwood. Metropolitan Transportation Planning: Process Reform. TRB, Transportation Research Record 582, 1976, pp. 1-13.
3. B. Hudson, M. Wachs, and J. Schofer. Local Impact Evaluation in the Design of Large-Scale Urban Systems. Journal of American Institute of Planners, Vol. 40, No. 4, July 1974.
4. J. Meyer, J. Kain, and M. Wohl. The Urban Transportation Problem. Harvard Univ. Press, Cambridge, Mass., 1966.
5. J. H. Boyd, N. Asher, and E. Wetzler. Evaluation of Rail Rapid Transit and Express Bus Service in the Urban Commuter Market. Institute for Defense Analyses, Arlington, Va.; U.S. Department of Transportation, final rept., 1973.
6. K. Bhatt. A Comparative Analysis of Urban Transportation Costs. Urban Institute, Washington, D.C., Working Paper 5002-5, 1975.
7. T. Keeler and others. Automobile Costs and Final Intermodal Cost Comparisons. In The Full Costs of Urban Transport, Institute of Urban and Regional Development, Univ. of California, Berkeley, Monograph 21, Part 3, June 1975.
8. R. J. Pozdena. A Methodology for Selecting Urban Transportation Projects. Institute of Urban and Regional Development, Univ. of California, Berkeley, Monograph 22, July 1975.
9. M. F. Reed, Jr. Comparison of Urban Travel Economic Costs. Highway Users Federation, Washington, D.C., Technical Study Memorandum 6, 1973.
10. V. Vuchic. Comparative Analysis and Selection of Transit Modes. TRB, Transportation Research Record 559, 1976, pp. 51-62.
11. E. Jantsch. Technological Forecasting in Perspective. Organization for Economic Cooperation and Development, Paris, 1967.
12. V. Vuchic. Place of Light Rail Transit in the Family of Transit Modes. TRB, Special Rept. 161, 1975, pp. 62-76.
13. T. Deen and D. James. Relative Costs of Bus and Rail Transit Systems. HRB, Highway Research Record 293, 1969, pp. 33-53.
14. S. Zwerling. Mass Transit and the Politics of Technology. Praeger, New York, 1974.
15. F. Lehner. Light Rail and Rapid Transit. TRB, Special Rept. 161, 1975, pp. 37-49.

Publication of this paper sponsored by Committee on Transportation Systems Design.

Analysis of Integrated Urban Public Transportation Systems

James H. Batchelder, Multisystems, Inc.
Brian C. Kullman, Cambridge Systematics, Inc.

Dramatic increases in transit patronage will require a major restructuring of present transit and paratransit operations to achieve integrated regional systems capable of changing as traffic increases and new markets are penetrated. The integration of new public transportation options such as dial-a-ride, jitney, and subscription bus with conventional mass transit promises significantly improved overall levels of service without increased total system costs. Integrated systems and expansion policies requires that the individual service and cost attributes of each system component must be modeled and the synergisms that result from various service combinations must be evaluated. Integrated system design is significantly more complex than the conventional bus routing and scheduling problem because of the increased number and complexity of available modes. This paper examines a case in which various service policies are evaluated, for parametrically varied demand levels, by using a combination of manual and automated procedures. Major conclusions are that significant economies of scale develop at relatively low levels of increased transit use and that major redesign of system operating policies is required to sustain desirable service levels and costs.

In the last 3 decades America has been sufficiently affluent to absorb the expense of using low-occupancy, small-capacity private vehicles to provide most urban transportation passenger services. In recent years, however, attention has been focused on shortcomings of this policy, such as pollution, congestion, decay of the central cities, lack of mobility for certain segments of society, and inefficient use of energy. As a result, interest in alternative policies has developed. A promising option is the expansion and integration of transit service in a form suitable to the new multinuclear urban environment.

Because different transit modes and operating policies are most advantageous under different conditions, an integrated regional system would consist of a variety of modes, each operating in its appropriate environment. There would be the kind of coordination among these various transit-service components that does not exist in current systems, and it would result in increased efficiency, service, and patronage.

An integrated regional system would be able to respond to both space and time changes in travel volume and patterns. Transit services in different parts of the region would change during the day in response to peak and off-peak travel and over the years in response to urban development and transportation policies. Current institutional barriers to the coordination of system components and to their operational responsiveness would be removed as part of the concept of integrated transit. Economies of scale in some parts of the system (derived from patronage increases) could benefit other parts and thereby increase the range of economically feasible components and enrich the total transit system as it evolves.

Recent studies have investigated the demand for transit service (1), changes in the urban environment and in travel patterns (2), and the operational characteristics of various modes (3), all of which are pertinent to an analysis of expanded and integrated transit. The following key issues, however, have not been fully addressed.

1. What is the full potential of integrated transit for offering high-quality, low-cost service?
2. Given an understanding of its potential, is integrated transit a cost-effective means of meeting the mobility and development goals of a region?

In an effort to fill this void, the U.S. Department of Transportation (DOT) has sponsored a study (4) that attempts to provide preliminary insights into these issues and to develop the basis of a sound methodology for exploring issues in policy and planning analyses.

MODELING APPROACH

In developing a set of models to analyze the impacts of a major diversion, the following criteria had to be considered:

1. Do the models reflect the reality and complexity of an urban environment?
2. Are they adaptable for the analysis of diverse urban areas?
3. Do they respond to changes in policy?
4. Do they provide the analyst with useful information for evaluation?

A typical urban area has a varied distribution of population, employment and activity centers, and transportation facilities, all of which have evolved in response to topography and changing social, economic, political, and technological forces (Figure 1). In order to capture this diversity and inject realism into the analysis, the travel patterns and street network of Rochester, New York, a medium-sized urban area, were used in the analysis. Rochester is typical of many American cities in that it has major employment centers outside the core, topographical constraints, and varying population patterns and highway development (the models developed are applicable to any urban area).

Although the study was a macroanalysis, a hierarchy of 135 zones, 32 districts, variable subregions, and 5 rings was developed to reflect the complexities of trip volumes and patterns. Zones, for example, were used to define regional networks and assign trips. Sample districts created from these zones were used in the analysis of local transit-service options. Rings and subregions were used to present results of aggregate market responses, such as the service levels provided to suburb-to-CBD transit patrons.

The trip data used in the analysis were based on 1970 peak and off-peak volumes and peak modal-split values provided by the New York State Department of Transportation. The purpose of the study was to analyze transit operations as the daily regional modal split increased from 5 to 60 percent. Because the scope of the study precluded supply and demand equilibria, transit patronage was varied parametrically over the 5 to 60 percent range to generate the four study cases shown below.

Case	Regional Modal Split (%) Daily	Peak	Off-Peak	Transit Peaking Ratio
1	5	10	3.5	3.6:1
2	15	25	10	3.2:1
3	25	40	20	2.6:1
4	60	77	54	1.9:1

As the regional modal split was increased, it was forecast that growth would occur nonuniformly; i.e., some markets would experience earlier or more rapid growth in transit ridership than others, primarily because of the relative ease of improving service in markets already served by transit. To prepare trip data for the cases given in the tabulation above, a modal-split transformation procedure was developed and applied to the base data on a district-interchange basis to produce a modal-split matrix for each case (4, Appendix A). Figure 2 illustrates the resulting aggregate peak-period penetration of selected markets by transit. For example, when the regional peak modal split is 40 percent, almost all morning peak trips from the city to the CBD are transit trips but only 15 percent of the trips destined for the suburbs are made by transit. To achieve peak modal splits greater than 40 percent, nontraditional markets must be heavily penetrated by transit.

Analysis of Integrated Service

As shown in Figure 3, transit services were modeled in two parts: (a) a regional network of fixed-route bus lines and (b) local transit services providing intrazone service and feeder connections to the line-haul network. Line-haul options were analyzed in terms of specific networks by using the processing and transit-assignment modules of the urban transportation planning system (UTPS) of the Urban Mass Transportation Administration (5). This approach, which was consistent with the use of real traffic data, reflected the interdependence of the transportation system and travel patterns. A range of local service options including doorstep, checkpoint, conventional fixed route, and route deviation were analyzed by using models expanded or developed for this study (8, 9). Unlike the regional network analyzed, the local transit models were based on typical districts with abstracted networks and trip distributions. The results of the separate analyses were combined to investigate the service potential and operating costs of integrated regional public transportation systems. Corresponding estimates of fuel consumption, pollutant emissions, and capital cost were developed by using recent DOT studies (4, Appendix B; 6; 7).

Network Design

Figure 4 shows three sets of options that exist in the design and operation of transit networks to serve increasing transit volumes and changing trip patterns as modal split increases. The first option, the basic configuration of the network (radial versus grid), is highly constrained by the roadway system. Within the basic configuration the spacing or density of routes as well as the outward extent of line-haul services can be varied. These options involve changes in the importance and extent of local transit services. A third set of options involves the connectivity of the network, which can be improved by providing transfer points among high-frequency trunk routes or by providing more direct service between points on the network.

Line-Haul Options

Several options for regional bus operations within a basic network were explored by using a route-based supply model in conjunction with the UTPS network models. The following options were examined: (a) trade-offs between service frequency and vehicle size, which have direct effects on level of service and both operating and capital cost; (b) introduction of express or skip-stop service; (c) use of suburban transfer points; and (d) use of exclusive lanes and other priority measures on expressways, arterials, and downtown streets. The evaluation of some of these options required, in addition to vehicle costs, estimates of conversion and operating costs for fixed facilities (such as exclusive lanes or the hardware required for prioritization schemes). Figure 5 shows the line-haul options in the context of a sample corridor.

Local Service Options

The full range of local service options, from fixed-route to fully demand-responsive service, was evaluated (Figure 6) by using a family of local service models. These models were designed to respond to varying levels and proportions of intrazonal, feeder, and intradistrict trips; varying locations of line-haul stations, transfer points between adjacent zones, and route or checkpoint route density within the service area; mixes of transit access modes, vehicle sizes, and load factors; and varying operating speeds. Within each typical service area examined, heuristic searches were made to identify optimal modes and operating policies based on estimated cost and level of service. The key trade-offs involved were those between vehicle size, walk distance, wait time, and average speed (as affected by circuity and start-stop cycles for boarding).

Dynamically routed services such as dial-a-ride were modeled based on computer simulations and validated by actual data from Haddonfield, New Jersey (4, Appendix C; 9). The models accounted for the impact on bus speeds of dwell times and number of stops per hour, fraction of dead time (percentage of time the vehicle is available for assignment when no demand exists), trip density (demands per square kilometer per hour), analyst-specified constraints on level of service, and fleet and vehicle size. Dial-a-ride was modeled as a coverage service in the off peak and as a supplement to either doorstep or checkpoint subscription service in the peak. The subscription service models, which were similar in framework to the fixed-route and route-deviation models, were extensions of work by Ward (8).

Cost Allocation

Operating costs were assigned to the local and regional service components based on the pro rata share for each service of total vehicle hours and kilometers. Overhead costs were estimated as a function of fleet size. Capital costs were allocated to local and regional service in proportion to vehicle requirements (or other capital equipment requirements) by time of day, which caused peak-hour services to bear the brunt of equipment costs. No attempt was made to estimate the marginal cost of off-peak transit labor; average labor costs were used over the full day.

Evaluation Criteria

The regional transit-service alternatives were evaluated on the level of service provided and the capital and operating costs of the system. The level-of-service measure was door-to-door travel time, which included estimates of access, egress, and wait times optionally weighted to form perceived impedance measures. Transit travel times were compared to existing automobile travel times on a zonal interchange basis in the form of bar graphs showing the percentage of transit trips in the region (or in specific markets) having travel times X minutes better (or worse) than their automobile alternatives. Because of the great variation in trip lengths and patterns, these measures were more meaningful than regional averages of transit travel time. Furthermore, in the absence of demand modeling, these service comparisons enabled the analyst to determine if a modal split and transit service assumptions used in an analysis approximated an equilibrium.

Nonuser impacts such as changes in fuel consumption and in emission levels of carbon monoxide, hydrocarbons, and oxides of nitrogen were also tabulated. Because no attempt was made to explicitly model the impacts of major diversions to transit on the level of service of the remaining automobile users, major additional benefits to automobile users are not included in the evaluation.

Results

The models described above were applied to a wide range of modal splits and system designs, and conclusions were drawn in the areas of economic performance, service levels, and prototypical system operating procedures. A brief summary of the conclusions follows; more detailed results are available elsewhere (4, 10).

Costs

Figure 7 shows the changes in transit-system operating and capital costs as the network is expanded and inte-

Figure 1. Study area.

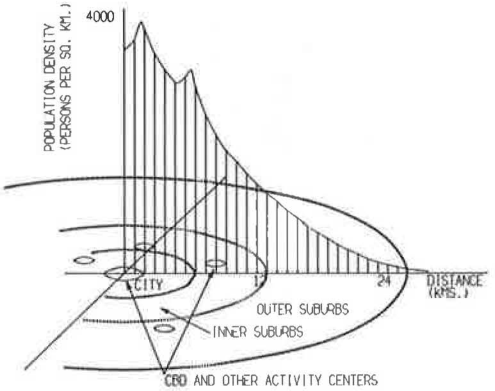


Figure 2. Morning peak trips by market.

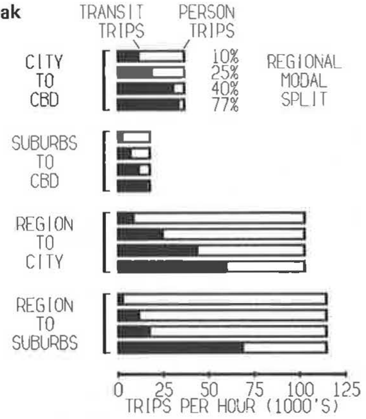


Figure 3. Model of transit service.

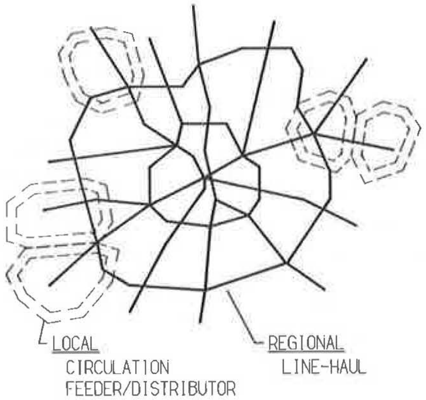


Figure 4. Network design options.

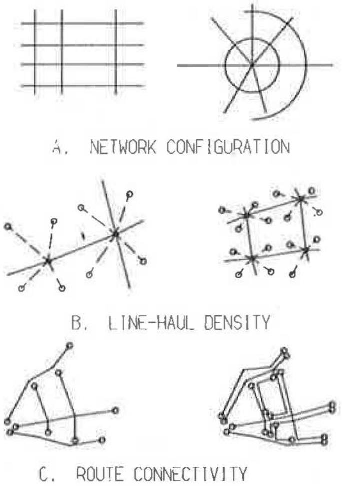


Figure 5. Line-haul service options.

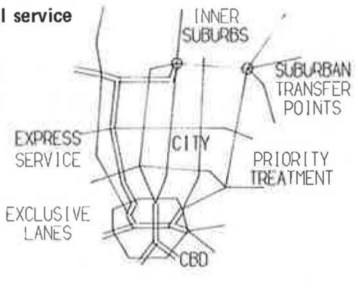


Figure 6. Local service options.

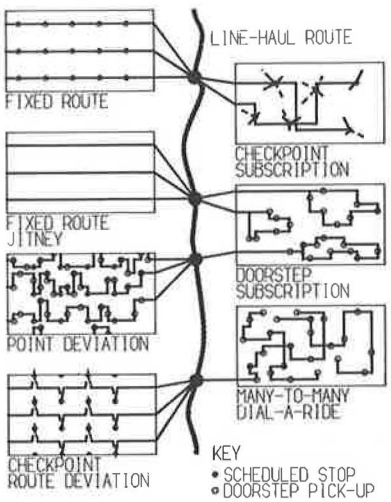


Figure 7. System costs versus modal split.

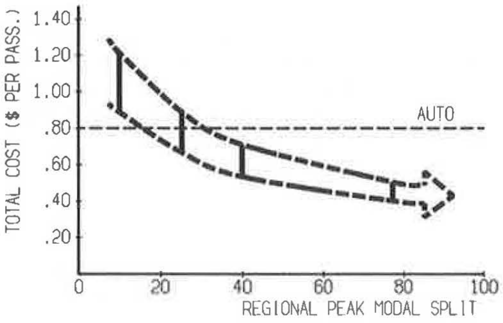


Figure 8. Alternative transit investment policies.

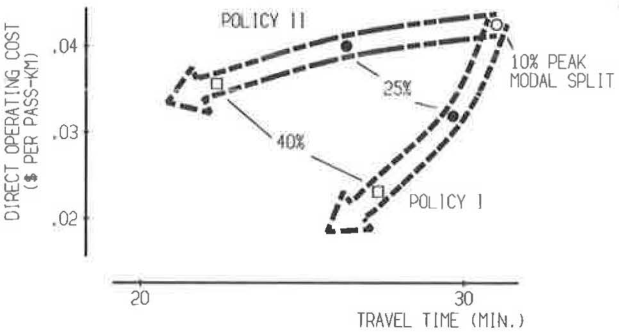


Figure 9. System levels of service.

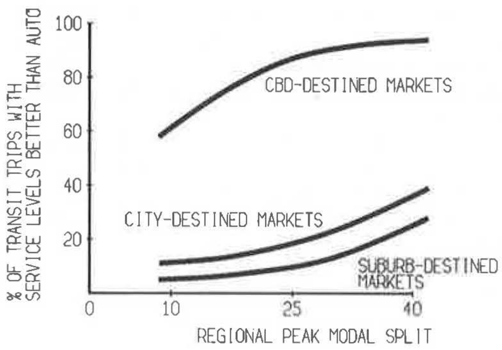
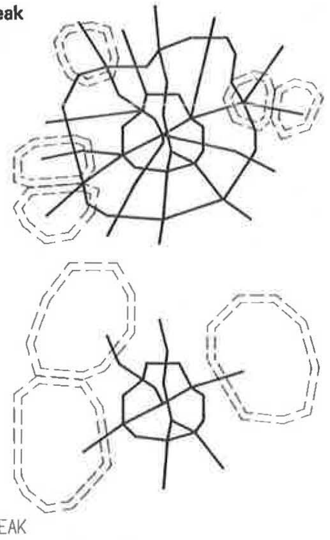


Figure 10. Daily peak and off-peak service configurations.



grated to serve increasing shares of peak-period trips. Significant economies of scale are apparent, contrary to the belief that the industry can only attract new riders by increasing average costs. The sources of these economies are

1. Increases in backhaul and other non-CBD-oriented trips to improve line-haul load factors;
2. A decline in the peak-to-base ratio as modal split grows;
3. An increase in transit travel speeds and reliability as facilities are dedicated, which results in improved line-haul vehicle productivity;
4. Use of larger vehicles as modal split increases to improve labor productivity without a decrease in the level of service; and
5. Provision of good local service by means of (a) low-cost, fixed-route operation at higher modal splits to reduce the unit cost of feeder service and (b) checkpoint and fixed-route services instead of doorstep services at low modal splits.

Figure 7 implies that modal split need not increase very much to produce economies of scale. In fact, such economies rapidly diminish after moderate modal shares are reached. There is not likely to be a threshold modal split at which direct benefits increase rapidly. The potential for transit-system cost and service benefits is greatest in the range of modal splits just slightly above current values. These benefits are the reverse of the disbenefits that in recent decades have accompanied decreased transit patronage.

Service Levels

The planner has a variety of investment options in trading system economies of scale for improvements in service. Figure 8 shows two policy options explored in the analysis of line-haul operations. In policy 1 the economies of large vehicles and dedicated facilities combine to achieve a 50 percent reduction in unit costs and a 15 percent reduction in travel time. In policy 2 potential economies in vehicle size are reinvested in additional, denser routes to enable a 33 percent reduction in travel time and a slight reduction in unit costs; the result is a fleet more easily adapted to flexible services in the off peak. It is not yet known whether options exist that would yield greater service improvements at higher unit costs.

Figure 9 shows the distribution of service improvements achieved in policy 2 in comparison to automobile travel. It is clear that, although significant service improvements can be achieved in all parts of the region, transit service cannot match the automobile for many interchanges within nontraditional markets.

Operating Policies

The key to providing integrated transit service is a range of operating policies that adapt to the changing travel environment. Not only do trip patterns vary over the day but transit trips are also on the average much shorter (between 15 and 30 percent) in off-peak periods. The results of combining the previously discussed attributes of the local transit system with shorter trips and greater dispersion of travel patterns in off-peak periods suggest that the following scenario is likely to be highly effective in providing efficient, high-quality transit service at ridership levels resulting from major diversions to transit (Figure 10).

During the peak period an extensive regional line-haul system is operated with a mix of medium and large vehicles. Travel oriented toward major activity centers including the CBD is served by express vehicles operating on dedicated rights-of-way. Arterial services within the city receive the benefits of signal and lane prioritization schemes. Major transfer facilities are established in the inner suburbs and thus the travel time for long trips not oriented to major activity centers is reduced. The line-haul system is fed by a system of fixed-route and subscription buses operated in small local service areas. During off-peak periods a radically different transit system is operated. The regional line-haul system is reduced in both scope and density. In contrast, the local service areas grow to accommodate the majority of trips within their boundaries. The longer off-peak trips are served by coordinated transfers either between these expanded local service areas or to the basically radial line-haul network. The local service is operated in either demand-responsive or route-deviation fashion depending on the travel densities encountered. Inner city districts continue to rely on the line-haul system of local service.

AREAS FOR FUTURE RESEARCH

The range of options examined in this preliminary analysis does not do justice to the rich variety of alternatives available to transit planners. For example, this study focused on highway-based modes; light and heavy rail technologies, shuttle-loop transit, and other automated-guideway options could not be considered. Such capital-intensive options might result in economies of scale continuing to be derived well beyond a daily modal split of 25 percent. At the other end of the range of options, local transit services such as dial-a-ride and point-deviation bus service were provided by using paid-labor alternatives. Options using in-kind labor, such as car- or van-pooling, are likely to improve system efficiency, especially during peak commuting periods.

A major concern is, of course, the need to determine whether the transit service levels provided can sustain the assumed modal splits. However, before a major effort is made to examine the demand side of integrated regional public transit, more research is needed to develop a better understanding of the full service and economy potential of integrated transit operations.

ACKNOWLEDGMENTS

The research reported here was sponsored by the Research and Development Policy Analysis Division of the U.S. Department of Transportation. Jerry Ward and Norman Paulhus of that organization and Daniel Roos, Nigel Wilson, Larry Englisher, Kenneth Sobel, and Lea Tsai of Multisystems, Inc., deserve special credit for their contributions to the scope and direction of this research. The figures in this paper were produced by GRAPHITI, an interactive computer graphics system developed by Multisystems, Inc.

REFERENCES

1. P. S. Liou, G. S. Cohen, and D. T. Hartgen. Application of Disaggregate Modal-Choice Models to Travel Demand Forecasting for Urban Transit Systems. TRB, Transportation Research Record 534, 1975, pp. 52-62.
2. J. Ward and N. Paulhus. Suburbanization and Its Implications for Urban Transportation Systems. Office of the Secretary, U.S. Department of Transportation, April 1974; NTIS, Springfield, Va., PB 231 819.
3. P. Benjamin and others. Urban Transportation Alternatives—A Macro Analysis. Transportation Systems Center, U.S. Department of Transportation, Dec. 1974; NTIS, Springfield, Va., PB 238 775.
4. J. H. Batchelder and others. Operational Implications of a Major Modal Diversion to Transit. Multisystems, Inc., and Office of the Secretary, U.S. Department of Transportation, April 1976; NTIS, Springfield, Va., PB 255 921.
5. UTPS Reference Manual. Urban Mass Transportation Administration, U.S. Department of Transportation, June 1975.
6. Characteristics of Urban Transportation Systems. De Leuw Cather and Co.; Urban Mass Transportation Administration, U.S. Department of Transportation, May 1975.
7. Highway Investment Analysis Package—Users' Guide. Multisystems, Inc.; Federal Highway Administration, U.S. Department of Transportation, March 1976.
8. D. Ward. A Theoretical Comparison of Fixed Route Bus and Flexible Route Bus Feeder Service in Low Density Areas. Transportation Systems Center, U.S. Department of Transportation, March 1975; NTIS, Springfield, Va., PB 240 808.
9. Development of a Method for Estimating Patronage of Demand Responsive Systems. Cambridge Systematics, Inc., and Multisystems, Inc., in preparation.
10. K. Sobel and J. Batchelder. The Evolution of Integrated Transit, Three Parables. Multisystems, Inc., and Office of the Secretary, U.S. Department of Transportation, June 1976; NTIS, Springfield, Va., PB 257 160.

Publication of this paper sponsored by Committee on Transportation Systems Design.

Application of a Large-Scale Dual-Mode Simulation to Milwaukee County

Timothy J. Heintz, Marquette University

Monte Carlo simulation, a technique for evaluating cost versus service trade-offs in the design of a dual-mode transportation system, is applied to a large-scale system for Milwaukee County. Model formulation, experimental design, analysis of alternative design configurations, and possible refinements of both the model and the methodology are considered. Experimental results show that, although there is an apparent interaction between system design and operations, within a reasonable range system performance is relatively insensitive to changes in operating conditions. Operating differences resulting from the size of the dual-mode network are noted. A reduction in dual-mode station capacity produced little effect except increased station congestion. Although problems associated with computer resource requirements did exist, it is concluded that methodological improvements would help make this type of simulation useful.

In recent years there has been an increasing interest in new and innovative concepts in mass transit, such as dial-a-ride (1, 2), personal rapid transit (PRT) (3), and dual mode (4). The development of these systems has created a need for new tools and techniques that will help transportation planners to compare alternative system designs. This paper presents an application of one such technique, Monte Carlo simulation, to a dual-mode transportation system.

A dual-mode system consists of vehicles that can be manually operated for collection and distribution of passengers on the city streets and can also be operated as line-haul vehicles on high-speed and automated guideways. Work has been done on the physical design of such systems (5). Software to support the planning of dual-mode facilities in specific urban areas is relatively scarce. One study performed in Milwaukee County (6) applied conventional traffic simulation techniques to a dual-mode network but did not fully consider the dynamics of a more sophisticated demand-responsive system. It has been suggested that Monte Carlo simulation could be used to monitor the activity of dual-mode buses (7). This paper presents the results of applying a similar model to traffic data in Milwaukee County.

The previous Monte Carlo simulation model assumed that the automated guideway operated in a single loop and that no transferring of passengers was allowed at dual-mode stations. The simulation model used in this study differs in the following major ways:

1. Any general type of guideway network configuration can be handled.
2. Once a bus is loaded onto the guideway, one intermediate stop is permitted to service passengers waiting at a station.
3. Buses collecting passengers in an off-guideway service area may pick up passengers who have different destinations and deposit them in passenger queues at the station.
4. Buses do not leave the guideway at the central business district (CBD) station location.
5. Station queuing disciplines allow limited parallel processing of buses.
6. Bus and operator requirements are no longer inputted constraints but are computed by simulation.

In addition to discussing the results of applying the dual-mode simulation to real data, this paper also discusses (a) the assumptions of the dual-mode transit system being simulated, (b) the methodology and the model used in the simulation, and (c) refinements in both the simulation model and the procedure used in applying it. The main objective of the study was to evaluate the practical aspects of applying the methodology to a large-scale system, but an attempt is also made to draw some conclusions from the simulated results.

DUAL-MODE TRANSIT SYSTEM

The dual-mode configurations used in this study were completely demand responsive. The Milwaukee metropolitan area was divided into a number of local service areas, each of which was serviced by a single guideway station. It was assumed that trip demands within the service area would be generated randomly and that small 20-passenger buses would be dispatched from the guideway station, on a call basis, in one of two passenger pickup modes: (a) a general pickup mode in which passengers waiting the longest time would be serviced regardless of final destination and (b) a dedicated pickup mode in which only passengers with a common destination would be assigned to a bus. Once passengers from the local service area arrived at the station, they might have to wait at the station for a bus to be dispatched to their final destination or they could continue directly onto the guideway by using the same bus. For a bus to be loaded onto the automated guideway, it would have to be assigned another local service area as its final destination. This bus, however, could also be assigned, either at time of dispatch or en route, one intermediate stop for the purpose of handling passengers waiting at the station. Each station in the network would serve as a holding area for both buses and operators; the operators, of course, would only be used to drive the local, off-guideway buses.

All stations in the dual-mode network except the CBD station would be operated in the same way. Local stations, however, could vary in the number of buses they could handle at one time. Within the CBD, buses were not allowed to leave the guideway but proceeded in a loop, stopping at each of a number of passenger drop-off and pickup points. Travel time through the CBD was fixed up to a certain limit. Once a capacity point was reached, headways for CBD buses were reduced by an amount proportional to the excess volume. In the simulation, the average travel time for off-guideway buses in service areas outside the CBD is a function of the number of passengers riding in a particular bus and the service area itself. Random variations around this average travel time are allowed.

SIMULATION MODEL

Design

The simulation model was developed by representing the status of the dual-mode system on the computer, by identifying events that cause changes in this status, and then by having the computer deal with these events. For the model described here eight events, given in the table below, were used. The possible effect of each event on the status of a bus is also given.

	Bus Status	
Event	Before	After
Trip demand	—	—
Release of new service area bus	—	In general pickup mode In dedicated pickup mode
Release of new guideway bus	—	Loading onto guideway
Arrival of bus at station	On guideway link	On new guideway link Unloading off guideway
Arrival of general pickup bus	In general pickup mode	Loaded onto guideway Removed from system
Arrival of dedicated pickup bus	In dedicated pickup mode	Loaded onto guideway
Bus loaded onto guideway	Loading onto guideway	On guideway link
Bus ready at station	Unloading from guideway	In general pickup mode In dedicated pickup mode Loading onto guideway

The status of the dual-mode system was represented in the computer by counts of passengers who had specific origins and destinations and were waiting in each service area and in each station. Counts were also maintained of buses at various locations in the dual-mode network.

The following example illustrates how the simulation handles various events. A general pickup bus arrives at a station designated i. The count of general pickup buses servicing area i is decreased by one and, if passenger demand is sufficient to warrant continuing the bus onto the guideway, the count of buses waiting in station i is increased by one. At the same time passenger queues at the station that represent people waiting for destinations other than those assigned to the continuing bus may be increased. On the other hand, passenger queues for the destination that will be serviced by the bus are decreased to zero or by as much as the capacity of the bus allows. Also at this point in the simulation, the time at which the continuing bus will be loaded onto the guideway is computed, and statistics are updated. A similar process occurs for the other events.

A number of assumptions are made in determining the timing and sequencing of these events. The times between successive demands are assumed to be generated from an exponential probability distribution, where the demand rate of each area equals the reciprocal of the total number of trips originating in that particular area. The destination for each trip is computed from a discrete probability distribution in which the probabilities are determined by dividing each element in a row of the trip table by the row sum. The on-the-road travel time for each bus is estimated separately for each service area and inputted into the simulation by specifying minimum and maximum travel times. These figures are then assumed to represent the three standard deviation limits of a normal probability distribution and are thus used to compute mean and variance. The mean of this distribution is then adjusted upward during the simulation by 15 s for each passenger being picked up and 9 s for each passenger being dropped off. Service-area buses are allowed to operate in the general pickup mode, in which a bus may be assigned a number of final destinations, and in the dedicated pickup mode, in which a bus collects only passengers having a single destination station. Trips that both begin and end in a single service area were not handled by this simulation.

When a bus from the service area arrives in a non-CBD station, when a new bus is released directly to the guideway, or when a bus stopping for an intermediate stop is ready to load passengers, the time at which this bus will enter a particular guideway link is computed. It is assumed that the bus will approach a loading ramp that has a capacity to process a limited number of buses in parallel. If the ramp is not congested, a fixed time is allowed for the bus to process passengers and accelerate onto the guideway. Otherwise, delays caused by congestion are assessed. In the simulation a bus can bypass the loading area if there is no need for that bus to pick up or drop off passengers. At this point the guideway-bound bus is also assigned a final destination and possibly an intermediate stop by means of a set of operating rules that will be discussed later. An intermediate stop may be preassigned when conditions permit, or an option may be enforced that allows a bus to decide, depending on demand, whether or not to stop as it approaches an intermediate station on the guideway.

Once a bus is on a guideway link, the time at which the bus arrives at the next station is computed from link distances and a speed factor inputted at the beginning of the simulation. The actual link to which the bus is assigned is determined from a table that provides the next link on the path from any station to any other station. These paths may be, but do not necessarily have to be, minimum travel-time paths.

When a bus arrives at a station, a check is made to see if that bus should stop or continue. If it continues, the next link and travel time are determined as described above. If it stops at either the final destination or an intermediate station, a delay time is determined in a manner similar to that used to describe buses leaving a station. If a destination station is in the CBD, this process is slightly modified in that the model assumes that the bus circles around a loop and stops at each of a series of equally spaced platforms.

In monitoring bus activity the simulation program computes total time and average number for buses in each of the following locations in the dual-mode system:

1. Buses traveling within the local service area,
2. Buses leaving the service area that stop at the station for passenger pickup or drop-off,
3. Buses leaving the service area that bypass the station and are loaded directly onto the guideway,
4. Buses entering the service area that stop at the station for passenger pickup or drop-off,
5. Buses entering the service area that bypass the station,
6. Buses on non-CBD guideway links, and
7. Buses on the CBD loop.

These computations are also made for buses that are in the process of making an intermediate stop; that process would overlap with items 2 and 4 above.

Average time and number were computed for passengers waiting both at a station and within the service area, and data were maintained for average bus loading, which is expressed as a fraction of the total utilized capacity for all the active buses in the system. The simulation also monitored the maximum number of operators and buses used during the collection period. In addition, a number of counters were incremented each time a certain type of activity occurred. This information is useful in checking the face validity of the simulation and in identifying potential problems in the operation of the dual-mode network.

These data were collected on an aggregated, system-wide basis. A routine that could collect similar information for each station could easily be written, but such detailed data collection would be costly in terms of computer time and storage. The best strategy would be to use the more aggregated program for experimentation purposes and then use the more detailed program on a one-shot basis.

Control of Simulated Operations

A number of operating decisions are made during the course of the simulation, including decisions on when and where to dispatch buses and on whether or not to allow intermediate stops. Such actions are controlled within the model by a set of simple decision rules of the following form: If x is at or exceeds level y, then do z. The major concern in determining parameter values was to set them at levels that would provide an efficient operation for each alternative design. This requires some experimentation with the operating parameters before other types of comparisons are made. The main purpose of the experimentation, or model tuning, is to avoid comparing a poorly operated configuration with a well-operated one. It should also be possible to make limited conclusions about what the important operating factors are (e.g., bus loadings, dispatching delays, intermediate stops, or station delays).

Table 1 gives a list of eight parameters used as a basis for making operating decisions. The parameters control the event sequence at three different points in the simulation. First, a decision must be made concerning when and in what pickup mode buses should be dispatched into the service area. This decision depends on either the number of passengers waiting for a particular destination or the passenger wait times. When a bus at the station is ready to distribute passengers within the service area, it checks the queues of service-area passengers waiting to be transported to various destinations. If the number of the largest queue is greater than or equal to the value represented by the parameter SAN, then that bus will, while dropping off current passengers, pick up passengers only in the largest queue, thus operating in a dedicated pickup mode. Otherwise, that bus will operate in a general pickup mode, collecting passengers from as many queues as capacity allows. The wait-time value SAT is used when there are no active buses available to serve a particular queue. Wait times are checked every 30 s. If a passenger has been waiting for pickup for a period longer than SAT, an empty bus is dispatched to handle passengers in that queue. If the total number of passengers waiting to be transported to the same destination reaches the bus capacity, an empty bus is immediately dispatched in the dedicated pickup mode.

A second set of decisions must be made when a bus arrives at a station from the service area. If a bus is operating in the dedicated pickup mode, the station queue containing passengers waiting to be transported to the assigned destination is checked and, if it is empty, the bus bypasses the station. In either case, if the total number of passengers in the bus is greater than the parameter STN1, the bus is not allowed any intermediate stops.

A general pickup bus will always stop at the station to discharge passengers and check the station queues. If the largest queue (including passengers delivered by the bus that just arrived) is less than the value of STN2, the bus is removed from the system. Otherwise, the bus continues onto the guideway having been assigned the same destination as that of the longest queue. Another check is made against STN1 to determine whether an intermediate stop should be allowed; then a second check is made of all queues having destinations along the path the bus will follow. If the number of passengers waiting for transport to intermediate stations is greater than or equal to the value of STN3, an intermediate stop is preassigned. If a sufficient number of station passengers are not available, the bus is assigned a single destination and is released to the guideway; an intermediate stop may be assigned in transit. The wait-time check against STT may trigger the release of an empty bus, which may then be assigned an intermediate stop as described above. When a station queue reaches the level of bus capacity, an empty bus is immediately dispatched to the guideway.

Finally, a decision must be made as to whether a bus approaching an intermediate station should stop. This decision, when allowed, is based on the number or wait time of waiting passengers who have the same destination as the bus. For an intermediate stop to be assigned in transit the number of waiting passengers must exceed the parameter GWN or the wait time must exceed GWT.

Varying the eight parameters given above would have some obvious general effects. For example, because the major trade-off in any transportation system is that between service, as measured by wait and travel times, and transit-system design efficiencies, as measured by bus and operator utilization and the size of transit facilities, lowering wait-time values may improve service measures but may also substantially increase bus and operator requirements.

APPLICATION

The simulation discussed above was applied to actual 1990 peak morning demand projections for Milwaukee County. Sketches of three alternative networks used in the project, as well as a more detailed discussion of procedures and data inputs, are available elsewhere (10). Station locations in the first configuration closely approximated those of the Allis Chalmers study (6). The two other configurations used a subset of the original 41 stations.

Data comparing the three configurations are given in Table 2. A few trips between station pairs, in which passengers would have to go considerably out of their way if they used the dual-mode system, were eliminated. The total number of trips eliminated for this reason was small compared to the total transit demand (Table 2). A total of 400 traffic zones were assigned to stations by examining a map and assigning zones in the general area of a station to that station. An estimate was then made of the minimum and maximum on-the-road travel times for buses within a station service area by examining the area interzonal and intrazonal travel times. This is the input into the simulation that most needs refinement. In this project, such estimates required considerable subjective judgment, but there are techniques available for dial-a-ride simulation (8, 9) that could provide a more objective and precise means of developing these data.

It was decided in this application to analyze the simulation under 1990 peak-load, steady-state conditions. This required first running the simulation until the dual-mode network was loaded with buses. The relative effects of varying a number of factors could thus be evaluated apart from time-dependent influences. In addition, because data could be collected over relatively short time intervals, computer time could be saved. The following approach was used.

1. Values for all the controllable factors were arbitrarily set (factor set refers to a specific group of experimental factor levels).
2. An initialization run was made in which statistics such as total times, average numbers, and average bus loadings were reinitialized and examined at 10-min intervals to see if they were relatively unchanged over two successive intervals.
3. After these steady-state criteria were met, data were collected for four additional 10-min intervals, the last two of which were used as the experimental observa-

tion and a replication of the experimental observation for the first factor set.

4. For different factor sets, the simulation was restarted at the point at which it was terminated in step 2. As in step 3, these subsequent factor sets were also run for four 10-min intervals, the first two of which were discarded because time had to be allowed for the effect of the experimental treatment to be realized. As discussed above, the next two data-collection intervals represented an observation with one replication.

The use of data derived from steady-state conditions should tend to produce results that are worse than one would expect in a real-life situation. During regular operation of dual-mode systems, demand would usually build to a peak and then taper off. Under these circumstances, the steady state simulated here would probably never be realized. However, since the objective of simulation is to compare alternative designs, these steady-state values should provide measures of the relative effect of various design and operating factors on overall system performance.

In determining specific values for each factor in a factor set, values first had to be established for the eight operating parameters discussed previously. Because using the same parameter values across all three networks might bias the results in favor of a certain type of configuration, and because the relation between system design and operation is of interest in itself, experiments were conducted that had the effect of "tuning" the operations of a system to a particular configuration.

The wait-time parameters SAT, STT, and GWT were arbitrarily set at 5 min. It was felt that, because of the traffic volumes involved in the simulation, there would in most instances be sufficient demand to trigger bus activity before obtaining a 5-min wait; these parameters would thus have little effect on the results. After the initial tuning procedure was completed, these maximum wait times were reset to 7.5 and 10 min to verify this assumption.

The other five operating parameters—SAN, STN1, STN2, STN3, and GWN—were simultaneously varied by using a quarter-fraction factorial design procedure that allowed measurement of all five first-order effects and the interaction effects between STN1 and STN2 and between STN3 and STN2. [For detailed discussion of this experimental procedure the reader is referred to Davies (11) and Myers (12).] After the initial experiment is completed, a series of new values can be determined for the operating parameters by fitting a regression plane to some measure of operating "goodness"; the new values should lie along a line that yields the maximum increase in this goodness measure. By conducting further experiments along this line (the path of steepest ascent), a series of improved simulation results should be realized until operating conditions that are at least closer to the best possible result are found. Further experimentation may or may not be desired depending on the new results.

The numerical values of the five parameters used in each initial experiment are given in Table 3. By using the numbers given in Table 3 rather than identical factor levels for each network, we were able to include in the experiments some preliminary runs.

There is probably no single statistic that can be used that is indicative of the goodness of a particular run. Instead, a somewhat arbitrary weighting of service statistics (measured by total times) and design statistics (cost) was used. Basically, for the service-related data, simple assumptions such as the following were made: (a) Waiting within the service area is more desirable than a wait in the station and (b) traveling on the automated guideway is more desirable than traveling in the service area or being processed through a station. An attempt was made to use some economic base for the design weight. A report prepared by Rohr Industries (5) was used to estimate annual vehicle and station cost. A trade-off factor that specified the relative weight given service and cost factors was arbitrarily set at a level that would balance these two criteria so that neither service nor cost would dominate the tuning procedure. Details on the formulation of these statistics—a procedure called the design effectiveness measure (DEF)—and results of a sensitivity analysis performed on these weights are given elsewhere (10).

When a set of satisfactory operating parameters is obtained, valid comparisons of the three networks can be made. Because the tuning procedures were conducted under conditions of excess capacity, an attempt was also made to observe the effect on the dual-mode system of reducing station size. Further experimental runs were made for different non-CBD station capacities. Additional runs could also be made for different guideway speeds, bus sizes, CBD capacities, and station delays.

SIMULATION RESULTS

All three simulations reached a steady state in 120 min. Therefore, time intervals ending at 130 and 140 min were used for initialization of the factor set, and data were collected in the two 10-min intervals ending at 150 and 160 min. Table 4 gives the results of the initial tuning experiments in Table 3 and summarizes a regression against the DEF measure. The fact that significant results were obtained in only two cases indicates either that the sensitivity of operating parameters was low in overall performance or that the statistical tests were not powerful enough to distinguish the differences. After detailed examination of the data, it was felt that results were stable enough statistically to discern some relationship but that more than a single replication was needed to obtain statistical significance. Because of computer time limitations, however, further runs could not be made.

This discussion attempts to relate the results given in Table 4 to selected items from the detailed output of the simulation. A more complete presentation of these data is available elsewhere (10).

The large negative coefficients for STN1 and STN2 in configuration 1 corresponded to a greater average number of passengers on a bus at the higher levels of STN1 and STN2, which allowed more intermediate stops (Table 1). In this case, system efficiency was improved by allowing intermediate stops, but the price paid was increased wait times. The net result, however, was a desirable lowering of the DEF. Similarly, the lower STN3 coefficients in the smaller configuration indicated that network operations should allow fewer preassigned intermediate stops. Negative coefficients for SAN and STN2 seemed to suggest that passengers could be allowed to wait longer at home or at the station in the two larger networks but that these wait times become more critical in the third, smaller network.

A path of steepest ascent was computed for all configurations and further observations were made. Because of low sensitivity in the DEF to variations in these parameters, the effort made here was not extensive. The final values of the operating parameters that were used in further experiments are given in Table 5. In every case, the DEF was only slightly lowered. However, this lack of sensitivity was not always observed. In a preliminary run, parameters STN1 and STN2 were set at the same level. This produced a situation in which very few intermediate stops were allowed and total bus requirements grew to such a point that the simulation had to be aborted.

As mentioned earlier, the delay-time parameters SAT, STT, and GWT were also increased from 5 to 10. For configurations 2 and 3 the DEF showed a slight increase, but for configuration 1 it decreased from 2.282 to 2.171. The assumption that the five activity-triggered parameters would dominate the delay-time parameters seemed to hold for the smaller networks but not for configuration 1. Results are thus reported for configuration 1 for conditions of 5 and 10-min maximum delay.

A comparison of bus activity across the three configurations is given in Table 6. Examining activity in service areas and stations reveals an expected pattern. Lower bus loadings from the larger networks result in a greater number of buses being dispatched to both the service area and the guideway. A lower percentage of buses are assigned dedicated pickup and a larger percentage are not allowed intermediate stops within the smaller network. The percentage of buses dispatched to the guideway that do make intermediate stops is drastically reduced.

In examining guideway link volumes, the capability of a single guideway lane to handle a specified number of buses is of interest. By dividing link volumes by link distances, the minimum spacing between the buses is varied (under 5-min maximum delays) from roughly one bus every 50.3 m (165 ft) for configuration 1 to one every 114 m (374 ft) in configuration 3.

Another means of comparing these networks was to develop a scenario for the typical trip within each configuration. This required the use of average time data, which, although collected by the simulation, were not reliable because of the short data-collection period. Estimates were therefore calculated by dividing average numbers by the corresponding activity counts (Table 7). As one would expect, service-area guideway travel times and service-area wait times decreased with increasing network size. The increased intermediate stops and the additional number of stations tended to produce higher station delays in the larger network. Because all passengers would not be stopping at the station or participating in intermediate stops, these two items were appropriately reduced in computing total trip time. Increased network size shortened typical waits by 3 to 4 min and travel times by 5 to 11 min.

All experimentation was done under conditions of ex-

Table 1. Operating parameters.

Parameter	Statistic	Decision Made When Statistic ≥ Parameter Value
SAN	Number of service-area passengers	Assign bus to dedicated pickup
SAT	Wait time of service-area passengers	Release new service area bus
STN1	Number of station passengers having certain destination	Assign bus to single destination only
STN2	Number of station passengers having certain destination	Continue local bus onto guideway
STN3	Number of station passengers having intermediate destination	Assign bus to intermediate stop
STT	Wait time of station passengers	Release new guideway bus
GWN	Number of station passengers having certain destination	Assign intermediate stop to bus on guideway
GWT	Wait time of station passengers having certain destination	Assign intermediate stop to bus on guideway

Table 2. Comparison of three dual-mode network configurations.

	Configuration		
Statistic	1	2	3
Service area			
Number of traffic zone assignments			
Average	9.3	17.8	31.1
Maximum	23	36	50
Minimum	4	6	19
Trip origins (non-CBD)			
Average	2660	4825	8114
Maximum	6258	10 934	15 592
Minimum	703	2293	4224
On-road trip time, min			
Average, all areas	15.8	18.7	24.0
Standard deviation, all areas	3.7	4.6	6.3
Maximum	36	40	50
Minimum	2	2	2
Guideway network			
Number of non-CBD stations	40	21	12
Number of links	122	74	42
Total link distance[a], one way, km	325	297	251
Average link distance, km	5.3	8	12
Total trips assigned to guideway[b]	107 884	102 825	98 876
Interstation routes not included	21	8	2
Interstation trips not included	1846	2817	625

Note: 1 km = 0.62 mile.

[a] Total link distance may include two or more links that occupy, in part, the same physical guideway.

[b] Total transit trips number 119 102. Interstation trips on excluded routes and trips within a single service area make up the difference.

Table 3. Experimental values of operating parameters.

		Factor Set							
Configuration	Parameter	1	2	3	4	5	6	7	8
1	SAN	18	16	16[a]	18	16	18	18	16
	STN1	12	14	12	14	12	14	12	14
	STN2	2	2	5	5	2	2	5	5
	STN3	5	5	5	5	8	8	8	8
	GWN	8	11	8	11	11	8	11	8
2	SAN	16[a]	14	14	16	14	16	16	14
	STN1	12	14	12	14	12	14	12	14
	STN2	5	5	8	8	5	5	8	8
	STN3	5	5	5	5	8	8	8	8
	GWN	8	11	8	11	11	8	11	8
3	SAN	18[a]	16	16	18	16	18	18	16
	STN1	13	15	13	15	13	15	13	15
	STN2	6	6	8	8	6	6	8	8
	STN3	3	3	3	3	5	5	5	5
	GWN	9	11	9	11	11	9	11	9

[a] Initial factor set.

Table 4. Regression analysis of initial tuning experiments.

Parameter	Configuration	Mean Factor Level	Distance Between Levels	Regression Coefficient ($\times 10^3$)	F-Ratio
SAN	1	17	2	-0.12	0.02
	2	15	2	-1.89	1.34
	3	17	2	1.77	1.49
STN1	1	13	2	-1.70	3.46
	2	13	2	-9.03	30.77[a]
	3	14	2	-2.34	2.64
STN2	1	3.5	3	-1.15	1.60
	2	6.5	3	-3.30	4.10
	3	7	2	0.51	0.13
STN3	1	6.5	3	-2.65	8.41[b]
	2	6.5	3	-1.49	0.83
	3	4	2	1.98	1.88
GWN	1	9.3	3	0.86	0.88
	2	9.5	3	-0.87	0.28
	3	10	2	0.74	0.27

[a] Significant at 0.01 level.
[b] Significant at 0.05 level.

Table 5. Final values of operating parameters.

Configuration	Parameter					DEF ($\times 10^6$)	
	SAN	STN1	STN2	STN3	GWN	Initial[a]	Final
1	17	15	4	8	9	2.295	2.282
2	16	17	7	7	10	2.242	2.196
3	15	18	6	2	9	2.600	2.579

[a] Parameter values for initial factor set are given in Table 3.

Table 6. Comparison of bus activity for three configurations.

Item	Configuration 1		Configuration 2	Configuration 3
	5-Min Delay	10-Min Delay		
Service Area				
Trip demand	19 197	20 476	18 308	16 987
Buses dispatched	1383	1222	1057	977
General pickup, %	18	6.4	26	40.8
Dedicated pickup, %	78.4	88.4	72	57
Station				
Buses to guideway	1383	1059	955	825
No stop allowed, %	18.1	32.1	60.9	80.6
Preassigned stop, %	58.4	51.5	30.3	16.5
Intermediate stops				
Number	942	651	340	152
Percent	68	61.5	35.6	18.4
Buses to CBD	228	232	223	204
Guideway				
Buses on highest volume link	64	60	86	74
Buses on next highest volume link	63	58	55	60

Table 7. Trip times for three configurations.

Item	Configuration 1		Configuration 2	Configuration 3
	5-Min Delay	10-Min Delay		
Travel time, min				
Service area	18.8	18.9	22.9	29.5
Station outbound-inbound	1.2	1.2	0.9	0.8
Guideway	12	12	10.8	13.1
CBD	7.9	7.9	7.9	7.9
Intermediate stops[a]	1	0.9	0.9	0.8
Wait time, min				
Service area	10.1	10.5	12.0	15.1
Station[b]	1.6	2.2	1.4	0.9
Total time, min				
Wait time	11.3	12	12.7	15.4
CBD travel	27	26.9	27.1	32.5
CBD total	38.3	38.9	39.8	47.9
Non-CBD travel	32.8	32.7	34.9	43.6
Non-CBD total	44.1	44.7	47.6	59

[a] Reduced by fraction of total trips making intermediate stop.

[b] Reduced by fraction of total trips stopping at station.

Table 8. Effects of station capacity restrictions.

Item	Station Capacity Factor				
	1000	2500	3500	4500	5500
Total ramps	255	132	92	67	55
DEF ($\times 10^6$)	2.196	2.207	2.206	2.216	2.229
Average number of buses stopping at stations					
Outbound	74.6	83.2	94.7	110.6	130.2
Inbound	23.7	26.4	28.5	32.0	34.4
Average times, min					
Outbound[a]	0.63	0.70	0.77	0.80	1.01
Inbound[b]	0.33	0.36	0.37	0.41	0.42
Intermediate stop	0.88	0.98	1.10	1.32	1.50
Average bus loadings	0.793	0.792	0.793	0.797	0.796
Bus requirements	3618	3641	3637	3652	3694
Operator requirements	2456	2465	2450	2454	2446

[a] Includes only buses at final destination station.

[b] Includes only buses at origin station.

cess capacity in station and CBD areas. Station capacity can always be manipulated by changing the number of ramps or the capability to process buses in parallel. To allow simultaneous changes at all stations, a capacity factor was used that simply states that there is at least one ramp for every X trip origins or destinations at a station. In the initial runs this factor was set at 1000 trips/ramp; it was then increased for configuration 2 from 1000 to 5000, in increments of 500. The effect on station capacity should be minimal until a certain point is reached. That point is difficult to estimate precisely, as can be seen from the selected data in Table 8; definitely poorer results were realized sometime after overall system capacity was reduced to between 92 and 55 ramps. In Table 8, the increased bus requirements appear to be roughly in agreement with the increased number of buses waiting in the station queue. In fact, capacity restriction appears only to affect station-related activity. The tabulated results show that both bus loadings and operator requirements remain virtually unchanged, outside of statistical variations.

SUMMARY AND CONCLUSIONS

Simulation Results

The main objective of this project was to demonstrate the applicability of the simulation methodology to a large-scale system. There was no attempt to make as thorough a study of the dual-mode networks as would probably be made if a specific design were to be recommended for implementation. But it is possible to draw some conclusions about dual-mode operations in Milwaukee.

The major items examined in this report were the interaction between operating policies and system design, the effects of network size, and the effect of restricting station capacity. The results were largely predictable. All the networks within a reasonable range appeared to be relatively insensitive to changes in operating conditions. Parameters affecting the number of intermediate stops seemed to be most critical. More intermediate stops and lower bus loadings were realized in the larger networks. No clear superiority was found between the 41-station and the 22-station networks, but the 13-station network (configuration 3) was definitely inferior. Effects of restricted station capacity were interesting in that, although the number of buses within stations increased by 40 percent, there appeared to be little or no effect elsewhere in the system.

Bus and operator requirements are important outputs of such a simulation but, because of steady-state conditions, the results reported here would be higher than expected in a real-world dual-mode system. The simulation did provide upper bounds for these data and did show that configuration 2 required fewer buses and operators than the other networks to service nearly the same demand levels.

Methodology

In designing large-scale, service-oriented transportation systems it is important to examine alternative system operations, to isolate critical factors and bottlenecks, and to identify the cost versus service trade-offs that may exist. The methodology described here is, for the most part, capable of meeting these objectives. The major drawbacks of the simulation procedures are the time required to prepare input data and the extremely high computer costs. There are, however, methodological refinements that can help to overcome some of these problems.

Substantial computer time savings can be realized through more efficient data-collection procedures. It would undoubtedly have been useful in this case to experiment with the time period for data collection. A shorter interval in data collection would tend to increase the statistical variance of the results but at the same time allow for more replications and thus result in better estimates. The 10-min intervals used in this simulation were probably longer than necessary.

One means of reducing total experimentation and thus computer time would be to integrate into the simulation procedure some analytical models—perhaps queueing models to generate station-related inputs. An iterative procedure has already been suggested (10) that would tie queueing into the simulation. Similarly, service-area travel-time inputs can be better estimated by integrating dial-a-ride simulations into this model. Further consideration might also be given to developing better ways of dispatching buses and how these could be incorporated into the simulation.

There was no attempt in this project to examine the non-steady-state behavior of the dual-mode system. Data on a whole day's operation would definitely be of interest and could be easily generated. In analysis of non-steady-state conditions, data collection must span an entire day or at least an entire a.m. or p.m. peak demand period. It would take much longer to obtain a sample result under these conditions; most of the initial experimentation would thus have to be done under steady-state conditions.

The work described here represents an exercise whose main purposes were to see what type of results were obtainable from a large-scale simulation and to examine the practicality of the methodology. It is important to note that, in applying such a methodology to the real design of a dual-mode system, the objectives must be much more specific and the simulation program must be able to meet those objectives. The scope of such a project must be precisely stated and the project plans carefully formulated.

ACKNOWLEDGMENTS

This report was prepared as part of a program of research and training in urban transportation at Marquette University sponsored by the Urban Mass Transportation Administration, U.S. Department of Transportation. The conclusions are not necessarily those of the sponsoring agency.

REFERENCES

1. Demand-Responsive Transportation Systems. TRB, Special Rept. 136, 1973.
2. Demand-Responsive Transportation. TRB, Special Rept. 147, 1974.
3. A. J. Sobey and J. W. Cone. The Case for Personal Rapid Transit. HRB, Highway Research Record 367, 1971, pp. 70-90.
4. L. M. Cole, ed. Tomorrow's Transportation—New System for the Urban Future. U.S. Department of Housing and Urban Development, 1968.
5. Dual Mode Transit Systems. Rohr Industries, Inc.; Urban Mass Transportation Administration, U.S. Department of Transportation, Sept. 1974.
6. Milwaukee County Dual Mode Systems Study: Socio-Economic Evaluation. Urban Mass Transportation Administration, U.S. Department of Transportation, Vol. 3, Dec. 1971.
7. T. J. Heintz and A. Chatterjee. A Simulation Model for Planning Dual Mode Bus Systems. High Speed Ground Transportation, Vol. 10, No. 1, Spring 1976, pp. 31-46.
8. K. W. Heathington, J. Miller, R. R. Knox, G. C. Hoff, and J. Bruggeman. Computer Simulation of a Demand-Scheduled Bus System Offering Door-to-Door Service. HRB, Highway Research Record 251, 1968, pp. 26-40.
9. L. L. Howson and K. W. Heathington. Algorithms for Routing and Scheduling in Demand-Responsive Transportation Systems. HRB, Highway Research Record 318, 1970, pp. 40-49.
10. T. J. Heintz. An Application of a Large Scale Dual Mode Simulation Model to Milwaukee County. Marquette Univ., 1976.
11. R. H. Myers. Response Surface Methodology. Allyn and Bacon, Boston, 1971.
12. O. L. Davies. Design and Analysis of Industrial Experiments. Hafner, New York, 1956.

Publication of this paper sponsored by Committee on New Transportation Systems and Technology.

Sampling Procedures for Designing Household Travel Surveys for Statewide Transportation Planning

John F. DiRenzo and Robert A. Ferlis, Peat, Marwick, Mitchell and Company, Washington, D.C.
Philip I. Hazen, Federal Highway Administration, U.S. Department of Transportation

This paper describes sampling concepts and techniques that can be used to design household travel surveys for statewide transportation planning. Emphasis is placed on defining survey objectives in terms of the level of precision desired in estimating key variables. The need to incorporate cluster sampling treatments for trip-related variables collected in household surveys is introduced and discussed. Detailed procedures are presented for computing the minimum sample sizes of household interviews needed to accomplish survey objectives at minimum cost. A simplified method is described to account for losses in precision because of clustering. Application of the sampling techniques to an actual state travel survey design illustrates the influence of alternative levels of precision and data stratification on survey sample size. The applicability of these sampling concepts to other areas of transportation planning is discussed.

Over the next 5 to 7 years, many states expect to expand the modal and geographic scope of their statewide transportation planning programs. Such expansion has resulted in states having to collect and analyze many different types of regional, corridor, and statewide data. Many state agencies that have limited familiarity with travel surveys have been or are likely to be faced with the proglem of designing and conducting such surveys.

A problem common to both statewide and urban transportation planning is the use of rules of thumb or available funding resources as a basis for estimating survey sample sizes. Estimated sample sizes based on such factors may bear little relation to the desired level of precision of survey estimates, and they generally do not account for important sampling issues such as clustering or stratification, which influence sample size and other sampling parameters. Within the transportation planning field, little research has been done to identify potential trade-offs between sample size (and survey costs) and the selection of desired tolerance levels, confidence levels, and geographic levels for which data are needed. These factors can significantly influence survey cost. In addition, many transportation planners have limited familiarity with or experience in applying statistical sampling procedures in the design of travel surveys.

This paper focuses on sampling techniques and concepts that can be used by transportation planners to design household travel surveys (e.g., home interview) for statewide transportation planning. Such surveys could include household surveys conducted on a statewide or regional scale or for a selected geographic subarea within a state. The specific objectives of this paper are to

1. Present an overview of sampling concepts applicable to the design of household surveys for statewide transportation planning,
2. Document statistical sampling techniques to be used in household surveys to determine the sample sizes needed for estimating such commonly used variables as trip generation rates and average trip lengths, and
3. Illustrate trade-offs between survey sample sizes and alternative levels of precision as well as the effect on sample size of stratifying survey variables by geographic area and socioeconomic characteristics of households within areas.

Sample size estimates developed for a sampling plan for a statewide household travel survey in Connecticut are used to illustrate the points noted in the above objectives (1). Although nonsampling biases can also impact the level of precision achieved in a survey, they are treated in detail elsewhere (2) and thus are not discussed here. The research effort on which this paper is based also developed sampling procedures and survey designs for roadside and modal (intercity bus and passenger train) surveys likely to be used for statewide transportation planning (2).

SURVEY DESIGN CONCEPTS AND ISSUES

Because of budget and data constraints and lack of familiarity with sampling procedures, rigorous statistical evaluation of sample sizes and alternative survey procedures often is not performed before a travel survey is conducted. In spite of such real-world problems, it is still important to develop survey designs and sampling procedures on a sound statistical basis. Application of valid sampling procedures can strengthen a travel survey program, particularly by providing a quantitative basis for evaluating trade-offs between the scope, precision, coverage, and cost of a travel survey. Application of statistical sampling procedures and concepts makes it possible to design travel surveys that, within available funding resources, provide data at the required levels of precision.

Preparation of a Survey Design

The first and most critical step in developing any type of travel survey design is to specify the survey objectives. Survey objectives must be clearly and specifically defined in the design process. One expert in survey design has suggested (3) that survey objectives should

1. Specify how the survey results will be used in the decision-making process;
2. Identify the variables of interest, the content and extent of the survey population, and the classification that will be used to analyze the results (e.g., trip purpose, socioeconomic groupings);
3. Identify desired or minimum levels of precision and the geographic areas for which such precision is to be maintained;
4. Specify how the variables are to be measured, coded, and processed; and
5. Identify how the data will be analyzed.

All of these procedures are necessary to develop a sampling plan for a travel survey. Inaccuracy or incompleteness in any of these areas could reduce the usefulness and possibly the precision of survey data.

Selecting a final set of survey objectives and a final sampling plan is likely to be an iterative process. One or more of the initial survey objectives may have to be modified to develop a feasible design, given the financial and staff resources and time deadlines of the study.

Factors Influencing Sample Size

The required sample size in a survey is related to (a) the desired level of precision of survey estimates, (b) the variance of characteristics of interest within the population, (c) the size of the population to be sampled, and (d) the procedure used to select the samples. The influence of the first three factors on sample size is shown by the following equation for estimating the sample size of a simple random sample without replacement (4):

$$n_0 = [t^2_{(1-\alpha/2)} s^2/d^2] / \{1 + (1/N)[t^2_{(1-\alpha/2)} s^2]/d^2\} \qquad (1)$$

where

n_0 = number of households to be sampled to estimate the mean of the sample at a specified level of precision;
$t_{(1-\alpha/2)}$ = Student's t-value at level of confidence $(1 - \alpha)$;
s = standard deviation of the sample observations about the sample mean;
d = acceptable difference (±) between the sample mean and the population mean (tolerance level); and
N = total number of elements in the population.

Level of significance (α) means that the sample estimate will fall outside the specified tolerance level with a probability (α). Assuming a symmetrical distribution of the sample estimate about the mean, this implies that an observation will lie above the range with a probability ($\alpha/2$). Therefore, for a given level of confidence ($1 - \alpha$, where α is the level of significance), the Student's t-value for $1 - \alpha/2$ should be used in Equation 1. This corresponds to a two-tail t-value.

The term level of precision here consists of the variables (d) and (α) in the above equation. For example, if a state wishes to estimate the average number of automobile driver trips per household for all households within the state, one possible level of precision is that the estimate be within ±10 percent of the population mean 95 percent of the time. The choice of a 95 percent confidence level indicates that, for the use to which the estimate will be put, a 1-in-20 chance of the mean trip-rate estimate from the sample lying outside a 10 percent tolerance level of the true value is acceptable. The level of precision specified for a survey estimate does not account for nonsampling errors that may affect survey estimates.

As shown in later sections, the selection of d and a confidence level, i.e., $1 - \alpha$, has a substantial influence on the required survey sample size, which is directly proportional to the level of confidence specified in the survey and inversely proportional to the acceptable error range of the sample mean. The variance (s^2) of sample observations about the sample mean also has a direct relation to sample size. As the variance of the characteristic to be sampled increases, the required sample size will also increase for a given level of precision.

Equation 1 also illustrates the important condition that, as the size of the population to be sampled (N) increases, its influence on the required sample size decreases and becomes negligible if N is large relative to n. Because the sampling fraction in most statewide household travel surveys is typically smaller than 1 percent, a simplified equation that consists only of the numerator [$n_0 = (ts/d)^2$] can usually be used with little or no loss in accuracy. The above relationships between sample size and tolerance levels, confidence levels, sample variance, and population size are applicable to sampling procedures other than those for a simple random sample.

Implicit in the equation is the specification of a geographic area for which the desired survey estimates are required, i.e., the areal unit of analysis. The selection of a geographic area for which data are to be obtained is likely to have a major impact on the number of samples required in the survey and thus on survey cost. For many variables, the sample size required to estimate the mean or the proportion of elements with a particular characteristic at a specified level of precision is likely to be similar at the regional and county levels to that required at the state level. This has potentially significant implications for the cost of conducting statewide surveys.

Sampling Techniques for Household Surveys

In many conventional statewide and urban household travel surveys, households in the sample are typically considered to have been selected by simple random samples. However, many of these surveys are, in total or in part, cluster samples. As noted by Kish (3), "sample elements are the units for which information is sought." In a cluster sample, each sample unit contains more than one sample element; in an element sample each sample unit contains only one sample element.

Table 1 gives a list of typical variables collected in household surveys and identifies the type of sample associated with each variable. In examples 1 and 2, the sample element (the unit for which information is sought) is the household. Specifically, the number of person trips produced and the number of automobiles owned by the household are of interest. The sampling units are the same as the sample elements, i.e., households, which indicates that element sampling procedures such as the simple random sample should be used to estimate required sample sizes. In example 3, the sample elements are person trips. Each sampling unit (household) thus potentially contains more than one sample element. Therefore, cluster sampling procedures should be applied to estimate sample sizes for trip-related variables. Examples 4 through 6 illustrate other variables that should be treated as cluster sampling problems.

Table 1. Variables in element and cluster samples.

Example	Type of Sample	Sampling Unit	Sample Element	Variable
1	Element	Households	Households	Person trips per household
2	Element	Households	Households	Automobiles owned per household
3	Cluster	Households	Person trips	Average length of person trip
4	Cluster	Households	Person trips	Proportion of person trips by purpose
5	Cluster	Households	Automobile driver trips	Average length of automobile driver trip
6	Cluster	Households	Automobile driver trips	Average automobile occupancy

Estimating minimum sample size for cluster samples is more complex. A factor commonly used to simplify the estimation of sample size for cluster or other complex samples is the design effect. According to Kish (3), the design effect (D) is "the ratio of the actual variance of a sample to the variance of a simple random sample of the same number of elements." This factor is calculated as follows:

$$D = \mathrm{var}(\bar{y})/[(1 - f)s^2/n] \qquad (2)$$

where

$\mathrm{var}(\bar{y})$ = variance of the sample mean calculated for a particular sampling procedure such as cluster sampling,
f = proportion of elements in the population that are sampled (called the sampling fraction),
s^2 = sample variance of a simple random sample (i.e., $s^2 = \Sigma(y_i - \bar{y})/(n - 1)$ about the sample mean, and
n = number of elements sampled.

The denominator in Equation 2 is the variance of the mean of a simple random sample.

The design effect provides a means of accounting for the effects of clustering on sample size. Kish (3) suggests the following approach, which uses the design effect and the sample size for a simple random sample to estimate sample size for complex problems such as cluster samples:

$$n = n_0 D \qquad (3)$$

where

n = number of elements to be sampled to estimate the sample mean at a specified level of precision in a cluster sample,
n_0 = number of elements to be sampled to estimate the sample mean at a specified level of precision for a simple random sample, and
D = design effect as defined above.

The significance of this concept is illustrated in the following example. Data collected in a Kentucky statewide household survey gave a mean trip length of 16 km (9.94 miles) and a sample standard deviation (s) of 34.4 km (21.3 miles). The required sample size to estimate mean trip length within ±10 percent (d) at a 90 percent level of confidence for a simple random sample was

$$n_0 = [t^2_{(1-\alpha/2)} s^2/d^2] = (1.645)^2\,(21.3)^2/[(0.10)(9.94)]^2 = 1250 \text{ trips} \qquad (4)$$

However, based on an analysis of survey results, the actual variance of mean trip length for a cluster sample was estimated to be 2.39 times as large as the variance of the sample mean trip length, assuming a simple random sample of trips (D = 2.39). Therefore, estimating the mean trip length at the same level of precision specified above would require sampling almost 3000 trips, as estimated below:

$$n = n_0 D = 1250 \text{ trips} \times 2.39 = 2995 \text{ trips} \qquad (5)$$

TECHNIQUES FOR ESTIMATING SAMPLE SIZES

The following discussion of simple random sampling and cluster sampling formulas presents formulas for two illustrative categories of variables that are of general interest to practicing transportation planners: (a) person-trip generation rates stratified by trip purpose, geographic area, and socioeconomic characteristics of households; and (b) average lengths of person trips stratified by purpose and geographic area.

Person-Trip Generation Rates

The general formula for estimating the minimum sample size of completed interviews needed to estimate the average number of person trips of purpose (p) per household at a desired level of precision, if a simple random sample of households is selected within each geographic area of interest (e.g., state, county, or traffic zone), is as follows:

$$n_p = [t^2_{(1-\alpha/2)}(s_p^2/d_p^2)]/\{1 + (1/N)[t^2_{(1-\alpha/2)} s_p^2/d_p^2]\} \qquad (6)$$

where

n_p = number of completed household interviews required to estimate the person-trip generation rate for trip purpose (p) for the geographic area of interest,
$t_{(1-\alpha/2)}$ = value of Student's t-statistic for level of confidence $(1 - \alpha)$;
s_p = estimated standard deviation of the person-trip generation rate for trip purpose (p) for the geographic area of interest,
d_p = acceptable error (or difference) between the estimated person-trip generation rate for trip purpose (p) and the true trip generation rate for purpose (p) for the geographic area of interest, and
N = total number of households in the geographic area of interest.

Average person-trip generation rates by trip purpose and corresponding standard deviations can be estimated by using the computer program XCLASS in the Federal Highway Administration (FHWA) urban transportation planning battery and the results of previously conducted household travel surveys. The total number of households (N) in each geographic area can be estimated from secondary sources such as the 1970 census.

Transportation planners often wish to estimate person-trip generation rates by trip purpose for households stratified by household income, automobile availability, or other socioeconomic variables. The formula for calculating required sample sizes to estimate such trip generation rates is essentially the same as above except that n_p, s_p, d_p, and N must be redefined as follows:

n_{ph} = number of completed household interviews with characteristic (h) (e.g., one automobile required to estimate the person-trip generation rate for trip purpose (p) for households with characteristic (h) for the geographic area of interest,
s_{ph} = estimated standard deviation of the person-trip generation rate for trip purpose (p) for households with characteristic (h) for the geographic area of interest,
d_{ph} = acceptable error (or difference) between the estimated person-trip generation rate for trip purpose (p) and the true trip generation rate for purpose (p) for households with characteristic (h) for the geographic area of interest, and
N_h = total number of households with characteristic (h) within the geographic area of interest.

The FHWA program XCLASS can again be used to estimate person-trip generation rates and standard deviations about the rates by trip purpose and type of household for each geographic area of interest.

Average Person-Trip Lengths by Town Class

As previously noted, the procedures for estimating required sample sizes for measuring average person-trip lengths at a given level of precision are analogous to but more complex than those for measuring trip generation rates. The formula for estimating the minimum sample size of households to estimate the average trip length for person trips of purpose (p) at a desired level of precision within each geographic area of interest is

$$n_p = D_p/\bar{x}_p \{[t^2_{(1-\alpha/2)}(s^2_p/d^2_p)]/[1 + (1/N)t^2_{(1-\alpha/2)}(s^2_p/d^2_p)]\} \quad (7)$$

where

n_p = number of completed household interviews required to estimate the average person-trip length for trips with purpose (p) for the geographic area of interest,
D_p = computed design effect for trip purpose (p),
$\bar{x}_p$ = person-trip generation rate for trip purpose (p) for the geographic area of interest,
$t_{(1-\alpha/2)}$ = value of Student's t-statistic for level of confidence $(1 - \alpha)$,
s_p = estimated standard deviation of the average person-trip length for trips of purpose (p) for the geographic area of interest,
d_p = acceptable error (or difference) between the estimated average and the true average trip length for person trips of purpose (p) for the geographic area of interest, and
N = total number of households within the geographic area of interest.

The above formula differs from the single random formula in that the design effort (D_p) compensates for the clustering of trips made by sampled households and the average person-trip generation rate for purpose (p) (i.e., $\bar{x}_p$) is included in the formula to estimate sample size in terms of households, not trips. The average trip length for person trips of purpose (p) and the standard deviations (s_p) about the average trip lengths can be estimated by using the XCLASS program.

The design effect is the ratio of the variance of the mean trip length, computed by using clustered sampling assumptions, to the variance of the mean (assuming selection of a simple random sample of trips). The design effect for a cluster sample of person trips derived from a simple random sample of households within each geographic area of interest is computed from the following:

$$D_p = \text{var}(r_p)/\text{var}(r_p)_0 \quad (8)$$

in which

$$r_p = y_p/x_p = \sum_j^n y_{pj} \Big/ \sum_j^n x_{pj} \quad (9)$$

$$\text{var}(r_p) = [(1 - f)/x^2_p]\,[n/(n - 1)]\left\{\left[\left(\sum_j^n y^2_{pj} - y^2_p\right)\Big/ n\right] + r^2_p\left[\left(\sum_j^n x^2_{pj} - x^2_p\right)\Big/ n\right] - 2r_p\left[\sum_j^n y_{pj} \times x_{pj} - (y_p \times x_p)\Big/ n\right]\right\} \quad (10)$$

$$\text{var}(r_p)_0 = \{(1 - f)/[x_p(x_p - 1)]\}\left(\sum_k^x y^2_{pk} - y^2_p\right)\Big/ x_p \quad (11)$$

where

D_p = computed design effect for purpose (p) for the geographic area of interest,
$\text{var}(r_p)$ = variance of the mean person-trip length for purpose (p) under cluster sampling assumptions,
$\text{var}(r_p)_0$ = variance of the mean person-trip length for purpose (p) under simple random sampling assumption,
r_p = average trip length for person trips of purpose (p) for the geographic area of interest,
y_p = total kilometers recorded in the sample for the geographic area of interest for person trips of purpose (p),
x_p = total person trips of purpose (p) recorded in the sample for the geographic area of interest,
y_{pj} = total kilometers recorded for all person trips of purpose (p) made by household (j) in the geographic area of interest,
x_{pj} = total number of person trips of purpose (p) made by household (j) in the geographic area of interest,
f = proportion of survey households sampled in the geographic area of interest,
n = number of households sampled, and
y_{pk} = total distance for trip (k) of purpose (p) in the geographic area of interest.

APPLICATION OF SAMPLING PROCEDURE

The sampling procedures were used to design a sampling plan for a statewide household travel survey for the Connecticut Department of Transportation (ConnDOT). The key variables to be estimated in the survey included (a) household person-trip generation rates stratified by town class, trip purpose, and socioeconomic characteristics of households and (b) average person-trip length stratified by town class and trip purpose. The term town class refers to the stratification of the 169 Connecticut towns into three classes based on residential development and transit use and level of service.

Rather than attempting to specify desired levels of precision to be achieved before determining sample size, ConnDOT suggested three tolerance levels (5, 10, and 25 percent) and three confidence levels (68, 90, and 95 percent) for which sample size estimates were to be developed. Sampling parameters were computed from a Connecticut statewide travel survey conducted in 1964.

Figure 1 shows, for a statewide simple random sample, the number of completed household interviews required to estimate the average number of person trips per household at various levels of confidence and tolerance. It can be seen in the figure that the choice of confidence and tolerance levels greatly influences the required sample size. The analysis also showed that far fewer completed interviews are required to estimate statewide person-trip generation rate than are required to estimate average person-trip length at comparable levels of precision. For example, only 216 completed interviews are required to estimate the true statewide person-trip generation rate within ±10 percent of the estimated rate at a 90 percent level of confidence, but 460 completed interviews are required to estimate the average statewide person-trip length at the same level of precision. These sample sizes represent, respectively, 0.02 and 0.05 percent of the

Figure 1. Simple random sample sizes for estimating statewide person-trip generation rate at various confidence and tolerance levels.

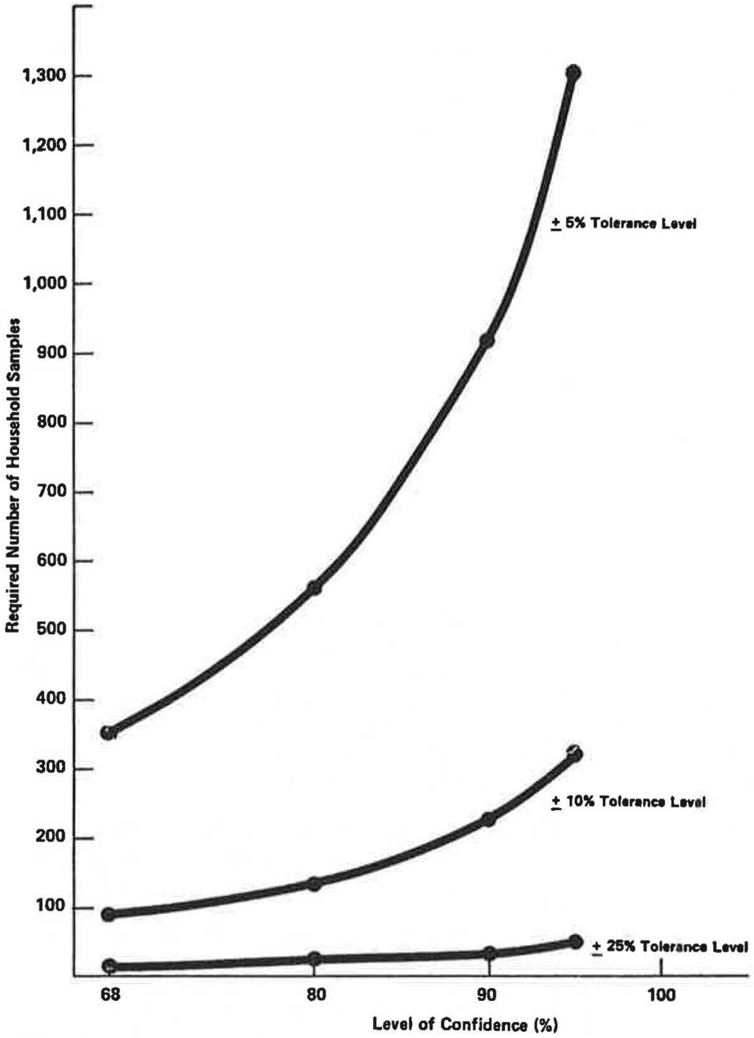


Table 2. Sample size required to estimate survey variables for households stratified by automobile ownership and tolerance levels.

Variable	Tolerance Level About Mean (%) ±5 Completed Interviews	±5 Total Sample	±10 Completed Interviews	±10 Total Sample	±25 Completed Interviews	±25 Total Sample
Home-based work person trips per household						
0 automobile	1 889	23 037	523	6 378	84	1024
1 automobile	660	1 473	165	368	30[a]	67[a]
2 automobiles	638	1 623	159	405	30[a]	76[a]
≥3 automobiles	379	4 922	95	1 234	30[a]	390[a]
Home-based nonwork person trips per household						
0 automobile	2 674	32 610	746	9 098	124	1512
1 automobile	1 114	2 487	279	623	46	103
2 automobiles	733	1 865	183	466	30[a]	76[a]
≥3 automobiles	607	7 883	157	2 039	30[a]	390[a]
Non-home-based person trips per household						
0 automobile	12 172	148 439	2693	32 841	509	6207
1 automobile	2 184	4 875	979	2 185	157	350
2 automobiles	3 935	10 013	1447	3 683	232	590
≥3 automobiles	928	12 052	327	4 247	52	675

Note: Data are for town class 2 at a 90 percent level of confidence.

[a] Minimum of 30 samples required.

Table 3. Sample size required to estimate survey variables for households stratified by household size and income.

	Household Income Class (1964 dollars)					
	0 to 4999		5000 to 6999		≥7000	
Variable	Completed Interviews	Total Sample	Completed Interviews	Total Sample	Completed Interviews	Total Sample
Home-based work person trips per household						
1-person households	720	8 000	58	2 762	281	16 529
2-person households	696	8 700	95	1 727	121	834
3- and 4-person households	291	6 326	130	1 831	107	421
≥5-person households	252	14 000	149	4 382	125	744
Home-based nonwork person trips per household						
1-person households	555	6 167	242	11 524	195	11 471
2-person households	338	4 225	271	4 927	170	1 172
3- and 4-person households	198	4 304	206	2 901	170	669
≥5-person households	197	11 500	203	5 971	151	899
Non-home-based person trips per household						
1-person households	2671	29 678	813	38 714	387	22 765
2-person households	1111	13 888	1052	19 127	452	3 117
3- and 4-person households	801	17 413	671	9 451	1829	7 201
≥5-person households	1288	71 556	918	27 000	464	2 762

Note: Data are for town class 2 at a 90 percent level of confidence.

estimated 933 050 households in Connecticut.

The table below and Tables 2 and 3 give data showing the influence of stratifying households by geographic area and socioeconomic characteristics. These tables, which were developed for suburban towns in Connecticut (town class 2), are based on an assumed 90 percent level of confidence and the indicated tolerance levels.

The following table shows the number of completed household interviews required to estimate the average person-trip generation rates and trip lengths by trip purpose within town class 2 at a 90 percent level of confidence (a minimum of 30 samples was required):

	Completed Interviews at Tolerance Levels of		
Variable	±5%	±10%	±25%
Person trips per household			
Total	883	221	35
Home-based work	712	178	30
Home-based nonwork	1098	273	44
Non-home-based	6084	1521	243
Average trip length			
Home-based work	1656	414	66
Home-based nonwork	2256	564	90
Non-home-based	6536	1633	261

Approximately the same number of completed interviews were estimated to be needed in each of the other town classes in the state. The sample size estimates for average trip length in each town class were developed on the basis of design effects computed from the 1964 Connecticut statewide household survey. Design effects for the three town classes ranged between 1.7 and 2.1 for home-based work trips, between 2.8 and 3.7 for home-based nonwork trips, and between 2.9 and 3.8 for non-home-based trips.

Tables 2 and 3 (1) give sample sizes of households for estimating person-trip generation rates for town class 2 for households stratified by automobile ownership and household size and income, respectively. These tables give both the estimated number of completed interviews for households having given socioeconomic characteristics and the total number of households that would have to be sampled to locate the required number of households having the indicated socioeconomic characteristics. For example, if 200 completed interviews were required to estimate a person-trip generation rate for households of five or more persons in a particular income group, and if such households represented 10 percent of all households in this town class, a total of 2000 households would have to be randomly sampled to locate 200 households having the desired characteristic.

These tables illustrate two of the important considerations in designing a household travel survey.

1. The stratification of households into detailed geographic or socioeconomic strata may require that, if households are randomly sampled, a large number of households (i.e., total samples) be contacted to locate households having the desired characteristics. Screening households may help to reduce survey costs if specific types of households must be sampled.
2. The sample size estimates also show that substantial numbers of completed interviews are required in each data stratification. Depending on the approach, substantially more than 30 completed interviews may be required for each data stratification if person-trip generation rates are to be measured at ±5 or ±10 percent tolerance levels at a 90 percent level of confidence.

CONCLUSIONS

This paper presents an approach that may be used to determine the sample size needed to achieve specific objectives in a household travel survey. The following points are of particular importance.

1. Computation of minimum sample size should be based on the attainment of specific survey objectives. These objectives should be translated into the desired level of precision to be achieved in estimating individual survey variables.
2. Stratifying survey variables by geographic region or household characteristics can result in a substantially larger sample size. In many cases the minimum sample size needed to develop a statewide estimate will be approximately the same as that for a single subarea.
3. Trip-related survey variables must generally be treated by using cluster sampling procedures in a household survey. The design-effect correction factor should be used to account for the impact of clustering in computing sample size. Failure to use this procedure in determining sample size may result in travel

information that is insufficient to achieve survey objectives or in unnecessarily high survey costs.

4. Although the sampling approach described was developed in support of statewide transportation planning needs, these procedures are equally applicable to urban and regional household travel surveys.

ACKNOWLEDGMENT

This paper is based on the results of research sponsored by the Federal Highway Administration, U.S. Department of Transportation.

REFERENCES

1. Sampling Plan for the Connecticut Statewide Household Travel Survey. Peat, Marwick, Mitchell and Co.; Federal Highway Administration, U.S. Department of Transportation, April 1976.
2. Guidelines for Designing Travel Surveys for Statewide Transportation Planning. Peat, Marwick, Mitchell and Co.; Federal Highway Administration, U.S. Department of Transportation, March 1976.
3. L. Kish. Survey Sampling. Wiley, New York, 1967.
4. W. G. Cochran. Sampling Techniques. Wiley, New York, 1963.

Publication of this paper sponsored by Committee on Transportation Information Systems and Data Requirements.