TRANSPORTATION RESEARCH RECORD 818

Design of Public Transport Services

TRANSPORTATION RESEARCH BOARD

COMMISSION ON SOCIOTECHNICAL SYSTEMS
NATIONAL RESEARCH COUNCIL

NATIONAL ACADEMY OF SCIENCES
WASHINGTON, D.C. 1981

Transportation Research Record 818
Price $5.60
Edited for TRB by Naomi Kassabian

modes
1 highway transportation
2 public transit

subject areas
12 planning
13 forecasting

Library of Congress Cataloging in Publication Data
National Research Council (U.S.). Transportation Research
Board. Meeting (60th: 1981: Washington, D.C.)
Design of public transport services.

(Transportation research record; 818)
Reports presented at the 60th annual meeting of the
Transportation Research Board.
1. Local transit–United States–Congresses. I. Title.
II. Series.
TE7.H5 no. 818 380.5s [388.4'068] 81-22280
[HE4451]
ISBN 0-309-03259-8 ISSN 0361-1981 AACR2

Sponsorship of the Papers in This Transportation Research Record

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

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

Committee on Traveler Behavior and Values
David T. Hartgen, New York State Department of Transportation, chairman
M.E. Beesley, Nils Anders Bruzelius, Katherine Patricia Burnett, Melvyn Cheslow, Ricardo DePaul Dobson, Michael Florian, Thomas F. Golob, David A. Hensher, Frank S. Koppelman, Peter S. Liou, Jordan J. Louviere, Daniel McFadden, Arnim H. Meyburg, Richard M. Michaels, Robert E. Paaswell, Shalom Reichman, Earl R. Ruiter, Peter R. Stopher, Antti Talvitie, Mary Lynn Tischer, Peter L. Watson

Public Transportation Section
George Edward Gray, California Department of Transportation, chairman

Committee on Paratransit
Ronald F. Kirby, The Urban Institute, chairman
Sandra Rosenbloom, University of Texas at Austin, secretary
Michael Albarelli, James A. Bautz, M. Douglas Birnie, R.P. Braun, Roy M. Coughlin, Lawrence N. Dallam, Frank W. Davis, Jr., Michael E. Dewey, Stanley G. Feinsod, Richard V. Gallagher, Gorman Gilbert, Christine M. Johnson, Roy E. Lave, Jr., Dan V. Maroney, Jr., Lew W. Pratsch, Barbara K. Reichart, Joseph S. Revis, Robert E. Samuels, Clarence J. Shallbetter, S. Ling Suen, Roger F. Teal, Nigel H.M. Wilson

James A. Scott, Transportation Research Board staff

Sponsorship is indicated by a footnote at the end of each report. The organizational units, officers, and members are as of December 31, 1980.

Contents

Authors of the Papers in This Record

Abkowitz, Mark, Rensselaer Polytechnic Institute, Troy, NY 12181
Furth, Peter G., Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139
Gelb, Pat M., DeLeuw, Cather and Company, 120 Howard Street, San Francisco, CA 94120
Lago, Armando M., Ecosometrics, Inc., 4715 Cordell Avenue, Bethesda, MD 20014
Mayworm, Patrick, Ecosometrics, Inc., 4715 Cordell Avenue, Bethesda, MD 20014
McEnroe, J. Matthew, Ecosometrics, Inc., 4715 Cordell Avenue, Bethesda, MD 20014
Newman, Debra A., SYSTAN, Inc., P.O. Box U, 343 Second Street, Los Altos, CA 94022
Rosenbloom, Sandra, School of Architecture, University of Texas at Austin, Austin, TX 78712
Sharfarz, David, Rochester-Genesee Regional Transportation Authority, 55 St. Paul Street, Rochester, NY 14604
Turnquist, Mark A., School of Civil and Environmental Engineering, Hollister Hall, Cornell University, Ithaca, NY 14853
Wilson, Nigel H. M., Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139

Setting Frequencies on Bus Routes: Theory and Practice

PETER G. FURTH AND NIGEL H.M. WILSON

Since most transit systems have relatively stable route structures and politically determined levels of subsidy, one of the main recurrent decisions the transit planner must make is the service frequencies to be provided on each route in the system. Current practical and theoretical approaches to this problem are reviewed and, in light of their seeming inadequacies, a new model for setting frequencies is developed. The model allocates the available buses between time periods and between routes so as to maximize net social benefit subject to constraints on total subsidy, fleet size, and levels of vehicle loading. An algorithm is developed to solve this nonlinear program that can be applied by using a small computer program or, simplified in some generally acceptable way, by using a pocket calculator. In a case study the model is shown to produce results quite different from the existing allocation, which suggests changes that are insensitive to the specific set of parameters and objectives. It is shown that the model can readily be applied to evaluate the impacts of an alternative vehicle capacity and to investigate the value of changing service policies.

The North American public transit industry has, in the past decade, emerged from a long period of stagnation and decline to become a major focus in strategic planning to deal with the energy problem. Increasing attention is being given to the problem of using the ever-mounting public resources being devoted to transit more efficiently. This attention has revealed an apparent enigma: Although there is a wealth of academic research on how transit planning should be done, methods in use in the transit industry are generally crude and dominated by the planners' experience and judgment, sometimes codified into simple rules of thumb.

In this paper, one important part of the short-range transit planning process is selected and used to investigate whether significant differences exist between current practice and reasonable theory (1). The topic is setting frequencies on bus routes, a problem that must be addressed, either explicitly or implicitly, several times each year by all transit operators. After a discussion of existing industry practice in setting frequencies, prior research is briefly reviewed. In light of the weaknesses identified in this prior work, a new model is proposed that accurately reflects the objectives and constraints with which the transit industry must deal. Finally, a case study of part of the Massachusetts Bay Transportation Authority (MBTA) system shows the differences between the actual allocation of buses and that suggested by the theory.

CURRENT PRACTICE

Methods used by schedulers to set frequencies on routes are generally poorly documented and seem to vary among operators. Typically, however, only a small number of rules of thumb have been used that can be overridden by the judgment and experience of the scheduler. The best way of assessing industry practice is to refer to the service standards that have been widely adopted by many operators in the past five years. Service standards cover a broad range of planning, operations, and management and (of interest here) usually include specific guidelines on service frequencies. These service standards are a result of both codification of existing rules of thumb and a statement of policy. As such they do not always accurately reflect decisions made by schedulers (and others) but are likely to include factors traditionally used in decision making.

Based on a survey of existing service standards (2), the most frequently used methods for setting frequencies are policy headways, peak-load factor, revenue/cost ratio, and vehicle productivity. Each of these is described briefly below.

Policy Headways

Policy headways are used by virtually all operators and serve as a lower bound on the frequency. Routes are categorized by factors such as orientation (radial or crosstown), function (line-haul or feeder), and location (urban or suburban); and each category is assigned a set of policy headways for each period of the day. Policy headways are most effective in systems that operate principally as a low-demand social service. However, in large cities, during peak hours, and whenever demand is high, policy headways lose their relevance and other methods must be used to assign headways.

Peak-Load Factor

The ratio of the number of passengers on board at the peak-load point to the seating capacity of the vehicle is widely used under heavier demand. A lower bound on frequency is based on maximum peak-load factors established by route category and time period. These factors are based on the physical capacity of the vehicle and on comfort and operational considerations.

Revenue/Cost Ratio

The revenue/cost ratio is often used to define an upper bound on the amount of service to be provided on a route. This ratio is a rough measure of efficiency and equity in the distribution of service and has the important advantage of being readily understood by both elected officials and the general public.

Vehicle Productivity

Either in the form of passengers per vehicle mile or per vehicle hour, vehicle productivity is also occasionally used to set upper bounds on the frequency. As in the case of the revenue/cost ratio, vehicle productivity is used to approximate the benefit/cost ratio of a specific service and to guard against inefficient allocation of resources.

Although service standards are an advance in the state of the art of transit planning, this brief review shows that they fall far short of ensuring that transit resources are allocated most efficiently (3). Specifically, the standards focus on upper and lower bounds for setting frequencies but say nothing about setting frequencies to maximize efficiency within these constraints.

To better understand how frequencies are actually established, a set of MBTA routes was analyzed and a set of empirical relationships tested by linear regression. The 17 routes analyzed all belong to the Arborway Garage of the MBTA and include a wide variety: radial and crosstown, high- and low-frequency, that serve affluent and poor neighborhoods. Results show that the midday frequencies are heavily constrained by the policy headways; only four routes have a higher frequency.

The following three empirical relationships for setting frequencies were tested:

Equal load factor (2-h peak):

$$Q = b(PLP)$$

$$Q = a + b(PLP) \quad (1)$$

Equal load factor (30-min peak):

$$Q = b(PLP)$$

$$Q = a + b(PLP) \quad (2)$$

Square-root rule:

$$\ln Q = a + b\ln(r/T) \quad (3)$$

where

- a,b = coefficients,
- Q = scheduled frequency (round trips/h),
- PLP = peak-load-point count (riders/h that pass peak-load point),
- r = ridership per hour (total boardings in both directions), and
- T = round-trip run time (min).

The most important results for the morning peak period are shown below (all estimates of coefficients are significant at the 99 percent level):

Empirical Relationship	Coefficients a	b	R^2
Equal load factor	-	0.024	0.85
(2-h peak)	1.54	0.020	0.93
Equal load factor	-	0.018	0.90
(30-min peak)	1.32	0.016	0.95
Square-root rule	0.43	0.72	0.82

It appears that existing frequencies are very well explained by setting the peak-load factor equal on all routes, particularly during the peak half-hour. The average peak load on these routes during the peak half-hour was 1.2, which is about 13 percent below the policy peak-load factor of 1.4.

This case study suggests that schedulers do follow a clear decision-making process, which revolves around the rule of an equal peak-load factor. These results are strikingly similar to those found by Morlok in Chicago (4); the important point is that in both cases the equal peak loads were significantly below actual bus capacity. As demonstrated in the next section, this fact makes the rule inefficient with respect to the optimization of passenger service.

PREVIOUS THEORY

The best-known theory for setting frequencies on bus routes is the square-root rule, which is based on the minimization of the sum of total passenger wait-time costs and total operator cost. In the general case when routes of different lengths exist, the rule states that the service frequency provided on a route should be proportional to the square root of the ridership per unit distance (or time) for that route (5).

Major weaknesses of the square-root rule, which explains its lack of acceptance by the industry, are that it does not consider bus capacity constraints and that it assumes that ridership is fixed and independent of the service frequency. Ignoring the capacity constraint means that on some heavily used routes not enough capacity will be provided (i.e., the solution is infeasible). The assumption of fixed demand means that the user benefits are limited to minimization of wait time, which is probably only a minor part of the public benefit of transit service.

A second, almost trivial, theory is that if the objective is simply to minimize operator cost, the frequencies should be set so that the capacity will equal the peak load on each route. If the system is at capacity, each route will have the same peak-load factors, but if the system is operating below capacity, efficiency arguments do not lead to equal peak-load factors.

Guinn (6) used a linear-programming approximation to allocate buses to maximize revenue subject to a fleet-size constraint. Although his objective and constraint set are too approximate for direct application, the model presented later in this paper uses the same general optimization framework. Scheele (7) proposed a more complex mathematical programming approach to determine optimal service frequencies in the long-run case in which the distribution of trips (but not total trip generation and production) is allowed to vary in response to the service provided.

Several models have been developed for the simultaneous choice of routes and frequencies (8-10). The frequency components of these models typically minimize passenger wait time subject to capacity constraints under an assumption of fixed demand. Recent work at the Volvo Bus Corporation (11) has resulted in a package for choosing routes and frequencies that has been successfully applied in numerous cities in Europe and elsewhere.

Most of these models and theories are designed for one-time application when the entire transit network is redesigned--by definition a major and infrequent undertaking. Furthermore, only one of these models has been applied frequently, and none has been accepted for use by transit operators. This is both because of their orientation to large-scale system change and because they are either complex and hard to use or crude and hard to believe. There is need for a model that accurately reflects the frequency-choice decision, that is simple enough in its data and application requirements to be used frequently by operators, and that focuses on small changes so it can be applied repeatedly over the years. Such a model is developed in the remainder of this paper.

PROPOSED MODEL

The fleet-allocation problem can be formulated as an optimization in which an objective function is maximized (minimized) subject to a set of constraints. Before the formulation is presented and discussed in detail, however, it is useful to illustrate the style of solution by using a simple example.

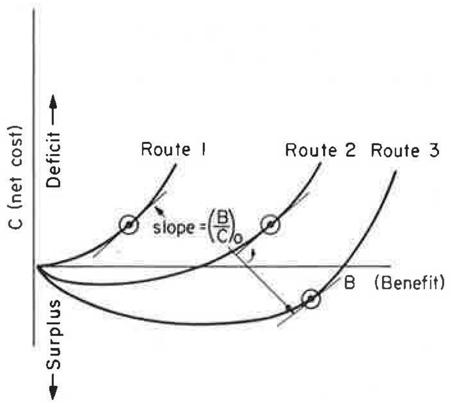
Suppose that a bus company operates three routes, charges a flat fare per passenger, and has allocated a fixed amount to cover the deficit that will result from providing the service. The single objective of the company is to maximize ridership by means of allocating buses, given that fares, routes, and operating speeds are fixed.

The problem can be viewed as a resource-allocation problem: How can the limited resources (subsidy) be allocated to maximize the benefit (ridership)? As shown in Figure 1, for each route a curve that relates net cost (deficit) to benefit can be obtained by varying the frequency of service on the route. At an optimal allocation, the ratio of marginal benefit to marginal cost should be the same for each route. Denoting the benefit of route i by B_i, its net cost by C_i, and its frequency by Q_i, this rule can be written as follows:

$$(dB_i/dQ_i)/(dC_i/dQ_i) = (dB_i dC_i) = (B/C)_o \quad (4)$$

As suggested in Figure 1, at the optimum some routes may be operating at a profit and others at a loss; however, the total benefit cannot be increased by

Figure 1. Efficiency in subsidy allocation.



shifting resources from one route to another. The optimum occurs when the marginal rate of return on each route is the same and is sufficient to exhaust the available subsidy.

With this simple example in mind we will now turn to the real problem, which will be formulated as a mathematical program. In the next two sections the objective of the optimization is defined and the constraints are specified.

OBJECTIVE

Increasing attention has been paid to the objectives of transit operators since 1975 when London Transport enunciated its objective of maximizing passenger miles (12). Defining objectives is an important step in developing good management practice in public transport agencies. Since in general transit has been recognized to be an important social service, presumably the general objective should be maximization of the social surplus. In the case of determining service frequencies while holding all other attributes of the system fixed, the objective includes two distinct components--consumer surplus and externalities associated with transit ridership.

It can easily be shown that the consumer surplus is the saving in wait time that accrues to system riders who would have been prepared to ride at lower frequencies (and thus endure longer waits). In the remainder of this paper, the mean passenger wait time will be assumed to be half the mean headway--based on the sample model of random passenger arrivals, regular headways, and buses not operating close to capacity. If the demand function is $r = r(h)$, where r is ridership and h is mean headway, the saving in wait time at headway h* is as follows:

$$h^* = \tfrac{1}{2}\int_{h^*}^{\infty} r(h)dh \qquad (5)$$

It can be argued that the major motivation for subsidization of transit service is not saving in wait time. Other, probably more significant, public benefits include mobility for those without automobiles; reductions in congestion, pollution, and energy use; and land use effects. These positive externalities are largely collinear with the ridership, and so a social ridership benefit can be defined crudely as being proportional to the number of riders. This marginal social ridership benefit would logically vary between ridership classes and time periods, which reflects the extent to which attracting different types of riders contributes to the social objectives of providing transit service. This term must be weighted to reflect the value of an additional rider as it relates to saving in wait time.

The objective function then consists of these two components--wait-time saving and ridership. One of the important questions that will be addressed later is how sensitive the service frequencies are to changes in the relative importance of these two objectives.

CONSTRAINTS

Four sets of constraints are included directly in the mathematical program--subsidy, fleet size, policy headways, and loading. Typically, an operator has a fixed level of subsidy available for the planning period (e.g., one year), and the solution that maximizes total benefit will inevitably exhaust the entire subsidy. Since we are concerned with short-range planning, the operator has a limited fleet to allocate, which may vary between periods of the day because of preventive-maintenance needs. As discussed earlier in this paper, two constraints currently used by many operators in setting frequencies are policy headways (which stipulate a maximum allowable headway) and peak-load factors (which specify the maximum load at the most heavily loaded point on the route). For any route during any period clearly only one of these two constraints can be binding, and so the mathematical program includes constraints that require that each headway satisfy the more binding of these two constraints.

In addition to these formal constraints included in the model, there is another set of constraints not included in the model, which can be dealt with externally. In general some services may be mandated for reasons other than social benefit as narrowly defined above; buses and subsidy should be set aside for these required services before the optimization problem is solved, and these services are simply added to the solution to produce the recommended set. Often only an integer number of buses can be assigned to a route (although interlining is a common means to circumvent this requirement); this constraint can be handled by adjustment of the final frequencies. Finally, some interdependencies between routes can be incorporated directly into the objective function. For example, if two routes should have the same frequencies, e.g., for timed transfers, they can be included as a single decision variable; this would help to reduce the size of the problem.

PROBLEM FORMULATION AND DISCUSSION

The problem can be stated as follows: Find the frequencies on each of a number of routes that maximize net social benefit subject to constraints on total subsidy, fleet size, and maximum headways. In the following formulation, headway is used as the basic decision variable.

Maximize:

$$Z = \sum_{j=1}^{P} D_j \sum_{i=1}^{N_j} \left[(b/2)\int_{h_{ij}}^{\infty} r_{ij}(u)du + a_{ij}\, r_{ij}(h_{ij})\right] \qquad (6)$$

Subject to the following constraints:

Subsidy:

$$\sum_{j=1}^{P} D_j \sum_{i=1}^{N_j} \left[(k_{ij}/h_{ij}) - F_{ij}\, r_{ij}(h_{ij})\right] = S_o \qquad (7)$$

Fleet size:

$$\sum_{i=1}^{N_j} (T_{ij}/h_{ij}) \leq M_j, \quad j = 1, \ldots, P \qquad (8)$$

Headway:

$$h_{ij} \leq x_{ij}, \quad j = 1, \ldots, P \quad i = 1, \ldots, N_j \tag{9}$$

where

- P = number of time periods,
- N_j = number of routes operated during time period j,
- D_j = duration of period j,
- b = value of wait time,
- h_{ij} = headway on route i during period j,
- a_{ij} = surplus marginal ridership benefit on route i during period j (marginal ridership benefit minus fare),
- r_{ij} = ridership on route i during period j (a function of h_{ij}),
- k_{ij} = operating cost per run on route i during period j,
- F_{ij} = fare on route i during period j,
- S_o = subsidy available,
- T_{ij} = run time (round trip) on route i during period j,
- M_j = fleet size during period j, and
- x_{ij} = maximum headway for route i during period j.

The objective function (Equation 6) can easily be shown to be equivalent to maximizing the wait-time savings plus the social-ridership benefit minus operating cost; this is the net social benefit. Equation 7 simply states that the operating cost minus the revenue must be equal to the known subsidy. Equation 8 is the fleet-size constraint, and Equation 9 constrains the headway to be less than the policy headway and the headway at which the loading constraint is binding.

This general formulation could be simplified or made more complex (for example, by defining classes of riders, each of which has a separate marginal benefit) in specific applications, but all important facets of the problem are included. Before the method developed to solve this mathematical program is presented, it is necessary to recognize and discuss perhaps the most important limitation of the model--the assumption of the independence of all routes in the system.

It is because both costs and benefits due to a headway on a specific route have been assumed independent of headways on other routes that the problem formulation is so straightforward, but this assumption is not always true, at least on the benefit side. In this model, ridership on a route depends on the headway of only that route, whereas, in general, ridership will also depend on the headways on competing and complementary routes.

When passengers have a choice among several routes, an improvement in service on one of those routes will divert riders from the other routes. Such route competition is less common in North America than in other parts of the world, in which an approach that directly considers route competition is called for (13). An improvement in service on one route can also raise the demand on another route when there is a large transfer volume between the two routes. Care must be taken, therefore, both in applying the model and in interpreting its results in situations in which strong route competition or complementarity exists.

THE ALGORITHM

Optimality (Kuhn-Tucker) conditions can be derived as a set of equations that relate headways to the other variables in the model; these equations then become the optimal decision rules for the operator. These optimality conditions are applied in the following step-by-step algorithm to determine the optimal set of headways by route and by period:

Step 1: Relax the fleet-size and maximum-headway constraints on all routes and for all time periods not yet constrained and solve the following set of equations for the headways, h_{ij}:

$$(b/\lambda) \cdot (r_{ij}/2)\, h_{ij}^2 - \left\{F_{ij} + [(a_{ij}/\lambda)(dr_{ij}/dh_{ij})]\, h_{ij}^2 - k_{ij}\right\} = 0 \tag{10}$$

where λ is determined to exhaust the available subsidy. If no routes violate their maximum-headway constraint, go to step 3.

Step 2: For routes and periods for which the maximum-headway constraint (Equation 9) is violated, set $h_{ij} = x_{ij}$. Compute the deficit incurred on those routes and reduce the available subsidy by that amount. Go to step 1.

Step 3: Identify time periods in which the fleet-size constraint (Equation 8) is binding. For each of these time periods solve the following set of equations:

$$(b/\lambda) \cdot (r_{ij}/2) \cdot h_{ij}^2 - \left\{F_{ij} + [(a_{ij}/\lambda)(dr_{ij}/dh_{ij})]\, h_{ij}^2 - (k_{ij} + w_j T_{ij})\right\} = 0 \tag{11}$$

where w_j is the shadow price of run time during period j and is determined to use all available buses.

Step 4: If no routes violate their maximum-headway constraint (Equation 9), go to step 5. Otherwise, for every route that violates its maximum-headway constraint, set $h_{ij} = x_{ij}$. Compute the number of buses required by all such routes in each period j and reduce the number of available buses in period j by this amount. Go to step 3.

Step 5: Compute the deficit incurred by the fleet-constrained time periods and reduce the available subsidy by this amount. Let $\lambda_c = \lambda$.

Step 6: Repeat steps 1 and 2 for the unconstrained time periods to find a new value of λ, which is λ_u.

Step 7: If $\lambda_u \simeq \lambda_c$, stop; otherwise set $\lambda = \lambda_u$ and return to step 3.

The theory behind this algorithm will not be presented in detail here (1). However, the computational burden of the algorithm is very small, since it consists basically of a sequence of one-dimensional searches that are performed very rapidly. Equations 10 and 11 can be solved very efficiently by using the Newton method (provided the demand function has continuous second derivatives), and values of λ and w_j can be found by making successive linear approximations.

CASE STUDY

The Arborway Garage of MBTA, which serves 21 bus routes, was chosen to illustrate the capabilities of the model. Fifteen of these routes were included in the analysis; the others were excluded for one of the following reasons: incomplete ridership data, highly irregularly scheduled runs, and interdependence of routes.

The most important data and assumptions made in the study are summarized here:

1. Two time periods were examined--the morning peak (7-9 a.m.) and midday (10 a.m.-2 p.m.). Scheduled headways were approximately constant during each of these periods.

2. Scheduled round-trip running times were used with a layover time of 25 percent of the run time. This is above MBTA's policy of 10-20 percent of run time for layovers but slightly below the average observed figure of 28 percent.

3. Costs per run were based on MBTA's figures of cost per vehicle mile and per driver hour.

4. The systemwide average revenue per bus ride of 18¢ was used.

5. The current deficit incurred on this part of the system was used as the available subsidy, and the current number of vehicles used in the morning peak was used as the fleet-size constraint in each period.

6. Policy headways of 30 min in the peak period and 60 min in the off-peak period were used based on current MBTA service policies.

7. MBTA uses two peak-period load-factor standards: For the 2-h peak period, the peak-load factor should be no greater than 1.2; for the peak half-hour it should be no greater than 1.4. Actual peak loads were estimated for both periods for each route based on peak-point counts taken over a three-year period. Only one of these load-factor constraints will be binding for each route. The off-peak policy load factor of 1.0 adopted by MBTA was also used.

8. Bus seating capacity of 45 was used for the load-factor constraints.

9. Current route ridership was taken from a 1978 on-board survey.

More-detailed discussion is warranted about the demand model and the operator's objectives. A binary logit demand model was used that has assumed coefficients for wait time taken from another study. Estimates of the base transit market share were also made based on mode-split characteristics of the Boston area. These assumptions implied wait-time elasticities of demand of -0.2 in the peak period and -0.5 in the off-peak period, which are within the range observed in other U.S. cities (14). It is hoped that advances in the state of the art of demand forecasting at the route level will soon obviate the need for such assumptions. In this case study, sensitivity analyses demonstrated that the results were very robust with respect to these parameters.

The model allows an objective function that consists of a weighted sum of total passengers and total passenger wait-time savings. The absolute coefficients of these terms do not have to be exogenously specified, but their ratio does. The initial ratio chosen implied a trade-off of one passenger for 12 passenger-min of wait time.

Table 1 shows the resource allocation between routes and between periods as suggested by the model compared with the current MBTA allocation. The results in terms of deficit, number of buses, and changes in wait-time and ridership benefits are given in Table 2. The most striking result is that only 59 of the 70 available buses are used in the peak period, and the peak period's share of the deficit declines accordingly. Only 44 percent of the total subsidy is allocated to the peak period by the model compared with 58 percent in the current system. The peak period is heavily constrained by capacity; nine routes operate at the maximum load during the peak half-hour. In general, the shorter routes have the smallest loads and so do not necessarily have the highest revenue/cost ratio. As expected, midday loads are much lower than those during the peak period.

Several factors contribute to this large shift in

Table 1. Frequency on case-study routes: actual and recommended.

Route	Actual Frequency (buses/h)	Actual Peak ½-h Load (passengers)	Actual Revenue/Cost Ratio	Recommended Frequency (buses/h)	Recommended Peak ½-h Load (passengers)	Recommended Revenue/Cost Ratio
Morning Peak						
21	5.0	63[a]	0.55	5.0	63[a]	0.54
24	4.0	42	0.50	4.0	42	0.49
25	5.0	35	0.37	3.7	44	0.47
28	3.0	54	0.50	3.6	48	0.44
29	13.3	57	0.60	12.0	63[a]	0.65
31	4.0	39	0.30	2.9	49	0.37
32	15.0	60	0.52	14.2	63[a]	o.55
35	5.0	53	0.32	4.0	63[a]	0.36
36	10.0	51	0.49	8.0	63[a]	0.60
37	5.0	54	0.42	4.2	63[a]	0.49
38	5.5	34	0.39	2.6	59	0.32
41	6.0	56	0.67	5.3	63[a]	0.74
46	2.0	29	0.63	3.1	21	0.46
50	3.3	59	0.39	3.0	63[a]	0.41
51	4.0	55	0.26	3.3	63[a]	0.29
Midday						
21	1.3	13	0.22	1.8	13	0.21
24	1.5	15	0.37	2.5	13	0.30
25	3.0	11	0.35	3.0	11	0.35
28[b]	0.0	-	-	0.0	-	-
29	5.0	38	0.85	5.6	35	0.77
31	1.5	13	0.58	3.4	9	0.40
32	4.6	28	0.60	4.7	28	0.58
35	2.0	33	0.48	3.2	26	0.38
36	2.0	40	0.57	3.5	30	0.42
37	2.0	24	0.44	3.0	20	0.36
38	2.7	15	0.20	2.0	17	0.23
41	3.5	24	0.66	4.2	21	0.58
46	2.0	7	0.18	1.5	7	0.20
50	2.0	20	0.28	2.3	19	0.27
51	2.0	21	0.23	2.0	21	0.23

[a]Capacity constrained. [b]Route 28 is not operated in the off-peak period.

Table 2. Deficit, number of buses, and change in wait-time and ridership benefits: actual and recommended.

Item	Morning Peak		Midday	
	Actual	Recommended	Actual	Recommended
Deficit ($)	2175	1651	1588	2112
Buses operating	70	59	27	32
Change in benefit for				
Wait time ($)	-	-177	-	607
Ridership ($)	-	-103	-	431

resources from the peak period to midday relative to the current system. First, reducing headways in the off-peak period, when base headways are higher, is more effective in reducing total wait time than applying the same resources to reduce headways in the peak period. Second, demand elasticity with respect to wait time is higher in the off-peak period, and so reducing headways in that period is more effective in increasing ridership.

Experiments that vary both the demand parameters and the objective function weights, described in the following paragraphs, consistently suggested shifting resources from the peak to the midday period. This leads to the strong recommendation that midday services be expanded at the expense of the peak periods. Such a shift could be expected to reduce costs by 10-15 percent and result in a reduction in the deficit incurred of about 20 percent.

Several experiments were run that varied the ratio of the wait-time value to the marginal ridership benefit from zero to infinity to study the sensitivity of resource allocation between routes and periods. The resulting allocations between periods were almost identical; changes in number of buses and share of the deficit allocated to the peak period were less than 2 percent over the full range of objective function weights. Variations between routes within the same period were of similar magnitude, which supported the finding that the relative weights given to the two objectives have little impact on the optimal allocation. This is to be expected, since any action the operator takes to decrease wait time will also tend to increase ridership, and vice versa.

Perhaps the greatest weakness in this case study is the uncertainty about the demand function and its parameters. To test the importance of this uncertainty, a set of experiments was run that varied the demand parameters to see whether resource allocation changed significantly. As the headway coefficient was varied from zero to 150 percent of its base value, the total variation in resource allocation between time periods was less than 5 percent, and the variation between routes did not exceed 10 percent. Similar lack of sensitivity to the assumed absolute and relative transit-market shares between periods and between routes was observed.

An important observation from these experiments is that the results obtained when demand is assumed to be inelastic differ little from those when a more realistic demand model is used. This lack of sensitivity is not altogether surprising, since providing the best service for current customers is usually a good way to attract new customers. When demand is assumed fixed, the solution algorithm becomes computationally much less complex, since Equations 10 and 11 can then be solved in closed form, which makes this procedure one that could be performed by using a programmable calculator or even manually.

The model was also used to explore one policy question, Would higher-capacity buses be beneficial on some routes? Striking results were found by increasing bus seating capacity from 45 to 53; the total benefit (the value of the objective function) increases by one-third. Only on three routes was the capacity constrained, the number of buses required in the peak declined from 59 to 55, and the peak period's share of the deficit declined from 44 to 38 percent.

This analysis shows the value of the proposed allocation model in policy terms and also suggests that in this case real benefits may accrue from using larger vehicles.

CONCLUSIONS

In this paper the allocation of buses to routes, one component of the short-range transit-planning problem, has been discussed. A model was proposed that treats the problem as a constrained resource-allocation problem. The objective was that net social benefit, which consists of ridership benefit and wait-time savings, be maximized subject to constraints on total subsidy, fleet size, and acceptable levels of loading. An algorithm was developed to solve the resulting mathematical program, which can be implemented on a computer or on a programmable calculator.

The case study of one garage of the MBTA system produced a number of important findings:

1. The best allocation of buses (and resources) is very robust with respect to the objectives and parameters assumed,
2. Existing rules of thumb used in the transit industry may not be as efficient as a formal model that uses a consistent objective, and
3. The proposed model can be useful in policy analysis, for example, in the development of service policies and vehicle procurement.

This study leaves a number of important topics for further research:

1. Better understanding of the objectives and constraints currently used by transit schedulers,
2. Relaxation of the assumption of route independence embodied in the model, and
3. Pilot implementation of these ideas in a transit agency.

More broadly, with the encouraging results obtained in this study, a new look at the role of more-formal methods for improving short-range transit planning seems badly needed.

ACKNOWLEDGMENT

The research reported in this paper was funded in part by a grant from the Urban Mass Transportation Administration of the U.S. Department of Transportation.

REFERENCES

1. P.G. Furth. Setting Frequencies on Bus Routes. Center for Transportation Studies, Massachusetts Institute of Technology, Cambridge, Res. Rept. CTS-RAMP-80-2, Feb. 1980.
2. J.P. Attanucci, L. Jaeger, and J. Becker. Bus Service Evaluation Procedures: A Review. UMTA, U.S. Department of Transportation, Rept. UMTA-MA-09-7001-79-1, April 1979.
3. R.F. Kirby and M.A. Green. Policies and Procedures for Transit Service Development. Traffic Quarterly, Vol. 33, No. 3, July 1979, pp. 413-428.
4. E.K. Morlok. Supply Functions for Public

Transport: Initial Concepts and Model. In Traffic Equilibrium Methods (M.A. Florian, ed.), Proc., International Symposium on Traffic Equilibrium Methods, Montreal, 1974, Springer Verlag, Berlin, 1976, pp. 322-367.
5. H. Mohring. Optimization and Scale Economies in Urban Bus Transportation. American Economic Review, Sept. 1972.
6. C.R. Guinn. Transit Optimization--Test 1 Analysis; Transit Optimization--Test 2 Analysis. New York State Department of Transportation, Albany, Preliminary Res. Repts. 2 and 12, 1967.
7. S. Scheele. A Mathematical Programming Algorithm for Optimal Frequencies. Institute of Technology, Linköping Univ., Linköping, Sweden, Linköping Studies in Science and Technology, Dissertation 12, 1977.
8. M. Lampkin and P.D. Saalmans. The Design of Routes, Service Frequencies, and Schedules for a Municipal Bus Undertaking: A Case Study. Operations Research Quarterly, Vol. 18, No. 4, Dec. 1967, pp. 375-397.
9. L.A. Silman, Z. Barzily, and U. Passy. Planning the Route System for Urban Buses. Computers and Operations Research, Vol. 1, 1974, pp. 201-211.
10. A. Last and S.E. Leak. TRANSEPT: A Bus Model. Traffic Engineering and Control, Vol. 17, No. 1, Jan. 1976.
11. O. Hasselström. A Method for Optimization of Urban Bus Route Systems. Volvo Bus Corporation, Göteborg, Sweden, working paper, 1979.
12. C.A. Nash. Management Objectives, Fares, and Service Levels in Bus Transport. Journal of Transport Economics and Policy, Jan. 1978.
13. A.F. Han and N.H.M. Wilson. The Allocation of Buses in Heavily Utilized Networks with Overlapping Routes. Transportation Research (in preparation).
14. Ecosometrics, Inc. Patronage Impacts of Change in Transit Fares and Services. Office of Service and Methods Demonstrations, UMTA, U.S. Department of Transportation, Rept. 135-1, 1980.

Publication of this paper sponsored by Committee on Traveler Behavior and Values.

Strategies for Improving Reliability of Bus Transit Service

MARK A. TURNQUIST

Four major classes of strategies for improving reliability of bus transit service are analyzed: vehicle-holding strategies, reduction of the number of stops made by each bus, signal preemption, and provision of exclusive right-of-way. The principal findings are that (a) strategies to improve service reliability can have very substantial impacts on overall service quality, including improvements in average wait and in-vehicle time as well, and (b) the best strategy to use in a particular situation depends on several factors, but service frequency is the most important. For low-frequency services (less than 10 buses per hour), schedule-based holding strategies or zone scheduling is likely to work best. For midfrequency services (10-30 buses per hour) zone scheduling or signal preemption is likely to be most effective, although headway-based holding can also work well if an appropriate control point can be found. In high-frequency situations (more than 30 buses per hour), an exclusive lane combined with signal preemption should be considered.

The concept of service reliability has come into increasing prominence in recent years as an important characteristic of the quality of service provided by transportation systems. A basic definition of reliability, as the term is used here, is the variability of a system performance measure over time. The focus is on stochastic variation in performance rather than on more-traditional engineering concepts of probability of component or system failure. The level-of-service measure most clearly subject to variation is travel time, and this variability is often described in terms of nonadherence to schedule.

Service reliability is important to both the transit user and the transit operator. To the user, nonadherence to schedule results in increased wait time, makes transferring more difficult, and causes uncertain arrival time at the destination. The importance of some measure of reliability to trip-making behavior has been emphasized in several attitudinal studies. For example, Paine and others (1) found that potential users ranked "arriving when planned" as the single most important service characteristic of a transit system. This finding has been substantiated in further studies by Golob and others (2) and by Wallin and Wright (3).

In addition to its importance to transit users, unreliability in operations is a source of reduced productivity and increased costs for transit operators. This is due to the need to build substantial slack time into timetables in order to absorb deviations from the schedule. This leads to reduced use of both equipment and personnel. The recent report by Abkowitz and others (4) provides an excellent summary of the major issues in transit-service reliability from the perspectives of both the user and the operator.

In light of the current need for more cost-effective public transportation in urban areas, it is important to understand the sources of unreliability and to investigate the potential of several alternative control strategies to improve both the quality of service provided and the productivity of the equipment and the personnel in the system.

The research on which this paper is based has had four major objectives:

1. Investigation of the sources of service-reliability problems in bus transit networks,
2. Identification of potential strategies for improving reliability of service,
3. Development of models to allow these strategies to be analyzed and evaluated, and
4. General evaluation of the relative effectiveness of these strategies.

Experiments with a network-simulation model to investigate sources of reliability problems are summarized in the next section. The following sections provide analyses of four major classes of strategies for improving service reliability: vehicle holding; methods for reducing the number of stops made by each vehicle, which include increasing stop spacing and zone scheduling; signal-preemption; and exclusive rights-of-way for buses. The paper concludes with practical implications of the results.

This paper is a summary of findings from a two-year project and is intended to highlight the major research results. Readers interested in additional details on model development and test results should refer to the two larger reports from this project, that by Turnquist and Bowman (5) and that by Turnquist (6).

SOURCES OF UNRELIABILITY AND IMPLICATIONS FOR CONTROL STRATEGIES

One of the objectives of this research has been to focus on the ways in which network characteristics affect schedule reliability and hence the level of service experienced by the users. A set of experiments has been conducted to examine two relationships that seem to be of primary importance--the effects (a) of factors that contribute to the tendency of vehicles to bunch together as they travel and (b) of network configuration as exemplified by grid versus radial networks.

The first relationship to be considered has previously been addressed by Vuchic (7) by using a deterministic model to explore the propagation of schedule disturbances along a transit line. This model attempts to explain the pairing of successive vehicles, or bunching, in terms of the arrival and boarding rates of passengers at stops. The conclusion reached is that the most effective means of controlling these schedule disturbances is to reduce boarding times. This work extends that research by including the effect of "batch" passenger arrivals from connecting routes and, more important, the variability in link travel times.

Grid and radial networks represent fundamentally different patterns of service. They will result in different trip routings, different lengths of trips on the network, and different transfer characteristics. Thus, it is vital to contrast the levels of service reliability offered by the two types of network structure.

In order to reach conclusions about the two major relationships indicated above, a set of experiments was designed that involve five factors: (a) frequency of service (buses per hour), (b) coefficient of variation of link travel time, (c) demand/capacity ratio (total passenger miles per hour divided by available "space" miles per hour--both seated and standing--on all vehicles), (d) route density (miles of two-way route per square mile), and (e) network orientation (grid or radial). Frequency of service was assumed to be the same for all routes, and the coefficient of variation in link travel time was the same for all links in the network.

The experimental design and details of the experimental results have been discussed at length by Turnquist and Bowman (8) and will not be repeated here. However, a summary of the major findings of the experiments is as follows. The experiments have indicated how vehicle bunching is related to frequency of service, level of demand, and the variability of link travel times. In particular, these results illustrate the importance of reducing variability in link travel time in an effort to prevent bunches from forming. This represents an extension to the results of Vuchic (7), which placed primary emphasis on the demand/capacity ratio and boarding times.

It is clear from the experimental results that service reliability is much more sensitive to frequency of service than to route density. This implies that there are substantial reliability impacts of the trade-off between operating fewer routes at higher frequency or more routes at lower frequency, given a limited amount of vehicle resources. Traditionally, this trade-off has been evaluated by using simplistic models of expected passenger wait time and the accessibility of transit service to users. However, this work has shown that service reliability is also an important factor in this trade-off and should be included in the evaluation.

This research has several practical implications for transit operators who are attempting to improve the level of service provided to passengers. First, the presence of large variability in link travel times can substantially reduce the benefits that result from increasing frequency of service, due to the tendency of vehicles to bunch together along the route. In such cases, it is well worthwhile to investigate techniques for reducing this travel-time variability.

The influence of transfers on level of service points out the need to pay special attention to the on-time arrival of vehicles at major transfer stations. This is especially true for radially oriented network structures. As a rule, providing excess slack time in the route schedule is to be avoided, since it tends to increase travel time and reduce vehicle productivity. However, when a large number of passenger transfers can be aided by creating enough slack time to assure successful connections, allowing a short delay may be highly beneficial.

In summary, the major sources of reliability problems in transit service are bunching of vehicles and poor connections at transfer points. In a broad sense, then, the major objectives of control strategies are to keep bunches from forming (or to break them up after they have formed) and to ensure that scheduled arrival times at transfer points are met. At a more detailed level, deviations from schedule, which lead to bunching and poor transfer connections, can be traced to excessive variability in either link travel times between stops or dwell times at stops. Therefore, potential control strategies should be focused on reducing one or both of these sources of variability.

This investigation has concentrated on four general classes of strategies: (a) vehicle holding, (b) reductions in the number of stops served by each vehicle, (c) modifications to traffic signal settings and operation, and (d) provision of exclusive rights-of-way for transit vehicles. Such a classification provides a useful framework for discussion of many individual strategies and a comparison of their relative effectiveness in particular situations. The following sections provide discussions of each of these classes of control strategies.

VEHICLE HOLDING

Vehicle-holding strategies attempt to prevent bunches from forming and serve to break up bunches that may already have formed. When enacted at major transfer points, such strategies can also be useful in ensuring that schedule connections are made.

Two important subclasses of strategies can be distinguished. One type is oriented toward holding vehicles to a particular schedule, and the second is focused on maintaining constant headways between successive vehicles.

Schedule-Based Holding

A schedule-based holding strategy is nothing more than creating checkpoints or time points along a bus route and insisting that no vehicle leave a time point before its scheduled departure time. This is probably the simplest form of schedule control possible and is practiced (at least in theory) by many transit operators. Theory and practice often differ, however, because of lack of enforcement.

The keys to successful implementation of a schedule-based checkpoint strategy are (a) to have a schedule to which vehicles have a reasonable chance to adhere and (b) to enforce the rule of no early departures from the checkpoint. It is important that the mean arrival time of buses at the checkpoint be approximately the scheduled time. If the schedule is unrealistic, so that vehicles are consistently late, this strategy will have little or no effect, since the control actions directly affect only those vehicles that are ahead of schedule. On the other hand, it is inadvisable to have a schedule so slack that almost all vehicles are early, since delaying all these vehicles to meet the schedule of the slowest vehicles imposes penalties on a large number of passengers and reduces overall vehicle speed and productivity.

A schedule-based holding strategy can be particularly useful on suburban routes or in other instances in which headways are quite large. When service is relatively infrequent, passengers tend to learn the schedule and coordinate their arrival at the bus stop with the scheduled arrival time of the bus so as to minimize wait time. In such cases, adherence to schedule by the buses is very important in provision of quality service to the passengers.

Headway-Based Holding

When service is quite frequent, we might expect headway-based holding strategies to be effective. If service is frequent enough so that passengers may be assumed to arrive randomly in time at a given bus stop without regard to the schedule of service, the average waiting time E(W) has been derived by Welding (9):

$$E(W) = [E(H)/2] + [V(H)/2E(H)] \quad (1)$$

where E(H) is the expected headway between successive vehicles and V(H) is the variance of headways.

It is clear from Equation 1 that making the headways more regular (i.e., reducing the variance) will tend to reduce average waiting time. This is the motivation for headway-based control strategies. In general, the objective of control is to minimize a weighted sum of wait-time savings due to reduced headway variability and expected delay due to the holding strategy. Three basic categories of strategies may be distinguished.

One type has been referred to by Jackson (10) and by Turnquist and Bowman (5) as the "prefol" policy because it splits the difference between the preceding and following headways for each vehicle. This policy requires a prediction of the arrival time of the following vehicle. Automatic train-control systems could provide train location in rapid transit applications. For bus systems, location may be determined by automatic vehicle-monitoring (AVM) technologies. A projection of its speed to the control point would also be required.

A less-reliable but much less-expensive prediction of the following headway would be its statistical expectation. This suggests an alternative control policy, which will be referred to as the single-headway policy. It is dependent only on the known current headway and previous hold. Detailed mathematical development of both prefol and single-headway strategies may be found in a report by Turnquist (6).

Both the prefol and the single-headway policies are likely to be more effective than the third category, the so-called threshold-based holding strategies. The strategy "hold until the headway reaches a minimum threshold" has often been suggested and modeled in the literature (5, 11-13). Simulation work has indicated that this strategy tends to delay too many vehicles too long, which increases the average headway and sometimes actually lengthens passenger wait time.

Note that a prefol implementation would not increase the average headway, since no vehicle is held past the arrival of its follower. (Technically, the average headway is increased by a small amount if the last bus is held past the end of the period.)

One of the most important aspects of the analysis of any headway-based holding strategy is to identify those situations for which it is likely to be effective and those situations for which it is not. The relative benefits of holding depend on three factors: (a) the coefficient of variation of headways, (b) the correlation coefficient between successive headways, and (c) the proportion of total passengers who must ride through the control point.

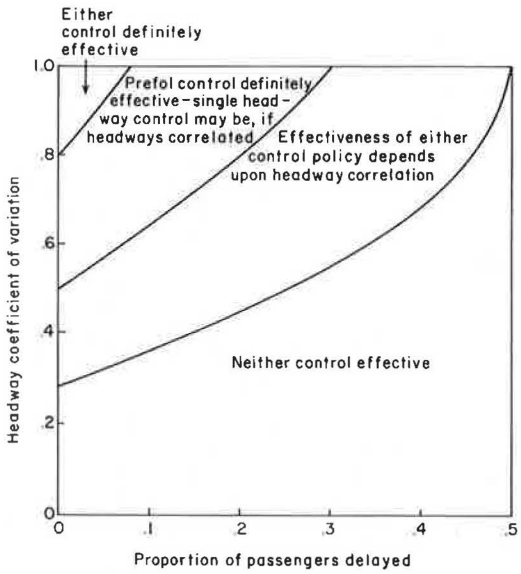
Control of headways will make the greatest reduction in total delay when headways alternate (i.e., short, long, short, long, etc.). This happens on routes on which vehicles are influenced substantially by the operation of the vehicle in front of them (successive headways are highly correlated). For example, this would tend to be the case when loading delays are relatively more important than traffic congestion in determining overall vehicle operating speed. Routes on which pairing or bunching is prevalent would be of this type. In such a situation, holding a vehicle to lengthen a short headway also serves to reduce the long one that follows. Thus, the variance of headways is reduced by a greater amount for a given delay to the held vehicle than if a short headway might be followed by another short headway.

Figure 1 illustrates sets of values of the headway coefficient of variation, headway correlation, and proportion of passengers delayed for which headway-based holding could reduce average passenger delay by at least 10 percent. By examining the two extreme cases of independent headways and perfectly correlated headways, we can bound the regions of effectiveness for a class of headway control strategies.

Note that the region for which the single-headway strategy produces definite benefits is much smaller than that for the prefol strategy. In general, the single-headway strategy is less effective than the prefol strategy because it uses no direct information about the following headway. However, the difference between the strategies diminishes as the correlation between successive headways becomes stronger because the predictability of the following headway is increasing.

It should also be emphasized that the effectiveness of headway-based controls is dependent on identification of an appropriate control point along the route. It is wise to control a route at a point at which there are relatively few people on the vehicle and relatively many waiting to board at subsequent stops. Generally, this means that the control point should be located as early along the vehicle's route as possible. However, it is also generally recognized that reliability problems worsen as one proceeds along a route. If dispatching at the route

Figure 1. Comparison of regions of effectiveness for prefol and single-headway strategies.



origin is effective, the headways will be reasonably regular at the early stops along the route, which implies that the coefficient of variation will be small. At stops further along the route, however, the coefficient of variation in headways will tend to be larger. Thus, the decision of whether to implement a control strategy is tied to identification of a logical control point along the route.

STOP REDUCTION AND ZONE SCHEDULING

Since a substantial portion of bus travel time is spent decelerating for stops, standing to allow boarding and alighting of passengers, waiting to reenter the traffic stream, and accelerating, reduction of the number of stops made by each vehicle is one way to improve travel time. In addition, since the variability of stop dwell time is a major source of deviation from schedule, reducing the number of stops should improve reliability. This research has examined two different ways in which to accomplish a reduction in the number of stops that each bus makes. The first is increasing stop spacing by eliminating some stops along a route, and the second is zone scheduling.

Increasing Stop Spacing

Increasing the spacing between stops is clearly one way to reduce the number of stops that must be made by each vehicle. The major disadvantage of increased stop spacing is that accessibility to the route is diminished. Passengers must walk further, on average, to get to a bus stop. This cost must be weighed against the improved travel time and reliability in order to arrive at optimal stop-spacing decisions.

Very little work has been done in this regard for bus operations. Vuchic (14) and Vuchic and Newell (15) have considered such problems for rapid transit lines, but reliability improvements were not among their measures of performance. Mohring (16) discusses optimal stop spacing for urban bus routes, but in the context of a very simple model and with no attention to reliability of service.

In order to test the effects of stop spacing more carefully, a series of simulation experiments have been run that use as a test network the Reading Road corridor in Cincinnati. This network is shown in Figure 2. The simulation tests reflect a morning peak period, and changes in stop density were made along a 7.1-km (4.4-mile) section from Clinton Springs Avenue to Government Square (downtown).

In the base case (which reflects existing operations), there are 36 stops in this section; the average stop spacing is 0.20 km (0.12 mile). For the tests, 17 of these stops were eliminated, which resulted in an average stop spacing of 0.37 km (0.23 mile). Five replications of each configuration were run and the averages over these replications compared.

The results show that average passenger speed over the system increased from 14.1 km/h (8.8 mph) to 14.5 km/h (9.0 mph). This change, although in the right direction, is not statistically significant at any reasonable level, however. The SD passenger speed was unchanged at 5.3 km/h (3.3 mph).

Reducing stop density also appears to have made small reductions in both the mean and SD (or variance) of waiting time. Mean waiting time was reduced from 7.5 min to 7.2 min and SD from 7.7 min to 7.0 min. However, as in the case of average passenger speed, these changes are not statistically significant.

Thus, the simulation results with respect to reduced stop density are not particularly encouraging. However, closer inspection of the simulation output showed that a major reason why eliminating stops had such small effects was that buses were still being slowed by traffic signals. Because of the signal settings, they could not take advantage of the potential reductions in travel time along the route; they simply spent more time in queues at traffic lights. In an attempt to rectify this, changes in both stop density and signal operation were made simultaneously. These results were more encouraging and are discussed in greater detail later.

Zone Scheduling

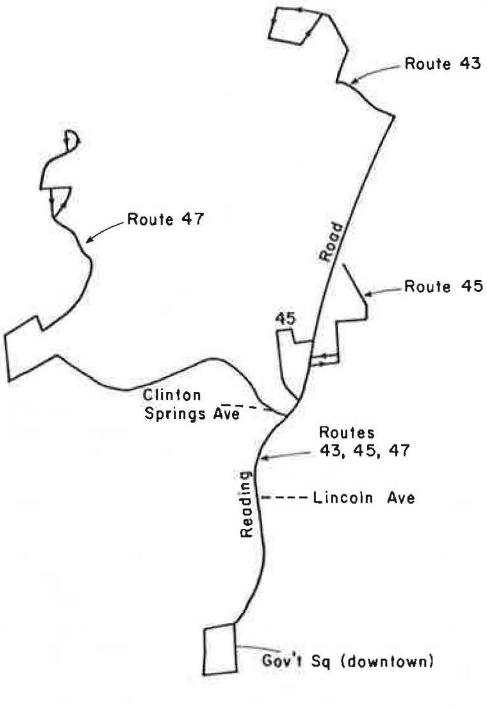
An alternative way of reducing the number of stops each vehicle must make without increasing overall stop spacing is to divide a route into zones. Each zone is a set of consecutive stops that has a subset of all the buses on the route allocated to it. An inbound bus dispatched from the outermost stop in its zone makes stops to pick up or let off passengers within its zone only; it runs nonstop to the route terminus after passing the inner zonal boundary. On its outward journey, the bus may provide local service all along the route; may travel express to the innermost stop of its zone, at which point it would again begin to offer local service; or may travel express all the way to the outer terminus of its zone and then begin another inbound run. Zones must overlap so that passengers bound from a stop in one zone to a stop in a different zone other than the route terminus can transfer.

Zone scheduling can improve both average bus speeds and reliability in two ways:

1. Average in-motion time and variability can be reduced by the nonstop service offered for a portion of each bus's run under a zone-scheduling scheme, and
2. The number of stops each bus makes can be reduced, which will lessen both average bus dwell time and variability in this time.

A dynamic-programming model has been developed to study the impact of zone scheduling on both service reliability and average wait and in-vehicle time. Service reliability is measured by the variance in

Figure 2. Reading Road corridor in Cincinnati.



passenger trip time summed over all passengers that use the route. The elements of the model are described in detail by Jordan and Turnquist (17).

The general conclusions from test applications of the model are as follows:

1. Substantial improvements in average trip time and reliability are attained through zone scheduling relative to all-local service;
2. While maintaining the improvements in average trip time and reliability, substantial decreases in a route's bus-fleet size can be made as a result of the improved productivity of all vehicles;
3. Average trip time is improved simultaneously with reliability under a zone-scheduling scheme; and
4. The major portion of reliability improvements can be attained by a very simple zone structure.

These results appear to be relatively insensitive to changes in important model parameters, at least in cases examined to date. This tends to increase the level of confidence in these results.

Our analysis has demonstrated that zone scheduling can be a very effective way in which to improve the quality and productivity of urban transit service. It should be emphasized that the attractiveness of zone scheduling in a given situation will depend greatly on the relative express and local speeds attainable, the relative variability in travel times on express and local links, and the proportion of total route ridership that is destined for (or originates at) the route terminus.

CHANGES IN TRAFFIC-SIGNAL OPERATION

Several authors, including Welding (9) and Jackson (10), have emphasized the importance of variability in travel times between stops as a source of bus reliability problems. A major portion of this variability arises from delays at controlled intersections. This section examines signal preemption as a method of reducing the impact of signalized intersections on average delay and the variability of delays.

Several studies have reported experience with preemption strategies at isolated intersections or on intersections along major streets that have light cross traffic. Such applications have shown substantial savings in mean transit time at insignificant costs to automobile users (18, 19). It must be noted that most previous researchers have focused on preemption as a means of reducing average delay at intersections and hence improving average speed of the buses. For the most part, service reliability benefits from reducing the variability of delays have not been considered explicitly. In order to include this measure of effectiveness, simulation experiments that use signal preemption have been performed in this project, again by using the Reading Road corridor in Cincinnati (Figure 2) as a test case.

As in previous tests, five replications of the preemption strategy were made with the simulation model. The pooled average of these five runs is then compared with a similar average from the base case.

One of the major effects of the preemption strategy is that bus travel times over the section that uses signal preemption are reduced by approximately 3.5 min (out of a scheduled 24 min), which implies an average speed increase from 17.7 km/h (11.0 mph) to 20.7 km/h (12.9 mph), or about 17 percent. The SD of travel time was reduced by approximately 0.5 min out of a total of 2.7, a reduction of 18 percent.

A second major effect is on passenger wait time. Average wait time is reduced by 0.6 min, from 7.5 min to 6.9 min, a reduction of 8 percent. The SD of wait time is also reduced, from 7.7 min to 7.0 min, or 9 percent. All the changes to both travel times (or speeds) and waiting time are significant at the 90 percent confidence level.

Signal preemption thus appears to offer significant potential for improving both average speed and reliability, with concomitant effects on both mean and variance of waiting time. As a further test of this strategy, a second experiment was conducted that combined signal preemption with reduced stop density, as discussed earlier.

The results of this experiment were a small (but insignificant) further increase in average speed and a small decrease in average wait (also insignificant) compared with the use of preemption alone. However, the SD of wait time decreased to 6.5 min compared with 7.0 min for preemption alone and 7.7 min for the base case. This further reduction in the variability of wait time is statistically significant and constitutes the major observed impact of combining signal preemption and reduced stop density.

RESERVED BUS LANES

In congested areas, traffic-stream delays account for a substantial portion of total transit travel time. Measures that remove the bus from these delays will reduce travel time and improve reliability. There is considerable empirical evidence from the United States, Europe, and Australia that reserved lanes can improve both average transit speeds and reliability. Additional simulation experiments conducted in this project have been designed to examine the effectiveness of reserved lanes together with signal preemption.

As a test case, the Reading Road corridor was again used as a basis but with substantial modifications. Bus operations are not heavy enough now to justify a reserved lane; there are only about 12 buses per hour. This would not be a very effective test of reserved-lane strategies intended for areas of much higher activity. To obtain a test case, the bus frequencies and passenger-arrival rates were multiplied by a factor of 5, which resulted in average headways of about 1 min and loadings comparable with those of the present case. Other elements of the corridor were left unchanged. Thus, our test is over a corridor 7.1 km (4.4 miles) in length that has a total of 36 stops and 32 signalized intersections. The reserved lane was specified to be a curb lane (the one that went with the traffic flow).

Table 1 summarizes the major results from testing the reserved lane alone and those from the reserved lane in combination with signal preemption. The addition of the lane itself results in a small reduction in average travel time, but this is not statistically significant. The reduction in the SD of travel time, from 5.1 min to 4.4 min (14 percent), is statistically significant at the 95 percent level. Reductions in mean and SD of waiting time are also statistically insignificant. Thus, the major impact of adding the reserved lane in this case appears to be a reduction in the variability of travel time, a direct improvement to service reliability.

Combining the reserved lane with signal preemption is noticeably more effective. The changes in passenger wait time are still insignificant, but the changes in both mean and SD of travel time along the corridor are highly significant. The estimated reduction in average travel time is 17 percent, and the reduction in the SD of travel time is 18 percent.

Two additional aspects of these experiments should be noted in order to aid interpretation of the results. First, the effects of the signal preemption on cross traffic have not been analyzed in detail. The expected value of delay to cross traffic has been included in the settings of the signals used in the preemption study. However, more-thorough analysis would require a more-detailed traffic-simulation model. The second point is that the vehicle traffic levels (both in the main direction and in the cross direction) assumed for these experiments are relatively light. Main-direction volumes are the heaviest and are in the range of 350-400 vehicles per lane per hour. Most situations in which reserved lanes would be considered are likely to have heavier traffic volumes as well as heavy bus volumes. Thus, the benefits of removing the buses from the mixed traffic scheme are likely to be greater than those measured in these experiments. In this sense, these results are likely to be conservative.

Thus, the combination of a reserved lane for buses and signal-preemption capability appears to be a potentially effective method for improving both average travel time and reliability in situations that involve very heavy bus movements.

CONCLUSIONS AND IMPLICATIONS

The most important determinant of the appropriate strategy for reliability improvement in a given situation seems to be frequency of service on the route or routes in question. For low-frequency situations (less than 10 buses per hour), checkpoint control (schedule-based holding) is likely to be the most effective strategy, provided that an appropriate schedule is constructed and adherence to schedule at checkpoints is enforced. In some low-frequency situations, zone scheduling may also be effective if most passengers are destined for (or originate at) one terminus of the route. The presence of an expressway roughly parallel to the route also makes this strategy more effective.

In medium-frequency situations (10-30 buses per hour), the most effective strategies are likely to be zone scheduling and signal preemption. If the origin-destination pattern of passengers is suitable and an express facility is available, zone scheduling is likely to be the best choice. If these conditions are not met, signal preemption on the local facility used by the buses should be considered. Headway-based holding can also be useful if an appropriate control point can be found along the route.

For high-frequency situations (more than 30 buses per hour) an exclusive lane together with signal preemption if the road is an arterial should be considered. Experience from several demonstrations of bus lanes and modeling results from this study and others indicate the effectiveness of such a strategy in improving both average travel time and reliability for buses.

Whereas these recommendations provide general guidelines for transit operators and planners in selecting service-improvement strategies, the most valuable product of this research is the battery of models developed for analyzing a number of strategies in any particular situation. These models include the analytic formulations for developing holding strategies, the dynamic-programming model for designing zone-scheduled systems, and the computer-simulation model for detailed analysis of many possible strategies. By using these tools, the transit operator or planner can design and test a service-improvement strategy appropriate for his or her particular situation.

It should be noted that effective implementation of service-improvement strategies need not imply the installation of expensive AVM equipment. Although such equipment is clearly beneficial in implementing headway-based holding strategies, for example, there are many other potential strategies that are likely to be just as effective (or perhaps more so) and that require substantially less investment in hardware.

Finally, it is important to emphasize the need for cooperation between transit-operating authorities and municipal departments responsible for streets and traffic signals. Many of the strategies for service improvement described in this paper would require agreement and joint action on the part of both agencies for effective implementation. In order to reach the point of acting together, it is important that they begin to plan together. Communication and agreement on overall goals at an early stage are vital to the success of many of the strategies that seem to be most effective in improving service reliability in transit systems.

Table 1. Summary of test results for reserved-lane strategies.

	Travel Time (min)		
Measure	Base Case (no reserved lane)	Reserved Lane Only	Reserved Lane and Preemption
Average bus travel time	25.2	23.8	20.8
SD of bus travel time	5.1	4.4	4.2
Average wait time	0.9	0.8	0.7
SD of wait time	1.3	1.2	1.1

ACKNOWLEDGMENT

This research was supported by the Office of University Research, U.S. Department of Transportation. This support is gratefully acknowledged, but the research findings and opinions are solely mine and do not necessarily represent the position of the U.S. Department of Transportation. I would like to thank several individuals, including Mark Abkowitz, Larry Bowman, William Jordan, and Steven Blume, for assistance in this research. Finally, Keith Armstrong of the General Motors Urban Transportation Laboratory in Cincinnati and the staff at Queen City Metro have also been very helpful in making data available from their system.

REFERENCES

1. T.F. Paine, A.N. Nashe, S.J. Hille, and G.A. Brunner. Consumer Conceived Attributes of Transportation--An Attitude Study. Department of Business Administration, Univ. of Maryland, College Park, 1967.
2. T.F. Golob, E.T. Canty, R.L. Gustafson, and J.E. Vit. An Analysis of Consumer Preferences for a Public Transport System. Transportation Research, Vol. 6, No. 1, 1972, pp. 70-71.
3. R.J. Wallin and P.H. Wright. Factors Which Influence Modal Choice. Traffic Quarterly, Vol. 28, 1968, pp. 271-290.
4. M. Abkowitz, H. Slavin, R. Waksman, L. Englisher, and N. Wilson. Transit Service Reliability. Transportation Systems Center, U.S. Department of Transportation, Rept. UMTA-MA-06-0049-78-1, 1978.
5. M.A. Turnquist and L. Bowman. Control of Service Reliability in Transit Networks. Office of University Research, U.S. Department of Transportation, Rept. DOT/RSPA/DPB-50/79/5, 1979.
6. M.A. Turnquist. Strategies for Improving Bus Transit Service Reliability. Office of University Research, U.S. Department of Transportation, June 1980.
7. V.R. Vuchic. Propagation of Schedule Disturbances in Line-Haul Passenger Transportation. UITP Revue, Vol. 28, 1969, pp. 281-284.
8. M.A. Turnquist and L.A. Bowman. The Effects of Network Structure on Reliability of Transit Service. Transportation Research, Vol. 14B, No. 1/2, 1980, pp. 79-86.
9. P.I. Welding. The Instability of Close Interval Service. Operational Research Quarterly, Vol. 8, 1957, pp. 133-148.
10. R.L. Jackson. Evaluation by Simulation of Control Strategies for a High Frequency Bus Service. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, TRRL Rept. 807, 1977.
11. A. Barnett. On Controlling Randomness in Transit Operations. Transportation Science, Vol. 8, 1974, pp. 102-116.
12. R.L. Jackson. An Evaluation of Bus Control Systems by Simulation. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, TRRL Rept. 712, 1974.
13. D. Koffman. A Simulation Study of Alternative Real-Time Bus Headway Control Strategies. TRB, Transportation Research Record 663, 1978, pp. 41-46.
14. V. Vuchic. Rapid Transit Interstation Spacing for Maximum Number of Passengers. Transportation Science, Vol. 3, 1969, pp. 214-232.
15. V. Vuchic and G.F. Newell. Rapid Transit Interstation Spacings for Minimum Travel Time. Transportation Science, Vol. 2, 1968, pp. 303-339.
16. H. Mohring. Optimization and Scale Economies in Urban Bus Transportation. American Economic Review, Vol. 62, No. 4, 1972, pp. 591-604.
17. W.C. Jordan and M.A. Turnquist. Zone Scheduling of Bus Routes to Improve Service Reliability. Transportation Science, Vol. 13, 1979, pp. 242-268.
18. K.C. Courage, C.E. Wallace, and J.A. Wattleworth. Effect of Bus Priority System Operation on Performance of Traffic Signal Control Equipment on N.W. 7th Avenue. UMTA, U.S. Department of Transportation, Rept. UMTA-FL-06-0006, 1977.
19. Mitre Corporation. Overview of Experimental Bus Priority Systems. UMTA, U.S. Department of Transportation, 1975.

Publication of this paper sponsored by Committee on Traveler Behavior and Values.

Ridership Response to Changes in Transit Services

ARMANDO M. LAGO, PATRICK D. MAYWORM, AND J. MATTHEW McENROE

Evidence on ridership response to changes in transit service is presented. Mean values and standard deviations of transit-service elasticities are presented for changes in headways, vehicle miles, in-vehicle and out-of-vehicle travel time, transfers, and seat availability. A review of the methods used in estimation of demand elasticity is presented as well as suggestions on how service elasticities can be used in joint transit-fare and service-level planning to improve revenues and ridership.

The demand for public transportation has traditionally been regarded as more responsive to changes in transit service (e.g., headways and bus miles) than to changes in transit fares. Although on the aggregate level this may be true, recent evidence shows that service elasticities vary considerably from one area to another by the time of day, type of route, service quality, and other classifications, which suggests that there may be situations in which patronage may be more responsive to fare changes than to service adjustments.

In this paper a summary of the current state of knowledge on the size of transit-service elasticities is presented compiled from demonstrations and demand models. In addition, suggestions are made about how service elasticities can be used in joint transit-fare and service-level planning to improve revenues and ridership.

APPROACHES TO ESTIMATING TRANSIT-SERVICE ELASTICITIES

Nature of Approaches to Demand Estimation

Two broad approaches to estimating service elasticities may be distinguished. These approaches include (a) monitoring service changes and demonstration studies, or those that rely on data generated either by a practical demonstration of an actual change or by monitoring an actual change in service levels, and (b) nonexperimental approaches, or those that rely on a data base either devoid of an actual change in service levels or in which actual changes are part of historical trends.

Approaches in the first category include the monitoring of transit-service demonstrations and individual service changes such as those that use monthly data series. The nonexperimental approaches generally include (a) the conventional time-series analysis of annual transit operating statistics, (b) aggregate direct-demand and modal-split models based on cross-sectional data, and (c) disaggregate behavioral mode-choice models based on cross-sectional data. All the nonexperimental approaches have in common the fact that the data base does not contain an actual service change and also that the data base is not generated with the objective of controlling for nonservice changes.

The demand-elasticity estimates presented in this paper from demonstrations and selected service-monitoring studies were calculated by using a midpoint elasticity formula (1). The demand-elasticity estimates from demand models are point elasticities.

Methodological Note on Special Problems of Cross-Sectional Models

In interpreting transit-demand elasticities, some problems are posed by overreliance on elasticity estimates developed from a cross-sectional data base that contains no service change. One cannot rely on elasticity estimates from cross-sectional studies to provide accurate estimates of annual changes in patronage in response to service changes because they reflect a different type of behavior from that implicit in time-series analysis. This difference between time-series and cross-sectional models arises because the residuals from both models cannot be assumed to belong to the same underlying population. In general, cross-sectional estimates represent behavior that, for lack of a better term, economists have labeled "long-run structural adjustments" (2), although it is possible that cross sections taken at a time of rapid growth or of cyclical change could also reflect short-run annual adjustments such as those characterized by time-series relationships. Although cross-sectional models have advantages in forecasting structural changes in demand, dynamic annual-change-type responses cannot be estimated with any degree of confidence unless supporting time-series information is available to establish a systematic relationship.

Another problem is that some recent work on disaggregate behavioral models has departed from McFadden's (3) original contribution and, as a consequence, as shown by Oum (4), some of these models (a) impose many rigid a priori conditions on the elasticities and cross elasticities of demand, (b) result in estimates of elasticities that are not invariant to the choice of the base or modal denominators, and (c) possess severely irregular and inconsistent underlying preference or utility structures. Moreover, an estimation problem arises whenever simultaneous mode choices concern more than two modes. Theil (5) and Nerlove and Press (6) argue that biased coefficients result when simultaneous choices--such as the choices that involve more than two transport modes--are estimated by using single-equation estimation techniques such as the maximum-likelihood approaches currently used by transportation mode-choice modelers.

HEADWAY ELASTICITIES

Public transportation headway elasticities vary considerably, due in part to the characteristics of the route in question, but the mean aggregate values show a remarkable similarity. The evidence shows that average bus and commuter-rail headway elasticities are equivalent and the mean value for all service hours is -0.47 ± 0.17 (16 cases). (The standard deviations presented in this paper measure the variation of the respective groups of means taken from the studies and do not represent a measure of confidence in the particular means.)

Bus Headway Elasticities

The information on bus headway elasticities summarized in Table 1 from a report by Ecosometrics, Inc. (7), from the Detroit Grand River Avenue demonstration (8), from the Chesapeake/Norfolk commuter-route demonstration (9), from the Boston bus headway demonstration (10), from the Madison circular-route demonstration (11), and from the demonstration in Stevenage, Great Britain (12) shows that, although the mean bus headway elasticity based on data from monitoring service changes is -0.47 ± 0.21 for all service hours, each elasticity value appears to depend on the route characteristics and on the level of service before headway adjustments are made. As shown in Table 1, headway elasticities depend on the previous level of service for both peak and off-peak periods. During the peak period, headway elasticities are -0.58 for low-service routes. These values exceed by more than 110 percent the elasticity values of -0.27 ± 0.14 for high-service routes. The same is true during off-peak periods in which the highest elasticities, which have a mean value of -0.71 ± 0.11, predominate among low-service routes.

With regard to differences in headway elasticities by time of day, off-peak elasticities are appreciably higher than peak-period elasticities. In the Chesapeake/Norfolk demonstration of 1965-1967 (9), the off-peak elasticities were more than 50 percent above the mean peak elasticity of -0.57. The same is true of the 1962 Detroit Grand River Avenue demonstration (8), in which off-peak elasticities were almost 100 percent above the peak-hour elasticity of -0.13. The limited evidence on weekend headway elasticities indicates that these values are similar to the off-peak weekday elasticities. However, the data from the 1975 Madison (11) and 1962 Detroit Grand River Avenue demonstration (8) show that the bus headway elasticities on Sunday were larger than those on Saturday.

Commuter-Rail Headway Elasticities

Analysis of the commuter-rail headway elasticity values [Table 2 (7,10)] shows the mean elasticity for all hours to be -0.47 ± 0.14, which is congruent with the mean headway elasticity value obtained for bus service. Furthermore, most of the generalizations made for bus headway elasticities are confirmed by similar experiences with commuter-rail elasticities. As presented in Table 2, the commuter-rail elasticities, which were estimated from the five-corridor demonstration in the Boston area in 1962-1964 (10), show an aggregate mean off-peak elasticity of -0.65 ± 0.19, approximately 82 percent above the mean peak elasticity value of -0.38 ± 0.16. The comparison of peak with off-peak elasticities for the Lowell and Reading corridors, which had approximately identical headways for both periods, reinforces the conclusion that off-peak ridership is more responsive to service improvements, since the off-peak period elasticities were 70-76 percent higher than the peak-period elasticities in these corridors.

Table 1. Bus headway elasticities by service level and time period.

Service Level[a]	Peak Hours Mean	Peak Hours SD	Peak Hours No. of Cases	Off-Peak Hours Mean	Off-Peak Hours SD	Off-Peak Hours No. of Cases	Weekends Mean	Weekends SD	Weekends No. of Cases	All Hours Mean	All Hours SD	All Hours No. of Cases	Aggregate Value Mean	Aggregate Value SD	Aggregate Value No. of Cases
High	-0.27	±0.14	2	-0.19	±0.09	3	-0.22	-	1	-0.25	-	1	-0.22	±0.10	7
Medium	NA	NA	NA	-0.49	±0.20	3	-0.43	±0.16	3	NA	NA	NA	-0.46	±0.18	6
Low	-0.58	-	1	-0.71	±0.11	3	NA	NA	NA	-0.51	±0.20	6	-0.58	±0.19	10
Aggregate value	-0.37	±0.19	3	-0.46	±0.26	9	-0.38	±0.17	4	-0.47	±0.21	7	-0.44	±0.22	23

[a]Levels of service classified as follows: high, <10-min headways; medium, 10- to 50-min headways; low, >50-min headways.

Table 2. Commuter-rail headway elasticities by service level and time period.

Service Level[a]	Peak Hours Mean	Peak Hours SD	Peak Hours No. of Cases	Off-Peak Hours Mean	Off-Peak Hours SD	Off-Peak Hours No. of Cases	All Hours Mean	All Hours SD	All Hours No. of Cases	Aggregate Value Mean	Aggregate Value SD	Aggregate Value No. of Cases
Medium	-0.38	±0.16	5	-0.46	±0.09	2	-0.41	±0.09	4	-0.41	±0.13	11
Low	NA	NA	NA	-0.78	±0.10	3	-0.69	-	1	-0.76	±0.10	4
Aggregate value	-0.38	±0.16	5	-0.65	±0.19	5	-0.47	±0.14	5	-0.50	±0.20	15

[a]Levels of service classified as in Table 1.

VEHICLE-MILE ELASTICITIES

In this section, aggregate vehicle-mile service elasticities are discussed, whether they relate to frequency, route length, route density, or service-hour changes. In fact, little is known about differences in elasticities among these components of vehicle miles.

Although most work in estimating vehicle-mile service elasticities has been developed from cross-sectional and time-series studies, two important studies that monitored the effects of individual fare and service changes were performed for the cities of San Diego and Atlanta. In San Diego, Kemp (13) and Goodman, Greene, and Beesley (14) developed vehicle-mile elasticities by using least-squares regressions of time-series data over the 40-month period, during which service expanded by approximately 80 percent. The aggregate vehicle-mile elasticity varied from +0.75 to +0.85. In Atlanta, where more service was available and where service expansion occurred over a much shorter period of time, Kemp (15) estimated a vehicle-mile elasticity of +0.30.

The results from transportation demand-modeling efforts that use nonexperimental data confirm the San Diego and Atlanta results that transit demand response is inelastic to variations in vehicle miles. The mean service elasticity for all 28 cases analyzed in Table 3 (7) is +0.61 ± 0.31, a value slightly larger than the mean elasticity obtained from studies of headway variations. As shown in Table 3, vehicle-mile elasticities during the peak period are found to be only half the value observed during off-peak hours. Again, this indicates the varying ridership responsiveness at different levels of service. The mean bus-mile elasticity of +0.64 is twice the elasticity of +0.30 observed for rapid-rail service. This observation must be tempered by the lack of cases for rapid-rail service.

TRAVEL-TIME ELASTICITIES

Perhaps the most important factor that affects public transportation ridership is travel time. Unfortunately, measuring ridership response to total travel-time changes as well as to changes in trip-time components is a difficult task. In contrast to the previous sections on service elasticities, there has been scant experimentation with travel-time variations.

In-Vehicle Travel-Time Elasticities

The only travel-time elasticities available from bus-monitoring studies are estimates of ridership response to in-vehicle travel-time improvements obtained from bus priority demonstrations in three cities--Seattle, Miami, and Boston. As shown in Table 4 (7,16-18), the aggregate elasticity from the demonstration data is -0.35 ± 0.21. However, the aggregate elasticity is dominated by peak-period elasticities, which make up 90 percent of the observations.

The results of the 1970 Seattle Blue Streak demonstration (17) can be used to analyze the differential effects of time periods on the in-vehicle time elasticities. Seattle's peak-period reverse-commute service elasticity of -0.55, although smaller than the off-peak value of -0.83, is 25 percent larger than the travel-time elasticity of -0.44 obtained in the peak direction.

The estimation of in-vehicle time elasticities from mode-choice models results in much higher estimates than those from demonstrations. The results of 12 cross-sectional models reviewed by Ecosometrics, Inc. (7) show mean elasticities of -0.70 ± 0.10 (two cases) for rapid rail and -0.68 ± 0.32 (seven cases) for bus--estimates twice the size of the values from demonstrations. Although slightly smaller than the mean, McFadden's (3) bus and rapid-rail in-vehicle travel-time elasticities (-0.46 to -0.60) are relatively similar and relatively close to the demonstration elasticities. Talvitie (19) shows a large mode-choice elasticity for bus service of -1.10; however, his elasticities greatly exceed those observed from the demonstration projects and consequently are suspect.

In 1977, Hepburn (20) analyzed the commuter-rail routes that served the London metropolitan area during the period 1966-1971. The in-vehicle travel-time elasticities he obtained were -0.49 for routes shorter than 25 miles and -0.86 for routes longer than 25 miles.

Table 3. Vehicle-mile elasticities from nonexperimental data by mode and time period.

	Peak Hours			Off-Peak Hours			All Hours			Aggregate Value		
Mode	Mean	SD	No. of Cases	Mean	SD	No. of Cases	Mean	SD	No. of Cases	Mean	SD	No. of Cases
Bus	+0.33	±0.18	3	+0.63	±0.11	3	+0.69	±0.31	17	+0.64	±0.30	23
Rapid rail	+0.10	-	1	+0.25	-	1	+0.55	-	1	+0.31	±0.19	3
Bus and rapid rail	NA	NA	NA	NA	NA	NA	+0.77	±0.27	2	+0.77	±0.27	2
Aggregate value	+0.27	±0.19	4	+0.54	±0.20	4	+0.69	±0.30	20	+0.61	±0.31	28

Table 4. In-vehicle time elasticities by time period.

	Elasticity			
Time Period	Mean	SD	No. of Cases	Bus Priority Project
Peak	-0.29	±0.13	9	Miami I-95, Seattle Blue Streak, Boston Southeast Expressway
Off-peak	-0.83	-	1	Seattle Blue Streak
Aggregate value	-0.35	±0.21	10	All the above

Out-of-Vehicle Time Elasticities

All the evidence regarding out-of-vehicle time elasticities comes from nonexperimental data estimates, mainly from mode-choice models. The mean elasticity of total out-of-vehicle time is -0.59 ± 0.15, a value in general agreement (in spite of the fact that its value is derived from only three studies) with the headway elasticity values estimated earlier. It is reasonable to expect headway and out-of-vehicle time elasticities to be similar, since wait and transfer times (the major components of out-of-vehicle time) are equal to half the headway when very frequent transit service is provided or when the schedule is unknown and passengers arrive at transit stops at random.

The evidence on component out-of-vehicle time elasticities (i.e., walk-, wait-, and transfer-time elasticities) is mixed, especially in relation to in-vehicle travel-time elasticities. The value of out-of-vehicle time has been estimated by several investigators--for example, Quarmby (21)--to be two to three times greater than the value of in-vehicle time. A mode-choice model estimated for Stockholm and other Swedish cities by Algers, Hansen, and Tegner (22) resulted in relative values of waiting times that were 3-12 times the in-vehicle travel-time values. This study also indicated that the relative waiting-time value will increase rapidly as headways are increased, a finding that corresponds to the earlier conclusion that the absolute value of headway elasticities is directly proportional to the level of service.

The walk-time elasticities estimated by Pratt and DTM, Inc. (23) for Minneapolis-St. Paul are very small, as shown in Table 5. The value for all work trips is -0.26, or half the in-vehicle time elasticity; for nonwork trips, the walk-time demand elasticity is -0.14. Passenger demand on bus routes that lead to the central business district (CBD) was estimated by Pratt to be less elastic to changes in walk time than the demand on non-CBD-oriented routes.

The study also shows that wait-time elasticities are only slightly larger than walk-time elasticities. As a rule of thumb for planning headways and route density, transit planners equalize the average wait time at a bus stop to the average walk time to the stops (24). This allocation of buses to routes suggests that wait- and walk-time elasticities are equivalent, as confirmed by the Pratt model.

As shown in Table 5, a wait-time elasticity for Montreal (25) that is twice the size of the in-vehicle time elasticity is presented; however, just the opposite is shown for San Francisco (3) and the results for Minneapolis-St. Paul (23) suggest that the difference is dependent on the trip purpose. McFadden's transfer-time elasticities for peak-hour service in San Francisco are higher than the comparable first-wait-time elasticities. Note also that although the rail transfer-time elasticity is greater than the values observed for bus service, the opposite is true for first-wait time. The inconsistencies in Table 5 point out the need for controlled demonstrations of transit service on this subject.

TRANSFERS

By using mode-choice estimation models, Algers, Hansen, and Tegner (22) discovered that the overall cash value of a transfer was 30 percent higher than the cash fare per trip and corresponded to approximately 24 min of door-to-door travel time. Thus, passengers appear to be willing to pay more than twice the base fare to avoid having to transfer. Their model showed that the value of avoiding a transfer was greater for bus than for rail, primarily because of the higher potential discomfort in transferring from buses.

In one of the few studies to focus on transit demand and the number of transfers, Pratt and DTM, Inc. (23) estimated a transfer elasticity of -0.59 in their nonwork mode-split model for Minneapolis-St. Paul. This value is much larger than the wait-time and transfer-time elasticities estimated from the same three-mode choice model (-0.24 and -0.17, respectively); this confirmed the previously mentioned studies that showed that avoidance of transferring is more important to the user than the time spent waiting for a bus.

SEAT AVAILABILITY

The importance of seat availability for transit users has been documented in several studies. For example, Algers, Hansen, and Tegner (22) attempted to quantify the value of getting a seat by introducing a dummy variable into their logit mode-choice models to test the hypothesis that those who do not get a seat value their travel time more than those who get a seat. They found that the trip value for individuals who do not have a seat was 40-75 percent higher than the travel-time value for people who have a seat.

As part of a service-improvement demonstration between Vancouver (Washington) and Portland (Oregon), sponsored by the Urban Mass Transportation Administration (UMTA) Office of Service and Methods Demonstrations, seating capacity on TRI-MET's Line 5 was increased by more than 40 percent by adding a trailer bus to six peak-period runs (26). The increase in ridership attributable to the availability of seating resulted in an elasticity of

Table 5. Comparison of in-vehicle time and component out-of-vehicle time elasticities.

Type of Elasticity	Montreal Bus and Rapid Rail	San Francisco Bus (two-mode)	San Francisco Bus (three-mode)	San Francisco Rapid Rail	Minneapolis-St. Paul Bus (work)	Minneapolis-St. Paul Bus (nonwork)
In-vehicle time	-0.27	-0.46	-0.60	-0.60	-0.52	-0.12
Out-of-vehicle time						
Walk	NA	NA	NA	NA	-0.26	-0.14
Wait	-0.54	-0.17	-0.19	-0.12	-0.32	-0.21
Transfer	NA	-0.26	-0.29	-0.66	NA	NA

+0.65. We calculated this elasticity from the data presented by Systan, Inc. (26) and assumed that the percentage of passengers seated was equal to the probability of getting a seat. This relatively high value is approximately 40 percent larger than McFadden's bus in-vehicle time elasticity.

INTERACTIONS OF TRANSIT FARES AND SERVICE LEVELS

The service elasticities presented in this paper and fare elasticities presented elsewhere (7,27,28) indicate that transit demand is inelastic to both fares and services. Consequently, independent variations of fares and services will not by themselves increase both revenues and patronage at the same time. For example, an increase in service--without a corresponding fare change--will probably not result in revenue increases large enough to cover the extra costs of the service improvement because the proportional change in patronage is less than the proportional changes in service.

Aggregate service elasticities (measured in vehicle miles), however, are twice as large as aggregate fare elasticities, which suggests that passengers are more responsive to service changes than to fare changes. On the aggregate levels, this is true. However, because both fare or service elasticities vary considerably from one area to another and by the time of day, type of route, and other classifications, this generalization is not always true. For example, by using the data presented in this paper, the mean bus headway elasticity on routes that have less than 10-min headways is -0.19 during off-peak hours. The average off-peak fare elasticity for bus service, however, may be only -0.35. Since the service elasticity is so low, a transit operator cannot hope to increase ridership and revenues substantially by further headway improvements. If headway adjustments are contemplated, then they should be reduced and the operating-cost savings should be applied either to other corridors that have relatively poor service or to the same route in the form of a fare reduction.

Patronage losses associated with attempts to increase revenue can be minimized by increasing fares only for users who exhibit small demand elasticities, such as commuters. The service saved as a result of reduced demand, albeit small during the peak period, could be applied to routes that have relatively poor service and result in further revenue increases if the patronage gained by the service adjustment is greater than the patronage lost due to the fare increase. Since the marginal cost per vehicle hour of operation during off-peak periods is at least 30-50 percent lower than that during the peak period (29,30), the cost savings due to the reduction in peak service should be applied to off-peak routes that have infrequent service, which would make possible a further gain in total ridership and revenues.

If the disaggregate fare and service elasticities are known for a particular transit market, the ridership or revenues generated by a particular action or set of actions could be improved by manipulating both the fare and the service levels. If, for analytical purposes, the revenues generated by an improvement in service are assumed equal to the additional costs of providing that service (i.e., a situation in which operating costs break even) and if the fare and service elasticities are not numerically equivalent, then transit ridership can be increased with no net effect on revenues by proper fare and service adjustments. These adjustments will in turn cause the demand elasticities to change if the elasticities are assumed variable and dependent on the respective fare and service levels. Opportunities for further ridership increases will cease when the fare and service elasticities are equal (31,32). Thus, when the service elasticity for a particular market is larger than the fare elasticity, a transit agency should raise fares and use the revenues produced to finance service improvements. Conversely, if the fare elasticity is larger than the service elasticity, then fares should be decreased and the revenue loss covered by the cost savings of a simultaneous service reduction.

As an example, Table 6 presents two fare- and service-adjustment strategies to increase total bus ridership with no change in net revenue, based on disaggregate fare and service elasticities. For convenience in analysis, the model assumes a situation in which operating costs break even, so that revenue-cost considerations can be deemphasized (32), and aggregate fare and headway elasticities are -0.35 and -0.47, respectively; adjustment factors presented by Ecosometrics, Inc., are applied (7).

The two strategies presented in Table 6, however, are not the only fare- and service-adjustment options available for increasing patronage. The peak to off-peak cross-subsidy scenario described earlier is an example of such an alternative. Whatever service-adjustment decision is made, the premise on the extent to which transit riders are willing to pay more for improved service or trade one service attribute for another must be based on the disaggregate fare and service elasticities.

In spite of the obvious need for more analysis of the interactions between fares and services, most of the demand approaches, whether from monitoring demonstrations or the more sophisticated mode-choice models, explicitly ignore the possibility of analyzing fare and service interactions by assuming constant-elasticity models (i.e., assume the interactions to be zero). These constant-elasticity models should be deemphasized in favor of variable-elasticity models that have interaction effects, such as the translog models (33).

SUMMARY

This paper has shown that transit demand is service-inelastic. Evidence of this less-than-proportional response of changes in patronage to changes in transit service is provided by the fact that all demonstration studies and modeling efforts reveal service elasticity values less than 1.0.

As we have shown elsewhere (7), service elastici-

Table 6. Example of bus-fare and service interaction strategies.

Bus Headway Level	Peak Period[a]		Off-Peak Period[a]	
	Disaggregate Service Elasticity	Strategy	Disaggregate Service Elasticity	Strategy
Frequent (<10 min)	-0.15	A	-0.26	A
Medium (10-50 min)	-0.31	B	-0.54	B
Infrequent (>50 min)	-0.39	B	-0.68	B

Note: A = finance a fare reduction with the cost savings from a service reduction; B = finance a service improvement with the revenue from a fare increase.

[a]Fare elasticities of -0.21 for peak and -0.48 for off-peak periods.

ties are generally larger than fare elasticities, which suggests that passengers are more responsive to service changes than to fare changes. However, because service elasticities vary considerably from one area to another and by the time of day (with off-peak elasticities 50-100 percent higher than those observed during the peak), type of route, service quality (with larger elasticities in low-service areas), and other classifications, this generalization is not always true. Fare elasticities, for example, may be larger than service elasticities when bus headways of less than 10 min are present. The differences in disaggregate fare and service elasticities may present transit operators with opportunities for ridership and revenue improvements.

Finally, this paper has noted a general consistency of headways, bus miles, and in-vehicle time elasticities from service demonstrations and inconsistencies in results from mode-choice models, particularly in out-of-vehicle time values such as walk-, wait-, and transfer-time elasticities.

ACKNOWLEDGMENT

This research was funded by UMTA, U.S. Department of Transportation. The study was prepared for the Pricing Division of the Office of Service and Methods Demonstrations within UMTA's Office of Transportation Management and Demonstrations. Opinions expressed in this paper are ours.

REFERENCES

1. A. Grey. Urban Fares Policy. Heath, Lexington, MA, 1975.
2. E. Kuh and J.R. Meyer. How Extraneous Are Extraneous Estimates? Review of Economics and Statistics, Nov. 1957.
3. D. McFadden. The Measurement of Urban Travel Demand. Journal of Public Economics, Vol. 3, 1974.
4. T.H. Oum. A Warning on the Use of Linear Logit Models in Transport Mode Choice Studies. Bell Journal of Economics, Vol. 10, No. 1, 1979.
5. H. Theil. On the Estimation of Relationships Involving Quantitative Variables. American Journal of Sociology, Vol. 76, 1970.
6. M. Nerlove and S.J. Press. Multivariate Log-Linear Probability Models for the Analysis of Qualitative Data. Center for Statistics and Probability, Northwestern Univ., Evanston, IL, 1976.
7. Ecosometrics, Inc. Patronage Impacts of Changes in Transit Fares and Services. UMTA, U.S. Department of Transportation, 1980.
8. Grand River Avenue Transit Survey. Department of Street Railways, Detroit, MI; Housing and Home Finance Agency Demonstration Grant Program, Final Rept., 1963.
9. Chesapeake Mass Transportation Demonstration Project. Wilbur Smith and Associates, New Haven, CT, 1969.
10. J.F. Maloney and Associates. Mass Transportation in Massachusetts. Commonwealth of Massachusetts Mass Transportation Commission, Boston, 1964.
11. K. Hicks. A Study to Determine Consumer Reaction to Weekend Service Changes and to Evaluate the Downtown Transit Information Center Demonstration Program. Department of Transportation, Madison, WI, 1979.
12. P. Mullen. Estimating the Demand for Urban Bus Travel. Transportation, Vol. 4, No. 3, 1975.
13. M.A. Kemp. Reduced Fare and Fare-Free Urban Transit Services--Some Case Studies. Urban Institute, Washington, DC, 1974.
14. K.M. Goodman, M.A. Green, and M.E. Beesley. The San Diego Transit Corporation: The Impacts of Fare and Service Changes on Ridership and Deficits, 1972-1975. Urban Institute, Washington, DC, 1977.
15. M.A. Kemp. Transit Improvements in Atlanta--The Effects of Fare and Service Changes. Urban Institute, Washington, DC, Rept. 12/2-2-1, March 1974.
16. J.A. Wattleworth, K.G. Courage, and C.E. Wallace. Evaluation of the Effects of the I-95 Exclusive Bus/Carpool Lane Priority System on Vehicular and Passenger Movement. Univ. of Florida, Gainesville, Rept. II-2, 1978.
17. A.M. Voorhees and Associates, Inc. Blue Streak Bus Rapid Transit Demonstration Project. U.S. Department of Transportation, 1973.
18. J.H. Dupree and R.H. Pratt. A Study of Low Cost Alternatives to Increase the Effectiveness of Existing Transportation Facilities--Results of Case Studies and Analysis of Busway Applications. U.S. Department of Transportation, 1973.
19. A. Talvitie. A Direct Demand Model for Downtown Work Trips. Transportation, Vol. 2, 1973.
20. D.R.C. Hepburn. Analysis of Changes in Rail Commuting to Central London, 1966-1971. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, TRRL Suppl. Rept. 268, 1977.
21. D.A. Quarmby. Choice of Travel Mode for the Journey to Work: Some Findings. Journal of Transport Economics and Policy, Vol. 1, No. 3, 1967.
22. S. Algers, S. Hansen, and G. Tegner. Role of Waiting Time, Comfort, and Convenience in Modal Choice for Work Trip. TRB, Transportation Research Record 534, 1975, pp. 38-51.
23. R.H. Pratt and Associates and DTM, Inc. Development and Calibration of Mode Choice Models for the Twin Cities Area. Metropolitan Transportation Council, Minneapolis-St. Paul, MN, Nov. 1976.
24. F.V. Webster. Urban Passenger Transport: Some Trends and Prospects. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, TRRL LR 771, 1977.
25. M. Gaudry. An Aggregate Time-Series Model of Urban Transit Demand: The Montreal Case. Transportation Research, Vol. 9, 1975.
26. Systan, Inc. Vancouver Price and Service Improvement Demonstration: Phase I Report. Transportation Systems Center, U.S. Department of Transportation, Cambridge, MA, 1980.
27. A.M. Lago, P.D. Mayworm, and J.M. McEnroe. Transit Ridership Responsiveness to Fare Changes. Traffic Quarterly, Vol. 35, No. 1, 1981, pp. 117-142.
28. M. Kemp. Some Evidence of Transit Demand Elasticities. Transportation, Vol. 2, 1973.

29. G.I. Thompson. Analysis of Urban Public Transportation Policy. Stanford Research Institute, Menlo Park, CA, 1971.
30. H. Mohring. Optimization and Scale Economies in Urban Bus Transportation. American Economic Review, Vol. 62, No. 4, 1972.
31. I.G. Black. Reduced Fares or Increased Frequency? Univ. of Warwick, Warwick, England, Working Paper 34, 1976.
32. F.V. Webster. Some Implications of the Results of Elasticity Determinations: Factors Affecting Public Transport Patronage. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, TRRL Suppl. Rept. 413, 1977.
33. L.R. Christensen, D.W. Jorgenson, and J.L. Lawrence. Transcendental Logarithmic Utility Functions. American Economic Review, Vol. 65, No. 3, 1976.

Publication of this paper sponsored by Committee on Traveler Behavior and Values.

Early Responses to Taxi Regulatory Changes in Three Cities

PAT M. GELB

Taxi regulatory changes and preliminary responses to them in San Diego, California; Portland, Oregon; and Seattle, Washington, are discussed. The full effects of the regulatory and industry changes are being evaluated. Each city relaxed its entry restrictions in some way; all provided for increased latitude in rate setting, but the specific provisions have varied. The impetus for regulatory revision was generally similar—to transfer the responsibility for regulating entry and establishing rates from the city government to the marketplace. The regulators hope to produce a greater range of improved taxi services by increasing competition and providing for flexible rate structures. Implementation of the new regulations and the earliest responses in terms of local industry size and rate structures are the main topics here. Preliminary analysis suggests that these first responses relate to conditions in the local setting. Problem areas identified during the implementation phase are highlighted, and a number of transferable implications that suggest themselves to other regulatory entities are presented. Findings of the analyses of the effects of the regulatory changes on the supply of and demand for taxi services are anticipated soon.

This paper reports on taxi regulatory changes in San Diego, California; Portland, Oregon; and Seattle, Washington. The implementation and effects of these changes are being evaluated by De Leuw, Cather and Company under contract to the Transportation Systems Center (TSC) of the U.S. Department of Transportation as case studies under the Urban Mass Transportation Administration (UMTA) Service and Methods Demonstration program. Each city adopted its new taxi regulations during 1979, so sufficient time has elapsed to permit identification of the early responses in terms of industry and rate structures while the analysis of operating and ridership data proceeds.

The impetus for regulatory change was similar in each city. Local regulators had experienced difficulties in administering their taxi regulations. In one case, alleged misconduct in the approval of a rate increase precipitated a citywide scandal that ultimately involved the indictment of every city council member. The regulators also began to doubt that the existing code provisions offered any guarantees of a balance between supply of and demand for taxi services or between operating costs and rates of fare. Population ratios were insufficiently sensitive to demand, whereas the data required to demonstrate the need for rate increases were difficult to interpret, costly to assemble, and required the regulators to rely on documentation supplied by the regulated service providers. Concepts like percentage rate of return on invested capital and ratios of overall operating costs to revenues appeared simply to guarantee that taxi rates would go up with costs.

The regulators also doubted that the existing laws served to preserve adequate levels of service. Financial difficulties had plagued the local industries during the 1970s, but one city rate analyst had demonstrated that taxi ridership had declined with each recent rate increase and asserted that rising fares produced a net loss in revenues. Some of the existing regulations inhibited taxicabs from serving a wider transportation market by preventing shared riding, fixed-route services, or differential pricing. Limited entry was charged with contributing to monopoly values in taxi licenses and suppressing competition, which impeded the very kinds of pricing and service innovations that these regulators saw as essential to the salvation of a declining industry.

The following sections describe the regulatory revisions and industry characteristics before and after the changes in each city. The responses to date across sites as well as of some of the problems that have arisen during and since the implementation phase are both discussed. The final section summarizes some transferable implications that have been found for other regulatory entities.

Evalution is in progress of the full effects of the regulatory changes on taxi operators in terms of trips per shift or fare or lease revenues or on taxi riders in terms of taxi availability or response times. At this writing, the collection of operation and ridership data was nearing completion in San Diego and had just begun in Seattle. (The Portland case study is a lower-level monitoring effort.)

SAN DIEGO

Regulatory Changes

The taxicab regulatory revisions adopted in San Diego have two major elements: (a) effective January 1, 1979, the previous ceiling on taxi permits was removed and entry was opened at a specified rate of new permits per month to independent owner-operators as well as to companies; (b) beginning August 1, 1979, competitive pricing, by which operators could charge individual rates up to an established maximum, replaced the citywide standard rate of fare. The council authorized issuance of 6 new permits per month between January and July 1979 and 15 permits per month from July 1979 to early 1981. It set the maximum rate of fare for exclusive-ride service at $1.50 drop and $1.50/mile. No maximum was imposed for fixed-route service, which was to be charged on a per-capita basis. Operators are required to file all rates with the San Diego paratransit office.

As of October 1980, the San Diego council removed the maximum rate and voted to permit customers and drivers to bargain for rates below the operators' filed rates for all types of services. Thus an operator's filed rate effectively becomes an individual maximum rate.

Additional code changes removed the previous certification requirement of public convenience and necessity and included specific code categories for all paratransit modes; changed the applicable regulatory fees; codified an appeals procedure for denial, suspension, and revocation of permits; and reassigned various regulatory responsibilities. The major features of the regulatory changes are summarized in Table 1.

Industry Characteristics

Prior to regulatory revision, the San Diego taxi industry was dominated by a single large operator, Yellow Cab, which held 280 (68 percent) of the total 411 licenses in 1978. Some 62 independents had obtained licenses in the wake of the 1976 Yellow Cab bankruptcy; 59 of these were still operating in 1978. The remaining 72 licenses were held by nine relatively small fleet operators with 5-15 taxis each; two of these were county-licensed fleets that held three limited city certificates.

Available evidence indicates that the number of outstanding licenses had exceeded the previous population-based ceiling of one license per 3000 residents at least as early as 1969. There was a waiting list of 230 applicants for new taxicab licenses, which included some 30 applications from existing operators. Most were independents; many had originally signed the list as part of an organized strike action against Yellow Cab during 1976 and continued to advocate open entry. The multicertificated operators opposed it; then they argued that the average number of trips per shift had declined since before 1976 and that demand was not sufficient to sustain additional suppliers.

Applicants have continued to file for new San Diego permits at a steady pace since open entry was established. During 1979, 75 taxicabs were put into service, which increased the license total by 18 percent. There were 629 San Diego taxicabs by the end of 1980, which raised the 1979 license total by 30 percent and produced an increase of 53 percent over the 1978 level. There were 180 firms at the end of 1980 compared with 69 firms in 1978. The number of permits held by fleet operations had more than doubled, whereas the number of independently held permits (in companies with one or two taxis) had more than tripled. Table 2 shows the distribution of permits by operator type before and after open entry.

License transfers (sales) have continued under open entry because of the long wait for new permits obtained through the applicants' list. Nine transfers were recorded during 1979 and 27 during 1980; six companies also went out of business during 1980 for which there is no record of a transfer. All the transfers have been from independent owner-operators; 15 were members of the original group of independents licensed in 1977. A few so-called transfers were consummated as soon as the new permit was obtained by applicants who apparently were not com-

Table 1. Major taxi regulatory changes, San Diego.

Category	Past	Revised (1979)
Entry requirement	Public hearing (city council); certification of public convenience and necessity; council resolution; limit on total permits according to population ratio established by council policy	Permit process (city manager or paratransit office); rate of permit issuance determined by council policy, currently set at 15 new permits/month
Right of appeal on denial, suspension, or revocation of permit	Not specified	Written appeal to city manager within 10 days; procedure on appeal requires hearing
Taxicab fare rate	Standard rate; public hearing on operator petition; council resolution	Rates by type of service; maximum rate (council resolution) for taxi services, up to which operators may charge individual rates, removed October 1980; public hearing (operator petition); necessity to file rates (paratransit office); bargaining below filed rates
Shared-ride rate	Meter to be reactivated after first passenger's destination	To be charged on per-zone basis
Equipment and specifications	Taxi meters required; identifying color scheme required; driver and vehicle identification number to be displayed; two-way radio communication required by council policy	Same, except two-way radio communication/dispatching capability required by code of all taxicabs operating under permits or certificates granted after October 31, 1976, and exterior rate posting required
Public liability	Insurance by authorized carrier required; minimum amounts to be set by council	Self-insurance permitted with council approval; minimum amounts to be set by city manager

Table 2. Distribution of taxi permits by operator type before and after open entry, San Diego.

	Certificates, Dec. 1978		Permits			
			Dec. 1979		Dec. 1980	
Item	No.	Percent	No.	Percent	No.	Percent
Operator type						
Large fleet (Yellow Cab Company only)	280	68	280	58	281	45
Midsize fleet (companies of 3-28 taxis each)	72	18	93	19	161	25
Two-taxi firms	0		10	2	50	8
One-taxi firms	59	14	104	21	137	22
Total	411		487		629	
Net change since 1978				+18		+53

mitted to beginning a taxi business and who may have applied for the permit with the intent of selling it.

With the exception of one operator who held limited permits and who died during 1979, no preexisting San Diego fleet operation has failed or decreased its permits since open entry. Indeed, all the original fleets have added permits, either by direct application to the city or by acquisition of independent permits. Nine new fleets have emerged that have 3-18 permits each, which accounts for 60 new permits in all. One county fleet has obtained three city permits. Five of the six companies that went out of business during 1980 were originally independents. Of the original group, 39 (66 percent) were still operating at the year's end, however; 12 of these had burgeoned into fleet-type or two-taxicab firms.

Changes in Fare Structure

Before variable pricing, San Diego taxi rates were established through procedures common to many municipalities throughout the country. Operators petitioned the council for changes, and the council evaluated their requests on the basis of operator information supplemented by its own notions of fairness. The San Diego council has enacted changes on an overall average of once every three years between 1956 and 1977. The average 3.5-mile trip fare had risen 146 percent from $1.40 in 1947 to $3.45 in 1978.

The city's prerevision standard rate was $0.80 drop, including the first 1/7 mile, and $0.70/mile. This was supplemented by a $0.30 gasoline surcharge in effect prior to variable pricing. As of August 1979, operators could file individual rates up to the maximum of $1.50 drop and $1.50/mile and refile as often as they wished. (Rates must be posted on the exterior of the taxi.)

The multicertificated firms had sought a rate increase prior to the new legislation. Although early in the public-discussion phase the independents claimed that they could make a profit at the pre-1977 rate of fare, more independents than fleets filed higher rates under the new maximum. To date, independents have filed more rate changes than fleet operators have, and both the highest and the lowest rates have been those of the independents. The most common San Diego rates are $1.20 drop and $1.00/mile (filed by the major fleet and others) and $1.00 drop and $1.00/mile (filed by the largest association of owner-operators and others). These rates represent an average 35 percent increase over the prerevision rate for a 3.5-mile trip.

The city council further modified its rate regulations as of October 1980 by lifting the maximum rate of fare and providing for operators' filed rates to act as an individual maximum under which bargaining would be permitted. This provision legalized several practices that had existed informally. It was common practice for passengers to bargain with drivers for fares lower than the posted rate; some rates were allegedly filed high with this in mind. Also, an association of independent owner-operators has offered discount scrip to its customers as a promotional gimmick. This sort of innovation, although it was precisely what the regulators sought to encourage, was illegal prior to this latest code revision.

A new airport taxi-rate policy has been established that is expected to restrain any sudden rise in city taxi rates. Owing to the queue problems attendant on variable pricing at the airport, the port district voted to limit taxi rates to a range plus or minus 20 percent of the weighted average of all city taxi rates. Operators who file city rates outside this range would therefore have to possess a dual-metering capability to continue their airport operations. In protest against this policy, a few operators filed (and some reportedly attempted to charge) exorbitant rates in an attempt to influence the weighted average. (It should be noted that the operator costs involved in changing rates--for recalibrating the meter and changing the exterior decal--are not negligible. Operator estimates of these costs vary from $50 to $75 per taxi.) Airport taxi problems will be discussed in more detail later.

SEATTLE

Regulatory Changes

Effective June 1, 1979, the city of Seattle removed its previous numerical limitation on taxi licenses and opened entry continuously to qualified applicants. There is no waiting list for new licenses. Vehicle safety and equipment standards were raised, and the groundwork was laid for an ongoing process of stiffening taxi-driver qualifications. The standard rate of fare was replaced by open rate setting by which taxi operators may file individual rates and change them as many as four times a year. Rates must be filed with the Department of Licenses and Consumer Affairs (DLCA); there is no maximum rate. Minimum operating requirements were eliminated as was the requirement that operators maintain trip sheets.

The 1979 legislation realized taxi-licensing reforms that had been pending since 1974. Milestone interim legislation had been adopted in 1977, which permitted contract rates and established taxi license reciprocity with King County, a measure that eased entry to the lucrative airport market for city-licensed taxis. The major features of the past and the revised regulations are summarized in Table 3.

Industry Characteristics

Prior to regulatory revision, Seattle had 240 regularly licensed taxicabs for the 1978-1979 license year. In addition there were 68 licenses involved in litigation following their revocation by the city for failure to meet the minimum operating requirement and the subsequent appeal of this action by the operators. These taxicabs were allowed to operate without a license pending the outcome of the appeal. There were also 92 King County reciprocal licenses, which made a total of 400 taxicabs, including those with the disputed licenses. Finally, there were 29 standby licenses issued without fee, for use in the case of a disabled vehicle or temporary loss of a regular license. These were not counted as regular licenses.

The majority (73 percent) of the 400 licenses were held by three large service companies of individual member-owners. There was one large fleet-type operator who had 27 licenses (7 percent), whereas the rest were held by many small firms and independents not affiliated with any association. Two of the service companies had a structure similar to that of a cooperative, in which members held stock in the corporation and shared the costs of dispatching and other joint services. The third service company's cabs were held primarily by one owner, who rented service to smaller firms.

The Seattle license ceiling was frozen at the existing number of permits in February 1977 as part of the interim legislation. Prior to 1977, licenses were limited by a ratio of one per 2500 population. Available evidence indicates that this ceiling had been exceeded due to the "grandfathering in" of taxicabs that were operating under licenses granted prior to September 1966. On the other hand, some 50 licenses that had not been renewed could have been issued for the 1978-1979 license year. The large numbers of revoked and renewed licenses indicate that the actual supply of Seattle taxicabs was well below the permit ceiling.

Comparing Seattle taxicab licenses before and after open entry is complicated by the fact that new licensing rules and categories apply to the postrevision estimates. The categories of county-reciprocal and standby licenses no longer exist, whereas the elimination of the minimum operating requirement removes the distinction that affected the disputed licenses. At the close of the first license year following open entry (September 1980), there were 522 outstanding licenses. The service companies' share of total licenses had dropped to 62 percent, and the number of independent firms that had one or two taxis each and were not affiliated with any association had risen to 81. Fifteen firms had 3-10 permits each, and the one large fleet operator (who concentrated on the airport) held 35.

By February 1981, halfway through the 1980-1981 license year, there were 497 licenses. The service companies' share of the total was relatively unchanged at 61 percent, whereas the large fleet operator had failed to renew more than a third of its city licenses. (Reports were that it was consolidating its operations in the county.) The smaller fleets retained 83 licenses (17 percent), whereas unaffiliated owner-operators who had one or two taxis accounted for 87 (another 17 percent). These data are summarized in Table 4.

The open-entry legislation prohibits transfers of licenses in Seattle, so there is no incentive for license holders or companies to retain the license after termination of a business. Such licenses are not routinely turned back to the city, however; many cancellations are only discovered through nonrenewals during the next license year. Thus, recorded cancellations tend somewhat to underrepresent taxi-business failures in Seattle. There were 16 cancellations during the 1979-1980 license year; these included representatives of all types of Seattle operations. The 1980-1981 files include 11 cancellations to February 1981; nine were from two of the major service companies.

Changes in Fare Structure

Prior to regulatory revision, Seattle taxi rates were established in a manner similar to those in San Diego and other municipalities. Between 1932 and 1974, the council had approved a fare increase every five to seven years; rates were subsequently increased in 1974 and 1976. The 1976 increase was enacted as a temporary measure and extended every six months through the adoption of open rate setting. This increase provided for a \$0.90 drop charge including the first 1/7 mile and \$0.70/mile. The cost for the average 3.5-mile taxi trip rose from \$2.35 in 1970 to \$3.25 in 1979 (up 38 percent). (Note that the Seattle consumer price index rose 90 percent over the same period.)

Since variable pricing went into effect, the most common Seattle rates have been \$1.00 drop and \$1.00/mile (two of the major companies charge this rate) and \$1.00 drop and \$1.20/mile, charged by the largest service company. Some 22 percent of the small fleets and independent operators are currently charging significantly more than the most frequently charged rate, however; the daytime exclusive-ride rates now vary from \$1.20 drop and \$0.90/mile to \$2.00 drop and \$1.50/mile. All the Seattle service companies and many of the independents have filed discounts for elderly passengers, and some offer lift-equipped vehicles and nighttime service at a premium. Whether all these rates are actually used is not verifiable with current data sources.

Table 3. Major taxi regulatory changes, Seattle.

Category	Past	Revised (1979)
Entry requirement	Licensing required; numerical limit on total licenses (frozen at 1977 level)	Licensing required; no limit on total licenses; "Seattle taxicab plates" as defined to be issued with each license; license fee, \$60 (replacement plates, \$15)
	Holders of valid King County licenses may obtain city license for \$25 and vice versa; fee for first jurisdiction's license, \$100; total for both licenses, \$125 (adopted in 1977)	Joint licensing suspended prior to adoption of open entry by county and not yet reinstated
	Minimum operating requirement of 10 miles/day, 230 days/year	Minimum operating requirement removed
Public liability	Insurance to specified limits from an insurance company required	Insurance limits increased to those required by state law; city not required to be named as additional insured; self-insurance permitted
Rate regulation	Standard rate of fare as established by city council; contract rates may differ from standard rate	Open rate setting; rates to be filed with DLCA director and must be meter-based; changes permitted up to four times per year; contract rates may differ from filed rates; zone-based fares for shared riding provided January 1981
Other requirements	Trip sheets to be kept for each shift operated and maintained on file for five years	Trip-sheet requirement removed; exterior rate posting required February 1981

Table 4. Distribution of taxi licenses by operator type before and after open entry, Seattle.

	Licenses by License Year					
	1978-1979		1979-1980		1980-1981[a]	
Item	No.	Percent	No.	Percent	No.	Percent
Operator type						
Service companies (more than 30 cabs each)	292	73	323	62	305	61
Large fleet (one firm)	27	7	35	7	22	4
Small fleets (3-10 taxis each)	81	20	83	16	83	17
Two-taxi firms			18	3	16	3
One-taxi firms			63	12	71	14
Total	400		522		497	
Net change since 1978-1979				+31		+24

[a]To February.

Sea-Tac Airport recently adopted a ceiling for airport taxi rates 10 percent above the weighted average of all King County taxi rates. Although the ceiling is administered more liberally than it is in San Diego (operators are permitted to round awkward per-mile amounts to the next $0.10), the rule has produced similar responses. That is, a few operators have filed high bogus rates with the county in an attempt to influence the average. The airport switched to the median rather than the average rate for computing the ceiling and now threatens to return to the old, exclusive-franchise approach if the new rules prove infeasible. (Airport issues are taken up again later in this paper.)

PORTLAND

Regulatory Changes

Portland's regulatory changes were both less dramatic than those adopted in Seattle and San Diego and less long-lived. Portland adopted three successive waves of revisions during 1979 and 1980. Effective March 21, 1979, the previous population-based ceiling on taxi permits was removed and entry opened to new operators on the basis of specified service standards and a finding that the public interest was served by the addition of a new supplier. Unaffiliated independents were effectively excluded, however, since the new law required that applicants operate sufficient taxis to provide citywide service, and this was interpreted in practice as no fewer than 10 taxis. Authority for determination of operator qualifications and other taxi regulatory functions was vested in a new official, the taxi supervisor, whose responsibilities consolidated functions formerly divided among the council, the chief of police's office, the business license division, and the traffic bureau.

The March 1979, regulations codified flat rates for shared-ride trips between the airport and downtown and wholly within downtown in addition to the council-established maximum for exclusive-ride service. The new regulations also permitted operators to develop discount or other contractual rates for special groups or services.

In October 1979, the Portland council increased the maximum rate for exclusive-ride service from $1.00 drop including the first 1/9 mile plus $0.90/mile and $0.25/extra passenger. The new maximum was $1.00 drop including the first 1/12 mile plus $1.20/mile and $0.50/extra passenger. All operators filed this rate within a matter of weeks, although one continues to offer a 10 percent discount to elderly and handicapped patrons. The flat rate for trips between the airport and downtown was also increased from $3.00 to $4.00/person. (One individual can still elect to pay the three-person minimum of $12.00.) The downtown flat rate was subsequently eliminated, evidently from lack of use. City staff report that a variety of contract rates is currently available.

Industry Characteristics

All outstanding taxi permits in Portland had been held by companies or associations of owner-operators. In 1978 these included Broadway Cab, which held 113 permits; Radio Cab, which held 102; and New Rose City Cab, which held 11. The total of 226 permits was less than the 253 that would have been permitted under the prerevision ceiling of one license for every 2900 residents. Moreover, some of these permits were unused. Permit holders who wished to leave the industry would typically sell their license back to the association, which would hold it until a buyer was found. Since the licenses reportedly accrued monetary value under entry restrictions, few were ever returned to the city.

The Portland taxi industry was apparently depressed. Owners of taxis driven by hired drivers reportedly operated at an average loss, whereas owner-drivers did only slightly better, receiving no return on their capital investment and lower wages than those paid to hired drivers. In keeping with these conditions, there was no waiting list for new taxi permits and few requests for permits following open entry.

One new firm, Portland Cab, entered the local industry during open entry and obtained seven licenses in May and five additional licenses in December 1979. One of the existing firms also acquired three new licenses. These changes brought the total of outstanding licenses to 241 by the end of March 1980.

Further Regulatory Revisions

On April 24, 1980, the Portland council adopted subsequent regulatory revisions that restated some of the previous criteria for entry and reasserted the council's authority for issuing new permits and for other taxi regulatory matters. This step resulted from difficulties in establishing the supervisor's authority. The three existing operators had appealed Portland Cab's second group of five licenses. Although the appeal was eventually dropped, it demonstrated the vulnerability of the supervisor's discretionary authority to continual challenge and moved officials to submit the revised draft to local industry review. The city also sought to clarify ambiguities it has perceived in the language of the March 1979 law.

This revision listed the factors to be considered in qualifying new entrants; this list included the adequacy of existing public transportation and the need for additional service (the burden of proof is on the applicant), the current ratio of taxi licenses to population, the current use patterns of existing taxicabs, and the commitment of the applicant to the local area. The minimum number of taxicabs required for citywide service was increased to 15, 10 of which must be operational at all times. Last, on June 3, 1980, the council stiffened the requirements for taxi drivers' permits and issued a form for operator submission of financial and operating data on a monthly basis.

City staff acknowledge that their recent revisions may be considered by some as a step backward from the 1979 legislation. But they also maintain that the newest changes encourage entry by minimizing the threat of an appeal and clarifying the application requirements. Interest continues to lag, however. The city's newest company, Portland Cab, obtained three additional permits in February 1981. Another operator applied for four permits at about the same time but at this writing has failed to submit the required documentation to support its application.

It should be noted that the Portland reregulation did not take place in response to an avalanche of permit applications but because of problems in the 1979 law. In fact, Portland's experiment with open entry elicited little immediate response, probably owing to the low profitability of the local industry and the exclusion of unaffiliated owner-operators.

DISCUSSION OF FINDINGS

Early Responses to Open Entry and Variable-Rate Setting

The early responses to open entry and variable-rate setting in these three case studies have seemed to vary with conditions in the local taxi industry and setting. Where the taxi industry appeared relatively healthy and a long-standing list of applicants for new permits had existed prior to open entry, as in San Diego, there was a steady stream of new entrants into the local market. In San Diego, the current license total is 53 percent higher than the 1978 level and all types of operators have obtained new licenses. On the other hand, where there were indications of an oversupply of taxicabs and the industry was relatively depressed and where the entry criteria excluded unaffiliated independents, as in Portland, new entry has been understandably slower.

The Seattle case is more complicated, owing to the large groups of disputed and previously unrenewed licenses and the new categories of licenses after open entry. On the basis of 400 regular licenses during the 1978-1979 license year, there was a 31 percent increase in total licenses during the first year following open entry. License issuance for the first six months of the 1980-1981 license year shows a 5 percent drop in permits from the preceding year but is still 24 percent more than the 1978 level. The new licensees include the existing service companies as well as smaller fleet-type operators and numerous unaffiliated independents. Although San Diego shows its greatest proportional increases and turnover in unaffiliated independent operators, in Seattle neither entry nor exit is primarily limited to any particular type so far.

Variable-rate setting has provided some price competition, although the major operators in all three cities have tended to charge similar (although by no means the highest) rates. Where there has been no maximum rate, as in Seattle, or where the maximum was set relatively high, as in San Diego, a wide variety of rates for exclusive-ride service can be observed. In Portland, on the other hand, the maximum rate has evidently been set too low to allow for price competition under it.

Some regulators argue that, during this era of skyrocketing gasoline and insurance costs, taxi rates are lower under variable pricing than they would have been under continued standardization. As Table 5 shows, rates have risen an average of 35 percent for the three cities during this first period of 18-24 months. In comparison, standard rate increases were more frequent and precipitous in all three cities during the six or seven years prior to variable pricing than they had been during the previous decades. Between 1976 and variable pricing, taxi rates rose 33 percent in San Diego, 15 percent in Seattle, and 45 percent in Portland. These changes average to an overall 31 percent increase for the two to four years immediately preceding variable pricing.

Problem Areas

Service Innovations

Although operators report that they are running much longer shifts under open entry than they previously did and there is more-aggressive marketing by some, operating practices have not changed dramatically. The code revisors' objective of achieving taxi-service innovations has yet to be realized. The San Diego paratransit office expended considerable effort to formulate a zone-based fare system and map along with informational brochures for use by local taxi operators. Although numerous San Diego operators filed zone-based shared-ride rates throughout 1979 and 1980, only one is reportedly close to offering non-fixed-route, nonsubsidized, shared-ride service even now. One of the Seattle service companies proposed its own zone-based fare system for shared-ride service. Implementing the service, however, required a new ordinance, since the 1979 law required all fares to be registered on a meter. DLCA spent several months discussing this operator's system as well as alternative proposals with local industry members. A codified approach to shared riding on an advance-reservation basis was adopted in late January 1981 for a nine-month trial period to go into effect during May. One Portland operator reportedly advertised shared-ride service but was unprepared to implement it. No other service innovations have been disclosed.

Airports

Variable pricing has run into problems at the San Diego and Seattle airports. Both airports have an underlying first-in, first-out principle of taxi-queue operation that militates against the incentive for competitive pricing. Moreover, airport taxi riders tend to be visitors who are generally unfamiliar with local taxi rates and trip distances. Although posted signs notify travelers that variable pricing is in effect (and, in Seattle, show average fares), it is still incumbent on the customer to choose an acceptable taxi.

At Sea-Tac Airport, this means rejecting the taxi that is sent up on request from the holding area in order to request another one. At San Diego International, it means waiting until the preferred cab reaches the front of the airport queue. Passengers have therefore been vulnerable to abuses, such as those perpetrated by operators who charge as much as 50 percent above the average fare, or to pressure to take the first taxi sent up. Independents have tended to concentrate on the airport business, and short-haul refusals have reportedly increased as taxi queues lengthen. Both of these kinds of problems raise enforcement needs and passenger complaints, whereas the dramatic increase in applications for airport taxi permits brought attendant administrative problems.

The port of San Diego imposed a six-month moratorium on airport taxi permits while it deliberated proposed solutions. The mayor established a task force of city council and port commission members to hasten the process and encourage dialogue between the two jurisdictions. Recontracting out the service as an exclusive franchise was considered and rejected. In July 1980, the port lifted its moratorium and raised the airport permit fee from $25 to $100. Late in September 1980, the port released its proposed solutions, which included the plus-or-minus 20 percent range on rates, a further increase in the permit fee to $200 annually, and adoption of the city's proposal for full-time starters to administer the taxi queues. These individuals are to be employees of the port, a proposition estimated to cost $125 000. Airport taxi-permit revenues were reportedly between $12 000 and $13 000 annually when the fee was $25 and are projected to be about $100 000 when the $200 fee is in effect. The port will also undertake to develop its own hearing and permit-revocation processes to back up its enforcement efforts. The new rules go into effect April 1, 1981.

Effective March 1, 1981, the port of Seattle also raised its airport permit fee from $100 annually to $90/quarter ($360/year) and implemented the plus 10 percent rate ceiling. In addition, all pick-ups (except as described below) are to be limited to the lower, deplaning drive. Passengers will no longer be able to request particular taxi companies from the airport dispatcher, although they retain the

Table 5. Changes in taxicab rates in three cities.

Item	Rate		Change (%)	Type of Charge
	Old	New		
City				
San Diego	\$1.10 drop + \$0.70/mile (standard)	\$1.19 drop + \$1.05/mile (weighted average)	+35	Weighted average for all operators
Seattle	\$0.90 drop + \$0.70/mile (standard)	\$1.00 drop + \$1.10/mile (weighted average)	+45	Weighted average for service companies and large fleets[a]
Portland	\$1.00 drop + \$0.90/mile (maximum)	\$1.00 drop + \$1.20/mile (maximum)	+26	All operators
Avg change			+35	

[a]Two of the service companies charge \$1.00 drop + \$1.00/mile; the third very recently changed its rate to \$1.00 drop + \$1.20/mile.

prerogative of rejecting the taxi sent up from the holding area. Passengers who wish to request a particular taxi will have to telephone the company and then move themselves and their baggage to the upper, enplaning drive to be picked up. It remains to be seen how these procedures will affect airport taxi operations. It should be noted, however, that both airports responded to current problems by adopting retrictions that reinforce the first-in, first-out approach rather than accommodate price competition.

Public Information

A major factor in realization of the benefits that the regulatory code revisors predicted from the new regulations is the public's awareness of the changes and how to take advantage of them. San Diego staff have attempted to assist operators with promotional materials and have described the regulatory changes to citizens' groups. Seattle's efforts have been limited to airport informational signs and a quarterly list of local taxi rates. But city resources were never budgeted to provide for a full-scale public information program, although both administrations recognize the need for such efforts. While open entry and variable pricing have been amply reported in the local press, these accounts tend chiefly to sensationalize the controversy between taxi operators and city administrators, polarizing them as adversaries.

Interjurisdictional Issues

Open entry altered the reciprocity between jurisdictions in the Seattle area. The city of Seattle and King County had had an agreement since the 1977 interim legislation by which a taxi licensed in one jurisdiction could obtain a license in the other for a reduced fee. This legislation was advocated by the regulatory revisors to ease entry to the lucrative airport market for city operators prior to proceeding with open entry and the other changes. When King County chose to retain its license ceiling and public convenience and necessity requirements for one year following the city's adoption of open entry, however, this reciprocity was interrupted. Open entry became effective in King County in June 1980, but reciprocal licensing was not officially restored, although the license fees were adjusted downward to the amount required to obtain both licenses under reciprocity. During the year-long interim and continuing through this writing, operators who wish to do business in both jurisdictions must still obtain each jurisdiction's license independently.

Administrative Issues

Although regulatory revision has released the city councils from some of the chores of taxi regulation, the early phases of it seem to demand an increased commitment in administrative and law-enforcement time. Prior to regulatory revision, city staff were principally engaged in preparing rate recommendations and other data analyses for the city council. Since the changes, staff have assumed the relatively unfamiliar and ongoing functions of qualifying applicants and issuing taxi permits, obtaining and recording rate filings, and undertaking numerous liaison activities to implement the new regulations. As the influx of applicants and operators continues, inspection and enforcement personnel have needed to devote additional time to assigning color schemes, inspecting vehicles, monitoring operations, and investigating compliance. Numerous small operators are more difficult to communicate with than a few large operators. Moreover, with the increasing transition to lease-type operations, some of the city staffs and veteran fleet operators have asserted that responsibility for driver behavior is being transferred from company management to the city. (Data are being collected to provide for estimation of the recovery costs of these administrative efforts in regulatory and license fees.)

There is also some question whether the changes have reduced the councils' involvement in taxi regulation. Relaxation of entry restrictions was extremely controversial and protracted in all cases. The Portland council was involved in a rate review, a court appeal, and a new set of regulatory changes within one year after its initial revisions. The San Diego council has considered proposals to lift the monthly permit limit, remove the maximum fare, do away with financial-reporting requirements, and simplify color schemes as well as to establish a member task force to help resolve airport problems. Seattle's 1979 regulatory revisions supplemented initial changes adopted in 1977; efforts continued throughout 1980 to codify more strict for-hire driver certification requirements, establish a zone-based fare system, and achieve exterior rate posting; these measures were finally adopted early in 1981.

Transferable Implications for Other Regulatory Entities

Achieving the regulatory changes has taken much time, both in terms of elapsed time and involvement of regulators, city staff, and local operators. Subsequent rounds of revision and promulgation of rules have been seen in all cases. Implementation has required that city staff assume additional or unfamiliar administrative tasks. The local taxi industries also spent considerable effort in analyzing or opposing the changes, providing supporting data, and testifying at public hearings.

Abundant resources may be required to provide public information for travelers of the potential benefits represented by taxi fare and service competition. Operator liaison efforts are also needed to smooth implementation and encourage development of service innovations.

Increasing numbers of airport taxi operators and variable pricing have exacerbated problems at airports. The fist-in, first-out taxi-queue principle wakens the incentive for price competition while hindering the patron's ability to respond to lower prices. Modifications to airport taxi regulations have resulted, which include rate ceilings, a moratorium on permits, increased permit fees, and restrictions on taxi movements, whereas some port officials threaten a return to the exclusive-franchise approach. More interjurisdictional cooperation prior to implementing the new regulations might have prevented some of these problems.

ACKNOWLEDGMENT

Grateful acknowledgment is due many people for their cooperation in DeLeuw, Cather's data-gathering and analysis efforts for the case studies reported in this paper. Elaine Balok, previous paratransit administrator; Carol Boland, current paratransit administrator; Victoria Whelan, project coordinator; and John Kay, project assistant--all of the city of San Diego--were generous with their time in providing essential information, as were San Diego city council members Larry Stirling, Leon Williams, and Fred Schnaubelt. The help of other city staff from the paratransit office and San Diego police department should also be acknowledged, as should that of the San Diego Unified Port District, the Harbor Police, the San Diego Association of Governments, and the sheriff's division of licenses.

Considerable assistance and essential information were provided by Regina Glenn, director of the Department of Licenses and Consumer Affairs; Jean Schiedler-Brown, license code revisor; Dave Setterholm, operations supervisor; and Ed Wood, project coordinator, city of Seattle; as well as Robert A. Marr, assistant superintendent of operations, port of Seattle. Seattle City Councilman Randy Revelle offered aid in assembling the historical background on taxi regulatory reform. Pam Holley, taxicab supervisor, city of Portland, has similarly offered much information and her help in describing local regulatory changes and the Portland taxi industry.

Members of the local taxi industries have also been generous in providing needed information. Although the individual operators and drivers with whom I have spoken are too numerous for specific mention, special thanks are due the following individuals, whose acknowledgment may serve to indicate my gratitude for the cooperation of the industry at large: Dan Fox and Larry Johnson, Coast Cab; Jim Barret, Graytop Cabs; Tex Hedrick, Red Cab; Bob Delacott, Jim Evans, and Larry Haver, Co-op Cabs; Mike Marguet, Independent Cab Owners Association; John Harper, San Diego Cab; Wes Walton, Farwest Cabs; Bob Ruddy, La Jolla Cab; and Bill Hilton, Yellow Cab of San Diego.

It should be noted that, although the research is being conducted under the auspices of TSC, neither TSC nor UMTA necessarily supports the conclusions presented in this paper, which are mine.

Publication of this paper sponsored by Committee on Paratransit.

Retrospective View of Dial-A-Ride Service in Rochester, New York

DEBRA A. NEWMAN, DAVID SHARFARZ, AND MARK ABKOWITZ

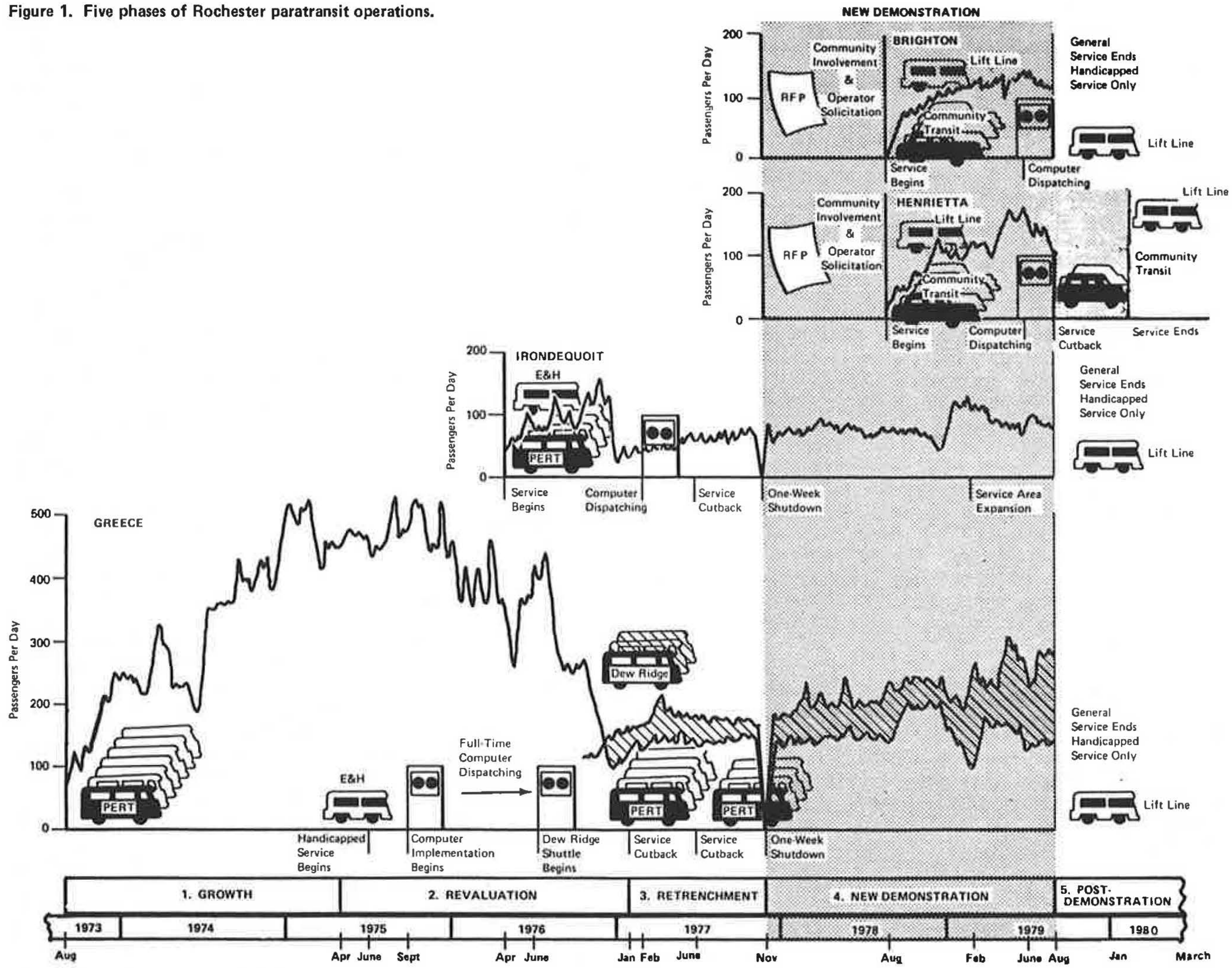
For one year, the Rochester-Genesee Regional Transportation Authority (RGRTA) offered dial-a-ride service to the general public in four suburbs under two different institutional arrangements. The public operator, Regional Transit Service, and a private operator, Paratransit Enterprises, each provided service in two communities. They also provided demand-responsive service to the elderly and the handicapped throughout Rochester. This unique arrangement was part of the Rochester community transit demonstration, an outgrowth of the earlier Rochester integrated transit demonstration, both projects funded by the Urban Mass Transportation Administration Service and Methods Demonstration program. The community transit demonstration was specifically designed to test cost-effective demand-responsive transit strategies. RGRTA sought competitive bids from paratransit operators and asked communities to fund a share of the operating deficits for postdemonstration services. Thus, the demonstration made it uniquely possible to compare service levels, ridership, and costs for public and private dial-a-ride that served both the general public and the elderly and the handicapped. By the end of the demonstration, three of the four communities found that they could not afford to continue paratransit services by using local subsidies. One town, however, developed an innovative funding strategy and supported dial-a-ride services for five additional months. By 1980, no general market dial-a-ride services were operating, although the cost-effectiveness of private operation was successfully demonstrated. Today, RGRTA supports privately operated paratransit services for the elderly and the handicapped throughout the county. The activities of the demonstrations are reviewed and implications are derived that may be useful to others considering implementing demand-responsive transit service.

The history of paratransit in Rochester can be divided into five phases: (a) growth (August 1973 to April 1975), (b) revaluation (April 1975 to January 1977), (c) retrenchment (January 1977 to November 1977), (d) new demonstration (November 1977 to August 1979), and (e) postdemonstration (August 1979 to May 1980). The timing of each of these phases and the key activities in the four Rochester suburbs most affected by the paratransit operations are shown in Figure 1.

GROWTH

The initial growth period lasted from service initiation in the suburb of Greece in August 1973 until early 1975. During this period, the Greece service area expanded several times, fixed-route bus services were eliminated, the demand-responsive vehicle fleet nearly doubled, operating hours were extended, and dial-a-ride ridership grew steadily. Work and school subscription services were offered in addition to the basic dial-a-ride service. All services were operated by the Regional Transit Service (RTS), the major operating subsidiary of the Rochester-Genesee Regional Transportation Authority (RGRTA), and were advertised under the acronym for personal transit (PERT).

Figure 1. Five phases of Rochester paratransit operations.



REVALUATION

By April 1975, when the original Rochester integrated transit demonstration began, the system was entering a period of transition and revaluation. Although expansion continued by the introduction of services for the elderly and the handicapped, the extension of service into the suburb of Irondequoit, and the implementation of computer control in Greece, several serious operating problems developed. Operating costs proved to be much higher than predicted, whereas demand was lower. In addition, passenger resentment grew over the replacement of fixed-route services with flexibly routed dial-a-ride services. The vehicle fleet, which consisted of several different models of small buses and vans, proved to be very unreliable, and service was further disrupted by the introduction of computerized dispatching. In addition, management disputes arose between RGRTA (the system developer) and RTS (the system operator).

RETRENCHMENT

As these problems developed, PERT's service reliability deteriorated and ridership dropped. At the same time, RGRTA was confronted with a transit funding crisis that not only threatened PERT expansion plans but also jeopardized future local fixed-route services. RGRTA responded by cutting PERT services drastically as part of a retrenchment process and by developing alternative strategies for reducing the financial burden of the dial-a-ride program.

NEW DEMONSTRATION

During retrenchment, RGRTA also applied for and received a new demonstration project to test innovative and more cost-effective funding options and to complete portions of the first demonstration delayed by service delivery and computer development problems. Publicly operated services continued under the new community transit demonstration, while competitive bidding introduced private, lower-cost dial-a-ride services to the Rochester suburbs of Brighton and Henrietta in July 1978. Lift Line services for the elderly and the handicapped were also expanded throughout Rochester's metropolitan region by using the same private operator. PERT's ridership increased slightly, dispatching functions were transferred to a new minicomputer, additional vans were leased, and Irondequoit expanded service townwide. During the first few months, community transit's ridership increased rapidly and then continued to rise slowly in Brighton while it fluctuated in Henrietta. During the final six weeks of the demonstration, all dial-a-ride services operated under computer control.

POSTDEMONSTRATION

In the summer of 1979, each town evaluated dial-a-ride operations. RGRTA continued subsidizing paratransit for the elderly and the handicapped, but only Henrietta decided to fund dial-a-ride services during the postdemonstration period. In August 1979, PERT closed its offices, RGRTA negotiated an extended contract with Paratransit Enterprises that increased the hourly Lift Line service rate from $13.20 to $16.60, and Paratransit Enterprises moved into the Henrietta Town Hall. RGRTA reissued requests for proposals, and in December 1979, a new private operator, Beacon Transportation, Ltd., was selected at $17.70/h of vehicle service. As a result of higher costs and limited local travel needs, Henrietta discontinued dial-a-ride services, but privately operated Lift Line services continued to operate throughout the Rochester metropolitan region.

Today, under a revised and more clearly defined contract, Lift Line operates more smoothly. By comparing the first quarter of RGRTA's 1980-1981 fiscal year to the same period in 1979-1980, Lift Line services show a marginal decline in productivity and a 58 percent increase in ridership offset by a comparable increase in service supply. However, complaints have decreased dramatically, and the accountability of the contract vendor is improved from the perspective of RGRTA. The 11-vehicle system, of which an average of eight are operated daily except Sunday, transports approximately 5500-6000 persons/month at a productivity of just more than 2.5 passengers/vehicle-h.

LESSONS LEARNED FROM SECOND DEMONSTRATION

Because of the unique nature of the second demonstration, many of the experiences can be studied to derive implications that may be useful to other communities considering implementing demand-responsive transit services. These lessons are discussed in the following sections.

Local Involvment and Funding

In the first demonstration, transportation analysts made practically all service decisions and selected and implemented dial-a-ride services in Greece and Irondequoit without stipulating any local commitment. In the second demonstration, Monroe County could not continue to provide dial-a-ride support. RGRTA offered eight Rochester suburbs (not including Greece and Irondequoit) the opportunity of establishing a dial-a-ride service provided that 50 percent of the deficit be assumed locally at the end of a one-year demonstration period if service was to be continued. Of the eight suburbs, only two (Brighton and Henrietta) agreed to assume the service and the attendant funding responsibility. The other six suburbs were reluctant to participate because little local support was expressed, dial-a-ride ridership projections were low, and the local postdemonstration funding requirements were considered too great a burden. Officials in these communities feared that after the demonstration concluded they would have to eliminate dial-a-ride services, cut back other community services, or raise local taxes, none of which were politically attractive alternatives. Officials from Brighton and Henrietta subsequently designed their own dial-a-ride services.

Although the town involvement process was effective, these efforts were less than totally successful in generating the necessary operating funds. Given the poor revenue recovery rates experienced during the demonstration, both Irondequoit and Brighton considered the dial-a-ride service too expensive to justify local funding and voted against subsidizing the service when the demonstration ended. In both these suburbs, the dial-a-ride service carried far fewer residents than the extensive fixed-route service operated by RTS. Officials from Greece contended that they were opposed to the concept of local funding from the start; they felt that dial-a-ride service was too expensive, poorly operated, and served too few residents. Consequently, the prospect of continuing dial-a-ride service under a local subsidy was never brought to a formal vote in this suburb.

Only Henrietta--a growing community that had an increasing tax base; a growing number of elderly, student, and transit-dependent residents; and relatively little fixed-route service--voted to subsidize continuing dial-a-ride operations from August until December 1979. To ease the local financial burden, service levels were scaled back to two vehicles, and Henrietta leased Town Hall space to the paratransit operator. At the end of the year, when Henrietta faced another funding decision, significant increases in costs, loss of operator rental revenue, and decreases in demand, local officials decided to discontinue subsidizing dial-a-ride operations.

The implication of this experience is that asking local towns to participate in short-term transit programs is probably not a feasible solution to financing transit programs. Even though demand-responsive transit may be more efficient than fixed-route service in areas that have low population density and diffuse travel patterns, demand-responsive services will probably have much lower vehicle-productivity levels than those of the overall fixed-route system in any major urban area. Consequently, they are likely to lose in a local political battle for scarce transit resources. It is often easier and more dramatic to eliminate a costly demand-responsive transit program than to isolate the least-efficient components of a fixed-route operation.

Most suburban governments rely on local property taxes for support, which provides a fairly limited tax base. In addition, these communities are likely to be more economically conservative; they favor limiting public services in order to maintain low taxes. As in Rochester, town officials may also feel that transportation should be handled at the county or regional level. Other local suburban officials might thus similarly sacrifice a desirable dial-a-ride service today to avoid making a politically undesirable funding decision in the future.

Labor and Operator Selection

Under the first demonstration contract, all operating and mechanical work was restricted to Amalgamated Transit Union (ATU) members. When the plan for a new demonstration based on competitive bidding was disclosed, the ATU local recommended against signing a 13(c) certification for the new project because they feared an erosion of union jobs. The new demonstration appeared doomed until an eleventh-hour agreement was reached with the International ATU. The agreement preserved current union contracts and prohibited the dial-a-ride services from competing with existing fixed-route services. Apparently, this reversal stemmed from national union concerns that long-term interests might suffer if it was accused of forcing transit services to collapse.

A request for proposal (RFP) to provide flexible, affordable community transit services in Brighton and Henrietta and expanded services for the elderly and the handicapped was publicly advertised in local newspapers and sent to 17 potential bidders. A bidders' conference was held to answer any questions, and RGRTA staff was available for additional information. Despite these efforts, only three firms responded, and only two of these offered competitive cost estimates.

RTS did not submit a bid because it felt that its high union driver wage rates made it impossible for them to compete with private operators. Most of the other locally solicited operators were small, privately owned and operated taxicab companies. In general, they felt that the bid specifications were unclear, they were inexperienced in dealing with government agencies and public funding procedures, and they lacked the expertise needed to prepare a formal bid and the first-instance money required to set up the services. Most of the out-of-town operators solicited did not submit bids because they felt that Rochester, New York, was not within practical geographical distance from their existing management operations and that it would not be profitable to develop, hire, and establish on-site management and services for a one-year contract.

After the demonstration, RGRTA again publicly advertised and distributed RFPs for continued Lift Line and possibly dial-a-ride services. Although four operators expressed an interest in providing service, only one local operator attended the RGRTA bidders' conference and only two bids were received--one from Paratransit Enterprises, the demonstration contractor, and one from Beacon Transportation, Ltd., a local private ambulance company. RGRTA contracted with Beacon Transportation to provide Lift Line service in 1980 and 1981.

Other communities who wish to enter into contracts with private operators for paratransit services may face a similar limited choice of operators. Obviously, this will depend on the amount and quality of local talent available in or near the particular service area; larger cities typically have a wider variety. Small and medium-sized cities may simply not have any providers or only marginal providers that are not well respected by potential users.

As the Rochester experience has shown, there are currently very few paratransit firms that are capable of competing on a national scale. An increased number of qualified private operators may be interested in providing paratransit services outside their immediate locale now that Rochester has shown that it can be economically profitable for them to do so. However, Rochester's postdemonstration service award to a new local contractor may deter other paratransit operators interested in providing longer-term services.

Local operators should have a competitive advantage over out-of-town providers because they are already locally established and thus need not incur new overhead costs and because they are more familiar with prevailing wage rates and supporting services, such as insurance agencies and vehicle maintenance services. Local providers would also be more familiar with the local physical environment and with potential users of the service, so they would be preferred, particularly for more-personalized services.

RGRTA's use of competitive bidding to select a new paratransit operator also succeeded in lowering local operating costs. RGRTA paid the private contractor a very significant 45 percent less than it paid for local public paratransit operations. Paratransit Enterprises' lower operating costs were primarily due to lower driver wages and maintenance expenditures. As demonstrated by Paratransit Enterprises' willingness to bid for the continuation of operations in 1980 at comparable rates (considering inflation, the reduced 1980 operation, and the aging of the vehicles), it appears that the firm also profited from the experience.

As the Rochester project demonstrated, an alternative to creating a new transit operation or to assigning demand-responsive services to an existing transit operator would be to competitively solicit and contract with private taxi or other operators for the provision of service. This can significantly lower local paratransit operating costs. Another alternative for decreasing costs would be to pay demand-responsive employees lower wage rates than conventional transit employees within the same transit operation. Several other transit providers, including Cleveland, Kansas City, and Bridgeport, have successfully established lower wage classifications for their paratransit service employees. These alternatives, however, may be opposed by existing local transit workers' unions, who, fearing an erosion of their positions and status, may try to prevent either alternative from being implemented.

Safety and Productivity Incentives

To encourage safe, high-quality transit services, a safety incentive of 50 cents/h was added to the wage rates of all Paratransit Enterprises drivers who avoided accidents for four weeks. An analysis of driver rates indicates about 70 percent of all drivers received the additional payment; of the remaining 30 percent, the majority were new drivers. Although a number of external factors prevent any conclusive statements from being drawn, an analysis of the total number of collision accidents recorded by Paratransit Enterprises and PERT drivers indicates that no significant difference occurred. However, both paratransit operators had substantially fewer accidents than did Rochester's fixed-route services.

In addition to the safety incentives, the size of the vehicle, the number of service hours, the individual driver's training and experience, and the way in which the safety incentive is presented and implemented may all affect accident severity and frequency. Instead of viewing it as a safety incentive, drivers may feel penalized if they are docked 50 cents because of an accident, which would contribute to a lower accident-reporting rate. Others interested in developing and implementing incentive or disincentive strategies are advised to try to foresee and control for possible abuses and adverse impacts that might result.

A productivity incentive between 15 cents and 50 cents per dial-a-ride passenger was also devised to encourage the private contractor to manage resources wisely. This additional payment schedule only took effect when average daily productivities rose above 4 passengers/vehicle-h of service. Average dial-a-ride productivity was 3.5 in Brighton and 3.6 in Henrietta, which resulted in relatively few productivity payments.

Although PERT was not offered productivity incentives, dial-a-ride productivities averaged 3.4 passengers/vehicle-h in Irondequoit and a significantly higher figure of 4.3 passengers/vehicle-h in Greece. Thus, there appear to be other, more important factors that affect dial-a-ride productivities than incentive payments. Some of these factors include trip patterns, demand for individual trips, group and subscription service, service-area size, vehicle size, operating speed, service hours, and quality of service. Although no safety or productivity incentives were included in the postdemonstration private operators' service contract, both of these concepts may warrant further examination. Other communities should, however, be aware of the possible abuses and other determinant factors that affect operations.

Contracts and Contractors

From the initiation to the conclusion of dial-a-ride services in Brighton and Henrietta, a number of controversies developed between RGRTA and Paratransit Enterprises. Topics of contention included the accuracy of reported vehicle hours; ridership and service-quality data; driver uniforms, courtesy, and training procedures; vehicle maintenance; and availability of the on-site manager during operating hours. The record-keeping procedures of Paratransit Enterprises were the major bone of contention in these controversies. In trying to ensure that high-quality service was offered and because of the need to provide demonstration evaluation data, RGRTA insisted that contractual details regarding record-keeping and accounting procedures be rigorously observed. Paratransit Enterprises was often frustrated by the reporting requirements and level of detail imposed by the public authority. As a private organization, it would have preferred a more independent relationship in which it was responsible for supplying a specified level of service in return for a flat fee.

Having learned from this experience, RGRTA now requires more contractor accountability and specifies these requirements in the postdemonstration operator's contract. Changes included weekly random RGRTA inspections, financial penalties for noncompliance with the contract, withholding of payments until reports are completed, detailed maintenance schedule and responsibilities, elimination of safety and productivity incentives, and input into selection of the resident manager. The differences between these two types of contracts are highlighted in Table 1. RGRTA also selected a different locally based operator to continue providing Lift Line services. Because of these changes, RGRTA new feels that higher-quality Lift Line services are being provided.

From this experience, it is clear that if outside groups are required to assist in operations or management, they should have on-site decision capability and the authority to execute their responsibility. Most public transit agencies have developed their own standards of service, reporting procedures, and levels of accountability as a public operator. If these standards are to be met by outside contractors, all responsibilities and requirements should be detailed in the service contract and fully understood by both parties at the outset. Contract penalties or rewards may be included to encourage adherence, although considerable negotiation and compromise may also be necessary to make private contractors adhere to the standards of the public operator. Readers should also understand that detailed Rochester operator information was necessary partly because of the collection of demonstration-evaluation data. Such detailed records may not be needed at other sites.

Operating Effectiveness

Average dial-a-ride vehicle productivity in Irondequoit, Brighton, and Henrietta ranged between 3.34 and 3.60 passengers/vehicle-h. Although these levels are less than the productivity goal of 4-5 passengers/vehicle-h set by RGRTA and markedly lower than those of other dial-a-ride systems in the United States (the productivity levels of which commonly range between 4 and 8 passengers/vehicle-h), trip-demand densities were also relatively low in these three suburbs. Vehicle productivity in Greece was somewhat higher; it averaged 4.25 passengers/vehicle-h, largely due to the greater demand density that occurred there.

The cost of service provided by the private firm, Paratransit Enterprises, was significantly lower than the cost of comparable service provided by the public operator, PERT. The average operating cost per vehicle hour for Paratransit Enterprises, measured by payments made by the RGRTA, was $13.35. This ranged from $13.06 for Lift Line service to $13.53 for dial-a-ride service. Collectively, it was 45 percent lower than the average PERT operating cost of $24.47/vehicle-h during the same period. This striking difference in cost can be traced to the lower driver wages and maintenance costs paid by Paratransit Enterprises. Driver wages and benefits for the private firm were estimated to be less than half the rate of $12.62/vehicle-h paid to unionized PERT drivers. However, it should also be noted that maintenance costs were significantly lower because the vehicles were newer, many repairs were covered under warrantly, and minor maintenance was often deferred. Between August and December 1979, the Lift Line hourly service rate increased to $16.60 and in January 1980 it increased to $17.70.

Although operating costs were considerably lower in the areas served by Paratransit Enterprises, all dial-a-ride and Lift Line services required substantial per-passenger subsidies. This was because vehicle productivities were also lower than anticipated, so that the target revenue recovery rates of 25-29 percent established by Brighton and Henrietta

Table 1. Contract differences.

	Contract Period	
Category	Demonstration (1978-1979)	Postdemonstration (1980-1981)
Incentive	Driver safety and contractor productivity	None
Inspection	No provision	Weekly random by RGRTA
Noncompliance with contract	Terminate contract	Financial penalties
Maintenance	General requirements	Detailed schedule, responsibilities, and requirements
Resident manager	Contractor solely responsible for selection	RGRTA can approve or veto contractor's selection
Special fuel purchase	No provision	Tax-saving purchases through RGRTA
Vehicle assignment	No flexibility on number in service	Flexible peak/off-peak assignment
Payment	No provision for withholding payments	RGRTA may withhold payments until completed reports submitted
Insurance	$3 million liability	$2 million liability
Contract length	One year	Two years; mutually renegotiable clause after first year

were not met during the life of the demonstration. In Brighton, revenues averaged 16 percent of costs; the corresponding figure for Henrietta was 18 percent. Because of the higher cost of PERT operations, recovery rates were still lower in Greece and Irondequoit; they averaged 15 and 9 percent, respectively.

The Rochester experience has shown that higher productivities can be completely overshadowed by lower operating costs. Private operators typically pay lower wages and maintain stricter control over finances, which results in lower operating costs than those of most public transit providers. Overall operating efficiency is attained when the maximum transit output is secured for the least amount of resources expended. At the same time, however, passenger revenues must also be considered. Other jurisdictions may therefore find it advantageous to invite competitive bidding for transit services while also reevaluating and possibly modifying their fare structure.

Vehicle productivities for Lift Line, the paratransit service for the elderly and the handicapped, averaged 3.20 passengers/vehicle-h for the PERT-operated service in the northern quadrant and 2.58 for the service operated by Paratransit Enterprises in the other three quadrants. Productivity of Paratransit Enterprises varied by quandrant; it ranged from 2.25 to 2.91 passengers/vehicle-h, in inverse relation to service-area size. This experience is consistent with that reported by other target market dial-a-ride systems operated in the United States. Since the four quadrants served by Lift Line totaled 341 miles^2, which is a significantly greater area than that of most other U.S. systems reporting performance data, there is reason to believe that the policy of operating Lift Line on an informal schedule to help consolidate demand was successful in boosting productivity.

Thus it appears that demand-responsive services that operate in large service areas cannot be expected to achieve high levels of vehicle productivity. Nevertheless, demand-responsive services, especially those that offer wheelchair-accessible door-to-door service, can provide significant mobility improvements for such transit-dependent groups as the elderly and the handicapped. Wherever possible, advance reservations should be required and trips should be aggregated to serve the demand more efficiently. Since users are typically quite appreciative, services should be offered by existing transit organizations, contracted operators, or subsidized taxis, depending on available local services and needs.

Computer Dispatching

The introduction of computerized dispatching and scheduling in Greece in 1975 was a lengthy and frustrating experience. For more than a year, system operations were seriously disrupted by hardware breakdowns and inaccurate scheduling caused by software errors. However, these problems were solved over time, and the subsequent conversion to computerization in Irondequoit (February 1977) and in Brighton and Henrietta (June 1979) was accomplished smoothly and with minimum disruption.

Conversion from a time-sharing system to RGRTA's minicomputer in January 1979 was also accomplished without major problems, and the minicomputer operated much more reliably during its seven months of use. Less than one hardware breakdown occurred each week compared with an average of 17 breakdowns per month under the time-sharing operation. Despite the relative ease of this implementation, both operators and staff were hesitant about accepting and relying on the computer.

During the first demonstration, computer dispatching proved capable of generating high levels of service under low-demand conditions. But when the computer was being tested in Greece and Irondequoit, vehicle reliability improved significantly, which lowered vehicle productivity and raised service levels. Thus, much of this service-level improvement could not be related to computer dispatching.

During the final 1.5 months, computer dispatching was introduced in Brighton and Henrietta. Compared with that of manual dispatching, computerized system response time decreased by about 15 percent. At the same time, deviation in mean pick-up time increased considerably, although the variation decreased. In Henrietta, average ride time lengthened appreciably, and in Brighton no significant change in ride time was detected. In all cases, the service quality in Greece and Irondequoit, which had been operating under computer controls since 1977, was superior to the recently implemented computer services in Brighton and Henrietta. Although this suggests that computerized service quality may improve over time, these findings present a mixed overall picture of the effects of computerized versus manual dispatching on paratransit service quality.

Capital costs were more than $300 000; about $240 000 was spent on converting from time sharing to the minicomputer and on implementing the computer system in Brighton and in Henrietta. Operating costs totaled approximately $10 000. These costs translate into an additional $1.27 per dial-a-ride passenger, which substantially exceeds any reasonable valuation of the improvement in service quality. In much larger dial-a-ride systems, computerization costs per passenger might be lower, but it is also possible that many of the high-demand conditions that would warrant a large dial-a-ride system might be more efficiently served by a fixed-route bus system.

Although the ability to schedule and dispatch dial-a-ride services automatically has been demonstrated, the benefit of computerization over manual operations is questionable for a system the size of Rochester's. This is because only a minor improvement in service quality was achieved, and the Rochester dial-a-ride operation was too small for any labor reductions to be realized.

It appears that dispatching can be effectively handled by a single highly skilled dispatcher when a system has fewer than eight vehicles, when demands are highly concentrated, or when the number of trips per vehicle hour is low. But as fleet size and demand increase and demand patterns become more diverse, computers may prove superior to human dispatchers, since they can continue to quickly process and retrieve the additional information. The inevitable phase-in problems and traditional operator reluctance to implement computerization probably outweigh its potential benefits for smaller systems.

However, computerized dispatching can also support a more sophisticated management information system than a manually operated system is able to provide. Demand-and-supply data are tabulated automatically, which simplifies the record-keeping process. In addition, data that are expensive to collect manually, such as service-quality measurements and trip-tour and origin-destination information, are continuously available. Better estimates of pick-up times and rescheduling of trips around vehicle breakdowns are then possible. In addition, quick access to such information enables the order processors to interact with customers more effectively. Although these advantages would be realized in a system of any size, they become more significant in larger systems. Although Rochester's relatively small system was not able to test this concept effectively, a dedicated in-house computer was shown to improve system reliability compared with that of a time-sharing operation. It is also suspected that larger vehicle fleets and higher-demand systems could achieve even greater coordination and level-of-service improvements.

Although a different type of computer was tested in the second demonstration, significantly few implementation problems occurred. This suggests that previous experiences with computer hardware and software may be transferable to new applications and increases the importance of findings for others interested in testing different computer-dispatching systems. Another computerized dispatching system is now being tested in Orange County, California, in which a larger paratransit vehicle fleet and higher demand are present. This demonstration should provide additional understanding, coordination, and level-of-service and cost information.

FUTURE OF DIAL-A-RIDE SERVICES

Numerous lessons and experiences have been learned in Rochester from the different public and private paratransit operators and from the various operating strategies tested. Many experiences have been positive, whereas others have not been so successful. Hence, there are mixed attitudes toward dial-a-ride service as a transit mode, and the future of demand-responsive transit has not been clearly specified.

In general, there is probably less support for dial-a-ride service as a transit mode than for the use of private contracts for the provision of service. There is a strong feeling in Rochester that, in order to become more effective, mass transit must capture a larger portion of the modal split and increase the revenue/cost ratio. Although success by these standards may not be attainable without severe changes in economic conditions and cultural values, it has been made clear by the Rochester demonstration that paratransit operations are least capable of being effective by using these standards.

Today, no general market dial-a-ride service operates in Rochester, although it continues to be used effectively in the city of Batavia (population, 20 000) and in rural operations in neighboring Livingston and Wayne Counties. In the near future, general market dial-a-ride service is not likely to be restored to Rochester's suburban towns either, despite fixed-route transit's inability to provide intratown mobility. On the other hand, the use of dial-a-ride service to provide special user services in Rochester has continued, and there is widespread support for it.

Lift Line's demand-responsive service for the elderly and the handicapped has continued under a new private contract and has expanded throughout Rochester's Monroe County. It is now embraced as both efficient and effective when compared with the alternative of the mandated fixed-route accessibility required by Section 504 regulations. RGRTA's application for a waiver to these regulations on the grounds that Lift Line is more affordable, offers fewer operating problems, and provides greater mobility for senior citizens and disabled persons was denied in 1980.

A number of meaningful lessons were learned on the basis of activities during the second Rochester demonstration. They include lessons on funding, operator selection, contracts and contractors, operating effectiveness, and computer dispatching.

Funding

Asking local towns to participate in funding transit programs with which they have not been involved may not be a feasible solution to public transit financial problems. Since most towns rely on a limited tax base from property taxes, town officials are often fiscally conservative. This suggests that transportation issues may need to be handled at the county or regional level.

Operator Selection

Local operators should have a competitive advantage over out-of-town providers because they are already established on location and are familiar with local operating conditions. The use of competitive bidding should also help in keeping operating costs down. Consideration should be given to having the local public transit agency provide service if competitive rates can be established. If the public operator is not used, care should be taken to assure that public transit rights are not violated.

Contracts and Contractors

If outside groups are responsible for managing operations, they should have on-site decision capability. All responsibilities and requirements should be detailed in the service contract and fully understood by both parties at the outset.

Operating Effectiveness

Private operators may be able to offer service at significantly lower costs than those for comparable service provided by the public operator. Other jurisdictions may therefore find it advantageous to invite competitive bidding for transit services while also reevaluating and possibly modifying their fare structure. Demand-responsive services that offer wheelchair-accessible door-to-door service can provide significant mobility improvements for this transit-dependent group. Whenever possible, advance reservations should be required and trips should be aggregated to serve the demand more efficiently.

Computer Dispatching

As fleet size and demand increase and demand patterns become more diverse, the benefits of computer dispatching are more apparent. Although the ability to automatically schedule and dispatch dial-a-ride service was demonstrated, the benefit of computerization over manual operations is questionable for a system the size of Rochester's. However, a dedicated in-house computer was shown to improve system reliability compared with that of a time-sharing operation.

The lessons learned in Rochester, both from the successes and the failures, have, it is hoped, helped clarify the role of demand-responsive transit services in urban transportation. These findings and those from ongoing demand-responsive transit demonstrations should be used in determining the direction of future paratransit policies and programs. More-detailed information on these demonstrations is contained in a set of evaluation reports submitted to the Transportation Systems Center (1-5).

ACKNOWLEDGMENT

The research for this paper was conducted by SYSTAN, Inc., under contract to the Transportation Systems Center of the U.S. Department of Transportation. Through this contract, SYSTAN, Inc., has evaluated various Urban Mass Transportation Administration (UMTA) Service and Methods Demonstrations, including

the two dial-a-ride demonstrations in Rochester, New York, discussed in this paper. RGRTA received UMTA Section 6 project funds to test these innovative strategies and to collect the necessary data to conduct this evaluation.

REFERENCES

1. R.E. Lave and M. Holoszyc. The Rochester, New York, Integrated Transit Demonstration--Volume 1: Executive Summary. SYSTAN, Inc., Los Altos, CA; UMTA, U.S. Department of Transportation, Rept. UMTA-NY-06-0048-78-1, March 1979.
2. R.E. Lave and M. Holoszyc. The Rochester, New York, Integrated Transit Demonstration--Volume 2: Evaluation Report. SYSTAN, Inc., Los Altos, CA; UMTA, U.S. Department of Transportation, Rept. UMTA-NY-06-0048-78-2, March 1979.
3. R.E. Lave and M. Holoszyc. The Rochester, New York, Integrated Transit Demonstration--Volume 3: Appendices. SYSTAN, Inc., Los Altos, CA; UMTA, U.S. Department of Transportation, Rept. UMTA-NY-06-0048-78-3, March 1979.
4. M. Holoszyc and D.A. Newman. Implementing the Rochester Community Transit Service Demonstration. SYSTAN, Inc., Los Altos, CA; UMTA, U.S. Department of Transportation, Rept. UMTA-NY-06-0048-79-1, May 1979.
5. D.A. Newman and M. Holoszyc. Evaluation of the Rochester Community Transit Demonstration. SYSTAN, Inc., Los Altos, CA; UMTA, U.S. Department of Transportation, Jan. 1980.

Publication of this paper sponsored by Committee on Paratransit.

Barriers to Coordination: Irrational or Valid Objections?

SANDRA ROSENBLOOM

Part of a larger study that focused on coordination of transportation resources in programs designed for the elderly is presented. The study attempted to determine the conditions under which local agencies and providers resisted coordination attempts and to evaluate the validity of their objections rather than simply styling them as "barriers." The study sought to identify the situations in which coordination models might offer more benefits than costs to participants and the conditions under which local agencies correctly and incorrectly assessed these outcomes. Legitimate objections to transportation coordination that actually arose in the 30 sites visited or surveyed and in recent research are identified and ways in which coordination proponents can objectively appraise those objections and, when appropriate, overcome them are suggested. When analysts and planners are certain that coordination in any community is the most sensible and efficient long-run approach to transportation delivery, they must be willing to provide time, money, and professional resources to convince local participants of this outcome and to help agencies cover costs. In addition, planners and analysts must recognize and address the very legitimate concerns that human-service agencies have about the quality of transportation services they wish delivered to their clients.

There is growing public concern over the unnecessary duplication of local transportation services and the fragmented nature of many types of human-service transportation projects. Two key U.S. Department of Transportation programs--Section 18 (rural operating assistance) and Section 16(b)2 (needs of the elderly and the handicapped) of the Urban Mass Transportation Act of 1964, as amended--have mandated a coordinated and cooperative approach to transportation delivery in programs that use their funds. The 1978 amendments to the Older Americans Act of 1965 reflect the concern of Congress with the efficient use of existing community resources in providing transportation services to the elderly; the act mandates a coordinated approach to transportation delivery (Federal Register, Vol. 45, No. 63, March 31, 1980).

The consolidation or coordination of transportation services at the local level is increasingly being seen as a way to reduce unnecessary duplication and to obtain economies of scale. Coordination can use existing resources more effectively and can capture the potential offered by underused vehicle and staff capacities. Analysts have identified several theoretical models of such coordination; the literature reports the experiences of some of the more successful or notable experiences in coordinated transportation services (1-7).

Most discussions of transportation coordination assume, first, that there is a great deal of service duplication and abundant potential for greater vehicle use at the local level (5). Second, they assume that service coordination is a desirable and meritorious idea in almost every context. Because of these prevailing beliefs, many analysts and observers have styled all objections to or reservations about coordinated service delivery as "barriers." They often imply that such objections are never rational or realistic or are always extremely protective of traditional modes.

This paper reports on part of a large study of the coordination of transportation resources in programs designed for the elderly or funded by the Administration on Aging (AOA). Thirty selected planning service areas (PSAs) (a geographic unit defined by AOA) were visited or telephoned to evaluate the operational experiences of local transportation providers and their responses to proposed coordination projects.

This study attempted to determine the conditions under which local agencies and providers resisted coordination attempts and to evaluate the validity of their objections. This study also sought to identify the situations in which different coordination models offered more benefits than costs to participants and the conditions under which local agencies correctly and incorrectly assessed these outcomes (7).

Although the literature is full of complex models and potential coordination arrangements (3-5,8) it is only necessary to identify four broad classes of coordination models here. Each model may include variants thought to operate and behave in a similar manner. [The AOA study itself developed a more comprehensive typology, which is too detailed for the needs of this paper (7).] The first model is nonservice coordination, which includes a mutual or cooperative agreement for any activity other than direct provision of transportation service (for example, joint purchasing of vehicles, joint dispatching services, and joint maintenance programs).

The second model is user-side coordination, that is, any arrangement that permits the client or user to pick an existing community transportation carrier. The third major model class is joint service coordination, which is any mutual or cooperative agreement between providers, public or private, and agencies for the coordinated delivery of services. The fourth class is purchase-service coordination, which is any agreement between agencies and providers, public or private, for the purchase of service or more capacity in vehicles or other resources.

All four major models described above can be combined or developed separately. What is often styled a "brokerage" can fit any of these four models or combinations of them; the term "broker" is and has been used to describe a number of different models of coordination. In general, and certainly for our purposes, there is no need to differentiate a brokerage as a separate model of coordination (5).

COORDINATION AND COST SAVINGS

Coordination brings benefits by reducing the redundant use of resources (such as duplication of the same or similar vehicle trips) or by increasing the efficiency or productivity of service delivery (3,5). However, it is important to be specific about what local agencies would actually want from participation in coordination and how coordination models work to meet these objectives. Moreover, it is equally necessary to identify the costs that are incurred as these mechanisms operate.

All four coordination models can meet the needs of local participating agencies if they provide one or more of the following benefits:

1. Reduce the resources (time or money) devoted to any cost component of service delivery (e.g., savings through joint purchase of vehicles or reduction in administrative costs by contracting for service),
2. Reduce total resources (time or money) devoted to transportation,
3. Reduce unit transportation costs,
4. Increase the amount of service delivered to existing clients,
5. Increase the number of clients provided equivalent service, or
6. Increase the quality or level of service (however defined) to existing clients.

Few local agencies are interested in overall efficiency or effectiveness if it does not translate into one of these objectives for them. In many ways, this is a checklist; local agencies would have to see one or more of these benefits to consider participating.

Many objections to coordination arise from a realistic appraisal of the costs and risks of coordination. If an agency decides that risks or costs are too high or the expected benefits are too small, these costs become barriers. Many so-called barriers to coordinated service delivery are a complex set of interactive responses and thoughtful concerns about the often-significant changes that are expected of a local agency participating in a coordinated transportation system.

Proponents of coordination often proceed from the assumption that because coordination can save money or increase efficiency, local agencies should be willing and eager to participate. Refusals to participate are not seen as rational responses to local coordination attempts.

This study and some very recent results from major demonstrations sponsored by the Office of Human Development Services (HDS) suggest that cost savings are far from a simple issue (1,7,8). First, there is strong evidence that, in the short run at least (one to three years), many coordination attempts have not saved money or appreciably increased the quality of service delivered to clients. Second, when there are cost savings, they may not accrue to the agencies that are being asked to coordinate but rather to larger governmental units or funding sources. This undoubtedly lessens the willingness of local agencies to participate in a program, since it saves them nothing.

Third, even when there will be demonstrably lower costs of service, there may be initial start-up costs, such as radio purchases and driver training. Local agencies may simply be unable (or perhaps just unwilling) to incur those costs in the expectation of future savings.

Fourth, agencies that have usable vehicles are in a different position than those without such vehicles. There is a great deal of pressure on the former group to calculate the true costs of the provision of service by using the full costs of donated vehicles, etc. This is undoubtedly a valid policy concern, but it does not change the fact that, realistically, agencies with free resources will not give them up unless forced to do so. Moreover, the more crucial point (often missed by proponents of coordination) is that such vehicle fleets and donated resources generally cannot be converted into cash, even if they can be given a dollar value. Whether they are forced to recognize the costs of those resources or not, many agencies cannot use them in any other mode of operation.

Only in a fifth category, one in which the agency will immediately save money or markedly increase service quality, can local objections be styled as irrational or "turf-protecting". Even in this case, it is not realistic to expect agencies and their staff to immediately abandon the ways in which they have traditionally delivered transportation services.

The ways in which each of the four cost-saving situations discussed above can create barriers to the implementation of local coordination efforts will be examined, and methods by which proponents of coordination can act to overcome barriers in each of the four cases will be identified. Last, the circumstances in which such solutions are appropriate will be discussed.

Are There Any Cost Savings?

In early 1980, a study was published of the results of the first two years of five major transportation-coordination demonstrations conducted for HDS (1). That report found that only one of the five sites was able to reduce costs after coordination and that two sites incurred increased costs after coordination. Only one system substantially increased the efficiency of service delivery and two increased the quality of service delivered. Even those sites that decreased some costs more than offset those cost reductions by significant cost increases in other areas.

The authors of the study stated early in the report (1, pp. 3, 4):

> Coordination does not necessarily lead to more efficient or effective transportation operations. In general, coordination and the number of riders served increased but costs per unit of service also increased, even after adjusting for inflation. Total cost savings were almost nonexistent.

They also stated that the best selling point for coordination was that it saved money (1, p. 10). In fact, this was generally not the case for these demonstration projects. They concluded that it is only under very special circumstances that coordination costs less.

The AOA study reported in part here concluded that the kind of redundant services for which coordination would obviously reduce duplicative costs rarely existed (7). In the 30 areas visited, the study team found very few examples of actual redundancy in direct service delivery. Very few providers were operating their vehicles along the same or similar routes for the same type of clients at approximately the same time of day.

What the study did find was the opportunity for increased productivity and use of resources. For example, a local nutrition project for the elderly might use its vehicles for only a few hours in the middle of the day to transport meals, whereas the local cerebral palsy agency was using similar vehicles only in the morning and evening peak hours to carry handicapped people to sheltered workshops. The AOA study also found some opportunity for increased efficiency in joint nonservice activities; the most hopeful is joint purchase of insurance (as in Oregon). However, these opportunities for more-productive use of existing resources might not lead to any discernible cost savings for potential participants.

Nothing in the discussion above proves that coordination cannot lead to cost savings or meaningful service improvements. It has been noted that if previous coordination efforts had been made correctly, they might very well have achieved measurable benefits and savings. Yet, given the history of several well-publicized coordination efforts, local agencies that express concern or reluctance to coordinate service are not necessarily behaving irrationally.

Existing coordination documentation suggests that, simply because a system currently operates inefficiently or underuses its existing capacities, one cannot jump to the conclusion that coordination will increase its efficiency or productivity. Moreover, existing evidence suggests that, even if there were increased productivity and efficiency, they might not lead immediately to cost savings for participating local agencies. It is questionable whether local agencies would be interested in such efficiencies if cost savings or service improvements were not immediately forthcoming.

Who Obtains the Cost Savings That Do Exist?

It is important to note that the potential benefits from any coordination model will differ as the agency perspective differs (4). State and federal funding sources should and do have different perspectives on what constitutes efficiency and cost savings; these views are rarely shared by local agencies (4). The state may wish to minimize the number of vehicles awarded in a certain area and may view a coordination effort as an ideal way to achieve that objective. In many cases there is no benefit to any given local recipient in foregoing a vehicle and being forced to coordinate in order to save the state money (4,5).

Overall or systems savings or areawide increases in efficiency are rarely a motivating factor for any given local agency. Simply because analysts find opportunities to optimize the entire human-service transportation network in a community does not mean that any given agency sees such a proposal as beneficial. Analysts and the taxi industry have charged that the total costs of direct provision by social-service agencies are far higher than alternatives would be. Recent research has found that this is often not true (4). However, even in the instances in which it is true, such comparisons are based on the total costs of service, including expenses not borne by the local agency (e.g., vehicles obtained by means of grants, volunteer drivers, and hidden-overhead subsidies). Although public analysts at the regional, state, and federal levels should make their decisions on these costs, local agencies do not. Local agencies will make their decision to participate in a coordination model based only on an evaluation of the out-of-pocket cost savings (as they understand those costs) offered by a coordination program.

Start-Up and Additional Costs Associated with Coordination

The five HDS coordination demonstrations incurred significant start-up costs and continuing additional costs. None of the projects succeeded in reducing direct costs, and one site actually increased maintenance costs (1). The HDS report noted, "For... potential cost savings in transportation expenditures to be realized from coordination, substantial planning and administrative expenditures are necessary" (1, p. 128). The report concluded, "The overriding theme emerging from the coordinated transportation demonstration program is that coordination is a more costly, complex, difficult, and time-consuming process than had been imagined. The process of coordination is arduous and does not end with initial accomplishments" (1, p. 5).

In addition to administrative costs, other costs may increase after coordination. The vehicles used by many social-service agencies in direct transportation provision are in marginal maintenance condition. Their continued use by only one agency may not create serious maintenance or capital-acquisition problems. If these vehicles become part of a coordinated fleet, however, they may quickly experience serious maintenance problems (1). Thus, the vehicles are a positive resource only to the original agency; they are a cost or a negative factor for most service-coordination models.

Smaller agencies sometimes work with volunteer networks that cannot be easily accommodated within an overall coordination model. By joining a coordinated system, an agency may break down these volunteer networks. The system may also have to hire people to provide the driver and escort services formerly volunteered.

The limited operational experience and the marginal maintenance condition of their vehicles may require social-service agencies that enter service-coordinated schemes to expend considerable resources. Their drivers may not be trained to deal with different types of clients (such as the retarded), and their vehicles may not be equipped to deal with the special needs of new travelers (such as those confined to wheelchairs). Their vehicles may not be compatible with the needs of an overall system (such as radio dispatching), and they may be incapable of operating additional hours or miles without extensive repair and increased routine maintenance.

These are not trivial costs, and they can represent a significant proportion of any agency's transportation budget. Most agencies would want to see significant and measurable changes in the cost or quality of service delivered to their clients as a result of such expenditures.

Agencies That Have Vehicle Fleets

Agencies that currently provide all or some of their transportation services directly in their own vehicles or in staff cars are generally in a different position than those that do not. Most local agencies that have existing fleets will only be willing to engage in service-coordination models (i.e., the joint or coordinated delivery of services) or nonservice models (e.g., joint maintenance). These are the only models that will allow them to use their own resources in such a way that their costs are lower or the quality of service improves. Even then they may require significant inducements to change their current operations.

In Houston, the local Area Agency on Aging (AAA) refused to allow their relatively large vehicle fleet (28 vehicles) to take part in a coordination effort. Finally, the coordinated system offered to carry the clients of AAA to congregated meals for a smaller unit cost than AAA thought it was currently incurring. In addition, the coordination system bought service from AAA for other clients during traditional periods of nonuse of AAA vehicles. In short, AAA reduced their own transportation costs and made money on their formerly idle vans. Only with such strong incentives were they willing to participate.

Agencies currently purchasing all their transportation services from local providers are already involved in one model of coordination. They may be more willing to switch to another coordinated effort (e.g., a system coordinated by a social-service system) because they can see whether they are getting lower costs or better services. Since they currently pay a clearly defined price, costs and benefits are seen easily. The Cerebral Palsy Association in Pittsburgh, for example, was willing to pay the coordinated system there, ACCESS, more money than they had previously paid private carriers because ACCESS lowered the amount of time that their own staff devoted to the administration of the contracted service. The current model of transportation service delivery followed by a local agency will have a significant impact on the type of coordination model it can and will consider.

BARRIERS TO COORDINATION

The previous discussion has shown that coordination efforts sometimes do not bring appreciable cost or service benefits. In other cases, the benefits that do occur do not accrue to the participating agencies but rather to their sponsors or other community agencies. In still other cases, operational cost savings can be wiped out by the high start-up costs associated with participation in a coordinated system.

Given these situations, it is not sensible to style all objections to coordination efforts at the local level as barriers or to see them as irrational or uninformed. Local objections to coordination must be seriously considered and appreciated.

Having an understanding of why local agencies rationally resist coordination efforts in their own self-interest does not mean that barriers to coordination cannot or should not be overcome. Rather, such an assessment of the actual incidence of costs and benefits suggests policies and programs that might realistically address the genuine problems and concerns of local agencies.

When local coordination efforts promise real systemwide cost reductions or service improvements either immediately or in the future, proponents can act to overcome objections and barriers. However, coordination proponents will only be successful in overcoming barriers to coordination if they recognize the differing perspectives of local agencies and the perceived self-interest of potential participants.

Many (although certainly not all) barriers to coordination arise from realistic appraisals of the costs and risks of participation. In such cases proponents can only overcome barriers if they can safeguard against the risks involved or change the incidence of costs and benefits. The following discussion centers on five classes of barriers to coordination found in the AOA study. The discussion suggests feasible and appropriate strategies to overcome these barriers.

Benefits Do not Accrue Directly to Participating Agencies

Agencies resist coordination because of their own perception that the costs and risks are not worth the expected benefits. Before addressing whether these are indeed barriers, it is necessary to first identify whether agency calculations of costs and benefits are correct. If agencies have not correctly calculated their cost savings, they can be trained and assisted to do so.

Often agencies have correctly assessed their own financial and service patterns. A local agency may be optimizing its own resources by using drivers trained under the Comprehensive Employment and Training Act and the "free" vehicles it received through grants. One or more funding sources may, however, see that granting more vehicles to various providers in the community would not be efficient. Such funding sources may feel that the best use of their resources lies in a coordinated effort by local agencies.

In this case, it is useless to try to overcome these "irrational barriers." Coordination proponents must recognize that few agencies will willingly give up free resources or accept communitywide cost accounting. Most agencies, particularly those that have existing vehicle fleets, would be foolish to give up their own resources.

These barriers are not irrational, and proponents are trying to convince the local agency to act against its own best interests. The funding source or relevant state and federal agencies must change their rules and requirements. Agencies should simply not be permitted to make short-term or highly individualized decisions at the expense of the efficiency of the entire community transportation system.

Physical Barriers

Many human-service agencies, particularly in rural areas, noted that there were real geographic barriers to coordination. Some agencies serve many counties. Some of the individual counties are themselves large and have low-density settlements and diverse origins and destinations. Many western urban areas can have very low-density settlement. Some rural agencies noted that their clients had no telephones and could not make easy use of either user-side or service-coordination models. Some providers noted jurisdictional problems; one county's vans could not carry another county's citizens.

These problems and their solutions are probably very site-specific. In some rural areas, social-service agencies provide scheduled, almost fixed-route service, which can be accessed by writing the provider a postcard, by flagging the vehicle down along the route, or by meeting it at an assigned stop at a specific time. Several Texas AAAs reported informal pick-up agreements along joint jurisdictional boundaries; a person would be carried to the county line by one provider and met there by another provider who was going into the urbanized area for medical services or shopping.

Services are so limited in rural areas that it seems likely that where rural providers are not considering coordination, they have some valid reasons. In urban areas, the situation is slightly more complex. If providers resist coordination efforts, they are in effect saying that the problems and costs created by coordinating supply (or maintenance or bulk-gas purchasing) over a large area that has a low-density demand are greater than any benefits they can expect from coordination. Before any attempt is made to overcome these barriers, it should be determined whether the assessment is correct.

Record-Keeping and Accounting Requirements

Many agencies interviewed, as well as the literature (2,3,9), report that the administrative and accounting problems that accompany all four types of coordination models can be significant.

A serious problem for the lead agency in a service-coordination model is that each participating agency and its funding source may demand different types of contractual arrangements for different time periods, different payment schedules, and requirements for different and not entirely consistent trip records. Some agencies feel totally unable to meet this myriad of requirements. Other agencies find themselves unable to get the kind of contracts and records that they need from the facilitator of the coordination effort.

However, many of the difficulties encountered by these agencies reflect their inexperience and failure to understand how transportation systems work. There is some evidence that agencies establish reporting requirements out of ignorance or fear of unknown factors like unit-cost measures and performance indicators. Many agencies simply fear monthly variations in costs (5).

Some local agencies have displayed interest in standard transportation-reporting measures and in vehicle and system productivity. The Houston coordination system held a one-day training workshop for 43 local social-service agencies, including two AAAs and several small human-service transportation providers. The session covered how to set up books and calculate the advantages of alternative contract arrangements.

With such record-keeping assistance, some agencies will better understand the kinds of records and books that they really need for their own information and for the auditing requirements of their funding source. Such assistance may allow lead agencies or coordination facilitators to provide appropriate and complementary records.

In Pittsburgh, the local AAA is purchasing transportation for medical trips for elderly clients from the coordination system there, ACCESS. AAA knew what had been spent in previous years and wanted a simple agreement; the agency wanted to give ACCESS all its transportation money and be assured that all elderly clients who wanted transportation for medical care would get it. This was unacceptable to ACCESS.

ACCESS had its consultant monitor the trip records of AAA's clients for the previous year and calculate the average trip length and average unit cost. They then offered a comparable unit-cost figure (plus an inflation increase) to AAA. AAA still is not quite sure what the figure represents, but a clause has been built into the contract that allows it to stop purchasing services if the total amount spent starts to approach the annual amount set aside for transportation. This is an interesting arrangement, which shows that the facilitator of a coordinated system was willing to help educate AAA and that AAA was willing to take what it perceived as a risk.

Of course, some record requirements are not so readily addressed. A number of respondents at all levels of government reported that certain federal programs, particularly Title 20, were a nuisance to administer (2,3,7). Others reported that state auditors often imposed severe and very limiting requirements on local contractors for fear of conflicting federal audit decisions. Several states and the U.S. Department of Health and Human Services (HHS) have recognized this problem and have established a seven-state consortium; Michigan has the lead agency to develop a model uniform state recording system.

Perceived Statutory or Regulatory Requirements

It is still commonly believed that vehicles purchased by using Title 3 AOA funds may not be used to transport noneligible elderly and nonelderly clients. A corollary is that Title 3 funds cannot be used to purchase wheelchair lifts or radio equipment.

This inaccurate portrayal of AOA policy may have come from the state level down rather than from the agency level. There is evidence that several state AAA units have declared that this is indeed AOA policy, perhaps out of ignorance or perhaps because it gives the state greater control.

Many local AAAs feel that they cannot use Title 3 Older Americans Act monies in coordination efforts if there are any elderly citizens that have unmet transportation needs or if the level of service delivered to the elderly after coordination is in any way inferior to that previously delivered. It is, of course, extremely unlikely that all the transportation needs of the elderly in any community have been met. It is always possible that the level of service will deteriorate after coordination, if only slightly. It is necessary to clearly and definitely explain to local agencies that Title 3 funds may be used directly or indirectly in coordinated systems. It should still be noted that the misconception was expressed by a number of social-service agencies, even in states in which Title 3 monies had already been used for coordination projects. Since this is so pervasive a belief, AOA's congressional coordination mandate might be served by the issuance of some policy guidelines on this topic. The guidelines should clearly explain the permissible uses of AOA funds and the circumstances under which varying coordination methods are possible. Can, for example, Title 3 funds be used for user-side subsidies?

It would also be extremely useful if the AOA were to consider establishing standards on permissible variances in service quality. Local AAAs could then consider how much reduction in the level of service they are willing and able to accept for their clients in order to achieve cost reductions.

Service-Related Features

Many social-service agencies have norms and ideas about how their clients ought to be treated and the quality of service that they require and deserve. Different agencies have different philosophical views about the role of transportation in the care of a client.

Many agencies that deal with the elderly and the handicapped adopt the case-management approach. They attempt to deliver all the services their clients need and try to be helpful to clients in all or most of their social-service activities. Therefore, such agencies provide transportation services directly to their clients. Direct provision in part ensures quality and in part maintains the overall relationship with the client. Some agencies reported that direct transportation provision showed the client that the agency really cared and that the client was important.

The case-management approach to transportation provision tends to be a very expensive model of service delivery and one that can limit how many people an agency can serve. But this normative model of service delivery has a great impact on the social-service community. Sometimes this normative model is linked with expressed fears that other transportation providers could not or would not provide the same intensity or level of care for their clients. Although this level-of-service variable can be expressed in terms of increased riding time, late pick-ups, and rude drivers, it is often expressed simply: "No one can care for our clients as well as we do."

Service-related responses have sometimes been dismissed by coordination proponents as "turf protection," but proponents must recognize that any agency's reluctance to use other community-transportation providers for their clients or to mix their clients with others generally reflects a strong concern for the client's welfare and dignity. Such a decision may be very expensive and not very efficient in the economic sense, but it cannot be characterized as irrational. Any attempt to change the transportation-delivery models chosen by local agencies in order to encourage coordination must address the real and underlying concerns of these agencies.

Most agencies have two key concerns about any type of service coordination. The first is the one they talk about freely; the second is part of their decision process, but they are less willing to articulate it publicly. First, many agencies are concerned that objectively measured indicators of service will move in unacceptable directions; for example, total waiting and riding time will increase and there will be late pick-ups and drop-offs.

The second concern is a related one; many agencies fear decline in far more subjective indicators of service quality. There are clear racial overtones to some of the resistance to coordination in southern and rural sites. This was complemented by the desire of many agencies to serve similar groups of the elderly--those from a cohesive ethnic or religious group or from a given neighborhood. There was real resistance to forcing the elderly to ride with children and strong resistance to mixing the elderly with the retarded or the severely handicapped.

Some agencies fear the breakdown of the volunteer network. Volunteers are important, not only in keeping costs low for social-service agencies but in maintaining a personal, hands-on service (9). Many systems use volunteers as escorts, not because the client or clients really need continued assistance but because it makes their clients feel better and more secure (4,7). Moreover, volunteers, although not continually available to either the agency or to individual clients, are often available for a special trip or particular purpose.

If agency participation in a coordinated mechanism breaks down the volunteer network or convinces volunteers that they are not needed, many specialized trip needs that are often imperfectly served by large or formal systems will no longer be served. It is difficult to know how much this thought consciously underlies service objections to coordination, but it is an issue about which analysts should have some concern.

Proponents too quickly dismiss the service concerns of local agencies as either ill-conceived or improbable. There is evidence from several sites studied by both HDS and AOA that some coordination models have reduced costs and increased efficiency by indirectly decreasing service quality. It may be possible to significantly lower the per-trip cost of transporting an elderly person to a doctor's appointment, for example, if a coordinated system has the capacity to group several comparable trips from or to the same geographic location. The elderly rider, however, would incur some--perhaps significant--increase in waiting and riding time and might also have to cope with an unfamiliar driver and ride with strangers and people unlike himself or herself.

Moreover, some coordination programs trade off one desirable service objective for another, sometimes in ways about which social-service agencies have concerns. There is a trade-off, for example, between high-quality, on-demand transportation services that can only be provided to a few very needy clients and a restricted reservation-type service provided to many more clients.

It is fairly easy for a small system that has its own vehicles and relatively few demands for service to be sure that it is delivering a fairly high quality of service. It would be hard for a coordinated system to provide an equivalent level of service in terms of a number of key variables (waiting time, on-board vehicle time, amount of privacy, etc.).

It certainly is doubtful whether any individual agency can or should be allowed to provide a superior service to a few clients at the possible expense of a lower-quality but more-comprehensive service system for many more travelers. But it must be clearly recognized that the feared change in service quality often voiced as an objection to coordination may be a reality.

The only feasible approach is to help local agencies to understand exactly what it costs them in time and resources to deliver transportation services to their clients in such a personal manner. In addition, it may be helpful to assist such agencies to calculate the increased number of trips they could provide to existing clients or the new clients they could serve within their current budget if they entered some form of coordination model. It is also worthwhile for coordination proponents to decide at what point any given agency should be allowed to "do its own thing," no matter how inefficient that may be, because it would cost too much to change that agency's behavior.

OVERCOMING BARRIERS BY RECOGNIZING HOW LOCAL AGENCIES WEIGH COSTS AND BENEFITS

Historical precedent is a major decision factor for most local agencies. Often they choose to continue their present mode of transportation delivery because it is acceptable and because it has "always been done that way." Being presented with a less costly or better solution is not a sufficient inducement for many agencies to abandon their current model if it is still satisfactory to them. Alghough this is often labeled "turfism," such a reaction is a common one, recognized in the literature of organization theory (5). People are understandably upset at being asked to change long-held beliefs and to reorganize service-delivery models.

There is evidence from a number of case sites that the proponents of coordinated services often made public their belief that local agencies were inefficient and ineffective (1,7). Such views often made local providers defensive; they were forced to develop reasons why they should not join or be forced to join a coordinated service. This defensive posture prevented them from seeing any potential benefits in coordination, and it tended to magnify the negative aspects. On the other hand, the response of system proponents also cut off opportunities for adjustments and resolution; many proponents refused to see any validity in agency concerns about service quality and personal approaches to client needs.

It is often easy (but not very useful) to quickly dismiss the stated objections of local social-service agencies because these objections are not "real" problems or because they were overcome in other communities. It is true that many objections are defensive ones and could be overcome with some persuasion. But even defensive objections are still real objections, and they must be dealt with. Very few of the objections of local agencies are without any basis at all; most stem from a minor problem that arose during the implementation of the coordination models.

All the previously identified objections are "real" ones. To a great extent they are susceptible to financial solutions; loans and start-up monies as well as driver-training and personnel-training courses could overcome many of the initial problems. Agencies can be trained to monitor and limit client trip making if they so desire, to take part in service coordination (joint maintenance, for example), and to keep the kind of records that would allow them to buy from or to supply services to a coordinated network.

It should be noted, if it is not immediately clear, that these strategies to overcome operational barriers to coordination all involve the commitment of resources (time, money, training skills) that must come from some other agency or service. Often some agency has to be willing to spend money to eventually allow for the saving of money. In Pittsburgh and Houston, the brokers that managed a service-coordination model provided assistance and, indirectly, funds to enable agencies to overcome their participation barriers.

Any local agency may find that because of increased overhead and administrative costs or high driver-training expenses, coordination is not cost-effective. If a regional plannning agency or social-service funding source believes that in the long run the community may be better served by the development of such a coordinated system, it may subsidize the local agency or in some way cover its additional costs.

The strength with which objections to coordination are advanced may decline as the agencies involved become more familiar with the coordination programs and less defensive. It is the unusual person who finds serious changes easy to make and easy to accept initially. As the agency staff think through a coordination model, they may become more open to suggestion if they are not further forced into defensive postures.

CONCLUSIONS

There are significant financial and psychological costs involved in implementing coordination programs in social-service delivery systems. Some of these costs are incurred directly by purchasing new vehicles or additional insurance or by setting up specialized record keeping. Other costs are incurred in overcoming the resistance and doubts of potential participants.

Sometimes coordination of transportation resources cannot be achieved without some diminution in the level of service delivered to agency clients and in the personal responsiveness of the service. Often these tangible and intangible costs are far higher than local participating agencies can or are willing to bear. More often than previously recognized, all economic and other costs are so high that they cast doubts on the cost-effectiveness of proposed coordination efforts.

Overall, when analysts and planners are certain that coordination in any community is the most sensible and efficient long-run approach to transportation delivery, they must be willing to provide time, money, and professional resources to convince local participants of this outcome and to help agencies to cover the costs that they cannot directly recover. And planners must recognize and address, to the greatest extent possible, the very legitimate concerns that human-service agencies have about the quality of services that they wish delivered to their clients.

REFERENCES

1. J. Burkhardt, S. Knapp, and M. Ramsdell. Coordinated Transportation Demonstration Results Evaluation of the Office of Human Development Services Transportation Demonstration Program. U.S. Department of Health, Education and Welfare, Feb. 29, 1980.
2. D.A. Cutler. Statutory and Regulatory Analysis of Incentives and Barriers to Coordinating Transportation Services for the Elderly and Handicapped. Ecosometrics, Inc., Bethesda, MD; U.S. Department of Transportation, Jan. 31, 1979.
3. Institute for Public Transportation. Statewide Study of the Feasibility of Coordinating or Consolidating Specialized Transportation Services. New York State Department of Transportation, Albany, Final Rept., Oct. 1978.
4. S. Rosenbloom and others. Cost-Effectiveness Measures for the Transportation Services for the Elderly and the Handicapped. UMTA, U.S. Department of Transportation, Oct. 1979.
5. A. Saltzman. Coordination of Transportation by Human Service Agencies: An Interorganizational Perspective. Technology Sharing Program, U.S. Department of Transportation, Jan. 1980.
6. Hindrances to Coordinating Transportation of People Participating in Federally Funded Grant Programs. U.S. General Accounting Office, Vol. 1, Oct. 1977.
7. Coordinated Transportation Services for the Elderly. Urban Institute, Washington, DC, Final Draft 1402-1, Sept. 1980.
8. HDS Transportation Initiative: A Report on Research Findings. Office of Human Development Services, U.S. Department of Health and Human Services, Jan. 1981.
9. Transportation Services: An Examination of Regulatory Problems in Medicaid and Social Service Programs. Office of Human Development Services, U.S. Department of Health, Education and Welfare, Jan. 1980.

Publication of this paper sponsored by Committee on Paratransit.