TRANSPORTATION RESEARCH RECORD 1064

Transit Planning and Operations: Issues and Options

TRB
TRANSPORTATION RESEARCH BOARD
NATIONAL RESEARCH COUNCIL
WASHINGTON, D.C. 1986

Transportation Research Record 1064
Price $8.20
Editor: Julia Withers
Compositor: Joan G. Zubal
Layout: Marion L. Ross

mode
2 public transit

subject areas
11 administration
12 planning
13 forecasting
14 finance
54 operations and traffic control

Transportation Research Board publications are available by ordering directly from TRB. They may also be obtained on a regular basis through organizational or individual affiliation with TRB; affiliates or library subscribers are eligible for substantial discounts. For further information, write to the Transportation Research Board, National Research Council, 2101 Constitution Avenue, N.W., Washington, D.C. 20418.

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data
National Research Council. Transportation Research Board.

Transit planning and operations.

(Transportation research record, ISSN 0361-1981 ; 1064)
1. Local transit—Congresses. I. National Research Council (U.S.). Transportation Research Board.
II. Series.
TE7.H5 no. 1064 380.5 s 86-21691
[HE4202] [388.4'068]
ISBN 0-309-04058-2

Sponsorship of Transportation Research Record 1064

GROUP 1—TRANSPORTATION SYSTEMS PLANNING AND ADMINISTRATION
William A. Bulley, H.W. Lochner, Inc., chairman

Urban Public Transportation Section
John J. Fruin, PED Associates, chairman

Committee on Public Transportation Planning and Development
Eugene J. Lessieu, Port Authority of New York and New Jersey, chairman
Carol A. Keck, New York State Department of Transportation, secretary
Paul N. Bay, Joby H. Berman, Daniel Brand, Dick Chapman, Chester E. Colby, John Dockendorf, David J. Forkenbrock, Hugh Griffin, Michael A. Kemp, David R. Miller, Robert L. Peskin, Gilbert T. Satterly, Jr., George M. Smerk, Donald R. Spivack, Patricia Van Matre, William L. Volk

Committee on Rail Transit Systems
Richard Miller Stanger, Los Angeles County Transportation Commission, chairman
John J. Bakker, Glen D. Bottoms, James A. Brown, Donald O. Eisele, Alex E. Friedlander, Howard Goode, Daniel L. Jones, Jr., Ata M. Khan, Snehamay Khasnabis, Shinya Kikuchi, Robert J. Landgraf, Thomas F. Larwin, Tom E. Parkinson, S. David Phraner, Peter J. Schmidt, Fred J. Silverman, Brian E. Sullivan, Edson L. Tennyson, Rush D. Touton, Jr., J. William Vigrass, Vukan R. Vuchic, Harold Wanaselja, Jimmy Yu

William Campbell Graeub, Transportation Research Board

Sponsorship is indicated by a footnote at the end of each paper. The organizational units, officers, and members are as of December 31, 1985.

1064

Contents

Evaluation of Competition in the British Local Bus Industry

IAN SAVAGE

ABSTRACT

The British local bus industry has been organized as a system of strictly regulated route monopolies for more than 50 years. Suggestions that this monopoly is undesirable have prompted a critical appraisal to determine the economically optimal market structure. Contained in this paper is an analysis that concurs with the common view that competitive stimulus can result in lower-cost operation. The analysis concludes that a reduction in cross-subsidy, caused by competition on the more profitable routes and timings, will be beneficial. This result arises because cross-subsidy currently disguises some loss-making services that are provided needlessly, and is also an economically inefficient way, vis-à-vis direct subsidy, of funding unremunerative bus services. However, the analysis concludes that the current United Kingdom government's solution to this, of permitting competition between bus companies "on the road," is also undesirable. This is because direct competition is liable to result in short-term waste and will not a priori lead to optimum provision in the long run. In addition, it can cause problems by severing demand- and supply-side linkages and increasing the chance of unacceptable driving and maintenance standards. Therefore, the institutional problem addressed in this analysis is how to obtain the long-run benefits without the costs of unfettered competition on the road. This would indicate that, in the bus industry, competition for the market, rather than in it, is required. The analysis concludes that for an effective potential competition in the bus industry to exist, a regulated system with low entry barriers such as franchising or contracting of services should result.

Internationally, there has been a general reduction in transport regulation in the last 10 years. For example, controls have been removed from the airlines in the United States and from long-distance coaching in Britain. However, the proposed total deregulation of local "stage-carriage" services in Britain is highly significant as, with the exception of Chile, no comparable change has occurred elsewhere. This paper, which is based on a Ph.D. thesis undertaken from 1981 to 1984, attempts to determine the optimal market structure for this industry. Its conclusions are somewhat at variance with current United Kingdom government policy.

HISTORICAL BACKGROUND

Following intense competition on local bus services in the 1920s, regulation was introduced in the form of the 1930 Road Traffic Act. In addition to quality controls on operators and vehicles, the act set up a protected monopoly on each route, using a licensing system administered by regional traffic commissioners.

The basis for the granting of the licenses had two profound effects on the structure of the bus industry. First, the protected monopoly was granted partially in return for an undertaking by the bus companies to provide unremunerative services out of the profits generated on other activities (known as cross-subsidy).

Second, if an operator was already operating a route, he would have priority if the license was challenged by a potential entrant. Amalgamations and takeovers of neighboring companies in the 1930s coupled with the priority for licenses resulted in a small number of large bus companies, each of which had a secure territorial monopoly. In recent years, it was suspected that these large companies, now all publicly owned, and together providing 92 percent of local bus miles, had been cosseted by the priority principle from effective competition; thus, inefficiencies had arisen and innovation had been stifled.

A Conservative government was returned in 1979 with a policy of encouraging a competitive atmosphere throughout the public sector, and the bus industry was no exception to this. The 1980 Transport Act, in addition to removing all quantity controls over long-distance (express) services, removed fare control and encouraged some competition of the direct on-the-road kind on local services. This was not a major relaxation of licensing, however, and widespread competition did not emerge.

However, the Conservatives were reelected in 1983 and, in October 1985, an act was passed that will eventually deregulate the industry. It was considered that only a complete deregulation would (a) allow free-testing of innovation, and (b) secure and sustain cost savings. The Conservatives thus proposed to remove the licensing system. However, the monitoring of the quality of operators and vehicles is to be retained and strengthened, to protect the public from any "foolish" behavior by operators. In addition, because of concern about the amount of money devoted to subsidy, the government proposes that

---

I. Savage, Transportation Consulting Division, Booz-Allen and Hamilton, Charles II Street, London SW1Y 4AE, England. Present address: Department of Economics, Northwestern University, Evanston, Ill. 60201.

public money only be used to sustain services on routes or at times of day that would not be provided in the free market. Competitive tender will be introduced for the allocation of such support. Finally, the large publicly owned bus companies are to be reorganized into smaller free-standing parts and transferred to the private sector.

The purpose of this paper is to determine the most optimal (in the economic sense) market structure appropriate to the stage-carriage bus industry. It is thus concerned with the form of regulation rather than with the issues of optimum subsidy levels or ownership. It will answer the following questions:

1. Is there a need for a competitive environment?
2. If so, should there be "unfettered" competition (similar to that proposed), and
3. If unfettered competition is not desirable, what requirements are there for an optimal market structure?

THE NEED FOR A COMPETITIVE ENVIRONMENT

Four arguments can be advanced for wishing to introduce competitive stimuli into the bus industry. They are

- Operating costs can be reduced,
- Demand and supply will be better matched,
- Innovation will be encouraged, and
- The industry is not a natural monopoly and competition is likely to be sustainable.

Lower-Cost Operation

By the 1970s, it was believed that the regulatory system had protected inefficient or high-cost operators. The introduction of a competitive stimulus can be expected to lead to cost reduction either by lower-cost entrants to the industry forcing out higher-cost ones, or by commercial pressures making existing operators become more efficient. The author undertook investigations to see if either of these was likely to occur following deregulation.

Existing Firms Becoming More Efficient

Some economic writers, such as Leibenstein, suggested that economic welfare losses attributed to inefficiency (or "X-efficiency" as it is known) in monopoly situations are greater than the resultant allocative efficiency, deadweight loss (1). He argued that when profits are high, or when there is no competitive pressure, slack working practices result.

The author attempted to observe the most likely source for X-efficiency gains within existing bus operators. Following studies of both a labor market (bus drivers) and a capital factor market (the market for buses), it was concluded that the former market had the most scope for an X-efficiency gain.

In the labor market, the author's investigations indicate that the competitive effect will be manifested in the productivity rather than the wage dimension. This is not surprising as wages are generally determined nationally, although work content is broadly under the control of local management. To test for this, the author undertook econometric analysis on data [for subsidiaries of the state-owned National Bus Company (NBC)] before and after the limited relaxing of licensing in 1980.

Analysis of the wage data did not identify any perceptible change following the new legislation. However, econometric results of investigations of productivity data gave indications, albeit not statistically significant, that there was room for improvement. This supports a wealth of descriptive analysis that indicates that productivity can be increased. In particular, there is a plethora of restrictions on the scheduling of driving staff to particular runs. As far back as 1967, the National Board for Prices and Incomes (2) had noted on the subject of scheduling constraints that "There is evidence . . . that the scope for negotiable change may well be considerable."

The author concurs with this and believes that statistically insignificant changes in productivity following the 1980 legislation were due to the general paucity of entrants to the market rather than the existence of limited potential to increase productivity. This suggests that greater X-efficiency reductions may only result from a market regime in which the threat of potential competition is more real and effective.

Lower-Cost Operators

In traditional economic theory, a benefit of competition occurs when a genuine lower-cost firm replaces a higher-cost one. In Britain, there are numerous independent operators (nearly 6,000) of which a proportion would wish to provide scheduled local bus services. The author investigated whether these operators were genuinely lower cost compared with the large, local bus companies.

A direct comparison of costs is problematic. For example, the comparison of operating costs for different operators on similar routes or timings is not meaningful. The interworking of routes or timings, or both, by bus companies means the level of cost on individual routes or timings depends on how they fit into a governed set of other operations.

Because of these economies of scope, route costs do not necessarily reflect the underlying differences in unit costs between operators. Thus, the cost of operation by two operators need not be ranked the same on all routes. However, it is argued that if sizable parts of networks were passed to independent operators, there would be a saving in resource cost.

Accepting this, the evidence that independent operators have cost advantages when they are small is considerable [for example, Tunbridge and Jackson (3)]. However, it should not be inferred that this advantage would persist if these operators gained a large local bus operating commitment. This involves additional costs of bus stations, inquiry offices, and bus stops as well as operating at times that are traditionally relatively expensive (e.g., evenings and Sundays, and the provision of high- and off-peak vehicle ratios). In addition, the increased company size may result in increased unionization or a change in labor union attitudes, or both. Nevertheless, a licensing system, based on longevity of operation and not level of costs, can preclude genuine lower-cost operators if they emerge. Recent evidence (4) has suggested that small private operators could be up to 20 percent cheaper than existing operators.

In summary, opportunities for reduced operating costs following deregulation do appear to exist. However, it will be noted that much of this reduction is due to a reduction in staff wages and conditions. Therefore, only part of the cost reduction will actually be a welfare gain to society, as much of the cost reduction will merely be a transfer from workers' to producers' or consumers' surplus. The actual split between transfer and social welfare gain will depend on the amount of passenger traffic generated as a result of the lower-cost operation being passed on to the consumer.

Better Matching of Demand and Supply

The 1930 legislation inherently encouraged cross-subsidy between services and times of day. Indeed, since the Second World War, the traffic commissions have, in the face of declining demand, explicitly tried to maintain the largest possible network by the use of cross-subsidy. However, it has been argued in recent years that cross-subsidy was both distorting individual bus markets and disguising services that were being provided needlessly.

It may be presumed that entrants to the stage bus industry, being primarily private companies, will seek to make a profit. They may thus be expected to attack the routes and timings of the existing network operators where they can make the most money. The abstraction of revenues from the profitable segments will lessen the amount of finance available for cross-subsidy. Thus, competition can be expected to reduce cross-subsidy. In the following section it is argued that this reduction in cross-subsidy will result in a more efficient allocation of resources.

The Definition of Cross-Subsidy

The definition of internal cross-subsidy is problematic. It exists because profits on some activities are used to support loss-making activities. It is therefore particularly important to define "profits" and "loss-making." This will depend crucially on the assumptions made concerning costs. For management purposes, the true definition of a cross-subsidized service must be when avoidable costs exceed avoidable revenues. Thus, the Ponsonby/Hibbs (5,6) test of "Would we be better off if we did not run service X?" would be the most appropriate. The problem of data has meant that, traditionally, a system of fully allocated costs and revenues has been used to identify cross-subsidy.

On this basis, certain characteristics of cross-subsidy have been identified by recent studies by the Institute for Transport Studies (ITS) (7), the MVA Consultancy (8), and Booz-Allen and Hamilton (9). The cross-subsidy between routes is widely recognized. The ITS work indicates that, generally, the interurban routes support rural and, to a lesser extent, urban routes. Cross-subsidy between times of day on individual routes is less well known and depends crucially on the allocation of costs adopted. The recent works have shown that the extent of cross-subsidy varies by location. However, the weekday interpeak and Saturday daytime periods have generally been identified as the main surplus generators, and the evening and Sunday periods are unremunerative. The financial position of the peak periods depends largely on the number of vehicles solely reserved for use at that time. A third type of crosssubsidy is between individual parts of a route. However, the data complexities have meant that none of the recent studies have tackled this.

The overall implication is that not only is there a transfer of surplus between passengers on different routes at different times of day, but there will also be a transfer between different person types and journey purposes. The recent studies have shown that cross-subsidy is not only widespread, but also, as the ITS and MVA work illustrates, can be more important than external subsidy in maintaining unremunerative activities.

A Critique of Cross-Subsidy

Internal cross-subsidy has been subject to a large amount of criticism. A particular criticism is that it can cause a misallocation of resources. This is because (a) passengers on remunerative activities are paying higher prices or receiving lower frequencies than they would if capacity were expanded to remove abnormal profit and (b) on some unremunerative activities, cross-subsidy is presently supporting a level of provision that does not reap sufficient consumer benefits to outweigh the resource costs. [Note that the distortion to efficient allocation of resources caused by cross-subsidy has been analyzed by Gwilliam (10).]

The implication, therefore, is that if competition on remunerative activities reduces the level of cross-subsidy, then, in these circumstances, there will be a better matching of demand and supply in all bus markets and, therefore, a more efficient allocation of resources. However, not all unremunerative activities reap insufficient consumer benefits to justify their existence. In these cases, the crucial issue becomes whether it is more efficient to financially support these services by raising abnormal profits on inherently profitable activities, or by direct payment from public funds.

The cost of raising public funds is not clear-cut, however, as any increased local authority support might come from a variety of sources. Browning (11) reviewed the shadow price of taxation and found it to lie in the region of 1.1, depending on the form of taxation used. This can be compared, on a purely allocative basis, with the welfare cost of raising abnormal profits on inherently profitable operations.

The distributional consequences are arguably the more important. Obviously, as a result of the relative numbers of people involved in the two scenarios, the burden of losses per person on the passengers in the subsector where finance for cross-subsidy is drawn is probably larger than the welfare losses of whatever taxation system provides the alternative. Therefore, if unremunerative activities are now provided by a general taxation system, then there would be a shift from raising money from (primarily) women on shopping trips to the community in general. It can be argued that this is certainly more equitable and may be better in terms of distribution.

The author concludes that on an allocative basis, it is not clear which (direct subsidy or cross-subsidy) is welfare-superior. However, the effects on cross-subsidized services cannot generally be used as an argument against competition, as activities that have higher consumer benefit than resource cost can potentially be funded by direct subsidy, which is liable to be preferable to cross-subsidy on a distributional basis.

In summary, the reduction in cross-subsidy as a result of competition can be seen as beneficial because (a) cross-subsidy is an opaque form of subsidy and can disguise loss-making services provided needlessly, and (b) on the balance of allocative and distributive arguments, direct subsidy is preferred to cross-subsidy.

The implication, therefore, is that competition will a priori increase the amount of direct subsidy required to maintain the current network. In Britain, deregulation is occurring at a time of great pressure on government expenditure. Therefore, some service loss might be expected unless subsidy is sufficiently increased. The United Kingdom government, however, has argued that the increased subsidy requirement, resulting from the loss of cross-subsidy, will be counteracted by a reduction in expenditure as operators' costs fall in the competitive environment. Whether this will be true or not will depend crucially on the amount of cross-subsidy in a particular local network and the magnitude and timing of cost reductions. A recent Booz-Allen and Hamilton study (9) graphically illustrated this trade-off. In the

County of Surrey, cross-subsidy was found to be relatively low (£1 of cross-subsidy to every £3 of external subsidy), yet cost reductions of from 10 to 15 percent were needed to counterbalance the effect on direct subsidy of the loss of cross-subsidy.

Undoubtedly, the high probability of withdrawal of loss-making services after deregulation can be seen as a negative argument for competition. However, the author does not concur with this because (a) the reduction in cross-subsidy is a desirable long-term objective of transport policy and (b) if there is a loss of service because external subsidy has not increased, then the tax-paying public have inherently shown their preference (via their elected representatives) on the extent of loss-making services to be provided. This contrasts with the present situation where the cross-subsidy is raised from some transport users who are often not identified, let alone consulted on their preferences on the size of the network.

Innovation

Academic researchers have not proved conclusively whether a monopolistic or competitive market structure produces more innovation (12,13). However, it is contended that, in this industry, it has been the form of monopoly (i.e., the issue of route licenses) which has meant that there has been inflexibility to experiment and innovate.

An objective of the 1980 legislation was the hope that innovation could be encouraged in rural areas where informal public transport would replace the fast-disappearing traditional service. However, experience in the 1980s suggests that these deeply rural (and deeply unprofitable) services would only be encouraged by local authority subsidy and not relaxation of licensing. Therefore, the author concludes that, if strict route licensing were relaxed, then innovation would be expected in urban rather than rural areas and would take the form of new links in the network, product differentiation (especially paratransit), and competition against the railways.

Although not all innovation will be an a priori benefit, it can be reasonably expected that competition will cause innovation in the industry. Indeed, in Britain, NBC is introducing high-frequency minibus services in many towns in the run up to competition. In the section "Optimal Innovation" (elsewhere in this paper), the author discusses whether unfettered competition actually leads to the optimal amount of innovation.

Sustainability

It is frequently argued that there is no a priori reason why the local bus industry should be a monopoly. Research indicates that first, the industry is not a classic natural monopoly, and second, competition is likely to be sustainable. A study of the publicly owned bus industry by Lee and Steedman (14) found few economies of scale relative to company size. It is now commonly accepted that the bus industry displays constant returns to scale in terms of bus miles produced. In terms of the classical definition, the bus industry is therefore not a natural monopoly.

In addition, where there are incentives to enter the market, competition appears likely to be sustainable, especially if traffic is heavy or if the capacity offered is small in relation to the existing operation. This is due to the nature of local bus competition with low entry and exit costs, free access to the market, and no prebooking. Qualifications are that there appears to be a need to regulate terminals to avoid monopoly returns to their owners, and also that competition needs to occur on enough fronts to stop predatory action against entrants.

The system of statutory monopoly with priority for (what became) large network operators is alleged to have led to inefficiency, stifled innovation, and cross-subsidy. There would thus appear to be strong and undeniable arguments (based on X-efficiency gains, the introduction of low-cost operators' greater control over the level of provision on unremunerative services, and encouraging innovation) for the introduction of a competitive market structure into the stage bus industry.

SHOULD THERE BE UNFETTERED COMPETITION?

The solution adopted in Britain to deal with the disadvantages of monopoly has been to encourage direct competition on the road. The author concludes that the unfettered competition has several serious disadvantages:

- Wasteful competition in the short run,
- Nonoptimum long-run price/frequency outcomes,
- Erosion of demand- and supply-side links,
- The existence of artificial monopolies,
- Uncertainty,
- Nonoptimal innovation, and
- Reduced levels of safety.

Wasteful Competition

In the 1920s, it was frequently argued that competition on the road was unnecessary and wasteful. This implied that the competitive benefits to consumers on the competed route--from reduced fares and waiting times and the consequent generated traffic--are outweighed by the additional resource costs involved in competition. The author undertook an analysis, using economic models, to determine under what conditions this argument holds.

To do this, a model was developed to help understand competitive market decisions. It was clear that actual competition in the bus industry will tend toward oligopoly (competition among few) rather than perfect competition (competition among many). In the case of oligopoly, the inappropriateness of existing theory meant that the author had to develop a game theoretical approach to the policy decisions made by the competitors.

Using this model, an analysis of possible decisions by operators indicated that two tactics would generally be favored in competitive situations. First, each operator would wish to time his bus to "headrun" the opposition, whereby an operator locates close in front of the opposition and takes all the traffic. This is a version of the well-known Hotelling (15) principle whereby competitors locate spatially close to each other. Second, there is strong pressure, when competition is based on a homogeneous product, not to let price differentials persist and, thus, matching of fares is noted. Bearing these points in mind, it is possible to analyze whether the move to oligopoly from a base monopoly fare and frequency combination will produce increased or decreased social welfare.

An analytical device was developed from an underlying bus route cost-and-benefit model [described in detail in Savage (16, Chapter 5)] in order to do this. A diagram can show the relationship between frequency offered (per period of time) and the social welfare level resulting for a given fare level. This is shown in Figures 1 and 2, in which fare level F2

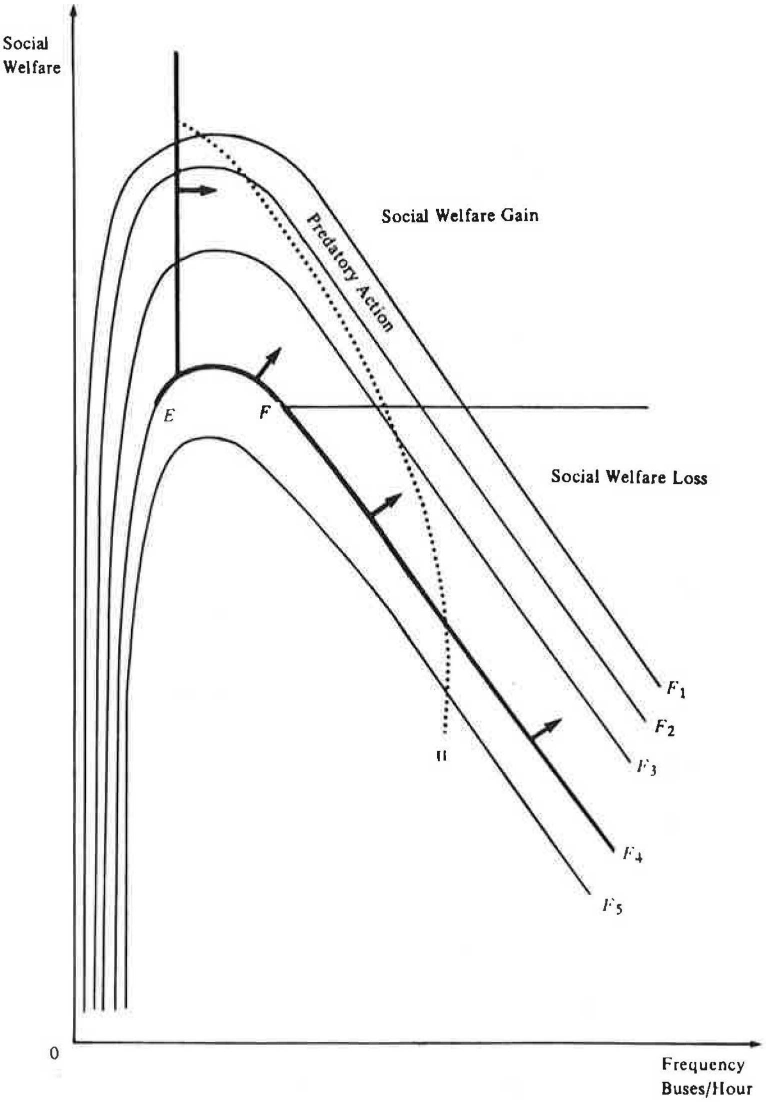

FIGURE 1 Wasteful competition—under-optimal provision.

is greater than fare level F1, and so forth. For a given fare level, additional buses at low frequencies produce an increase in social welfare as waiting times are significantly reduced and considerable traffic generated. An optimal level is then reached and, after that, social welfare declines as additional buses are put on. This is because the benefits of reduced waiting times are now much smaller (and the amount of generated traffic much less) and these are outweighed by the additional resource cost of the additional capacity provided.

The level of producer surplus (or profit) can also be represented in the diagram. This is shown by the broken contours. The most important of these is labeled Π_0 and represents the break-even position. All fare and frequency combinations outside of this contour represent a loss on the bus route. If the fare and frequency pair on a route is on the break-even contour (or, because of the indivisibilities, up to one bus per unit of time inside it), it would not be possible to expand capacity without incurring a financial loss on the route. Unless it would be taking predatory action, no bus company would be willing to move the route (and hence itself) into a loss-making position. The most favorable routes for entrants are those that generate a surplus. Thus, it can be expected that the routes on which competition is likely to occur are those on which the present fare and frequency combination is well within the break-even contour.

Oligopolistic competition is now introduced into the model. In the succeeding analysis, the following initial assumptions have been made:

1. Fare matching occurs;
2. The competitors have similar costs; and
3. Except when buses are full, the greatest advantage to he consumer accrues when buses are inserted equally between existing departures.

(Assumptions 2 and 3 will later be relaxed, however.)

To observe whether competition will bring a social welfare gain or not, it is necessary to look at two general cases. The first of these is where the monopoly frequency was originally less than the optimum, as it may be, particularly in some peak periods. This is shown in Figure 1. The monopoly fare and frequency combination is at point E. A feasible region for competition can be defined by applying the criteria that (a) fares cannot increase, and (b) frequency must increase by at least one bus per unit of time, as the competitor has to introduce some capacity. The representation of this in Figures 1

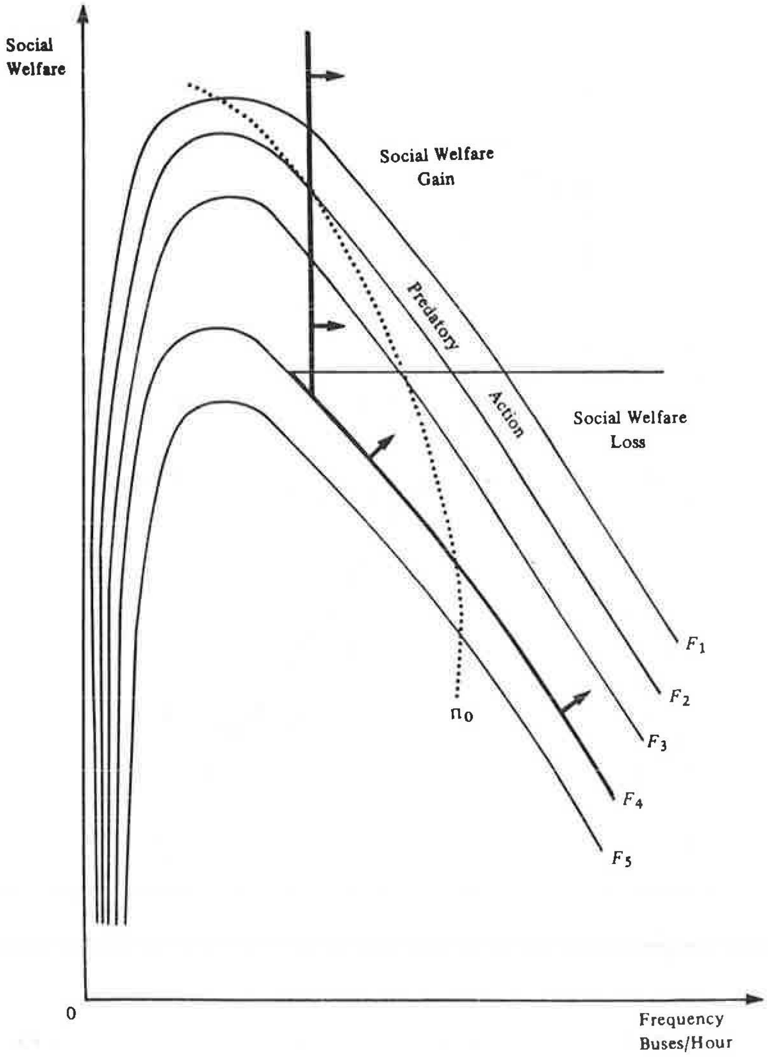

FIGURE 2 Wasteful competition—over-optimal provision.

and 2 will depend crucially on the horizontal scale adopted.

This is the area above and to the right of the bold line. The part of the area beyond the break-even contour represents fare and frequency combinations that would make the route unprofitable. Thus, fare cuts or frequency increases, which move the route into this region, depict predatory action on behalf of one of the bus companies. The area inside the break-even contour, however, represents fare and frequency combinations in which all firms are making a profit and, thus, oligopoly is more stable.

If a horizontal line is drawn through the feasible region at the same level of social welfare as point E, it is observed that all points above this line represent a welfare gain and all points below, a loss. In this particular case, it is noted that on the frequency and welfare function between points E and F, social welfare can be increased by introducing new capacity alone, without the need for reductions in fare. It is only in the case where monopoly fare and frequency are suboptimal, and competition takes the route to the optimal point, that oligopoly has been successful in moving a suboptimal monopoly resource allocation toward the welfare optimum.

However, in an industry with declining demand, a dynamic version of the model would have the frequency and welfare functions moving down and to the left. Attempts to maintain capacity in the face of declining demand would lead to the monopoly frequency being greater than the optimum (Figure 2). It is observed that the fare and frequency combinations in which a social welfare benefit, without losses (depicted by the shaded area), occurs is now much smaller. For a welfare gain, any increased frequency must be matched by a cut in average fare levels. However, for any given increase in competitive capacity, the entrant will maximize his constrained profit by pricing close to the existing fare. This is not compatible with moving to the shaded area. This rule remains valid regardless of how far point E may be from the optimum frequency.

When the assumptions on cost and timings are relaxed, it is observed, in the case of the entry of a lower-cost operator, that the area where a welfare gain can be experienced without financial loss increases marginally but does not alter the overall conclusion of the analysis. However, if, as has been observed, entrants have located themselves close to existing timings (known as "headrunning"), then society will gain little consumer benefit at the expense of additional resource costs. In this case, it is extremely unlikely that there would be any scope for social welfare gain, even if massive fare reductions were offered.

In conclusion, unless peak inadequacy is relieved or substantial traffic is generated, which, in practice, is unlikely, it would appear that in the short run the oligopolistic market structure will not cause a previously suboptimal monopoly resource allocation to converge on a welfare maximum. Furthermore, it is probable that an oligopolistic regime will lead to reduced social welfare and a waste of resources, particularly if the favored competitive tactic of headrunning is employed.

Nonoptimum Prices and Frequencies

In the preceding section, it has been concluded that the on-the-road competitive phase of the oligopolistic game will generally be wasteful. However, this wasteful interlude is likely to result in a return to monopoly, either by some of the competitors dropping out of the market, or by collusive agreements being reached. It is therefore appropriate to ask whether the long-run fares and frequencies resulting from competition will be optimal.

An analysis of which fare and frequency combinations will be chosen on the return to monopoly is difficult as, in practice, a monopolist can select one of many combinations to offer on a route. Nash (17) identified four likely areas for maximization management objectives as follows:

- Social welfare,
- Profit,
- Passenger mile, and
- Bus mile.

Apart from the specific welfare-maximizing policy, only passenger mile maximization--with passenger miles weighted according to their social function (18)--is a proxy for social welfare optimization. A profit-maximizing monopolist will not therefore select a fare and frequency combination consistent with an optimum allocation of resources. Indeed, it would appear that unless a welfare-maximizing management objective, subject to budget constraint, is adopted, there is no a priori reason why a monopolist will select an optimum allocation of resources in preference to any other fare and frequency combination.

Two conclusions can be drawn at this point. First, left to their own devices, monopolists are unlikely to provide socially optimal fare and frequency combinations. Second, it does not appear a prior that competitive interludes will necessarily improve matters as there would appear to be no reason why the competitive phase will necessarily influence the final fare and frequency choices.

In economic terms, where the final outcome is not welfare-superior to the precompetitive resource allocation, the intervening oligopolistic period--on the basis of the analysis of the section on Wasteful Competition will probably have been wasteful. Even if the intervening competitive phase does lead to a welfare-superior final outcome, there is likely to be a "pay-back" period in which the benefits of the new monopoly solution compared with the original one are cancelled out by the wastes of the competition.

Overall, different market structures can be judged according to whether they will converge on a social welfare-maximizing solution. However, the difference between the units of demand and supply in bus operation (meaning that operators can choose both the fare they charge and the output they produce) results in there being many possible fare and frequency combinations that satisfy any particular budget constraint. In neither of the market forms studied (monopoly and oligopoly) was there any reason why the social welfare-maximizing combination, rather than any other combination, would necessarily be chosen. In addition, the introduction of competition is not likely to make a previously inefficient monopoly allocation converge on the social optimum. In conclusion, it would therefore appear that to obtain the optimal allocation on a route, it is better to use a policy that would encourage a monopolist to act in a socially efficient way rather than a policy of unfettered competition on the road.

Demand and Supply-Side Links

Competition will be expected to occur only on the remunerative parts of existing networks. It is therefore quite likely that networks will be broken up. This may have undesirable consequences if there are linkages, either on the demand or supply side, between routes. The author undertook an analysis to try to identify whether any such links exist.

On the supply side, some linkages are inevitable in an industry producing multiproducts (i.e., routes). These links--described as "economies of scope"--occur when advantages can be made from complementarity in production. In the bus industry, this occurs not only when these are joint costs (i.e., management) but also when vehicles are interworked between services. Where this occurs, there may be localized natural monopolies arising from the economies of scope (19,20). If price is divergent from costs (i.e., for the purpose of cross-subsidy), then a regime of indiscriminate competition can result in breaking of the natural monopoly and a loss of the cost-saving complementarities.

On the demand side, there are often complementarities of revenue. This classically occurs in the case of feeder routes. The feeder route may make a loss by itself, but it may generate a more-than-compensating increase in revenue on a trunk route. This may also be the case when a bus company provides some services unprofitably in, for example, the evenings because it knows this will have a positive overall effect by making the service more attractive to the rider. In both cases, these unprofitable but commercially viable (in a network context) services can only exist by a monopoly operator being able to realize jointness in demand by making financial transfers between services. In an era of uncontrolled competition, these financial links will be severed--as profits are competed away--and services only provided commercially on the basis that revenues they generate elsewhere might become unviable and, therefore, endangered.

A similar argument also applies to the increasing trend toward prepurchased system- or zone-wide tickets. Although these ticketing systems are not a priori necessarily a benefit, the fragmentation of networks in a competitive regime is likely to reduce the ability to continue to offer these schemes. The author was unable to undertake any analysis on this point.

Artificial Monopoly

The recent work in identifying cross-subsidy in bus operations (discussed previously) has indicated that the profit incentive does not exist in many parts of the bus industry. Therefore, even if entrants have relatively low costs, competition will not be seen on much of the current network (i.e., artificial monopolies exist). The United Kingdom government is proposing to overcome this problem by instigating a system of specific operating subsidies for individual routes, allocated between operators by competitive tender. Therefore, on the unremunerative services, competition will be encouraged for the market rather than in it.

However, the cross-subsidy analysis also reveals that, of the profitable services, many are financially marginal in nature. With the prospect of increased competitive capacity leading to reduced load factors, which, in turn, leads to operating losses, some currently commercial services will also not witness competitive activity. With the absence of incentives to enter the market in a proportion of the local bus industry, it can be assumed that unfettered competition will not allow the effects of competitive stimuli to be fully felt.

Uncertainty

An international collaborative study (21) highlighted the importance of service reliability in determining public transport demand. A concern raised with unfettered competition is that the short-run, intensely competitive phase will feature relatively frequent changes in operator timetables and fare scales. This will lead to uncertainty and could have a damaging effect on the overall level of patronage. The author, while concurring with this view, cannot personally bring any concrete evidence for debate.

Optimal Innovation

The preceding discussion (see section on Innovation) concludes that moving away from the existing controlled license system would encourage innovation in the industry. It is clear that innovation that produces increased social welfare is desirable; however, it is not clear whether unfettered competition will necessarily result in only optimal innovation. The author investigated this point.

Whatever form innovation will take, it is likely to impinge on existing services in one form or another. Therefore, the proper way of evaluating innovation is to compare the original service with the innovated service running exclusively. The analysis splits innovation into two types. The first is where the innovation is welfare-inferior to optimal provision by the existing service, but can compete because the existing service is currently inefficient. Entry of this type is likely to not only cause short-run losses of on-the-road competition, but could, if successful, lead to a nonoptimal method of provision. It would have been preferable if the existing operator had been initially encouraged to adopt a more socially desirable output and price combination.

The second case is where the innovation is commercially viable, and operating exclusively would be welfare-superior to the optimal provision by the existing service. In these circumstances, it is desirable that the innovated service, at least partially, replace the existing one. However, competition on the road might lead to the innovation not coming to fruition (because of the financial dominance of the existing operators), or, even if successful, the competition during the innovation's introduction is likely to be wasteful in social welfare terms.

In conclusion, it would appear that unfettered competition is not an effective sorting device for use in selecting the most beneficial innovations in the bus industry.

Safety

In the 1920s, unruly competitive driving practices and suspect maintenance initiated public interest in regulation. Although vehicle engineering and general road safety have improved considerably in the intervening period, the prospect of renewed competition has provoked many safety concerns.

Safety concerns can be divided into two aspects. The first is road safety, about which the author concludes that there is a possibility of unruly driving practices as a result of competition on the road. This arises from the polarizing of timings attributed to headrunning, which leads to racing and blocking of stops.

The second aspect is the quality of operators. A comparison of small operators, who might constitute the entrants to stage operation, and large network operators, indicates that there are no grounds for believing that there is any difference in accident rates (22).

However, a survey by the author (23) indicated that the smaller firms tend to have a much higher number of faults on their vehicles. The survey related to the Yorkshire area for 1983. The number of faults (officially graded as "prohibitions" and "notices of defects") detected by government examiners was tabulated by operator fleet size. The analysis is given in Table 1.

TABLE 1 Prohibitions and Defects per Million Vehicle Kilometers Tabulated by Operator Fleet Size

Fleet Size	Number of Operators	Faults
1	131	6.5
2	63	5.6
3	61	6.4
4	54	3.4
5	29	1.4
6–9	71	2.1
10–14	35	2.5
15–19	16	1.7
20–49	7	3.1
50+	11	0.7

The figures indicate that a typical 1-vehicle firm has more than 9 times as many faults per vehicle-kilometer as compared with a large operator, while a comparable figure for a 10- to 14-vehicle fleet operator is about 3.5 times as many as the large operator. What becomes clear is that there is a continual (and statistically significant) decline in the number of faults as fleet size increases.

Because the public cannot readily determine the quality of operators, drivers, and vehicles, there would appear to be no case for lessening the quality regulatory controls. Indeed, if a change in market regime leads to smaller operators undertaking stage work, there would be a case for more vigilance on the part of regulators. This would particularly be the case when fierce competition reduces financial returns to operators, who may then be forced to make economies in their maintenance.

The United Kingdom solution, which inherently encouraged direct competition on the road, does not appear to be the most optimal way of dealing with the disadvantages of monopoly. Thus, liberalization, or total removal, of licensing does not provide the answer. Unfettered competition has the following serious disadvantages:

- Direct competition on the road is likely to lead to a short-run social welfare loss on the route, as consumer benefits are outweighed by the additional resource costs; in addition, oligopolistic competition does not necessarily produce a long-run optimum resource allocation;

- Some jointness of demand may be broken and thus endanger services (i.e., feeder routes) commercially justified as a result of contributory revenues;
- Financial dominance of existing operators may impede the introduction of beneficial innovation;
- Some local economies of scope may be lost;
- Some nonbeneficial innovation might be introduced and could, if successful, lead to a nonoptimal service provision;
- Chance of unruly driving practice is increased;
- Integration between services may be lost and public goodwill may be endangered by a bad operator; and
- Uncertainty may arise.

In addition, the existence of artificial monopolies means that competition is unlikely in some parts of the present system and, thus, the full competition stimulus may not be felt.

REQUIREMENTS FOR AN OPTIMAL MARKET STRUCTURE

Given the problems of unfettered competition, a more nearly optimal solution to the disadvantages of the existing monopoly has to be found. In this section, the features of an ideal market regime are identified.

Direct Competition to be Avoided

The disadvantages of competition on the road, particularly the short-run welfare losses, the dangers from unruly driving practices, and the possible introduction of nonbeneficial innovation, indicate that a route monopoly system would be preferable.

No Priority System

The problem with route monopolies is how to allow for a control of costs, and also ensure that the monopolist maintains socially efficient fares, frequencies, method of operation, and reliability. Recent work (24,25) has indicated that the threat of potential competition can be as effective as actual competition in achieving these objectives. The problem in this industry is how to make the threat of competition real, yet preserve route monopolies. The solution would appear to be that any route monopoly should not be for perpetuity, as has been the case since 1930, but should be renewable after a certain period of time.

A system would have to be devised to decide between rival operators when route monopolies come up for renewal. Some options are (a) where a controlling authority sets socially optimal fares and frequencies and invites tenders on the basis of cost (known as "contracting"), and (b) where firms tender a proposed cost-fare-frequency combination, from which the controlling authority chooses the most optimal (known as "franchising"). Mackie (26) describes both of these systems. The optimal length of the contract-franchise would have to be determined with regard to the depreciation of capital (the most important being vehicles) to make bus operation attractive to operators.

This system will have the desired effects in that the competing tenders for the franchise and the determination of the contract terms will influence operators to act in a socially efficient way. This may include innovative routes and methods of operation, and the introduction of low-cost operators attributed to the implicit cost competition in the tendering process. In addition, a short-period, contract-franchise system will mean that the threat of potential competition, when the routes are next put up for tender, will encourage monopoly incumbents to maintain efficient management objectives, be reliable in operation, and also control X-efficiency. However, it may be necessary to word the contract-franchise in such a way (i.e., inflation-linked cost allowances) so as to maintain pressure on costs during its currency.

Recognition of Demand- and Supply-Side Links

Peacock and Rowley (27) present a solution to the problem of service tendering while preserving the benefits of natural monopolies or demand-side links, or both. They argue that localized groups of services, rather than individual services, should be the unit by which bus operations are put out for tender.

Unremunerative Services

It would be possible to put both profitable and unprofitable activities out to tender. In the latter case, routes would be tendered and evaluated on the basis of fares, frequencies, and the amount of revenue support required. This would mitigate against artificial monopolies, which would otherwise preclude competition on much of the present network.

Controlling Authority

There would need to be a controlling authority which, in addition to unbiasedly administering the contracting and franchising system, could also maintain goodwill and request through-fares and other integration policies. As a result of the need to make revenue support available for unremunerative activities (e.g., integration or other policies), the body to undertake this work would preferably have to be directly publicly accountable and able to raise public finance.

An additional task for a controlling authority, especially if a competitive stage-carriage market leads to more smaller operators, is to monitor the quality of operators, vehicles, and drivers. This need not necessarily be conducted by the contracting and franchising authority previously described, although safety considerations must be input to the outcome of a tendering exercise. At present, the government-appointed regional traffic commissioners do undertake such duties in the bus and coach market. Because local services are a minority of total coaching operations, it might be sensible to leave quality regulation of operators in their hands.

CONCLUSIONS AND POLICY PRESCRIPTION

A market regime has to be found that would give the benefits of competitive stimuli without the disadvantages of direct competition. Baumol (25) and others have argued that the benefits of competition can accrue from potential and not actual competition. He stated, "The heroes are the (unidentified) potential entrants who exercise discipline over the incumbents."

The institutional problem is how to make the threat of potential entry effective (i.e., have low barriers to entry), but avoid direct competition. It would thus appear that in this industry, the optimal solution is competition for the market rather than competition in it. This would suggest that a system of competitive contracting or franchising of services should result.

This will bring the benefits of competitive stimuli, while avoiding the problems of the wastes of direct competition and the danger to the public posed by unruly driving practices. In addition, the authorities can monitor goodwill and safety standards, and request through-fares or other integration policies. The benefits of demand-side links, or localized economies of scope, can be realized, if necessary, by the controlling authority putting out to tender groups of, rather than individual, services. A competitive atmosphere can also be encouraged across all the network, by the controlling authority offering unremunerative services on a "negative tender" system, whereby services are allocated to operators on the basis of who requires the least subsidy.

ACKNOWLEDGMENT

Represented in this paper are the conclusions of a Science and Engineering Research Council (SERC)-funded Ph.D. dissertation entitled "Economic Evaluation of Competition on Stage Carriage Bus Routes." It was undertaken between October 1981 and September 1984, at the Institute for Transport Studies, University of Leeds, under the supervision of Professors K.M. Gwilliam and P.J. Mackie. The thesis was published in November 1985 under the title "Deregulation of Bus Services" by the Gower Publishing Company. The author would like to thank the SERC, his supervisors, and all those who made the project possible. The paper has been enriched by contract work performed as a consultant for Booz-Allen and Hamilton during the past year.

REFERENCES

1. H. Leibenstein. Allocative Efficiency vs. "X-Efficiency." American Economic Review, Vol. 56, No. 3, Aug. 1969, pp. 392-415.
2. Productivity Agreements in the Bus Industry. Report 50, Command 3498. National Board for Prices and Incomes, Her Majesty's Stationery Office, London, England, 1967.
3. R.J. Tunbridge and R.L. Jackson. The Economics of Stage Carriage Bus Operation by Private Bus and Coach Companies. Laboratory Report 952. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1980.
4. Guildford/Cranleigh Bus Study. MVA Consultancy, Plymouth, Devonshire, England, 1984.
5. G.J. Ponsonby. Transport Policy: Co-ordination through Competition. Hobart Paper 49. Institute for Economic Affairs, London, England, 1969.
6. J. Hibbs. Transport Without Politics. Hobart Paper 95. Institute for Economic Affairs, London, England, 1982.
7. Cross-Subsidy in Urban Bus Operation. Institute for Transport Studies, University of Leeds; National Bus Company, London, England, 1984.
8. The Impact of the "Buses" White Paper on Plymouth. MVA Consultancy, Plymouth, Devonshire, England, 1985.
9. The Budgetary Consequences for the [Surrey] County Council of the Transport Bill. Booz-Allen and Hamilton, London, England, 1985.
10. K.M. Gwilliam. Aims and Effects of Public Financial Support for Passenger Transport. 67th Round Table, European Conference of Ministers of Transport, Paris, France, 1984.
11. E.K. Browning. The Marginal Cost of Public Funds. Journal of Political Economy, Vol. 84, No. 2, pp. 283-298.
12. K.J. Arrow. Economic Welfare and the Allocation of Resources for Invention. In The Rate and Direction of Inventive Activity: Economic and Social Factors, National Bureau of Economic Research, Princeton University Press, Princeton, New Jersey, 1962, pp. 609-626.
13. H. Demsetz. Information and Efficiency: Another Viewpoint. Journal of Law and Economics, Vol. 12, No. 1, April 1969, pp. 1-22.
14. N. Lee and I. Steedman. Economics of Scale in Bus Transport. Journal of Transport Economics and Policy, Vol. 4, No. 1, Jan. 1970, pp. 15-28.
15. H. Hotelling. Stability in Competition. Economic Journal, Vol. 39, No. 1, pp. 41-57.
16. I.P. Savage. Deregulation of Bus Services. Gower Publishing Company, Limited, Aldershot, Hampshire, England, 1985.
17. C.A. Nash. Management Objectives, Fares and Service Levels in Bus Transport. Journal of Transport Economics and Policy, Vol. 12, No. 1, Jan. 1978, pp. 70-85.
18. S. Glaister and J.J. Collings. Maximisation of Passenger Miles in Theory and Practice. Journal of Transport Economics and Policy, Vol. 12, No. 3, Sept. 1978, pp. 304-321.
19. W.J. Baumol. On the Proper Cost Tests for Natural Monopoly in a Multiproduct Industry. American Economic Review, Vol. 67, No. 4, Dec. 1977, pp. 809-822.
20. J.C. Panzar and R.D. Willig. Economies of Scope. Papers and Proceedings, American Economic Review, Vol. 71, No. 2, March 1981, pp. 268-272.
21. F.V. Webster and P.H. Bly. The Demand for Public Transport--Report of the International Collaborative Study of the Factors Affecting Public Transport Patronage. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1980.
22. Full Time/Part Time PSV Driver Accident Study: Complete Report. Department of Transport, Her Majesty's Stationery Office, London, England, 1984.
23. I.P. Savage. Safety in the Deregulated Bus Industry. Traffic Engineering and Control, Vol. 25, No. 11, Nov. 1984, pp. 564-565.
24. W.J. Baumol, E.E. Bailey, and R.D. Willig. Weak Invisible Hand Theorems on the Sustainability of Multiproduct Natural Monopolies. American Economic Review, Vol. 67, No. 3, Jan. 1977, pp. 350-365.
25. W.J. Baumol. Contestable Markets: An Uprising in the Theory of Industrial Structure. American Economic Review, Vol. 72, No. 1, March 1982, pp. 1-15.
26. P.J. Mackie. Competition in the Bus Industry. Public Money, Vol. 3, No. 1, pp. 45-48.
27. A.T. Peacock and C.K. Rowley. Welfare Economics and the Public Regulation of Natural Monopoly. Journal of Public Economics, Vol. 1, No. 2, Feb. 1973, pp. 227-244.

Publication of this paper sponsored by Committee on Public Transportation Planning and Development.

Transportation Planning Process: The Case of the Chicago Region

CLAIRE E. McKNIGHT, ED J. CHRISTOPHER, and DAVID A. ZAVATTERO

ABSTRACT

In the 6-county northeastern Illinois region, efforts to involve the private transportation sector in the public transportation planning process have advanced further than in most other areas. The formation of the Metropolitan Transportation Association, the purpose of which is to promote private sector involvement in the planning process, is a unique event that has led to this region being in the forefront of this issue. Presented in this paper are the historical background of transportation in the region and a description of the evolution of private involvement in the planning process. Experience in the Chicago region indicates that there are several issues that need to be resolved, such as the role that private operators should have in the process, the organization of the private operators to ensure balance and equitable representation, and the organization of efforts to fund the private operators. The paper concludes with recommendations for other regions attempting or contemplating a public or private cooperative planning process.

There has been great interest in involving the private transportation sector in public transportation. Most of the emphasis has been on the private operation of public transportation, but there has also been a federal-level policy directive (1) and two recent Urban Mass Transportation Administration (UMTA) decisions (referred to in a letter from J.P. Ettinger of UMTA to a grantee on July 19, 1985) stressing the importance of private sector input to the transportation planning process. In the 6-county northeastern Illinois region [i.e., the Chicago standard metropolitan statistical area (SMSA)], efforts to involve the private sector in transportation planning have advanced further than in most other areas. Thus, a review of the efforts in this region can benefit other areas by presenting the prospects and problems of such involvement.

Reviewed in this paper is the involvement of the private transportation operators in the planning process for northeastern Illinois. In the first section, some important background material is presented including the history of public transportation in the area, followed by a description of the private sector operators and an outline of the transportation planning process as it has evolved in the region. This review shows that private transportation operators have been involved in mass transportation, public and private, throughout the area's history. However, the private operators involved in the early stages of the process were conventional mass transit firms (i.e., the operators of rail and fixed-route bus systems). Over the last century, these firms have been closely regulated, subsidized, and sometimes purchased outright by the public sector. Even when still privately owned or operated they have been closely identified with the public sector.

The discussion in this paper is focused on a different group of private transportation providers: the taxi firms, charter bus operators, and limousine companies. In the past, this group has tended to minimize its involvement with the public sector. In the last few years, this attitude has been reversed somewhat. The expansion of public transportation agencies into such areas as paratransit and special services has placed the public sector in competition with this group of private operators. As a result, these operators have viewed it as increasingly important that they, too, have input into the transportation planning process. It has been suggested that greater use of such private operators could improve transit services at a lower public cost (2).

Following the background information is a section that describes the organization of this "new" private sector and its efforts to become involved in the Chicago region's planning process. Contained in the next section is a description of the private sector's participation in the planning process and the activities associated with its involvement to date. Also included is a discussion on the issues that are yet to be addressed and the future prospects of private sector involvement. Finally, based on this experience, some recommendations and guidelines are offered for other regions integrating private operators into the transportation planning process.

HISTORY OF PUBLIC TRANSPORTATION IN THE REGION

Discussions concerning public responsibility for urban mass transportation in Chicago started in the mid-19th century. Still, the private sector built and operated in the street railways until the mid-20th century. By the 1930s, however, all the surface lines and elevated companies were in receivership (3). The public sector assumed the responsibility for operating mass transportation in Chicago with the formation of the Chicago Transit Authority (CTA) in 1945. Between 1947 and 1952, the CTA acquired all transit services in the city except commuter rail and formed a unified transit system throughout Chicago and its adjacent suburbs. The CTA has improved the system over the years, adding several rail lines and adjusting bus routes to reflect changes in travel patterns.

C.E. McKnight, Urban Transportation Center, University of Illinois, Chicago, Ill. 60680. E.J. Christopher and D.A. Zavattero, Chicago Area Transportation Study, 300 W. Adams Street, Chicago, Ill. 60606.

In the 1960s and early 1970s, the CTA began to experience financial difficulties as costs outstripped revenues. These problems eventually led to the formation of a regional organization with a greater capacity to solve them. The Northeastern Illinois Regional Transportation Authority (RTA) was formed in 1974 to plan, coordinate, and fund mass transportation in the six counties of the Chicago SMSA (see Figure 1). The RTA was granted taxing power and responsibility for setting fares, planning a coordinated transit system, and allocating subsidies among the actual operators. These operators included the CTA, several suburban bus operators, many of which were privately owned, and eight commuter rail companies, all of which were privately owned.

To solve the continued financial problems that resulted in a dramatic fare increase in 1981 and equally dramatic ridership losses, the RTA was reorganized in 1983. The new organization consists of an oversight board and three service boards or operating divisions: METRA, the Commuter Rail Division, became responsible for the commuter rail lines; Pace, the Suburban Bus Division, was made responsible for the bus and paratransit operations in suburban Cook County as well as in the five collar counties; and the CTA continued its responsibility for the system it operated before the reorganization of the RTA. All four entities (RTA, METRA, Pace, and CTA) have their own policy boards. Under the reorganization, the RTA's role is primarily to allocate subsidies and provide financial oversight of the three operating divisions.

Although the RTA, through the service boards, operates the majority of transit services, there has been a history of providing some of this service with private contractors. METRA has purchase-of-service agreements with five of the original rail operators, and Pace contracts out feeder bus routes, paratransit service, and some local bus systems. The CTA recently signed contracts with four private operators to pro-

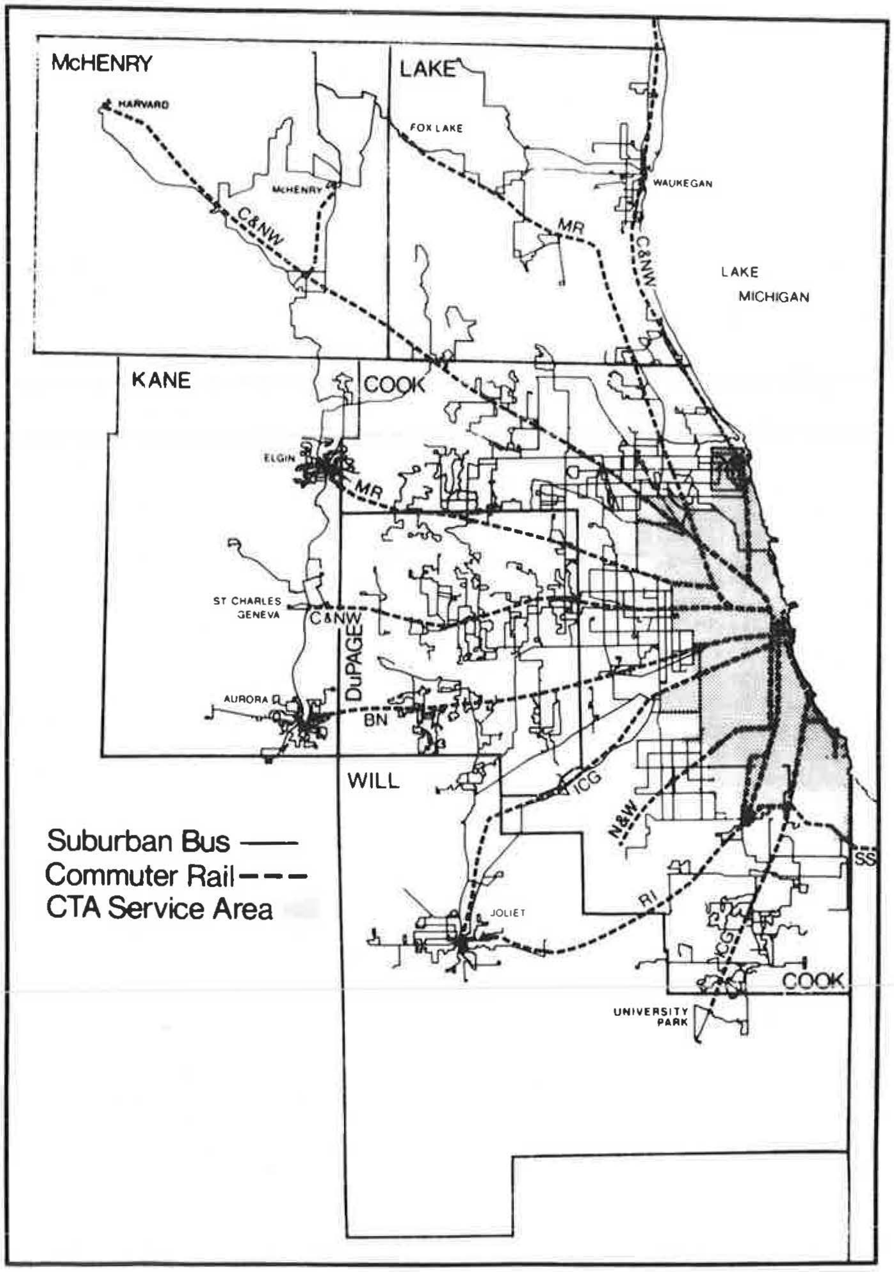

FIGURE 1 Chicago SMSA and transit services.

vide demand-responsive service for disabled individuals unable to access CTA main-line fixed-route service.

PRIVATE SECTOR TRANSPORTATION IN THE CHICAGO REGION

In the Chicago metropolitan area, the private transportation sector includes at least 150 taxi companies, 210 limousine companies, and 170 bus companies. Added to these is a large but less well-defined group of companies that offers a variety of services such as dial-a-ride and vanpooling. Finally,there are a number of organizations representing private sector labor including taxi and bus drivers. When taken together, these diverse groups, companies, and labor organizations make up the private sector. (Note that the best available estimate places the region's private fleet at over 8,000 vehicles, which contrasts sharply with about 3,000 buses and 1,800 rail cars operated by the three public operators.)

Private operators offer an array of services ranging from completely private service arrangements to contractual agreements with the public sector. For example, private taxi operators may be organized as associations, cooperatives, or closely held companies. Some operate a strictly street-hail business, while others offer telephone dispatch services or some combination of both. Common taxi driver arrangements include commission drivers, lease drivers, and owner-operators.

The livery and the bus sectors of the industry are also characterized by many companies offering a variety of services and working in all types of operating environments. The limousine industry is a significant resource in the Chicago region in terms of the number of companies, vehicles operated, and passengers carried. In addition to conventional livery service, the industry provides a major link between O'Hare Airport, which is located in Chicago, and the rapidly growing suburban markets, which are generating an increasing number of air passengers.

THE CHICAGO AREA TRANSPORTATION PLANNING PROCESS

Several agencies plan transportation in the Chicago region. Each of the three operating divisions of the RTA undertakes strategic and operations planning for their service areas, while the RTA plans for a coordinated transit system. The Chicago Department of Public Works, responsible for maintaining the city's streets and highways and building its rail guideways, is involved in transportation planning within the city limits. The Illinois Department of Transportation, the Northeastern Illinois Planning Commission, the Illinois Toll Highway Authority, and the planning and transportation departments for the individual counties and municipalities all do some transportation planning. When taken together, there are over 300 public entities responsible for different elements of the region's master plan.

These planning efforts are coordinated through the Chicago Area Transportation Study (CATS), which is the metropolitan planning organization (MPO) for the region. The MPO is the forum through which all the decision makers responsible for public transportation investments and operations cooperatively decide on mutually acceptable transportation plans and programs for the region. It is through this process that federally funded transportation programs are planned and implemented. As the MPO, CATS has the lead responsibility for preparing and endorsing the region's Long Range Transportation Plan, the Five-year Transportation Improvement Program, the Annual Element for Transportation Investment, implementation studies in support of engineering and construction activities, and the Unified Work Program for transportation planning.

CATS was established late in 1955 by the city of Chicago, Cook County, and the State of Illinois, in cooperation with the U.S. Bureau of Public Roads (later reorganized into the FHWA). Originally, CATS was financed by these four sponsoring governmental entities and acted toward them in an advisory capacity. CATS's purpose was to develop a unified transportation plan for the metropolitan area with a 1980 target year.

It was evident to the sponsors of CATS as early as 1957 that an expanding urban area with a shifting population required continuous changes in the transportation planning and policy process. With this in mind, the sponsors provided for a permanent planning agency capable of updating the region's plans while acting as planning staff to the supporting agencies and other local government entities. Once the original transportation plan was published in 1962 (7), CATS was established as a continuing agency for planning an coordinating the region's transportation system. (Chicago was the first region to receive certification for meeting the "c3" requirements.)

CATS is an unusual organization in that its direction is set by several committees (similar to boards of directors) representing other agencies and interest groups. Heading up the committee structure is a policy committee (PC). The first PC was comprised of representatives of the four sponsoring governments with the Executive Director of CATS acting as secretary. Since then, the PC has been expanded from 4 to 20 voting members with the Executive Director of CATS acting as secretary. Figure 2 shows the expansion of the membership of the PC from its formation in 1955 to the present. [Note that (a) the U.S. Bureau of Public Roads was reorganized into the FHWA in 1955 under the U.S. Department of Transportation; (b) Commuter Railroads were replaced by the Commuter Rail Division in 1984 after the reorganization of the RTA; and (c) Suburban Bus Operators were replaced by the Suburban Bus Division in 1984 after the reorganization of the RTA.]

Assisting the PC and providing day-to-day guidance of the various agencies responsible for planning are the responsibilities of the Work Program Committee (WPC). Over the years, the WPC has developed into a cooperative planning and programming process in which all the agencies with transportation planning responsibilities are involved. Disputes and inconsistencies between agencies and local jurisdictions are reviewed at this level. The WPC is currently composed of one member from each of the PC agencies plus representatives of the following agencies:

- The Chicago Department of Development and Planning
- The Northwestern Indiana Regional Planning Commission,
- The Division of Public Transportation, Illinois Department of Transportation,
- The Illinois Environmental Protection Agency,
- The Illinois Department of Transportation--District 1,
- The Illinois-Indiana Bi-State Commission, and
- The Chicago Area Transportation Study.

There is also a Unified Work Program (UWP) Committee, which annually recommends to the PC and WCP the transportation-related planning activities to be performed in the region. The UWP is mandated as part of the MPO regulations (8). Members of the UWP Com-

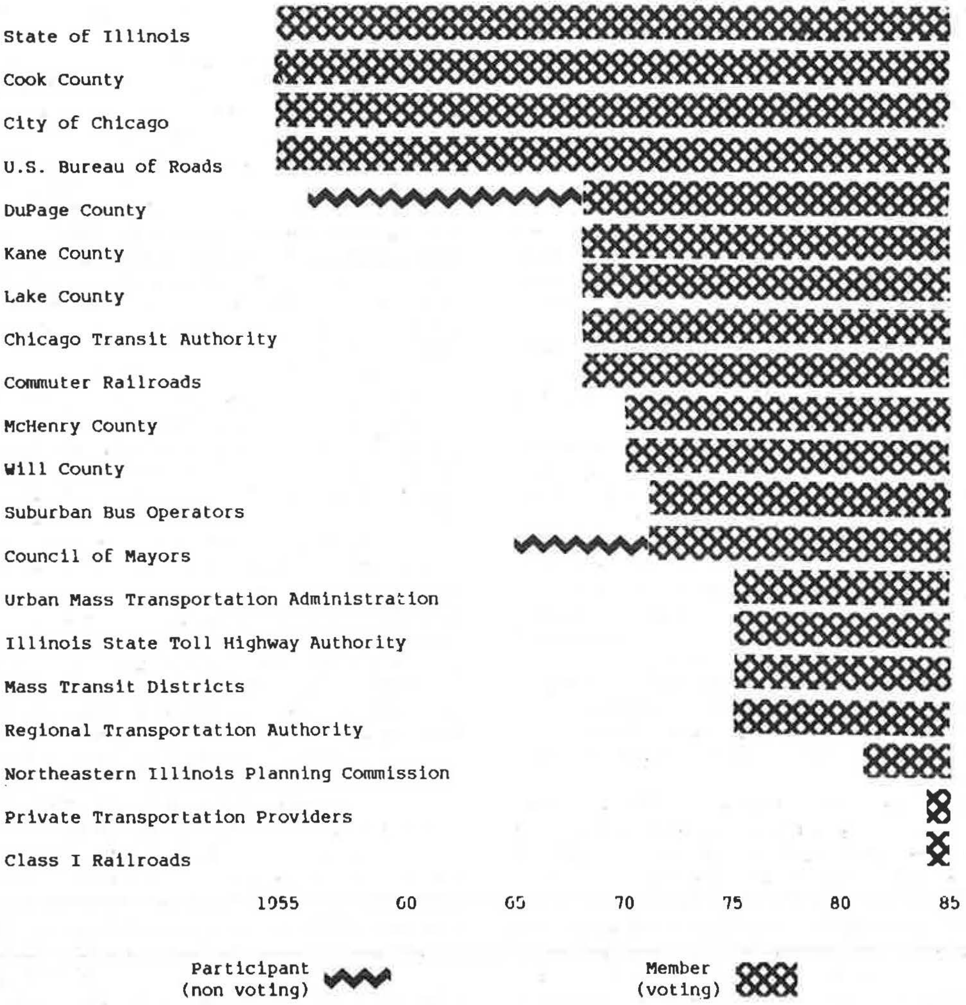

FIGURE 2 History of membership on the CATS PC.

mittee include one representative from each of the following agencies:

- The City of Chicago,
- The Regional Transportation Authority,
- The Chicago Transit Authority,
- The Council of Mayors,
- The collar counties (the five outlying counties of the SMSA),
- The Illinois Department of Transportation, and
- The Northeastern Illinois Planning Commission.

Traditionally, the seats on each of these committees have been held by staff from the member agencies responsible for highway and transit planning and programming. The WPC is supported by a series of standing advisory committees representing a broad spectrum of interests. Through the advisory committees, the planners, local mayors, operators, implementers, special interest groups, and general public provide the input needed to shape a coordinated transportation system. Although most of the advisory committees grew out of specific concerns and interests, many have continued as an ongoing resource, ensuring that decisions are made in a climate of full participation. The current six advisory committees to the WPC are

- Aviation,
- Freight Advisory,
- Transportation Operations,
- Air Quality Advisory,
- Mobility Limited Advisory, and
- Private Providers.

In addition to these advisory committees, several subcommittees and task forces have been assembled whose purposes are more short-term in nature. These groups are formed as needed to address specific transportation issues. The Private Providers Committee (PPC) was originally formed as such a group. However, it became apparent from its first year's work that continuous participation by private operators was desirable. Thus, the PPC has become a permanent advisory committee.

FORMATION OF THE METROPOLITAN TRANSPORTATION ASSOCIATION

The Metropolitan Transportation Association (MTA) is an organization of private transportation providers established to coordinate private sector activities and to inform the public sector about private transportation issues. The events leading to the organization of the private transportation operators in the Chicago region started when the owners of two private carriers, a taxi company and a bus company, approached the same consultant for advice on dealing with the public sector. Recognizing that the two businessmen had mutual interests, the consultant suggested that they meet and discuss these interests over lunch. Because they found their discussions worthwhile, other lunches were organized involving additional private operators. Over the following year, several informal lunch meetings were held, with six to eight carriers represented. At first, there was no specific objective for these meetings. However, the businessmen soon discovered that they had common problems and concerns.

The individual owners had differing motives for being interested in the public sector. Some owners felt that the public transportation agencies were encroaching on markets that had traditionally been served by the private sector. For instance, publicly owned dial-a-ride services were eroding the taxi business in several suburbs. Other owners were looking for new markets. A few of the owners had less self-interested motives. They were concerned about how the public sector provided, operated, and, in particular, paid for public transit service. The media gave extensive coverage to transportation problems including the financial crisis leading to the 1983 reorganization of the RTA. As a result, much of the general public believed that the public transportation agencies were mismanaged or overpoliticized, or both. Several of the owners and managers of the private transportation firms believed that they could provide better transportation at a lower cost.

Eventually, the informational lunches led to the creation of a formal organization. The Metropolitan Transportation Association (MTA) was incorporated in July 1982. Its purposes were to increase the public sector's awareness of the private sector and to provide information to the public sector on the private sector's capabilities. MTA dues are $100 per year for an associate member and $1,000 per year to become a director with voting privileges. The MTA started with 14 members and, in the last 2 years, has expanded to 20, 11 of whom are directors.

During 1983, because some members of MTA wanted to take a more active role in soliciting public contracts, they formed a second group called the Transit Service Corporation (TSC). TSC has eight members, all of whom are also members of the MTA. TSC's first project was to do a comparative cost study of several CTA bus routes showing the savings that the CTA could achieve by contracting for service rather than operating its own service (9). This study led to a TSC proposal to operate CTA's special services for the disabled. Since that time, CTA has signed contracts with four private operators, including TSC, to operate its special services.

In the meantime, MTA has continued as an informational organization. In April 1983, the MTA gave testimony before the Illinois House Transportation Committee concerning the private sector's role in the transportation planning process. The following April, MTA made a presentation at a locally sponsored forum on the same topic. In October 1984, an MTA member was a speaker at an UMTA-sponsored conference on private and public involvement. In November, the governor of Illinois appointed an MTA member to a task force on transportation for individuals with disabilities. Because MTA has acted since its formation as a unified voice for the private transportation operators, it has received the attention and respect of the public sector.

THE EVOLUTION OF PRIVATE SECTOR PARTICIPATION

In March 1983, the MTA petitioned the CATS PC for membership. The By-Laws Subcommittee of the PC discussed the appropriateness of the MTA being a member of the PC, possible duplication with other interests, and the overall makeup of the PC. Members suggested that if the MTA was to be considered an appropriate member, then other private interest groups, such as the Illinois Road Builders or the Chicago Motor Club, should also be considered. It was concluded that because the MTA represented service operators while the other organizations represented support to service operators, the MTA should be considered for membership. The Chicago Motor Club and other consumer groups were felt to have input to the process through the Regional Council of Mayors, which represents the general public on the PC.

The By-Laws Subcommittee agreed that there was a need for input from private providers at the advisory level and recommended that a meeting be convened of the private for-profit and not-for-profit nonrail transit carriers. The purpose of the meeting would be to explain the transportation planning process and the role of CATS as the MPO. The attendees would be asked to appoint representatives to the Mobility Limited Advisory Committee and the Transportation Operations Committee. Through these committees, the private providers would have input to the process. The appointed representatives would report back to the other private providers at quarterly meetings if desired. The Subcommittee felt this level of involvement would be appropriate for a 1-year trial period. After 1 year, private provider input and participation would be evaluated to see if further involvement, such as WPC membership, would be warranted. This approach is commonly used for agencies seeking membership on the PC (i.e., they are welcomed as observers for 1 year before becoming full voting members).

In May 1983, the PC discussed the Subcommittee's recommendations. One member supported putting the MTA on the WPC as well as the advisory committees. Other PC members were concerned that it would be inappropriate to put the private operators on a committee that they perceived as a governmental body. The PC finally voted to accept the Subcommittee recommendations, giving the private operators an advisory role, but not a vote, on the policy and planning decisions. It was also decided that representatives to the advisory committees should be chosen at a meeting of all private operators. Although MTA had initiated the effort to get private input into the planning process, the response of the CATS PC was to deal with all operators rather than a single association.

In September 1983, CATS convened the first meeting of the private transportation providers. The invitation list was assembled from regional telephone books. Invitations were sent to over 480 carries from the bus, taxi, and limousine industries. About 50 people from private transportation firms attended the meeting at which representatives from CATS and RTA made presentations. CATS staff explained the transportation planning process and a model taxi ordinance that they had developed. The RTA representatives explained how new service was planned, RTA policy on using private carriers, and the process for bidding fixed-route and paratransit services.

As the discussion progressed, it became apparent that many of the private operators felt that they still were not connected to the public planning process. They raised several concerns about public decisions and programs that were being implemented and the impact of these decisions on their businesses. One person questioned whether the transportation planners really knew what problems faced the private operators. Many expressed the opinion that they should have some formal link to the planning process.

To establish this link, a CATS representative proposed that the private operators begin functioning as an advisory committee within the planning process.

As a starting point, three activities were proposed as follows:

1. A Private Providers Steering Committee (the Steering Committee) to shape the direction and activity of the full committee. The full committee would continue to be composed of all the private operators.
2. Individual semi-annual meetings of all the operators and the full Steering Committee.
3. A quarterly newsletter speaking to and representing private carriers. The newsletter would provide a forum through which information affecting the carriers could routinely be disseminated. Suggested items for the newsletter were articles covering the RTA's solicitation of bids, legislative changes, and other issues of relevance to the carriers. A bulletin has since been added for items of interest that cannot wait for a quarterly publication.

It was clear from the first and subsequent meetings that there were many issues facing the private operators. Bus and taxi firms were concerned about the opportunity to operate public systems through contractual agreements, municipal regulations, government interference, and public sector competition. The livery firms had two overwhelming concerns: traffic congestion and regulations at O'Hare Airport, and conflicting municipal regulations. Initially, this division of interests led to the objection that the discussions and meetings only held the attention of a few operators.

At its first meeting in November 1983, the Steering Committee discussed the nature of the involvement that they, as private operators, should have in the planning process. In looking for some guidance or precedents, CATS staff reviewed federal laws and regulations and contacted MPOs in other regions. They found that published federal policies encourage the involvement of private operators in the process but leave the structure of this involvement to local officials. A telephone survey of nine MPOs discovered no private involvement at the policy level or on a permanent basis, although several regions had organized task forces involving private operators to address specific issues.

The discussions at the first meeting of the Steering Committee ranged over many topics, including the following:

- The nature of the influence that the Steering Committee might have in the planning process;
- Problems inherent in allowing local units of government to make spending decisions that are not cost-effective;
- Private operator avoidance of the mass transit market because of the bureaucratic requirements and red tape;
- Organization of the MTA and its efforts to influence spending policies given its ability to provide service at a substantial savings to the public;
- The interface between the Steering Committee and the MTA;
- Educating the policy makers and implementers to the problems of the limousine industry;
- Reducing the expenditure of tax dollars by using more private carriers;
- Lack of understanding of the issues facing private companies by the policy makers;
- Enforcement of the rules and regulations that are currently in existence at O'Hare Airport; and
- The need to operate by the dictates of a few government planners who do not really know what the issues or industries are all about.

In December 1983, the first issue of a newsletter entitled <u>Transit Dispatch</u> was published by CATS staff after review by the Steering Committee. The first issue and later ones present three types of articles: reports on the meetings and other actions of the Steering Committee; articles on issues of general interest to the private operators, such as explanations of the structure and purpose of the various public agencies or current legislation; and announcements of specific public transportation contracts that are up for bid or that are being considered for future letting.

Following its initial meeting, the Steering Committee has met on a monthly basis. Much of the agenda is informational. In the summer of 1984, the issue of private-operator representation on the PC was renewed. The reorganization of the RTA required that the membership of the PC be reviewed. This review offered an opportunity for reconsideration of a PC seat for the private operators. Several changes in the PC membership were made as a result of this review. First, the Suburban Bus Division of the RTA (Pace) took over the seat previously held by the suburban bus operators (most of whom had been taken over by the public sector). Second, the Commuter Rail Division (METRA) assumed the seat that had been held by the commuter railroads. Third, the Class I railroads, which retain ownership of the right-of-way used by METRA carriers, were allocated a seat. Finally, the private transportation providers received a seat on both the PC and the WPC, which included full voting privileges.

Although the PC could not implement these changes until its December meeting, the Steering Committee drew up a slate of representatives for both the PC and the WPC. The Steering Committee chose representatives who were familiar with CATS and the planning process in order to reduce the need for an educational period. It was recommended by the Steering Committee that the representatives be capable of putting aside the interests of their own firms and their specific industry (e.g., taxi, livery) in favor of the general interests of the private sector.

The slate chosen by the Steering Committee was presented at the semi-annual meeting of the Private Providers Committee held in November 1984. One operator objected that Chicago taxi drivers were not represented, indicating that at least some of the operators still had a sectarian attitude toward the private sector representation. It was pointed out that several drivers had been invited to participate on the Steering Committee, but none had responded. The slate was accepted by acclamation. The taxi drivers' groups were again invited to participate on the Steering Committee. Since that meeting, there has generally been more active participation by this segment of the private sector.

In December 1984, the private provider representatives assumed their positions on the respective committees. Reporting back to the Steering Committee, they noted the complexity of the issues being discussed and that the planning process covered a broader range of issues than was initially perceived.

In the spring of 1985, several changes were made to the structure of the Private Providers Committee and the Steering Committee. Initially, CATS staff had acted as chair and guided the development of the Steering Committee. Although much progress had been made, the CATS representative suggested that this organization was inappropriate, and the committee agreed. A chair and co-chair were elected from the operators, and a fixed schedule of monthly meetings was established. It was felt that this would strengthen recognition of the committee as a legitimate part of the planning process and provide a focal point for private representation within the industry.

SOME EARLY OUTCOMES OF PRIVATE INVOLVEMENT

The development of private participation in the planning process has been gradual, while the benefits of such participation are most likely to occur several years hence. Thus, a discussion of outcomes at this stage is preliminary. The major benefit has been that of educating both the public and private sectors about each other. More specific outcomes have been the provision of a forum for the operators to express their concerns, further organization of the private sector, and the expansion of private contracting options. These are briefly discussed.

The meetings of the private operators through the Private Providers Committee and the Steering Committee have provided an opportunity for a wide range of participants to discuss issues of broad interest. Many of these issues are complex and often difficult to identify. One product of the meetings has been a concerted effort by the private operators, with support from the CATS staff, to have a voice in the decisions concerning ground traffic at O'Hare Airport.

In May 1984, the Chicago Department of Aviation (DOA) developed a ground plan and a related city ordinance for the airport. Although the Chicago City Council held public hearings on the plan and the ordinance, the operators believed that their input had not been seriously considered or used by the planners in developing the program. The private transportation operators, particularly the liveries, felt that the DOA did not understand the problems of ground access and was ignoring the expertise and needs of the carriers. Interestingly, these accusations are similar to those made previously against the regional transportation planning agencies. Through the forum provided by the Private Providers Committee, discussions involving operators (many of whom operate competing modes at the airport) were undertaken on a neutral ground.

As the private providers were struggling to have a collective voice on the airport issue, a new Commissioner of Aviation was appointed. The new Commissioner had been a member of the CATS PC and was familiar with the committee planning process. One of his first actions was to establish a task force of all city departments and other interested individuals involved in ground transportation at the airport. This task force has worked closely with the Steering Committee and has been receptive to the private operators' concerns.

The second outcome of private sector participation in the planning process is the continuing organization of the private operators. As the private sector representatives attend PC meetings, they become more aware of the complexity of the issues being addressed and the need for specialized knowledge if they are to have meaningful involvement in setting policy and contributing to the plans being considered. Because of the time required to attend the requisite meetings, it is costly for any single company to participate in the planning process. Therefore, in November 1984, MTA presented a proposal to the UWP Committee to be funded by the region's UWP. Initially, this proposal called for public funding of staff for the MTA. The MTA felt that support for a staff person would allow better participation. As stated by a representative of MTA, the intent of the proposal was to ensure private input and guarantee coordination.

The MTA request for UWP funding resulted in considerable discussion about the use of public funds to support a private organization. Although the UWP Committee declined to provide funds, they did recognize the problem faced by the MTA. The CATS staff, at the direction of the Steering Committee, developed a cooperative project proposal to provide public funds to support private sector participation. After considerable negotiation among CATS, MTA, and the public agencies, a final proposal was prepared and submitted to the WPC and the PC in June 1985 for their endorsement (10). Following its acceptance by the PC, the proposal was submitted to UMTA for funding. Administrator Stanley announced funding for the "Private Initiatives" project at a July 1985 press conference in Chicago. For the first time, public funds would be used to support private sector involvement in the planning process. The project emphasized coordination of private involvement and also funded three technical studies aimed at developing and evaluating opportunities for increased use of private carriers to operate transit services.

The third outcome is the expansion of service contracting. Although it is not clear if the public transportation agencies are contracting for more service than they would have without the private sector participating in the process, it is clear that more carriers are being informed of such opportunities. At an early meeting of the Private Providers Committee, the chairman of the Suburban Bus Board (the policy board of Pace) expressed interest in developing innovative ways to meet the pressures created by reduced federal operating subsidies and encouraged the private operators to present to the board any ideas they had concerning suburban transportation alternatives. In addition, a CTA representative informed the operators of CTA's interest in contracting with private operators for the CTA's $4 million special service program for the disabled. Interested operators were requested to contact the CTA. As noted, the CTA ultimately did select four private contractors to operate this service. Clearly, these actions suggest an increased role for the private operators in the future.

There are, of course, many other emerging issues that will require the planners, operators (public and private), funders, and decision makers to work together. With the existence and actions of the Private Providers Committee and associations, the private sector is assured that it will have serious, as opposed to token, input to the issues and influence on the decisions affecting it.

ISSUES RAISED BY PRIVATE SECTOR PARTICIPATION IN THE PROCESS

There are several issues concerning the involvement of the private sector in the public planning process that have yet to be resolved. This section contains a review of some of these issues, which will be resolved gradually as the process evolves.

The most basic of these issues is the nature that private sector involvement should take. Few people would object to the private operators acting in an advisory role to inform the public transportation planners and providers of their needs and capabilities. An advisory role, however, frequently means limited participation in the decision-making process.

On the other hand, many people object to the private operators being in a policy-making or decision-making role. The representatives of the public agencies in decision-making positions on the PC and other MPO committees are either elected officials, appointees of elected officials, or the staff of appointees. The goal of these officials is to serve the public good. In contrast, the goal of the private representatives on these committees has not been clearly stated. Are they to represent only the interests of the private transportation operators, or are they to represent the interests of the public from the perspective of the private operators?

Another issue is the appropriate organizational structure for representing the private sector. As noted, these private transportation companies vary in many ways. There are school and charter bus companies, taxi companies, limousine companies, and several types of paratransit companies, all of which have differing problems and interests. Even within a particular mode, there are many different groups with opposing views. For instance, within the taxi industry, there are taxi companies, taxi associations, owner-operators, and labor organizations. Geographic differences are also important. For example, taxi regulations are set by the individual municipalities, sometimes causing conflict between the Chicago and suburban taxi and limousine operators over markets.

Finding a means to achieve a balanced representation of these interests and ensuring fair and continuous representation of all the operators is a difficult task. In many ways, the involvement of the firms and operators parallels the political involvement of the American electorate. Some companies are always concerned; some are concerned only at certain times; others are not concerned at all. The reasons motivating private sector concerns also vary. There are some who see this process as a way of capturing new or expanded markets through increased involvement; others who have a particular complaint about government regulations and encroachments on their markets; and the altruists who want to help public sector planners develop a better transportation system.

There are also problems of time and size. Attending the many meetings required to coordinate transportation planning in a large metropolitan region is time consuming. Business people tend to have tight schedules, particularly owners of small firms who tend to have little time to attend meetings. Thus, the large firm with an owner or manager who has the time may have a disproportionate voice in the process. This was a major concern of CATS when setting up the Private Providers Committee. On the other hand, MTA feels that the Private Providers Committee is, to some extent, redundant. From their point of view, the MTA may provide a more orderly and continuous representation of the private sector. In spite of these early differences, the MTA has continued to work closely with the Private Providers Committee and the CATS staff to the benefit of all the operators and the public planning process.

A public official who had worked with the private sector before either the MTA or the Private Providers Committee was established reported the difficulties in reaching a consensus among the various companies. The small firms would attend some of the meetings, perhaps taking strong stands on an issue, but then either they would stop attending or, at the last minute, withdraw their support for an action on which it had taken months to reach agreement. An organization such as the MTA provides a forum for the development of consensus or compromise where the various degrees of financial responsibility are given some weight. However, the current MTA membership is a small fraction of all the private firms in the region, and it is mostly biased toward the largest firms. It should be pointed out, though, that the most enthusiastic private operators are members of the MTA. Most of the seats on the CATS committees are held by MTA members. Thus, although MTA is not the official organization representing the private sector in the MPO planning process, it is perhaps the de facto representative.

A closely related issue is funding. CATS has spent a large amount of staff time in organizing committees to represent the private sector. To keep the small and less-involved firms informed, it publishes a quarterly newsletter and frequent bulletins that add to the cost. A simple, low-cost alternative would be to accept the first volunteer to represent the private sector in the process. But, as previously discussed, this approach would not provide balanced or equitable representation.

The funding issue also raises questions as to the appropriateness of using public funds to support private sector involvement in the process. In the formative stages of the region's planning process, the private operators paid dues to support their participation. However, with the availability of Section 8 planning funds, this policy has been relaxed. Obviously, the CATS' current involvement with, and staff support to, the private carriers is paid for from public planning funds. With the approval of the Private Initiatives proposal, some public monies will now be devoted to supporting private sector participation in the process.

CONCLUSIONS AND RECOMMENDATIONS

The northeastern Illinois region has taken the lead in private sector participation in the transportation planning process. From the experience in this region, it is clear that developing a process for involving the private sector is not simple nor can it be done overnight. Attitudes of both the public officials and the private sector need to change. There has been a basic mistrust on both sides that is gradually dissipating. As noted earlier, decisions on the appropriate role and representation of the private operators must be made.

In the Chicago region, the ice has been broken. Because the members of the public sector have met the individual private operators and found them to be conscientious and reputable business people, they have developed a greater respect for them and a better understanding of their problems. Although some public officials are still not receptive to private sector involvement, this is slowly changing. Similarly, the private sector representatives participating through the MTA or on the Private Providers Committee have begun to understand the public planning process. They have seen that the need for coordination and cooperation among many agencies requires a slower and more deliberate decision-making process than exists in their private business negotiations. They also have come to realize that participation in the process will not resolve all their problems. Conflicts over the appropriate actions (for instance, whether a particular service segment should be operated by a public agency or private firm) will continue. However, the ability of the PC to make these decisions based on the maximum information about the capabilities and needs of both sectors has increased substantially. The results of the technical studies to be done as part of the demonstration project should provide more information on the benefits and difficulties of private sector participation in specific planning projects.

The experiences in this region suggest the following recommendations for MPOs in other regions that are attempting to involve the private transportation operators in the planning process:

1. The MPO should form a committee of private operators as an advisory group to the MPO. Membership on the committee should be open to all private operators. Initially, the committee will act as an educational forum. The private operators will learn about the planning process and the public transportation agencies. The public agencies will learn that the private operators have legitimate concerns and that they can provide capabilities and resources to assist them in providing service. To this end, rep-

resentatives of public agencies should attend the meetings of the committee.

2. Task forces should be established to resolve specific transportation issues of interest to the private operators. The task forces should report to the private operators committee as well as the MPO on their deliberations or progress.

3. The MPO should consider providing support staff and resources to organize the private operators. To ensure serious commitment from both sides, the private sector should be required to match public sector funding after an initial start-up period.

4. After the private operators have become informed concerning the planning process and issues, the MPO should consider giving them a voice in the process. This voice should be at a policy level directly involved in the decision-making process in addition to the advisory level.

The private operators should also take an active role in developing participation in the process. To this end, they should (a) establish a region-wide industry association of all private operators that will focus on public and private interaction and (b) educate themselves on the public planning process, including the public officials and agencies responsible for planning and operating transportation, the specific functions of the agencies, the coordination and funding roles of each agency, and the goals and values behind the public process.

There is still a long way to go to reach a fully developed participatory process in Illinois as well as elsewhere. However, the time and effort devoted to that end should result in a more efficient and effective transportation system. This improved system will benefit the private sector transportation operators, the public transportation agencies, and the general public.

REFERENCES

1. Federal Register. Docket No. 83-D. UMTA, U.S. Department of Transportation, Oct. 22, 1984.
2. C. Lave. Urban Transit: The Private Challenge to Public Transportation. Ballinger Publishing Company, Cambridge, Massachusetts, 1985.
3. J.A. Tecson. The Regional Transportation Authority in Northeastern Illinois. In Chicago Bar Record, May-June, 1975.
4. J.A. Tecson. The Regional Transportation Authority in Northeastern Illinois. In Chicago Bar Record, July-Aug., 1975.
5. Chicago Area Transportation Study--Final Report, Vol. 1. Chicago Area Transportation Study, Chicago, Illinois, 1959.
6. Chicago Area Transportation Study--Final Report, Vol. 2. Chicago Area Transportation Study, Chicago, Illinois, 1961.
7. Chicago Area Transportation Study--Final Report, Vol. 3. Chicago Area Transportation Study, Chicago, Illinois, 1962.
8. Code of Federal Regulations. Vol. 40, No. 181. UMTA, U.S. Department of Transportation, Sept. 17, 1975.
9. C.E. McKnight and R.A. Paaswell. Cost Analysis: Selected Routes. Urban Transportation Center, University of Illinois, Chicago, Illinois, Sept. 1984.
10. Private Enterprise Initiatives for the Chicago Metropolitan Region. Chicago Area Transportation Study, Chicago, Illinois, June 1985.

Publication of this paper sponsored by Committee on Public Transportation Planning and Development.

Predicting Annual Transit Fare Revenue from Midyear Results

SAMUEL SEWARD, RICHARD P. GUENTHNER, and HASSAN KH. NASSER

ABSTRACT

Uncertainty about the future and the possibility of major changes are problems that face all transit systems. As a tool in facing these uncertainties, contained in this paper is an exploration of the possibility of predicting total annual revenues for transit systems based on the revenue collected part of the way through the year. Six Wisconsin cities representing three city sizes (large, medium, and small) were used as a case study. A single prediction along with the limits of the 95 percent confidence interval were computed. Also, the percentage errors (as compared to the actual annual revenues) were calculated and the distribution of these errors was examined. The findings revealed that the proposed method is applicable to the Milwaukee County Transit System because it results in a prediction of annual revenue that will have less than 5 percent error 95 percent of the time. This method was also found to be applicable to small Wisconsin cities. However, in the case of the only medium Wisconsin city, the percentage error was higher because of anomalies in some of the data.

After World War II, the United States transit industry suffered a general decline marked by dwindling patronage and increasing costs to a point where fares would no longer cover costs for service levels desired by most urban communities. The termination of private transit services in many cities resulted, and the federal government responded in 1961 with an aid program for transit. This aid program had evolved into a full spectrum of transit assistance programs by the mid-1970s. By 1980, the U.S. Congress was providing several billion dollars annually for transit. Also, a number of states started to provide capital and operating aid directly to local transit agencies to supplement federal matching grant programs.

All these funds and grants from federal, state, and local government might appear sufficient to ensure that transit operators are able to cover their operating costs. Unfortunately, this has not been the case. Transit operating ratios still declined through the late 1970s. Because of these trends and recently decreasing federal funds, transit systems have had to be increasingly concerned with predicting annual revenues in order to balance budgets.

The purpose of this paper is to test a methodology for using partial-year transit revenue as a means for predicting total current-year revenue for a transit system. By applying such a method early in the year, transit managers will have an indication of whether to expect a shortage, surplus, or balance of funds at the year's end. The methodology will also give budgeting personnel better information for developing their annual budgets, which typically are prepared during the middle of the preceding year.

For this analysis, past revenue data were collected from six Wisconsin cities: Milwaukee, Madison, Racine, Kenosha, Janesville, and Green Bay. These cities were chosen to represent Wisconsin's large, medium, and small cities. Some statistics for these cities are given in Table 1. This paper is presented in two sections. The first section describes the methodology that was used in making the predictions, and the second describes the results from applying the model.

UNCERTAINTY IN FORECASTING

One reason that transportation planning has limited influence in the policy process is the proven inaccuracy of the forecasts. Unfortunately, this limitation also applies to financial and budgeting forecasts for transit systems. The nature of forecasts is to be in error, and no one can ever eliminate all such errors. However, realizing this, concern should be focused on anticipating the errors and limiting both their size and their consequences.

Past research has focused on the sources of errors in forecasting and ways to reduce them (1). Although this is a worthwhile avenue to pursue, an insufficient amount of attention has been devoted to characterizing uncertainty and conveying useful information about it to decision makers. In characterizing uncertainty, researchers are primarily concerned with preparing realistic estimates of the likely range of key forecast values (2). Assessing the level of uncertainty in the final outputs is not easy, largely

TABLE 1 Selected Information Concerning the System Analyzed

System	Population	No. of Routes	1984 Vehicle-Hours	Fleet Size	1984 Operating Budget ($)
Milwaukee	965,000	71	1,617,000	616	63,493,000
Madison	225,000	21	337,000	194	14,909,000
Green Bay	141,000	16	84,500	29	2,559,000
Kenosha	94,000	7	63,600	29	1,717,000
Janesville	52,000	7	30,300	22	929,000
Racine	85,000	12	105,800	39	2,700,000

Note: From operating reports of respective transit authorities.

S. Seward, Milwaukee County Transit System, 1942 N. 17th Street, Milwaukee, Wis. 53205. R.P. Guenthner, Department of Civil Engineering, Marquette University, Milwaukee, Wis. 53233. H. KH. Nasser, Community Design Center, Inc., 340 West Brown Street, Milwaukee, Wis. 53212.

TABLE 2 Milwaukee Monthly Revenue ($)

Year	January	February	March	April	May	June	July	August	September	October	November	December
1976	1,424,372	1,413,677	1,580,086	1,460,876	1,415,669	1,354,835	1,308,211	1,306,352	1,513,938	1,556,177	1,439,284	1,429,220
1977	1,434,763	1,435,700	1,618,745	1,413,143	1,436,619	1,380,091	1,286,282	1,368,485	1,569,966	1,624,888	1,597,294	1,571,302
1979	1,676,124	1,599,742	1,750,953	1,604,397	1,596,898	1,520,803	1,494,836	1,565,873	1,620,860	1,781,055	1,662,921	1,548,043
1980	1,696,710	1,713,554	1,809,995	1,747,461	1,696,813	1,605,917	1,618,049	1,607,891	1,768,568	1,869,083	1,648,512	1,670,711
1981	2,084,294	1,998,171	2,238,437	2,052,699	1,940,304	1,915,990	1,908,101	1,839,716	2,010,404	2,103,085	1,943,993	1,880,787
1982	2,120,047	2,259,030	2,496,680	2,400,401	2,089,531	2,021,844	1,919,912	1,957,116	2,164,924	2,260,056	2,113,576	1,884,130
1983	2,139,726	2,259,448	2,493,455	2,183,944	2,177,069	2,065,629	1,958,554	2,005,853	2,183,183	2,187,442	2,188,989	1,804,155
1984	2,196,698	2,275,843	2,393,248	1,105,655	2,220,354	2,023,739	1,984,635	2,137,704	1,994,967	2,272,296	2,069,856	1,946,431

because of the multiplicity of inputs, estimates, assumptions, and the uncertainty associated with each. At least as important is finding constructive ways to convey the nature and significance of this uncertainty to the users of the forecasts.

Yet the difficulties that researchers have in dealing with uncertainty suggest that there are many risks involved in revealing this information. The risks come from the possibility of frightening the decision makers when the reality of uncertainty is spoken about openly.

One way to deal with these issues will be discussed in the next section. The attempt here is to understand the characteristics of errors to convey useful information about them. This has been done through studying the seasonal trends of transit revenues and, consequently, developing a methodology to predict these revenues, with a reasonable, understandable error. Also, although much of planning addresses the long range, the method presented here is applicable to the short range. This method is also simple and easy to understand.

METHODOLOGY

This study was primarily done for the Milwaukee County Transit System (MCTS), which is a mass transit organization owned by Milwaukee County, but managed and operated by Milwaukee Transport Services, Inc., a private, nonprofit organization. For the first 115 years of its history, the Milwaukee Transit System was able to provide reasonable service at acceptable rates at no cost to nonusers. The transit system was a taxpayer during this period and the system was supported solely by the fares. The Milwaukee Transit System continued as a private system many years longer than did most transit systems in other cities.

But, with a less dense community, more scattered riding requirements, and subsidized highway programs, the time was reached in Milwaukee when quality transit service could no longer be supported by the farebox alone. Beginning in 1975, the system has been subsidized by federal, state, and county funds. In fiscal year 1984, the total system operating cost was $63.5 million. Operating revenue (primarily fares) covered $30.5 million or 48 percent of that cost. The remaining cost was covered by $7.4 million in federal funding, $21.8 million in state funding, and $3.7 million in county funding. As can be seen, even though the farebox does not cover all the costs, its contribution is large enough that short-range predictions would be useful.

In this paper, the predictions have been based on monthly revenues from 1976 through 1984 excluding 1978 due to a strike that made that year's annual total inapplicable. These values are given in Table 2. Annual revenue predictions were made from each month's cumulative total during the 8-year period. The second step was to obtain a 95 percent confidence interval for these predictions to provide a clear range as to where the estimates will fall.

The prediction was computed by first finding the cumulative monthly revenue and cumulative percentage of the total revenue for each month for all 8 years. The resulting annual percentage values for each month were then averaged over the 8-year period, and used to predict the annual revenue. These averages, along with the standard deviations, are given in Table 3.

The predictions for each month during the 8-year period were determined from the average cumulative percent and the revenue through the month in question, by using the following formula:

$$\mathrm{PRDI} = (\mathrm{QDATA}/\mathrm{QAVE}) * 100 \qquad (1)$$

where

PRDI = the middle prediction,
QDATA = the cumulative monthly revenue, and
QAVE = the cumulative percent average.

In other words, these values represent the predictions that could have been made at that point in time. For example, at the end of May 1982, the annual revenue estimate would have been made from the January through May totals. To determine the accuracy of the method, each prediction was then compared with the actual annual revenue. The percentage errors between the two are given in Table 4.

To determine the precision of each prediction, a 95 percent confidence interval can be computed. The high and low limits of this interval represent the 95th percentile confidence interval of the bounds of the cumulative average. In other words, transit systems will be able to predict the total annual revenue

TABLE 3 Cumulative Percentages of Milwaukee Revenue by Year and Month (%)

Year	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.
1976	8.28	16.50	25.68	34.17	42.40	50.28	57.88	65.48	74.28	83.33	91.69
1977	8.09	16.18	25.31	33.28	41.38	49.16	56.41	64.12	72.98	82.14	91.14
1979	8.63	16.87	25.88	34.14	42.36	50.19	57.89	65.95	74.30	83.47	92.03
1980	8.30	16.67	25.52	34.07	42.36	50.21	58.13	65.99	74.63	83.77	91.83
1981	8.72	17.07	26.43	35.01	43.13	51.14	59.12	66.81	75.21	84.01	92.14
1982	8.25	17.05	26.77	36.11	44.25	52.12	59.59	67.21	75.64	84.44	92.67
1983	8.34	17.15	26.87	35.39	43.88	51.93	59.57	67.39	75.90	84.43	92.97
1984	8.57	17.46	26.80	35.02	43.68	51.58	59.33	67.67	75.46	84.32	92.40
Average	8.40	16.87	26.17	34.65	42.93	50.82	58.49	66.32	74.80	83.74	92.11
Standard deviation	0.203	0.378	0.591	0.841	0.906	0.964	1.040	1.105	0.894	0.723	0.537

TABLE 4 Percentage Error by Year and Month

Year	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.
1976	-1.40	-2.20	-1.82	-1.37	-1.22	-1.08	-1.03	-1.28	-0.70	-0.49	-0.45
1977	-3.67	-4.06	-3.24	-3.96	-3.62	-3.29	-3.56	-3.32	-2.44	-1.91	-1.05
1979	2.77	-0.01	-1.06	-1.46	-1.32	-1.24	-1.02	-0.57	-0.67	-0.32	-0.09
1980	-1.21	-1.16	-2.43	-1.68	-1.32	-1.20	-0.62	-0.51	-0.22	0.04	-0.30
1981	3.78	1.20	1.04	1.05	0.46	0.61	1.07	0.72	0.55	0.32	0.03
1982	-1.72	1.06	2.33	4.22	3.07	2.54	1.89	1.33	1.12	0.84	0.60
1983	-0.65	1.68	2.74	2.14	2.21	2.18	1.85	1.60	1.47	0.83	0.93
1984	2.10	3.48	2.44	1.06	1.75	1.48	1.43	2.02	0.88	0.70	0.32
Standard deviation	2.41	2.24	2.26	2.43	2.11	1.90	1.78	1.66	1.20	0.86	0.58
95 percent interval	5.70	5.30	5.34	5.74	4.99	4.49	4.21	3.93	2.84	2.03	1.37

and give a range in which the actual annual revenue will fall 95 percent of the time. Because the sample size is somewhat small (8 years of data), the normal distribution was inappropriate. Instead, the Student's T-distribution was used. The Student's T-distribution for 7 degrees of freedom (for 8 years of data) equals 2.365 (3). The resulting equations are

HI = QAVE + STUDT * QSSD (2a)

and

XLOW = QAVE - STUDT * QSSD (2b)

where

STUDT = 2.365,
HI = the upper limit of the mean standard error,
XLOW = the lower limit of the mean standard error, and
QSSD = the cumulative standard deviation.

The limits of each revenue prediction are

PRD2 = (DATA/HI) * 100 (3a)

and

PRD3 = (DATA/XLOW) * 100 (3b)

where DATA is the monthly revenue.

By using the mean and standard deviation for each column given in Table 4 and applying Equations 2 and 3, confidence intervals for the observed errors were computed. (These confidence intervals are given in Table 4.) These confidence intervals represent the expected range of the error of the predictions. For example, a prediction made using only information from January can be expected to be within 5.70 percent of the actual revenue 95 percent of the time.

Because the concept of the 95 percent confidence interval does not automatically have a great deal of meaning to a nonstatistician, the average error was also computed by separating the predictions from each month's data into positive and negative variations and then averaging those categories. Figures 1, 2, and 3 show these average errors as well as the largest negative and positive variations found in the predictions for each month. From this line, the range and direction of the variations can be seen.

It should be recognized that predicting revenues based on more limited data will reduce the accuracy of the results. Ideally, a prediction system such as this should be based on data from as many years as is feasible. By using the data in Table 3, the variation is shown to be much greater when a limited

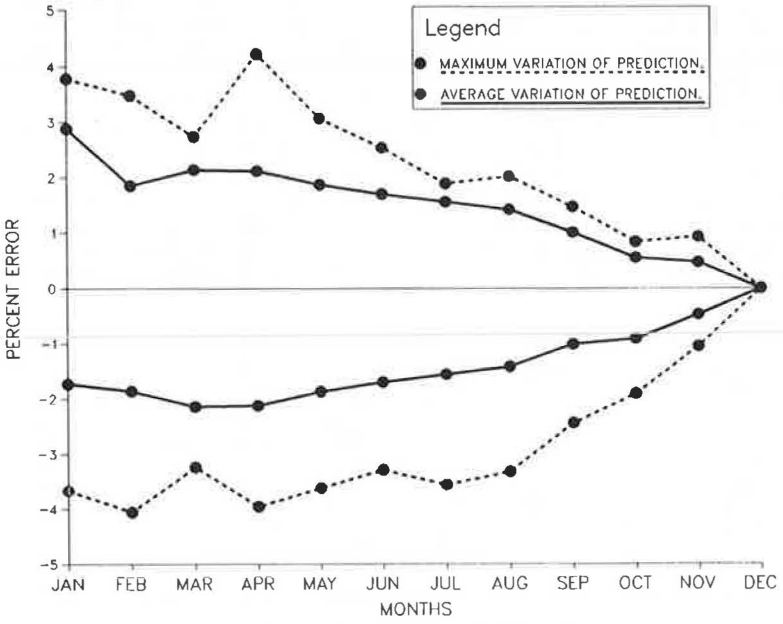

FIGURE 1 Largest errors in revenue predictions in Milwaukee.

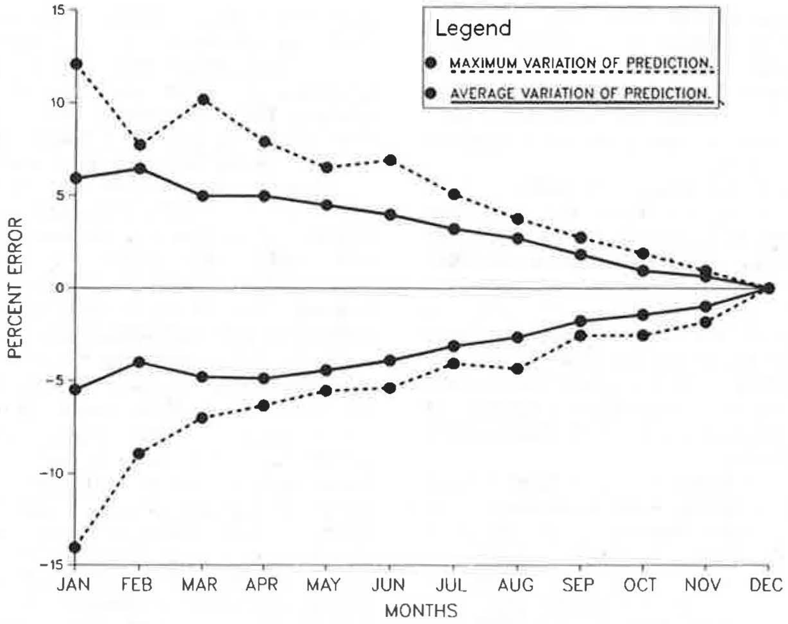

FIGURE 2 Largest errors in revenue predictions in Madison.

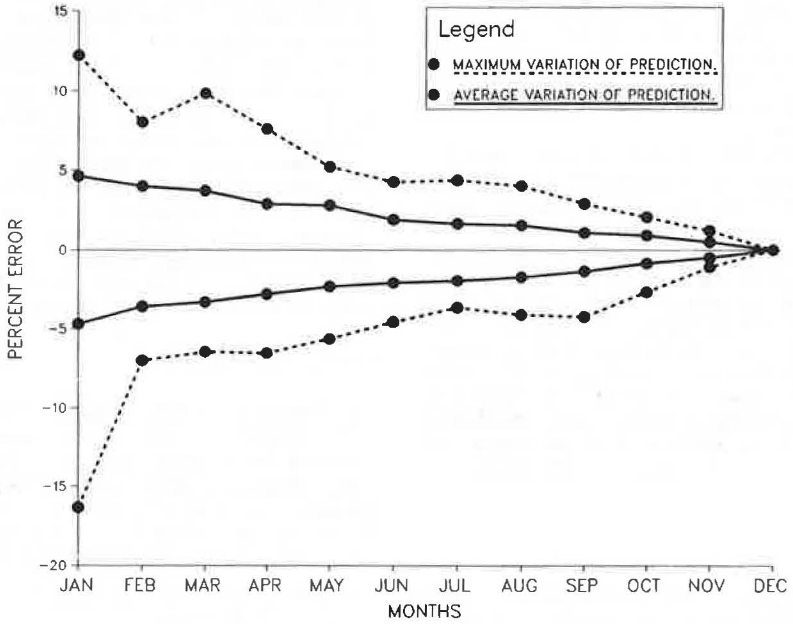

FIGURE 3 Largest errors in revenue predictions in small cities.

data base is used. For example, if a January prediction had been made in 1979 using only the 1976 and 1977 data, the prediction would have been 5.4 percent above the actual revenue. In comparison, the maximum error for a January prediction throughout the entire 8 years is only 3.78 percent.

DATA ANALYSIS AND RESULTS

Milwaukee

Predicting total annual revenue on the basis of the revenue collected in the first few months of that year is applicable to the MCTS. The average and maximum errors in predicting these revenues shown in Figure 1 indicate an acceptable range of predictions. Table 4 indicates that the highest negative error is -4.06 percent in February 1977 and the highest positive error is 4.22 percent in April 1982. These figures are good indications that the prediction will stay in the range of -5 to +5 percent error because the highest and lowest errors in the actual data, which include 1984 predictions, do not exceed the 95th percentile confidence interval.

In general, the revenue in January is about the same as that in February. Consequently, if the standard deviation of the cumulative percentage for January and February is less than twice that for January alone, more accurate predictions can be made. Examination of the standard deviation as given in Table 2 reveals that the cumulative percentage in

February is slightly more stable than that in January. A February prediction should be slightly better than one using information only from January. However, after April, the standard deviation is greater than four times that for January. The resulting predictions in April are not as accurate as those in January, February, and March.

Starting in May, the standard deviations have less-than-proportional increases from the previous months; consequently, the accuracy of the predictions starts to improve at that time. This indicates that no prediction should be made before April because the information during April is not sufficient to make the prediction. This can also be seen clearly by tracing the variation of error of the prediction shown in Figure 1. It appears that weather variations and the Easter holiday cause a significant amount of year-to-year variation (a high percent of Milwaukee's ridership is schoolchildren).

As can be seen from these numbers and from Figure 1, the best time (from a management standpoint) to make the prediction is during May or June. By that time, the 95th percentile error is within 5 percent. By then, the prediction is accurate and there is still time to react to it. This indicates that the method developed is highly applicable to the MCTS.

Madison

Unfortunately, Madison was an exception to the rule of consistent results because the error range was wide. As can be seen from Figure 2, the average errors do not enter the 5 percent range until March, and the limits of the maximum error found do not enter that range until July.

One important observation of the Madison results here is that the minimum worst possible errors occurred between 1974 and 1979; the maximum worst possible error occurred between 1981 and 1984. One reason for this is that changes in services must have occurred around 1980, such as adding new routes or increasing fares. Based on general information concerning the system, a large number of service and fare changes have occurred during the study period. Accordingly, no stable data were available for this analysis. One positive finding from the Madison results is that transit operators could have predicted their total annual revenue with a percent error of less than 6 percent during all years except 1982.

Small Cities

The results from all of the small cities were similar to each other. The dotted lines in Figure 3 represent the highest errors from the four cities for each month in question. The fluctuations of the graph are due to increasing services or a raise in fares. For example, a sudden increase of the percent error of January 1982 in Janesville was due to a fare increase from $0.35 to $0.50. Another increase had also been instituted in February 1981 from $0.25 to $0.35.

Another change that can cause variations is the addition of routes. This was the case for the Racine system, where two routes were added between June 1979 and April 1980. Also, in October 1982, peak-hour service was improved on four routes, with headways reduced from 30 to 20 min. These changes will affect ridership and, consequently, the monthly revenues.

With the exception of the first 4 months, the maximum errors stayed in the 5 percent range, which indicates that this model is applicable to Wisconsin's small cities. Also, with the exception of Janesville, the maximum and minimum worst errors for these cities varied between -6.73 in February 1979 in Green Bay and 10.47 in January 1984 in Green Bay. Unlike other small cities, Janesville's percentage error ranged from 12.23 percent in January 1984 to -16.35 percent in January 1981. Also note that these Janesville errors dropped to about one-half their magnitude in February to 6.32 and -6.00, respectively.

As was done for Milwaukee, the standard deviation and the 95th percentile confidence interval for the prediction errors were computed for Madison and each of the small cities. These are given in Table 5. Note that in Madison and Janesville, an error smaller than 5 percent could not be expected until the end of August. However, acceptable results could be obtained in the other small cities during the summer months between May and July. At this time, adjustments for fall could still be made. None of these results was as consistent as those in Milwaukee. However, wider ridership fluctuations can be expected in smaller cities. Also, because the small cities have much smaller operating budgets, an erroneous prediction of 5 percent would not be as much in terms of dollars as would one in Milwaukee.

USING THE FINDINGS

To demonstrate how future predictions can be made, the following predictions were computed using the revenue data from January, February, and March of 1985 obtained from the MCTS. The actual monthly revenues were $2,614,225 for January, $2,485,342 for February, and $2,637,972 for March. The calculations were done according to Equations 1-3 as follows:

Average cumulative percentage of January (1976 to 1984) = 8.3975,

TABLE 5 Standard Deviation by City and Month

	Month										
City	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.
Madison											
Standard deviation	7.43	5.75	5.57	5.21	4.61	4.18	3.45	2.88	1.98	1.31	0.86
95 percent interval	17.57	13.60	13.18	12.32	10.90	9.89	8.16	6.81	4.68	3.10	2.04
Green Bay											
Standard deviation	4.73	4.71	4.11	3.78	3.38	2.63	1.95	1.39	1.08	0.69	0.39
95 percent interval	11.20	11.13	9.71	8.94	8.00	6.22	4.61	3.29	2.55	1.63	0.75
Janesville											
Standard deviation	8.66	5.05	5.17	4.20	3.52	2.78	2.53	2.26	1.52	1.01	0.67
95 percent interval	20.50	11.94	12.21	9.94	8.32	6.58	5.98	5.35	3.59	2.39	1.58
Racine											
Standard deviation	4.27	3.20	2.43	2.10	2.33	2.16	2.19	2.01	1.88	1.44	0.66
95 percent interval	10.09	7.56	5.75	4.96	5.51	5.10	5.17	4.76	4.45	3.40	1.56
Kenosha											
Standard deviation	4.05	3.99	3.82	2.73	2.28	1.26	1.40	1.69	1.10	0.85	0.53
95 percent interval	9.58	9.44	9.04	6.45	5.39	2.98	3.31	4.00	2.60	2.01	1.25

Standard deviation = 0.2028, and
Student's t-distribution * Mean standard error
= 2.365 * 0.2028 = 0.4796.

Annual projection = January revenue/Average cumulative percentage = 2,614,225/0.083975
= $31,130,991.

The lower bound of the 95 percent confidence interval
= January revenue/(Average cumulative percentage
+ 0.004796) = 2,614,225/(0.083975 + 0.004796)
= $29,449,015.

The upper bound of the 95 percent confidence
interval = January revenue/(Average cumulative
percentage - 0.004796) = 2,614,225/(0.083975
- 0.004796) = $33,016,729.

In the following table, the rest of the results are given (in dollars) for February and March. The predictions are decreasing from the January predictions.

Predictions	January	February	March
Low	29,449,015	28,713,778	28,067,103
Middle	31,130,991	30,032,785	29,580,575
High	33,016,729	31,912,183	31,235,019

Because the final 1985 revenues are not finalized, these predictions cannot be tested. However, if the trend has not varied much from the past 8 years, it is expected that starting in May, accurate predictions should be available.

CONCLUSIONS

Predicting total annual revenue based on partial-year revenue is an important mechanism that can be used by planning and budgeting personnel in the transit industry. Revenue collected from regular transit operations is often the largest component of overall system revenue. In addition, while other income is predetermined, the ridership levels cannot be controlled. Thus, accurate predictions of annual revenue can be extremely valuable. A methodology to project transit revenues was developed based on the variations of revenue of large, medium, and small Wisconsin cities. A prediction based on the cumulative percent average of each month was developed in addition to two upper and lower limits for this prediction.

Analyzing the results obtained from these cities indicated that this mechanism is applicable, and that accurate short-term predictions could be generated by using it. The results of the data analysis indicate that reliable estimates, with less than 5 percent error, of future revenue for the MCTS could be predicted as early as May with 95 percent confidence. Similarly, reasonably accurate forecasts were obtained for Wisconsin's small cities with errors not exceeding 5 percent. Less accurate estimates were made for Madison as a result of the inconsistency of some revenue data (year 1982), and an apparent system change in 1980. Therefore, a practical model cannot be developed for the Madison system until a stable data base is available.

Applying this model will give budgeting personnel a clearer picture of their future revenue by giving them a clear date around when year-end shortfills or surpluses can be detected. In addition, the possible error accompanying this projection can be estimated.

Additional research should be conducted to improve the validity of the model by exploring the year-by-year fluctuation. The years with the greatest error are those that changed most from the previous year and if a version of the model can be developed that takes this variation into account, earlier and more accurate forecasts may be possible. This will encourage more operators to consider using it as an ongoing management analysis tool.

ACKNOWLEDGMENTS

Necessary data were provided by the Milwaukee County Transit System, Madison Metro, Kenosha Transit, Racine Transit, Janesville Transit, and Green Bay Transit. Partial funding was provided by a grant from the National Science Foundation.

REFERENCES

1. F.S. Koppelman. Uncertaintly in Travel Behavior Forecasting. In New Horizons in Travel Behavior, Lexington Books, Lexington, Massachusetts, 1981.
2. M. Godet. The Crisis in Forecasting and the Emergence of the Prospective Approach. Pergamon Press, Inc., New York, 1979.
3. E.R. Walpole and H.R. Myers. Probability and Statistics for Engineers and Scientists. The Macmillian Company, New York, 1972.

Publication of this paper sponsored by Committee on Public Transportation Planning and Development.

Multicriteria Evaluation of Alternative Transit System Designs

N. JANARTHANAN and J. SCHNEIDER

ABSTRACT

One of the most important but underdeveloped parts of the transit planning process is the evaluation of alternative designs. The results from evaluation studies provide a basis for decision making. Evaluation of alternative transit system designs is now even more important because of reduced public funds. A computer-based multicriteria method using concordance analysis is described and applied to evaluate alternative transit system designs. Development of objectives and criteria, normalization methods, and the use of relative important weights are presented. A nonlinear method of normalization technique that uses a logistic curve is introduced. The shape of this curve can be varied by the user. An application of the multicriteria evaluation methodology to five alternative transit system designs is presented to illustrate how the best design can be identified.

Evaluating alternatives is one of the major tasks in the field of planning. Decision makers must rely on evaluation results to determine the comparative performance of alternatives as measured against project goals. The increased complexity of today's problems in public transit planning has made evaluation and decision making a particularly difficult task. The problem has become more complicated recently because of reduced public funds, more acute political concerns, forecasts that involve more uncertainty, and heightened awareness of environmental impacts. In such situations, a systematic procedure for conducting an evaluation process can be useful and necessary. An evaluation method is a means by which the pros and cons of alternative plans can be described in a logical framework so as to assess their various net benefits.

Transit planning is one of many multiobjective problems that have conflicting goals. This means that better performance of one objective often cannot be achieved without negatively affecting other objectives. In addition to these inherent conflicts, the differing opinions of local government agencies, political groups, citizen groups, and system users have to be taken into account. In this paper, a recently developed multicriteria evaluation methodology is described and used to evaluate several alternative transit system designs. Development of objectives and criteria, normalization methods, and the use of weights to represent the relative importance of different criteria are discussed. An application of this methodology that identifies the best out of five alternative transit system designs for a hypothetical city is also presented.

EVALUATING TRANSIT SYSTEMS

A public system such as a transit system needs to be evaluated periodically to justify the public money spent to run it. Evaluation is also required to analyze alternative designs resulting from modifications or changes in the route network, demand patterns, or other schedule or service allocation changes. Previous literature on this topic includes a paper by Dajani and Gilbert (1) who have presented a framework for evaluating transit systems using performance measures and a ranking approach. Fielding, Glauthier, and Lave (2) defined three efficiency, four-effectiveness, and two system-level indicators to evaluate transit systems using operating and financial data. Both papers assume that the comparisons among transit agencies can be made by grouping them properly. Mundle and Cherwony (3) suggested a methodology consisting of uncontrolled and controlled comparisons of performance measures among transit agencies. This method overcomes the drawback of not totally reflecting the differences in operating characteristics or environment among transit systems. Holec, Schwager, and Fandialan (4) used Section 15 data to develop performance indicators and then used them in routine evaluations in Michigan. The evaluation involved identifying the variation in performance values of the system considered compared to others or to itself over time. Fielding, Babitsky, and Brenner (5) used Fiscal Year 1980 Section 15 data and factor analysis to select the seven best indicators. These seven indicators were recommended for the performance evaluation of a system over a time period or among different systems.

All the articles reviewed have used the magnitude of the performance measures directly to evaluate or compare the transit systems. But this is valid only when all the measures carry equal weight or the set has a dominant alternative that has better performance values in all the measures considered. Only Dajani and Gilbert (1) have discussed the problem of differing preferences between performance measures. An evaluation procedure needs to take into consideration different perspectives, such as the federal or local government, community, transit agency, citizen groups, and users. These perspectives give rise to multiple minimizing and maximizing objectives. An evaluation framework should be therefore be robust enough to include all multiple performance measures and multiple weight sets to represent different views about the relative importance of the performance measures. The evaluation method should be flexible enough so that the user can add and delete performance measures/criteria and weight sets depending on

---

N. Janarthanan and J. Schneider, Department of Civil Engineering, University of Washington, Seattle, Wash. 98195.

the purpose and objectives of the evaluation process. Based on these requirements, a multicriteria methodology called concordance analysis has been chosen for this study. The following sections briefly explain this recently developed multicriteria methodology.

MULTICRITERIA METHODOLOGY

Recent developments in the area of decision analysis have generated a number of methods for dealing with complex decision-making problems such as the transportation planning problem (6,7). Multicriteria analysis has been used in many fields successfully including transportation planning (7-13). In the literature, some authors have used the terms "multiobjective decision making," "multicriteria decision making," and multiattribute decision making" interchangeably. In this paper, the following definitions will be used, following Giuliano et al. (10). Objectives are the measurable targets representing the project goals and these objectives are made empirically operational in the form of criteria that are used in determining the extent to which the objectives have been achieved. The multiobjective problem is defined as a problem where there is more than one objective and the objectives cannot be combined in any way. Because of different viewpoints and values held by decision makers, there is generally no "best" alternative in any situation. What is best for one set of decision makers may not be best for another when conflicting objectives exist. The multiobjective method of evaluation seeks to identify the set of "best possible" alternatives, recognizing that different preferences exist.

There are two general categories of multiobjective problem-solving methods. One is based on whether the problem is conceptualized as continuous or discrete (7). The transit system evaluation problem in this study is a discrete problem. Discrete methods are simple and do not require extensive mathematical expertise and there are many methods available to use in solving them. The method selected for this study is concordance analysis and it is explained in Giuliano et al. (7).

The framework used for the evaluation of different alternatives for a transit system is shown in Figure 1. The decision makers in the process could be federal or local government officials, or both, transit system users, citizen groups, transit agency personnel, or other elected officials. The role played by these decision makers will vary at different stages of the evaluation process depending on many factors such as the local government's policies, the transit agency's policies, the involvement of citizens, and the political agenda in the region.

CONCORDANCE ANALYSIS

Concordance analysis is a multicriteria evaluation technique in which alternative plans are evaluated by a series of pairwise comparisons across a set of criteria. It is based on the Electre method developed originally in France. References to and discussions of the development of the Electre and concordance methods are presented in Nijkamp and Van Delft (14), Giuliano et al. (10), and Guiliano (15). The concordance analysis technique used in this research is an improved version of the program developed by Giuliano et al. (10). Improvements have been made in the normalization procedure by adding a nonlinear normalization method.

The first step in applying concordance analysis is to develop the project effects matrix. This matrix contains the performance values of all the criteria/attributes used to evaluate all alternatives. For an analysis with 'm' alternatives and 'n' criteria, the project effect matrix 'X' contains m x n elements. In general, the raw values are measured in different units. To make the various criterion scores compatible, it is necessary to transform them into one common measurement unit. The normalization procedure adopted in this study transforms each value in the raw project effects matrix so that all the normalized values are dimensionless and lie between 0 and 1, and so that the higher values are always better. Various types of linear and nonlinear normalization procedures are available. Two linear methods and one nonlinear transformation method are used in this study. The first method, magnitude-scaled normalization, uses

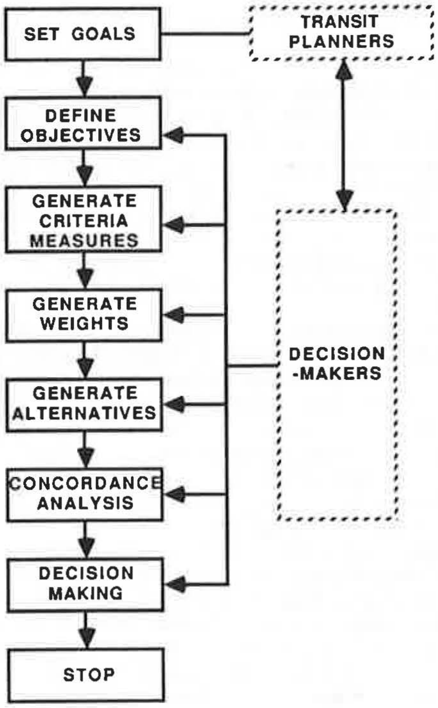

FIGURE 1 Framework for the evaluation of transit system improvement alternatives.

$$r_{ij} = x_{ij}/x_{ij}^*$$

where

r = normalized value,
x = raw project effect,
$i = 1, 2, \ldots, m$ alternatives,
$j = 1, 2, \ldots, n$ criteria, and
$x_{ij}^* = \max(x_{ij})$, for $j = 1, 2, \ldots, n$.

The advantage of this normalization method is that all outcomes are transformed in a linear way, so that the relative order of magnitude of the values remains the same. In the case of a criterion with a "less-is-better" objective the following is used:

$$r_{ij} = 1 - (x_{ij}/x_{ij}^*)$$

The second linear method, interval-scaled normalization, uses

$$r_{ij} = (x_{ij} - x_{ij}^{**})/(x_{ij}^* - x_{ij}^{**})$$

where x_{ij}^{**} is $\min(x_{ij})$ for $j = 1, 2, \ldots n$.

In the case of a criterion with a less-is-better objective:

$$r_{ij} = 1 - \{(x_{ij} - x_{ij}^{**}) - (x_{ij}^{*} - x_{ij}^{**})\}$$

The advantage of the interval method is that the scale of measurement varies precisely between 0 and 1 for each criterion. A possible drawback of this procedure is that the interval method does not lead to a proportional change in outcomes. The magnitude-scaled method is useful in normalizing a project effect matrix that will be analyzed by a weighted summation technique. The interval-scaled method is especially appropriate where a technique is used that performs a pairwise comparison of the criterion scores (16).

In the literature, a criterion with a more-is-better objective is called a "benefit criterion" and one with a less-is-better objective is called a "cost criterion." The same terminology will be used here. In the raw project effects matrix, higher values are better for a benefit criterion and lower ones are better for a cost criterion. In the normalized projects effect matrix, higher values for both benefit and cost criteria are better.

The third type of normalization used allows for nonlinear variation. In the transportation planning field, one cannot assume a linear form for all utility curves. Many of the criteria used in the analysis may behave nonlinearly and a nonlinear normalization technique is needed. This utility curve is assumed to have an S-shape (like the logistic curve) symmetric about its midpoint. This curve can have a linear portion in its middle. The shape of the curve can be defined differently for each criterion by changing the input parameters. The logistic curve used in this study is of the form

$$r = e^{G(x)}/[1 + e^{G(x)}]$$

where

r = the normalized value,
$G(x)$ = some function of the level of performance x, measured in performance units, and
e = the base of natural logarithms.

As shown in Figure 2, such a function describes an S-shaped curve, which is asymptotic above to the line $y = 1.0$ and below to the x-axis. In addition, over the range $y = 0.2$ to $y = 0.8$, it can be varied by the user to fit different criteria. A utility curve of this form implies a relationship between performance and normalized value that is linear over a certain range, while having an exponential decay near the upper and lower limits. It describes a relation in which the principle of diminishing rate-of-return applies to both high and lower levels of performance.

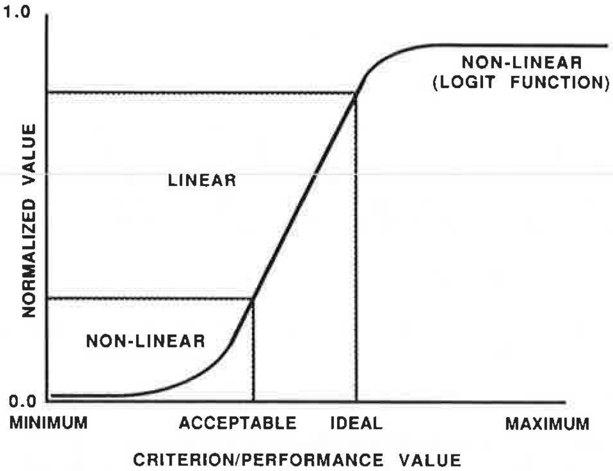

FIGURE 2 General form of nonlinear utility curve.

Although the general shape of the curve is the same for all decision makers and all performance measures, the exact relationship between performance level and normalized value must be determined by inputs that are used to calculate the G(x) function in the preceding equation. A simple linear form was chosen as follows (17):

$$G(x) = ax + b$$

To determine the values of a and b, the analyst is asked to specify two points on the curve. This is accomplished by specifying an ideal and an acceptable value for each performance measure in performance units. An ideal value is defined as the level of performance beyond which further improvement will bring only a minimal increase in the normalized value (i.e., the point of diminishing returns). The acceptable value is defined as the minimum (or maximum), which is the lowest (or highest) level of performance that will be tolerated. The user also has to define the corresponding normalized values for the ideal and acceptable values of the performance measure. With these inputs, the following three equations can be used to normalize a given project effects matrix. For $A_j \; \underline{x}_{ij} \; \underline{I}_j$,

$$r_{ij} = \{r(I_j) - r(a_j)\} * C/E + r(A_j)$$

For $x_{ij} < A_j$,

$$r_{ij} = e^{\ln(M/N)(C/E)+\ln(N)}/1 + e^{\ln(M/N)(C/E)+\ln(N)}$$

For $x_{ij} > I_j$,

$$r_{ij} = e^{\ln(M/N)(D/E)+\ln(M)}/1 + e^{\ln(M/N)(D/E)+\ln(M)}$$

where

r = the normalized value,
x_{ij} = the value of the jth criterion of the ith alternative, for which the normalized value needs to be determined,
I_j = the ideal value for the jth criterion,
A_j = the acceptable value for the jth criterion,
$M = \{r(I_j)/(1 - r(I_j)\}$,
$N = \{r(A_j)/(1 - r(A_j)\}$,
$C = (x_{ij} - A_j)$,
$D = (x_{ij} - I_j)$, and
$E = (I_j - A_j)$.

(Note that the same equations can be used for both benefit and cost criteria.)

WEIGHTS FOR THE CRITERIA

After normalizing the raw project effects matrix, the next step is to establish the relative importance or priority for the criteria included in the evaluation. The relative importance of criteria to one another is reflected by a set of weights. There are different techniques available to assign weights to the criteria (18-20). The assignment of weights to the set of project criteria is a critical part of any evaluation as it establishes the relative importance of each objective. These weights have a major effect on the final evaluation results. In some cases a slight variation of these weights can yield another ranking of the alternatives under consideration. In transit planning because decision making involves more than one interest group, it is generally useful to device one or more sets of weights for each group involved. The best alternative will be the one that ranks higher than the others for all of these weight sets. These weights have to be normalized using

$$w_j/\Sigma w_j$$

for $j = 1, \ldots n$, and summation over $j = 1, \ldots n$.

The other steps in the concordance analysis involve developing the concordance and discordance sets, calculating the concordance and discordance index matrices and dominance values, and ranking the alternatives. These steps have been discussed in detail in Giuliano et al. (10) and Giuliano (15) and are not repeated here because of space limitations.

CONCORD-NL (computer program)

A computer program has been written for the concordance analysis procedure discussed here. The program is an interactive program written in FORTRAN 77. It is operational on the Cyber 180/855 at the Academic Computing Center at the University of Washington. The program can handle 30 alternatives, 30 criteria, and 10 sets of weights. Data input can be made either interactively or from disk files. Alternatives can be added to or deleted from the analysis. The program allows the user to use any one of the three normalization methods for each criterion. If the nonlinear normalization method is chosen, the user must define the parameters of the normalization curve to fit the criterion's characteristics.

APPLICATION OF CONCORDANCE ANALYSIS

Concordance analysis was used to evaluate a set of alternative transit system designs developed for a problem assigned to senior students from the Civil Engineering and Urban Planning departments as a class project. They were asked to develop a high-performance transit system design for the 1-hr a.m. peak period. Each design team was required to maximize all benefit criteria while minimizing all cost criteria. The final (best) designs submitted by these students were evaluated using concordance analysis.

Students used an interactive graphic software system called the Transit Network Optimization Program (TNOP) to design their alternatives (21). TNOP can be used to generate many alternative designs quickly. It calculates values for various key performance measures that can be used for evaluation purposes. In addition to this, graphic maps and charts are also generated. TNOP can be used to design and calculate the performance of alternative fixed-route, fixed-schedule bus and rail transit systems. Through interactive graphic computing, TNOP allows the user to generate a wide range of design alternatives and to easily compare their performance characteristics.

The transit network and the origin-destination patterns are shown in Figure 3. This figure represents 97 nodes and 181 two-way bus links. The transit origindestination matrix contains 41,370 trips (for the 1-hr a.m. peak period). The students initially used another interactive graphic program called FLOWMAP to study the commuter flow patterns (22). FLOWMAP is operational at the University of Washington and produces a wide variety of origin-destination maps. It provides the user with the ability to examine the spatial pattern of the origin-destination data much more effectively than is otherwise possible. Based on this information, students used various modules of TNOP to define the transit lines, specify the service allocated to each route, assign the trips, look at the overview statistics of the system and routes, review the network loading and transfer patterns, and execute the timetable optimization subroutine. The performance values from the best designs developed by the students were used to conduct the concordance analysis.

EVALUATION OF ALTERNATE DESIGNS

The first step in the evaluation process is to select the objectives and the criteria that will represent the objectives. The selected objectives should be helpful in evaluating alternative transit designs and should also be able to represent the different

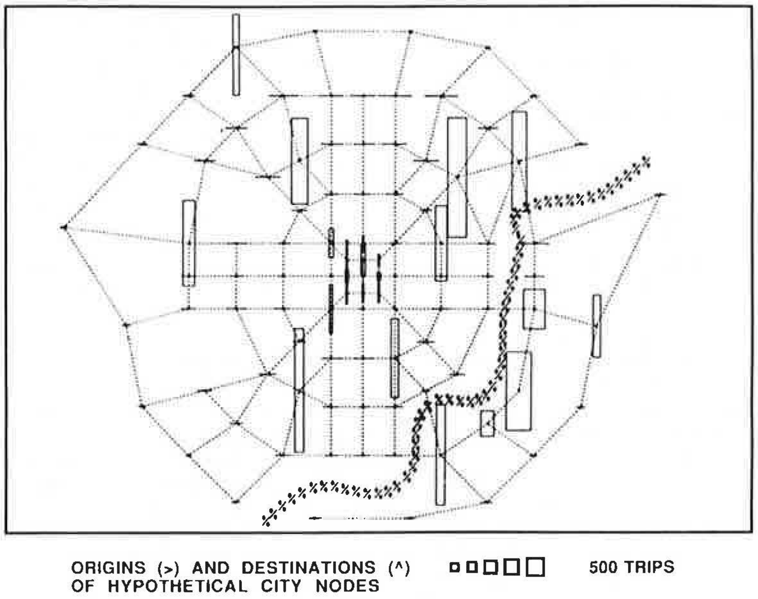

FIGURE 3 Origins and destinations of hypothetical city nodes.

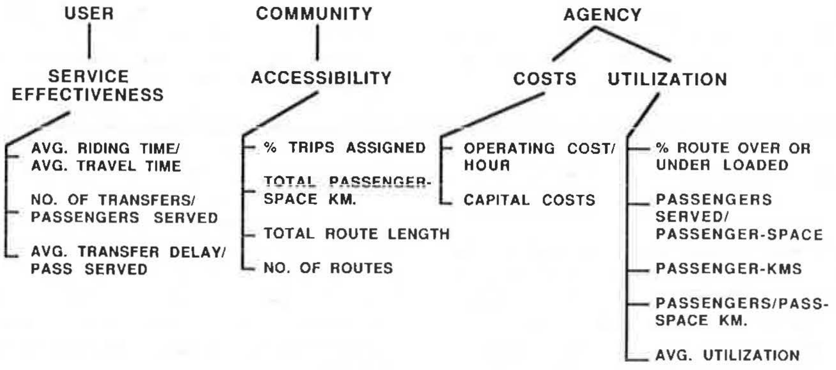

FIGURE 4 Objectives and criteria.

perspectives of the evaluations. After a careful review, the following objectives were chosen:

1. Minimize cost,
2. Maximize utilization,
3. Maximize accessibility, and
4. Maximize service effectiveness.

The next step is to select criteria that will represent the objectives chosen for the evaluation. Much care has to be exercised in this step because it defines the evaluation approach in detail. Figure 4 shows the objectives and criteria and the relation between them in a hierarchical tree form. A brief definition of the criteria used in this study is presented here as follows:

1. Operating costs per hour: These include administrative costs, operating and maintenance costs of vehicles, and crew costs.
2. Capital costs: These include all capital costs including vehicle cost and permanent structure cost.
3. Percentage of routes not within 10 percent of capacity: This includes the percentage of routes where the maximum load is not within 10 percent of the total capacity provided on a route for the 1-hr a.m. peak period.
4. Number of passengers served per passenger space provided: This is calculated by dividing the total number of passengers assigned in the system by the total passenger space service provided. The total passenger space is calculated by adding total capacity of seats and standees for all routes in a 1-hr period.
5. Passenger kilometers: The volume for each link is multiplied by the length of corresponding link, and the passenger-kilometers values for all links are summed.
6. Number of passengers served per passenger space kilometers of service: This includes the total number of passengers served over the total passenger space kilometers of service. Total passenger space kilometers are calculated by summing the product of total capacity of vehicles by the route length for all the routes.
7. System-wide average utilization: The average utilization on each route is first calculated by dividing the average volume on the links by the total capacity of the route. The total capacity is calculated by multiplying the total vehicle capacity by the number of trips made. The average volume of a route is determined by summing, for all links on that route, the product of the link volume and the link length and dividing the result by the line length.
8. Percent trips assigned: This is the total number of trips that could reach their destinations using a design over the total number of trips in the network that need transit service.
9. Total passenger space kilometers: This is the sum of the product of total capacity of each route by its length over all routes.
10. Total route length: This is the sum of link lengths included in each route, for all routes.
11. Number of routes: This is the total number of routes defined in the system.
12. Average riding time/average travel time: The riding time is the in-vehicle time. In TNOP, it is the product of the link volume values and link travel times, summed over all links and divided by the number of assigned trips. The average travel time is the summation of average riding time, average wait time, average transfer time, and average walk time.
13. Number of transfers per passengers served: This is the total number of transfers divided by the total number of passengers served.
14. Average transfer delay per passenger served: This is the total average transfer delay (in minutes) in the system divided by the total number of passengers.

Table 1 gives the project effects matrix for the five alternative designs included in the evaluation. By looking at the project effects matrix, it is difficult to get any idea as to which design is superior. Some of them are less expensive, but others are better in other respects. There is no single alternative in this set that dominates all the others for all criteria. Concordance analysis can be extremely useful in a situation like this. The next step is to choose the normalization methods for each criterion by choosing shapes for their utility curves. Table 2 gives the normalization methods and parameters selected for these 14 criteria. All the attributes using physical characteristics use linear methods and the others use the nonlinear method of normalization. Figures 5-8 show the shapes of utility curves for these criteria. Table 3 gives the normalized project effects matrix. In this matrix, higher values are better for all criteria. Still, given these data alone, it cannot clearly be determined which design is superior because the relative importance of the criteria differs.

OBTAINING WEIGHT SETS

To use concordance analysis, one needs to get a few sets of weighs together that will represent different decision makers' values and perceptions. For this study, a hierarchical comparison technique was used. This technique assumes that, at each branch of the tree, all the factors contributing to the worth of a higher level element have been identified. Decision makers are then asked to judge the relative importance of the contribution of each lower level element to the one above by dividing a constant sum among

TABLE 1 Raw Project Effects Matrix

	Alternative				
Number and Performance Measure	1 (100)	2 (200)	3 (300)	4 (400)	5 (500)
1 Operating costs per hour	27,712.000	28,980.000	33,014.000	35,081.000	30,132.000
2 Capital costs	1,576.000	2,242.000	770.000	2,359.000	1,777.000
3 Percent routes within 10 percent capacity	14.000	0.000	44.000	14.000	10.000
4 Passengers served/passenger space provided*	1.910	1.490	2.170	1.580	2.000
5 Passenger kilometers*	550 700.000	882 700.000	581 300.000	754 900.000	519 200.000
6 Passengers per passenger space kilometers*	0.029	0.026	0.030	0.018	0.031
7 Average utilization*	41.600	57.100	44.600	54.200	40.700
8 Percent trips assigned*	86.900	95.900	92.900	96.100	89.400
9 Total passenger space kilometers*	1 235 972.000	1 523 584.000	1 276 182.000	2 175 817.000	1 220 478.000
10 Total route length*	439.000	458.000	557.000	475.000	654.000
11 Number of routes*	7.000	10.000	9.000	11.000	10.000
12 Average riding time/travel time*	0.640	0.750	0.710	0.750	0.710
13 Number of transfers	0.630	0.510	0.580	0.610	0.520
14 Average transfer delay/passenger	2.510	2.060	2.080	1.060	2.410

Note: * = a more-is-better system; otherwise assume a less-is-better system; and ( ) is the design number.

TABLE 2 Normalization Method and Parameters

		Attribute Value[a]		Worth Curve Value[a]	
Number and Performance Measure	Method	Ideal	Acceptable	Ideal	Acceptable
1 Operating costs per hour	3	25,000.00	30,000.00	0.95	0.20
2 Capital costs	3	1,500.00	2,000.00	0.95	0.20
3 Percent routes within 10 percent capacity	3	15.00	25.00	0.80	0.20
4 Passengers served/passenger space provided	2	0.00	0.00	0.80	0.20
5 Passenger kilometers	2	0.00	0.00	0.80	0.20
6 Passengers/passenger space kilometers	2	0.00	0.00	0.80	0.20
7 Average utilization	3	55.00	50.00	0.80	0.20
8 Percent trips assigned	3	95.00	90.00	0.80	0.20
9 Total passenger space kilometers	2	0.00	0.00	0.80	0.20
10 Total route length	2	0.00	0.00	0.80	0.20
11 Number of routes	2	0.00	0.00	0.80	0.20
12 Average riding time/travel time	3	0.75	0.70	0.80	0.20
13 Number of transfers	3	0.25	0.50	0.80	0.20
14 Average transfer delay/passenger	3	1.50	2.00	0.80	0.20

Note: Method 1 is a magnitude-scaled normalization, Method 2 is an interval-scaled normalization, and Method 3 is a nonlinear normalization—logit curve.

[a]Applicable to Method 3 only.

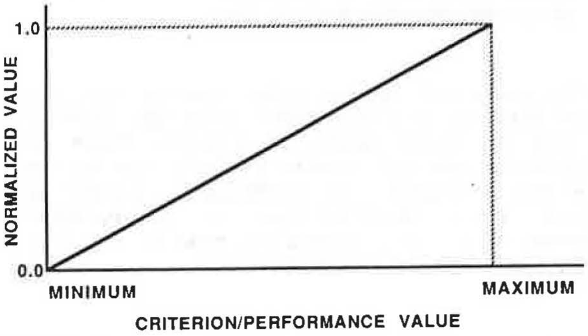

FIGURE 5 Structure of the utility curve used for passengers served per passenger space provided, passenger kilometers, passengers served per passenger space kilometers, total passenger space kilometers, total route length, and number of routes.

them. The process begins by weighting the highest level general goals and proceeding down the tree until the bottom level attributes are reached. Weights are calculated for each attribute by starting with the top value and forming a product of the values that appear at each branch as one progresses down the tree. Both the point division and weight computation may be conducted by working from the bottom to the top of the tree, if desired. Eight different sets of weights were developed using this method to represent different values and perspectives. The normalized weight set is given in Table 4. These eight sets of weights reflect four different perspectives as given in the following table:

Values and Perspectives Represented	Weight Sets
User	1,2
Transit agency	3,4
Community	5,6
Federal agency	7,8

RESULTS

Concordance analysis was applied to rank the five alternatives. The analysis included eight sets of weights. Table 5 gives the average dominance ranking, and the final ranking of the alternatives is as follows:

Rank	Alternative	Design Number
1	2	200
2	4 (not totally nondominated)	400
3	3 (not totally nondominated)	300
4	5 (not totally nondominated)	500
5	1 (not totally nondominated)	100

From the dominance ranking, one can find that alternative 2 got a total of 12.0 points (lower values are better because this number is the summation of the average ranking for all weighting sets) closely

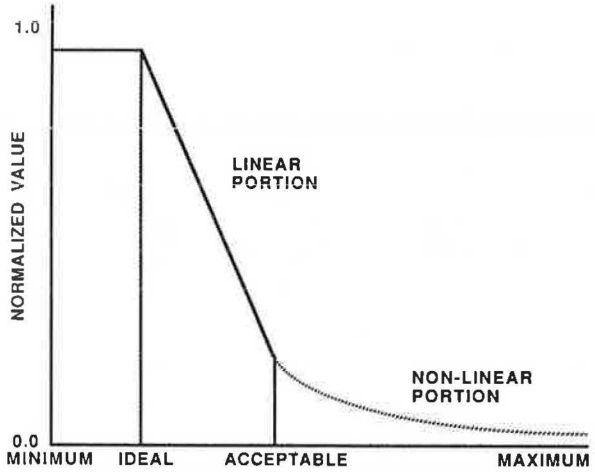

FIGURE 6 Structure of the utility curve used for operating cost per hour and capital cost per hour.

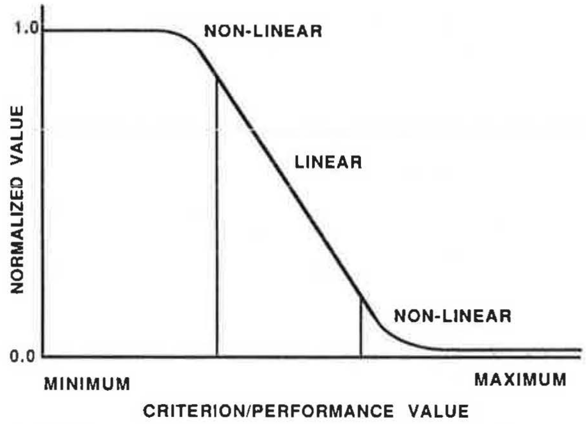

FIGURE 7 Structure of the utility curve used for percent routes not within 10 percent capacity, the number of transfers per passenger, and the average transfer delay per passenger served.

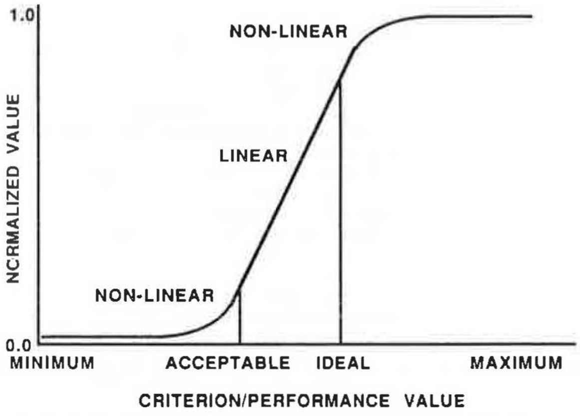

FIGURE 8 Structure of the utility curve used for system-wide average utilization, the percent of trips assigned, and the average riding time per travel time.

followed by 14.0 for alternative 4. Based on this, design 2 is ranked first and design 4 is ranked second. For the weight sets considered, alternative 2 is the only nondominated (superior) alternative. Alternative 2 is nondominated for all the weighting sets, whereas alternative 4 is nondominated for all sets except 3 and 4, which represent the values and perspectives of the transit agency. If weighting sets 3 and 4 are not considered, the results will be different, as given in Table 6, and the following table of the final rankings:

Rank	Alternative	Design Number
1	4	400
2	2	200
3	3 (not totally nondominated)	300
4	5 (not totally nondominated)	500
5	1 (not totally nondominated)	100

According to this ranking, alternative 4 ranks first with a total of 9.0 points and alternative 2 ranks second with 9.5 points. This illustrates the crucial importance of different perspectives and how they can influence the results of the analysis. In this analysis (without weight sets 3 and 4), both alternatives 4 and 2 are nondominated. Even though alternative 4 is better for many criteria that represent the objectives of the user and community, the cost of this design is the highest. Because of this, the heavy weights given by the transit agency for operating costs and capital costs prevent this alternative from being ranked first when the analysis includes all eight weighting schemes. These results illustrate how concordance analysis can be used to aid the evaluation of alternative transit system designs taking different values and perspectives into account in a more rigorous way than is the case with existing evaluation methods.

TABLE 3 Normalized Project Effects Matrix (more-is-better system)

	Alternative				
Number and Performance Measure	1	2	3	4	5
1 Operating costs per hour	0.645	0.377	0.018	0.003	0.182
2 Capital costs	0.908	0.030	1.000	0.011	0.633
3 Percent routes within 10 percent capacity	0.841	0.996	0.001	0.841	0.941
4 Passengers served/passenger space provided	0.618	0.000	1.000	0.132	0.750
5 Passenger kilometers	0.087	1.000	0.171	0.648	0.000
6 Passengers/passenger space kilometers	0.846	0.615	0.923	0.000	1.000
7 Average utilization	0.002	0.928	0.012	0.704	0.001
8 Percent trips assigned	0.043	0.868	0.548	0.880	0.152
9 Total passenger space kilometers	0.016	0.317	0.058	1.000	0.000
10 Total route length	0.000	0.088	0.549	0.167	1.000
11 Number of routes	0.000	0.750	0.500	1.000	0.750
12 Average riding time/travel time	0.009	0.800	0.320	0.800	0.320
13 Number of transfers	0.056	0.183	0.093	0.069	0.167
14 Average transfer delay/passenger	0.015	0.152	0.138	0.979	0.025

TABLE 4 Normalized Weights

Number and Performance Measure	Weighting Scheme 1	2	3	4	5	6	7	8
1 Operating costs per hour	0.028	0.030	0.125	0.150	0.088	0.053	0.072	0.120
2 Capital costs	0.042	0.030	0.125	0.150	0.038	0.053	0.048	0.180
3 Percent routes within 10 percent capacity	0.013	0.028	0.050	0.020	0.025	0.078	0.045	0.060
4 Passengers served per passenger space provided	0.039	0.028	0.050	0.060	0.025	0.020	0.027	0.030
5 Passenger kilometers	0.033	0.028	0.050	0.040	0.025	0.020	0.018	0.020
6 Passengers per passenger space kilometers	0.026	0.028	0.050	0.040	0.025	0.029	0.045	0.010
7 Average utilization	0.020	0.028	0.050	0.040	0.025	0.049	0.045	0.080
8 Percent trips assigned	0.090	0.075	0.075	0.135	0.125	0.180	0.100	0.090
9 Total passenger space kilometers	0.060	0.075	0.075	0.075	0.125	0.030	0.150	0.060
10 Total route length	0.090	0.075	0.075	0.060	0.125	0.030	0.150	0.060
11 Number of routes	0.060	0.075	0.025	0.030	0.125	0.060	0.100	0.090
12 Average riding time/travel time	0.200	0.175	0.063	0.070	0.150	0.120	0.040	0.060
13 Number of transfers	0.150	0.175	0.125	0.070	0.050	0.160	0.120	0.060
14 Average transfer delay/passenger	0.150	0.150	0.063	0.060	0.050	0.120	0.040	0.080

Note: Columns total 1.00.

TABLE 5 Average Dominance Ranking Using Eight Weighting Schemes

Weighting Scheme	Alternative[a] 1	2	3	4	5
1	5.00 (0)	1.50 (1)	3.00 (0)	1.50 (1)	4.00 (0)
2	5.00 (0)	1.50 (1)	3.50 (0)	1.50 (1)	3.50 (0)
3	5.00 (0)	1.00 (1)	3.00 (0)	3.00 (0)	3.00 (0)
4	5.00 (0)	1.50 (1)	2.50 (1)	2.00 (0)	4.00 (0)
5	5.00 (0)	2.00 (1)	3.50 (0)	1.00 (1)	3.50 (0)
6	5.00 (0)	1.50 (1)	3.50 (0)	1.50 (1)	3.50 (0)
7	5.00 (1)	1.50 (1)	3.50 (0)	1.50 (1)	3.50 (1)
8	5.00 (0)	1.50 (1)	3.00 (0)	2.00 (1)	3.50 (0)
Total	40.00	12.00	25.50	14.00	28.50

Note: (1) = nondominated and (0) = dominated.
[a]Includes concordance + discordance.

TABLE 6 Average Dominance Ranking Using Six Weighting Schemes

Weighting Scheme	Alternative[a] 1	2	3	4	5
1	5.00 (0)	1.50 (1)	3.00 (0)	1.50 (1)	4.00 (0)
2	5.00 (0)	1.50 (1)	3.50 (0)	1.50 (1)	3.50 (0)
3	5.00 (0)	2.00 (1)	3.50 (0)	1.00 (1)	3.50 (0)
4	5.00 (0)	1.50 (1)	3.50 (0)	1.50 (1)	3.50 (0)
5	5.00 (0)	1.50 (1)	3.50 (0)	1.50 (1)	3.50 (1)
6	5.00 (0)	1.50 (1)	3.00 (0)	2.00 (1)	3.50 (0)
Total	30.00	9.50	20.00	9.00	21.50

Note: (1) = nondominated and (0) = dominated.
[a]Includes concordance + discordance.

CONCLUSIONS AND SUGGESTED FUTURE RESEARCH TOPICS

Assessing alternative transit system designs is a typical multiobjective evaluation problem that has many conflicting objectives. Proposed in this paper is a simple but effective approach to solving such problems using concordance analysis. This technique can be applied to the evaluation of alternatives within a particular transit system or to the ranking of alternatives from different cities competing for federal funds. Concordance analysis could be a powerful aid to decision making for problems involving multiple objectives. It provides a logical but flexible approach. Application of this methodology to transit system evaluations allows for the inclusion of the different perspectives of multiple decision makers in the evaluation procedure in a well-defined scientific manner. Concordance analysis takes care of the problem of comparing different criteria measured in different units and having different importance ratings. This multicriteria evaluation technique is more transparent and easily understood because public and other groups can participate and express their opinions through weighting of criteria.

Further research is required in the areas of (a) criteria selection, (b) guidelines for selecting appropriate values to determine the shape of the nonlinear utility curves used to normalize the project effects data, and (c) weight definition and collection methods. Further research is also required to identify the basic objectives and measures that are most suitable for evaluating transit system alternatives is an alternatives analysis study. Depending on the problem, additional objectives and performance measures can be added. Because this is a difficult task and guidelines would be useful at every agency level, perhaps some region-specific sets of weights could be developed using hierarchical comparison methods to represent diverse but relevant perspectives. A large survey could be conducted among relevant decision-maker groups in different regions to get several sets of weights for different types of transit project evaluations. These results could be used by UMTA on a regional basis.

ACKNOWLEDGMENT

The computer program for the concordance analysis was written by Shieng-I Tung of the Civil Engineering Department at the University of Washington in Seattle. His efforts are gratefully acknowledged. (Please note that the authors extended his program to include the interactive and nonlinear normalization capabilities used in this study.)

REFERENCES

1. J.S. Dajani and G. Gilbert. Measuring the Performance of Transit Systems. Transportation Planning and Technology, Vol. 4, 1978, pp. 97-103.
2. G.J. Fielding, R. Glauthier, and C. Lave. Performance Indicators for Transit Management. Transportation, Vol. 7, 1978, pp. 365-379.
3. S. Mundle and W. Cherwony. Diagnostic Tools in Transit Management. In Transportation Research Record 746, TRB, National Research Council, Washington, D.C., 1980, pp. 13-18.
4. J. Holec, Jr., D. Schwager, and A. Fandialan. Use of Federal Section 15 Data in Transit Performance Evaluation: Michigan Program. In

Transportation Research Record 746, TRB, National Research Council, Washington, D.C., 1980, pp. 36-38.

5. G.J. Fielding, T.T. Babitsky, and M.E. Brenner. Performance Evaluation for Bus Transit. Transportation Research, Vol. 19A, No. 1, February 1985, pp. 73-82.
6. K. MacCrimmon. An Overview of Multiple Objective Decision-Making. In Multiple Criteria Decision-Making (J. Cochrane and M. Zeleny, eds.), University of South Carolina Press, Columbia, 1973.
7. J.L. Cohn and D.H. Marks. Review and Evaluation of Multiobjective Programming Techniques. Water Resources Research, Vol. 11, No. 2, 1975, pp. 208-220.
8. M. Aboul-Ela, A. Stevens, and F. Wilson. A Multiple-Criteria Decision-Making Methodology for Transportation Policy Analysis. The Logistics and Transportation Review, Vol. 18, No. 3, 1982, pp. 279-294.
9. T. Friesz, F. Tourreilles, A. Han, and J. Fernandez. Comparison of Multicriteria Optimization Methods in Transport Project Evaluation. In Transportation Research Record 751, TRB, National Research Council, Washington, D.C., 1980, pp. 38-41.
10. G. Giuliano et al. Making Transportation Corridor Investment Decisions Within a Multimodal Framework, Final Report. Institute of Transportation Studies, University of California, Irvine, Oct. 1983, 248 pp.
11. R. Neufville and R. Kenney. Multiattribute Preference Analysis for Transportation Systems Evaluation. Transportation Research, Vol. 7, 1973, pp. 63-76.
12. B. Roy and J. Hugonnard. Ranking of Suburban Line Extension Projects on the Paris Metro System by a Multicriteria Method. Transportation Research, Vol. 16A, No. 4, 1982, pp. 301-312.
13. D. Teodorovic. Multicriteria Ranking of Air Shuttle Alternatives. Transportation Research, Vol. 19B, No. 1, 1985, pp. 63-72.
14. P. Nijkamp and A. Delft. Multicriteria Analysis and Regional Decision-Making. Martinus Nijhoff Social Sciences Division. Leiden, The Netherlands, 1977, 135 pp.
15. G. Giuliano. A Multicriteria Method for Transportation Investment Planning. Transportation Research, Vol. 19A, No. 1, 1985, pp. 29-41.
16. H. Voogd. Multicriteria Evaluation for Urban and Regional Planning. Pion Limited, London, England, 1983, 367 pp.
17. J. Parsons. An Analytical Methodology to Aid Technical Decision-Making in the Transportation Planning Process. M.S. thesis. University of Washington, Seattle, 1973.
18. L. Thurstone. The Measurement of Values. University of Chicago Press, Chicago, Ill., 1967.
19. R. Eckenrode. Weighting Multiple Criteria. Management Science, Vol. 12, No. 3, Nov. 1965, pp. 180-192.
20. H. David. The Method of Paired Comparisons. Charles Griffin and Company, Limited, London, England, 1963.
21. N. Janarthanan and J. Schneider. Reducing the Energy Requirements of Suburban Transit Services by Route and Schedule Redesign. In Transportation Research Record 994, TRB, National Research Council, Washington, D.C., 1984, pp. 47-57.
22. J.B. Schneider. Mapping Origin/Destination Data: Now We Can "See" What's Going On Out There. ITE Journal, Dec. 1983, pp. 26-30.

Publication of this paper sponsored by Committee on Public Transportation Planning and Development.

Modeling MultiPath Transit Networks

DAVID B. RODEN

ABSTRACT

In analyzing transit investments, issues related to the distribution among access modes or competing routes are often critical to the evaluation. Presented in this paper is a method of transit path building that permits the consideration of multiple paths in mode choice and network loading. The technique is capable of subzone distributions at the access and egress ends of the trip as well as traditional mode-of-access distributions at the transit stops or stations. Included also is a description of the technique as installed in the Transportation Analysis Process (TAP) used by the North Central Texas Council of Governments (NCTCOG). The model builds the best, second-best, and third-best paths to each node in the network. Specific criteria related to access and egress links are used to select up to seven trip tables that are then loaded to the paths according to the same criteria. The model has been calibrated and used in regional and subregional planning applications.

The function of transportation planning models is to simulate travel behavior at a reasonable cost. The trade-offs between the complexities of human decision making and computer modeling have resulted in an established set of modeling constructs that adequately address most regional issues. These techniques have been embodied in the Urban Transportation Planning System (UTPS). There are many theoretical shortcomings in the UTPS package, but most professionals would generally agree that the complex programming and data processing required to resolve these shortcomings are not cost-effective. Consequently, the UTPS is a standard in the industry because it adequately simulates regional travel behavior at a reasonable cost and has technical support financed by the federal government.

As the UTPS process gained acceptance and widespread application, the planning emphasis shifted from regional to subregional issues. The alternatives analysis process created a need for comparative ridership forecasting in subarea planning studies. Large macro issues such as the major facilities in a long-range regional plan were replaced by often subtle and subjective distinctions among alternative technologies. To accommodate the demands for detailed forecasts placed on regional agencies by the federal government, planners turned to developing elaborate mode-choice models. In many areas, these refinements adequately addressed the important issues. In other areas, planners were less satisfied with the results. Adding an elaborate mode-choice model to the generalities and assumptions in the UTPS package seemed incongruent.

Presented in this paper are technical issues related to transit path building and a discussion on how these issues affect mode-choice and network-loading results. The ways in which various agencies have attempted to use the basic tools available in the UTPS package to improve the overall performance of modeling transit systems are described, and the advantages and disadvantages of these techniques are presented. The purpose of the paper is to demonstrate how a relatively simple improvement to the basic UTPS algorithm can overcome many of the problems associated with other techniques and can achieve that objective in a cost-effective and theoretically satisfying manner.

REGIONAL TRANSIT MODELING

In traditional modeling theory, transit trips are generated in a mode-choice model that evaluates the pertinent differences between the characteristics of a transit trip and a highway trip on a particular interchange. The trips are loaded onto a transit network to determine the ridership on a particular line. The UTPS process builds the best transit path between each interchange using the program UPATH. The characteristics of these paths are used in a mode-choice model to generate a trip table. This trip table is loaded to the transit network by the program ULOAD using the best path from UPATH.

There are several assumptions made by this technique that are worth noting. The first is that transit system capacity does not affect a traveler's path selection. Generally, this is a reasonable assumption. Line capacity is often well above ridership forecasts and most transit patrons have relatively few alternative paths available for their particular trip. If the programming difficulties and costs are considered, the decision to accept this theoretical shortcoming is understandable. The complexities of transit path building using trip segments (i.e., boarding and alighting pairs) do not lend themselves to the individual line-segment analysis required for capacity-constrained modeling.

A second assumption is that the characteristics of the best path are sufficient for mode-choice analysis. From a regional perspective, the level of detail required for adequate mode-choice analysis is primarily associated with line-haul characteristics. A regional zone structure is often aggregate enough to make transit access considerations impractical. The zones are so large that reasonable walk distances have little meaning as far as access or coverage concepts are concerned. The result is a regional mode-choice model calibrated with the explanatory variables available to a regional data base. From a path-building perspective, adequate line-haul information can be obtained from a best-path model.

The third assumption is that all transit trips between any two points will use the best path. The

---

D.B. Roden, JHK and Associates, 4660 Kenmore Avenue, Alexandria, Va. 22304.

logic behind this assumption has several dimensions. From a regional perspective, line-specific ridership is only an issue on large line-haul facilities such as busways, rail lines, or high-occupancy-vehicle lanes. The best-path loading will reasonably forecast large line-haul facilities. The access and feeder systems are generally ignored in regional modeling because they serve only a supporting role to the purpose of the study. If they are evaluated, it is at a line group or large area level of aggregation. At this level, the obvious inaccuracies of the individual components are averaged away and often show reasonable results.

From the perspective of cost, best-path loading is a practical reality. Path building--transit path building in particular--is an expensive endeavor. Transit path building is complicated to the critical influence of transfers on path selection. The decision of which link to take next is dependent on the mode and line of the current link. As mentioned earlier, this requires that transit path building be organized around trip segments rather than links. The relatively few links in a transit network are expanded into a large set of potential boarding and alighting pairs before path construction. The number of options that the program must consider is related to the permutation of the number of stops on each line. Therefore, it is highly desirable that transit paths be built only once and used directly in trip loading.

Even if cost is not an issue, there are few benefits in using a multiple path concept on a regional planning study. The only issue that may be important is the mode of access at line-haul stations. Concerns related to the forecasts prepared for several new rail projects in this country have focused attention on the mode-of-access assumptions. Even with a multiple path model, mode of access cannot be accurate on a regional zone structure without a relatively sophisticated concept of zone coverage in the mode-choice model as well as the path-selection process. Controlling the mode-of-access results in the network coding phase of the study is perhaps a more cost-effective solution to the problems that have occurred. An awareness of the bias that is created by inaccurate or inappropriate coding is a major step toward minimizing problems associated with the mode-of-access elements of a regional model.

For regional transit modeling, the traditional modeling systems, such as UTPS, achieve the primary objective of transportation modeling. The best-path algorithm can provide adequate results for the majority of regional issues at a cost that is compatible with the accuracy required for mode-choice and network-loading procedures.

SUBAREA PLANNING ISSUES

The purpose of subarea planning is to enable accurate forecasts to be made at a level of detail beyond that which is advisable from a regional modeling context. Subarea modeling is the basic process of developing a zone structure compatible with the level of detail of the network to be evaluated. Subarea types of analysis could be performed at the regional level if the network size and total number of zones could be cost-effectively processed. The problem is that the cost of computer processing is closely related to the square of the number of zones. If cost were not an issue, the difficulties in managing the space and core requirements for such large data bases or the human elements of error and limited comprehension make detailed regional planning inadvisable.

A natural outgrowth of these concerns is subarea planning. The problem size is limited so as to be manageable yet detailed enough to produce the needed results. In other words, the basic modeling process is applied on a smaller geographic area. For highway planning, this presents no major difficulties. Some adjustments must obviously be made to trip distribution relationships and correction factors, but this is exactly the purpose. Subarea planning affords the modeler the opportunity to refine regional relationships to more accurately address the area-specific characteristics. The objective is to produce a better forecast in the area of interest. Beyond model validation, there is no significant theoretical difficulty in using a capacity constraint procedure developed for modeling freeways and major arterials to forecast traffic on minor arterials and collectors. It may be desirable to modify volume-delay relationships on low-capacity facilities, but it is not theoretically necessary.

Unlike highway planning, transit planning at the subarea level is not theoretically compatible with regional modeling techniques. Subarea transit issues focus greater attention on the mode of arrival and local service elements of the system. The performance characteristics and distributional concerns of these subsystems are different from the line-haul characteristics of major routes. At the subarea level, it is no longer possible to ignore the submode access and coverage concepts. The mode-choice and path-building models need to incorporate these elements if they are going to be used to forecast demand for each component of the transit system.

In order for a subarea model to accurately estimate demand for transit subsystems, the implications of walking to or from a transit facility must be explicitly incorporated into the model. The two basic dimensions of the walk choice are walk distance and drive opportunities. These are more commonly called walk coverage and mode-of-access issues. Walk coverage is defined as those trips for which a reasonable walk (i.e., 0.5 mi) is available to and from the transit facilities. Mode of access is associated with the subchoice between walking or some form of driving such as park-and-ride, kiss-and-ride, and pool-and-ride.

Figure 1 shows a typical mode-of-access and coverage subsystem. The figure shows two feeder bus lines serving a rail facility. The zone in question has been connected to the network through a drive approach to station A and two walk links to nodes B and C. Unless significant bias factors are introduced into the path-building algorithm, the best path to the rail line will always use the drive approach to station A. The time to walk and wait for the bus, travel by bus, and transfer to the rail line will invariably be worse than driving to the station. A one-path model will evaluate the mode choice based on the drive access and will load all trips to this

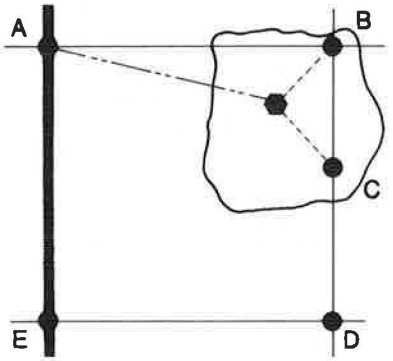

FIGURE 1 Subsystem example.

access link. If this example is typical of other zones in the vicinity of station A, the mode-of-arrival distribution will overestimate driving and underestimate feeder bus.

In the preceding example, the best path was selected irrespective of the characteristics of the zone. In all likelihood, the mode-choice model did consider the characteristics of the zone in choosing between the transit and highway paths. These characteristics may include such aspects as income, automobile ownership, or household size. If the mode-choice model is calibrated to consider the socioeconomic characteristics of the zone in combination with the access mode of the path, the overall demand for the path can be significantly biased. For example, assuming the zone represents a depressed area with little automobile ownership, the fact that the best access to transit is automobile-related may cause the mode-choice model to underestimate the transit demand from that zone to all paths using the rail line.

If the best path from the zone in Figure 1 includes the walk to node C and the bus line from node C to node D, the one-path process will load transit trips to this path. Here again, the path was selected irrespective of the characteristics of the zone. In particular, the mode-choice model is not informed about how many of the people in the zone can actually walk to node C. The assumption is that all households are within a reasonable walking distance of that location. If the zone is large, this assumption is incorrect. Figure 2 shows the coverage areas for the walk connections at nodes B and C. For all of the shaded area surrounding node C, the path of choice is the best path. For people living outside the coverage of node C but within the coverage of node B, the path of choice involves walking to node B and riding from node B to node C to node D. For the portion of the zone not covered by nodes B or C, the only path option is to drive to station A, ride the rail line to node E, and transfer to the bus serving nodes E and D.

The preceding example suggests that there should be at least three paths from the zone to node D. Each of these paths serves a different constituency and has a different probability of choosing transit. The same concept could easily be extended to the previous discussion related to access to the rail station A. The model should consider the drive to station A as well as the walk paths using nodes B and C. This would permit the mode-choice model to distribute the access among the drive and walk options based on the actual differences in the paths as well as the socioeconomic characteristics of the zone. The access distribution should then be loaded to the appropriate paths for performance analysis. The results of the model would, therefore, show a smoother and more logical generation and assignment of trips on the various access options and ultimately on the transit lines.

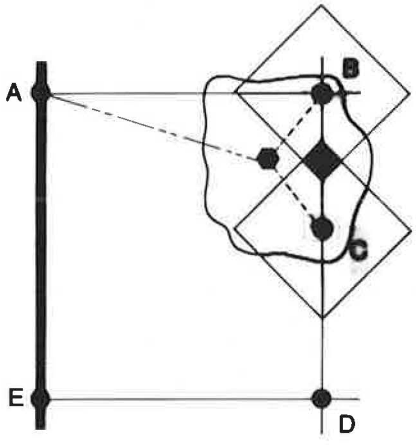

FIGURE 2 Walk coverage example.

MULTIPATH MODELING USING UTPS

Several techniques have been developed to perform transit subarea planning using UTPS models. Each of these techniques attempts to resolve the problem discussed in the previous section. In this section, each technique will be presented along with a discussion on its advantages and disadvantages. The critique will focus on two measures of effectiveness. The first is the ability of the procedure to address the shortcoming of the one-path approach. The second is the cost-effectiveness of the procedure. In other words, does the technique produce reasonable results on a reasonable schedule for a reasonable cost?

Zone-Structure Techniques

The first technique is perhaps the most obvious. It attempts to remove the need for multiple paths by increasing the zone detail in the vicinity of the facilities in question. This is a normal part of subarea planning. The difference is that the number of zones is not a function of the transit system as much as it is a function of the access issues. Each zone must, therefore, be small enough to reduce the walk options to a single choice. If a walk distance of 0.5 mi is assumed, the zones cannot exceed a 1-mi^2 area. In areas where parallel service exists, the zones must be divided so as to separate the access between the two lines.

The detailed zone-structure technique attempts to reduce the need for market segmentation and coverage considerations. It will smooth the assignment by providing more detailed and frequent access points. All people within the zone are by definition within walking distance of the access point, so no coverage analysis is needed. Those areas without walk access are provided drive opportunities or no access at all. The technique cannot resolve the distribution between drive and walk options for a particular interchange. It can only smooth the results by performing more frequent analysis. It is also rigid and time-consuming to construct. For this technique to work, the zone structure must be network-specific. Each network alternative would require a modified zone structure. The computer costs associated with path-building and mode-choice analysis for a large number of zones are exorbitant. The technique improves the results at the expense of time and computer resources.

Mode-Choice Techniques

The second technique is one that attempts to address all of the access issues within the mode-choice model. In this approach, the best path is modified before the mode split. The access portions of the path are stripped away according to various criteria. Only the line-haul characteristics of the path remain. The access alternatives are derived by a separate procedure and are evaluated alongside the line-haul characteristics by the mode-choice model. These access alternatives generally include identification of walk and feeder bus options and alternative park-and-ride opportunities. They also include an evaluation of zone coverage and average walk distances. These data are generally prepared by hand and are fairly detailed in nature.

The mode-choice model is provided with all of the basic data needed to conduct the detailed submode analysis. The model can be structured with several nests of drive, walk, or feeder bus options, the trips can be segmented into those with walk opportunities and those without, and detailed reports of the submode analysis at each zone can be produced. What the model cannot do is guarantee that the selected options were actually available for any particular trip. The zone-related access data are not easily correlated with the line-haul path. But perhaps more important, the results of this detailed analysis are never assigned to a transit network. All of the trips on a particular interchange are loaded onto the best path. In other words, the final result has a better estimate of transit trips, but the ridership on any particular line does not show how the trips were actually made. Only through complicated hand analysis is it possible to adjust some of the results to reflect the mode-choice distribution. The process requires considerable time to (a) prepare the access inputs needed for mode choice and (b) hand-adjust the network loading in exchange for a presumably better estimate of total transit demand. This procedure does not produce reasonable network results and the cost of time may be exorbitant.

Multiple-Path Techniques

Perhaps the most comprehensive approach is one that constructs alternative paths, uses them for mode-choice analysis, and loads the corresponding trip table to each path. This can be done with UTPS by selectively adjusting the parameters in UPATH to generate the desired path. The cost of running UPATH generally restricts multiple-path considerations to the best walk path and the best drive path. The characteristics of the two paths are used by the mode-choice program to distribute trips between walk and drive options and to improve the estimate of the automobile-versus-transit probability. The transit share is split into walk and drive trip tables to be loaded to the two networks by ULOAD. The two assignments are merged to produce the final result.

The use of several minimum paths performs particularly well at distributing trips between drive and walk options. It does not distribute trips among several drive or several walk options. The assumption is that all travelers can and will take the best path. For large zones or dense networks, the distribution among walk or drive options, or both, can be important. In fact, the alternative walk or drive paths may be more attractive than the opposite mode option. The distribution among walk paths is also coverage-dependent. The combination of coverage and path is necessary for a smooth and logically distributed assignment. Smaller zones can help to reduce these concerns but that raises the cost. This process is extremely expensive from a computer resource point of view. Increasing the number of zones would make it much more costly. If care is taken and enough time and computer resources are available, this method can work.

Postprocessing Techniques

A postprocessing technique is a way of adjusting the results to reflect access issues. It assists the hand adjustments that are necessary to smooth and rationalize the performance summaries and ridership estimates. In this approach, the access components of the paths are stripped after loading. The ridership is distributed among the alternative access options by mode- and distance-choice relationships derived from observed data. The access options are developed from the network data and hand-coded paths. The process is generally limited to transit stations because each zone and line-haul access combination must be addressed individually.

The postprocessing approach does not improve on the overall modeling process; however, other techniques could be used in conjunction with postprocessing to improve the overall results. By itself, there is no correction for access or coverage issues made to the estimate of total transit demand. If the model does adjust demand it will only adjust the access legs and not the line-haul legs. The fact that the process is zone-to-station-related makes it less practical for improving the local and feeder bus components of the transit system. The approach is, however, a relatively inexpensive solution to the station access issues faced by many studies.

A NEW APPROACH TO TRANSIT MODELING

The preceding concerns led the author to formulate a new approach to transit modeling. The approach that was selected resolved many of the problems previously mentioned and maintained the objectives of cost-effectiveness.

The approach takes maximum advantage of an aspect peculiar to transit path building--that of legs. Unlike highway paths where each subsequent link is independent of the previous link, transit paths are dependent. Because of this fact and the logic of a path-building program, the transit system is converted from links to legs. A leg is defined as a trip between a potential boarding and alighting sequence. By converting the network to legs, the path builder can assume that selecting a leg will require a boarding and thus a transfer. A transit path is a short sequence of legs constrained by the maximum number of transfers permitted.

The technique involves a traditional UPATH-like minimum-path-building exercise. As the best path is being built, alternate path information is stored. The key to the process is that the second- and third-best paths to any particular node are controlled by their association with zone connectors (i.e., mode-of-access alternatives). In other words, a path is only considered an alternative to the best path if it serves a different access or egress location (i.e., a different part of the zone) or a different mode of access. In this way, extraneous alternative paths are eliminated. Because the transit paths involve only a few legs, the computational efficiency is not compromised when checking the access link of a potential alternative. The assumption is that the leg or the previous leg must be a zone connector for consideration in the alternative path table.

The result of this technique is a series of alternative paths to intermediate nodes on the best path. This means that each realistic access location and mode serving a particular interchange is made available for consideration by the mode choice model and to the transit-loading program. The trip tables associated with the path alternatives are interchange-specific and, therefore, are appropriate for loading the logit distribution of the access alternatives of that interchange. The mode-choice model can address the distribution both between modes and among potential access points simultaneously as fully dependent alternatives.

Because egress options are also considered, a distribution of destinations within the zone is developed. In addition, the characteristics of the trip to the egress alternatives are included in the analysis. The egress alternatives may include line-haul paths different from that of the best path. In this way, multiple line-haul options are considered. Using the same argument on the various access locations reveals an additional source of multiple path alternatives.

It must be noted that no effort is made to force the consideration of all modes of access or line-haul options. This is not, however, a weakness of the technique, but a strength. In a technique that finds the best walk path and then the best drive path, the data about the second-best walk path and the second-best drive path are ignored. The mode-choice model will only compare the two best paths. When the second- and third-best paths are developed (regardless of mode), the truly bad paths are never selected and therefore are not considered by the mode-choice model. The prescreening of paths keeps the mode-choice model from assigning trips to unrealistic alternatives while, at the same time, concentrating the analysis of coverage and opportunity on all viable alternatives. The mode-choice model can, therefore, assign trips to more than one park-and-ride lot or more than one feeder bus line that is appropriate for the particular interchange.

This approach coordinates multiple-path and mode-of-access alternatives through path building, mode choice, and loading. It requires only one pass through the path-building algorithm and is therefore relatively inexpensive. It serves the needs of mode-choice modeling and produces realistic distributions of ridership profiles even with large zone sizes. The approach serves the needs of the planning community at a reasonable cost.

A MODEL DESCRIPTION

The previous modeling approach has been installed and applied. The latest version of the TAP developed by the NCTCOG includes the multiple-path transit networking techniques presented herein. The approach was developed in direct response to the needs of the Dallas Area Rapid Transit (DART) staff for accurate mode-of-access data at rail stations. The forecasts were performed on a regional forecast zone system of 800 zones. The access distribution within the large zones; between competing stations; and among walk, drive, and feeder bus modes was critical to the analysis. The model that resulted is described in the paragraphs that follow. It has been calibrated and applied with reasonable success.

The transit path-building algorithm used in the TAP model is a typical best-path technique. The minimum cumulative impedance path from each origin to all destinations is determined by the "bush" method of path building. The leg impedance is a function of travel time, distance, cost, waiting time, level of service, and a link-specific bias factor. Each impedance parameter varies by mode and is cumulative. The value of a transfer to a particular mode and the transfer costs are added to the impedance as the path is being built. Mode-to-mode transfer prohibitions and the total number of transfers are also considered during path building.

A transit path can typically be described with only three to five transit legs. The relatively few legs that represent the best path and the numerous alternative legs that serve the same path are used by the path-building program to construct up to seven alternative paths. The second- or third-best path to a node is used in conjunction with a set pattern of access and egress alternatives to define the alternative paths. The first path is the best path. The second through fourth paths are constructed from the set of second-best paths to nodes along the best path. The order of inclusion is

1. The first alternative path closest to the destination that is, or whose next leg is, a zone connector.
2. The next alternative path after the first mentioned in item 1 that is, or whose next leg is, a zone connector.
3. The first alternative zone connector at the destination.

The first two alternatives approximate a distribution of access links and the third alternative is an egress option. Paths five through seven use the same inclusion technique with the third-best paths to each node along the best path.

Figure 3 shows an example of the path-building logic. The best path is the drive connector (mode 2) path Z1-A-C-D-Z2. The first alternative path diverts at the last node with an alternative path whose leg or previous leg is a zone connector. In this example, the second path is Z1-B-C-D-Z2 because B-C is the last alternative path. The third path would be the walk connector (mode 1) path Z1-A-C-D-Z2 because this is the second-to-last alternative path whose leg or previous leg is a zone connector. The fourth path is the egress option Z1-B-E-Z2.

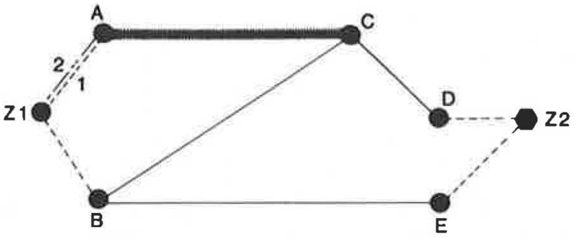

FIGURE 3 Path-building example.

To take this example to the next logical step, Figure 4 adds the third-best path options to the network shown in Figure 3. The best path is Z1-F-C-D-Z2. The second-best alternative path from the last node with an alternative path is Z1-F-C-D-Z2. This would be used as the fifth path. Because there are no other logical paths from node A, the sixth path would be missing. In other words, not all interchanges will have seven path options available to them. The seventh path would be the egress option Z1-B-G-Z2.

After the best and alternative paths are constructed, the path summary files and reports are generated. The node-and-mode string representing the best path is stored for path loading. The second- and third-best alternate branching nodes are saved as needed. From these three arrays, up to seven paths are reconstructed during path loading. The zone-to-zone summary files are also generated. The mode-choice model requires, at a minimum, the cumulative impedance and the access codes for each path. Access codes include the access mode and link number, the

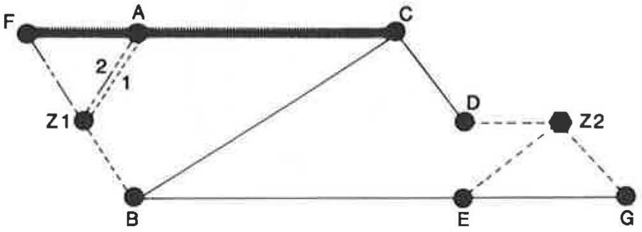

FIGURE 4 Additional path alternatives.

first transit mode, the last transit mode, the level of service, the principal mode (i.e., the mode with the greatest cumulative contribution to distance), and the number of transfers. The mode-choice model may optionally require in-vehicle travel time, distance, cost, or out-of-vehicle travel time skims. These data are available only for the best path.

The mode-choice model is an aggregate nested logit model with accessibility segmentation. Each origin-destination interchange is first evaluated at the transit and highway submode level. A combined utility is then used to determine the highway-versus-transit shares. The highway and transit share is then distributed among the appropriate submodes. For the purposes of this paper, the remainder of the discussion will focus on the implications of the multiple-path algorithm on the mode-choice and path-loading programs.

The accessibility segmentation process involves dividing the trips between the share of the zone that is accessible to transit by walk or drive and that which is only accessible by driving. The walk coverage is the sum of the coverage of each unique walk connector identified by the seven paths. A separate sum is made for local and express-mode first boardings. The access walk links are also summed independently from the egress walk links. The sum of the local mode coverage at the access zone is tested against the maximum allowable local coverage for that zone. The express mode at the access zone and the local and express egress coverage are likewise compared with the appropriate maximum coverages. If any of the maximums are exceeded, the coverage of each link using that particular access or egress class is factored down to the maximum. The resultant coverage for any particular interchange will not exceed the maximum by access or egress and local or express categories. The walk-access coverage is the maximum sum of the local and express options for each walk link. The egress coverage is the corresponding sum at the destination zone. The maximum number of trips on the interchange covered with walk access is the minimum of the walk access and egress coverage.

Figure 5 shows an example of the walk-access calculations. The path builder used three walk connectors in constructing the seven paths. The shaded coverage areas for nodes A, B, and C are summed as an estimate of total coverage. In this example, the coverage beyond the zone boundary and in overlapping areas should be subtracted from the estimate of total coverage. This is done by factoring the total coverage back to the maximum coverage permitted for the zone.

In the example shown in Figure 5, the best path used node B. Because the coverage for B is not larger than the total coverage for the zone, the transit utility for the best path will only apply to the area covered by B. The utility experienced by people traveling from A or C will be different from B. The weighted average utility is the utility of the path through B, weighted by the full area it covers, and the utilities of A and C, weighted by the total area minus the area of B, divided by two.

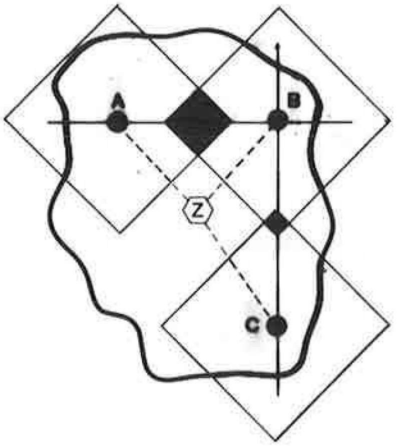

FIGURE 5 Walk coverage and utility.

The utility of the walk access is determined from a composite of the coverage and utilities of each access link. A comparison between the best single access link coverage and the total walk-access coverage is first made. If the single coverage is equal to the total, the walk utility is the utility of the best walk path. If the single coverage is less than the total coverage, the walk utility is based on the ratio of the single coverage to the total coverage. This ratio multiplied by the best walk-path utility is added to a percentage of the remaining walk-path utilities. The percentage is 1 minus the ratio divided by the number of additional walk-access links. This method attempts to capture a weighted average utility based on overlapping coverage.

The composite walk utility and coverage and the drive utility and coverage are compared with the composite highway utility in two parts. The first part represents that portion of the zone that has walk and drive options. The number of trips affected is equal to the walk coverage of the interchange. The second part represents that portion of the zone that has drive access but no walk access. The number of trips affected is equal to the positive difference between the drive coverage and the walk coverage.

The transit share from each segment is proportioned back to the appropriate paths according to its contribution to the segment utility. For the walk and drive segment, the trips are first divided among walk and drive paths accorded to the composite walk utility and drive utility. The drive share is added to the transit share from the drive-only segment to obtain the total drive share. The trips on each walk or drive link are distributed according to their share of the total walk or drive utility. Trips are also proportionally divided among the paths using a common link. The final result is a trip table for each path. Because the number of transit trips divided among seven paths on each interchange will generally be small, the trips are stored in hundredths of trips to avoid round-off error in the trip tables.

The seven trip tables from the various trip purposes are summed and loaded to the paths constructed by the path-building program. The trips are first posted on each leg of the corresponding path. The node-and-mode sequence of the best, second-best, and third-best paths to each node are traced according to the access and egress mode criteria previously discussed. Data regarding the node numbers, mode, previous mode, next mode, and volume are stored for each leg of each path. The one-way leg file is then merged with the legs of each path. Access and egress mode distributions are saved for each leg in the network. The access modes include walk, drive, bus, express bus, and rail. The egress modes include walk, bus, express bus, and rail. The result is a single-leg record of all lines of that mode with a distribution of boarding and alighting transfer activities. The combined leg is distributed to the line legs according to the proportion of each leg's service rating relative to the sum of the weight of all legs. The leg data are then summed and posted on each link of the line. Reports are generated that summarize the ridership in both directions on the link according to boarding and alighting activities at each node.

CONCLUSION

The technique for multipath transit network analysis as presented in this paper and as installed in the TAP is a significant improvement to the generally accepted algorithms. It provides substantially more data for mode-choice modeling and is capable of posting the results of that analysis on individual lines. It also handles the mode-of-access issues related to subzone distributions and competing access and egress locations. These improvements are cost-effective. The model applications developed for the TAP process are no more expensive to use than a single application of the UTPS counterparts. The technique achieves the objective of improved theoretical modeling at a reasonable cost of time and computer resources.

The opinions and viewpoints expressed in this paper are those of the author and do not necessarily reflect the viewpoints, programs, or policies of any federal, state, or local agency.

Publication of this paper sponsored by Committee on Public Transportation Planning and Development.

Estimating Cost Savings Attributed to Improvements in Railcar Reliability and Maintainability for the Chicago Transit Authority

W. DIEWALD and D. MUOTOH

ABSTRACT

The findings of an analysis of railcar fleet reliability and maintainability for the Chicago Transit Authority (CTA) and the development of cost models to assess the cost effectiveness of railcar rehabilitation and replacement program alternatives are presented. Data files and extensive discussions with CTA maintenance personnel provided the basic data on maintenance and operations; detailed cost data for each railcar series were also provided by the CTA. Reduction of the data yielded reliability-maintainability factors such as mean time between failures, mean time between maintenance, mean time between inspections, mean time to repair, mean time to maintain, and mean time to restore. Using this information and a previously developed modeling approach, models for estimating cost savings attributed to improvements in mean time to maintain and mean time between maintenance were prepared for the 2200, 2400, and 2600 Series of railcars for the CTA fleet. Models for estimating fleet capital cost savings as a result of improved railcar reliability and maintainability were also prepared. Specific suggestions for using these models in maintenance practice to estimate cost savings from alternative actions were presented.

The authors recently completed a project for the Chicago Transit Authority (CTA) that was aimed at answering a number of questions regarding current CTA railcar maintenance practices and evaluating alternative programs that include overhauls, rehabilitation, and replacement (1). As part of the project, the authors carried out an analysis of CTA fleet reliability and maintainability to establish the cost effectiveness of rehabilitation and replacement program alternatives; this aspect of the project is reported on in this paper.

RAILCAR PERFORMANCE EVALUATION

Transit properties generally collect the same basic types of information relating to transit vehicle operation and maintenance (2). These include data on revenue service incidents, periodic inspections, and maintenance activities. Vehicle maintenance data are also generated in the same basic manner at most properties: a vehicle problem is reported in revenue service or is discovered during maintenance, the

---

Lea, Elliott, McGean and Company, 600 West Service Road, Suite 320, Dulles International Airport, P.O. Box 17030, Washington, D.C. 20041.

vehicle is repaired, and information concerning the maintenance action is recorded. The range in the depth of detail pursued and the methodology used by the properties in recording this information, however, vary considerably.

Problems that occur in revenue service are either communicated orally by the train operator to central operations (i.e., "central" as such, or the tower having jurisdiction over the particular line) or, with some properties, recorded by the train operator onto a form that is passed on to maintenance. In the case of oral communication, the problem description is retransmitted by central operations to maintenance via telephone or computer terminal, and the information is transcribed onto a maintenance form to initiate a repair activity.

Although a preliminary indication of a suspected service vehicle failure may be recorded on an incident report, this information is based on observation and not equipment tear-out or repair. For this reason, this reported failure data cannot be used to determine actual equipment reliability without the associated maintenance data describing the repair.

Major deviations among properties begin at the reporting of primary repair data, that is, recording what was done to and on the vehicle to fix it. Some transit properties, for example, provide only a narrative summary of the repair activity, which is also recorded onto their respective incident forms. Other properties provide narrative data to describe each defect found and the repairs made. Defective parts or assemblies are identified by part number. One property's narrative defect data are later coded on the same form to assign a single entry from a "Fault Code Table" that best describes the mode of failure. Another property goes one step further similarly encoding the repair data. Other properties use a similar approach except that they deal exclusively with codes.

With few exceptions, the repair of components following removal from a vehicle cannot be related to the revenue incident through the existing data collection methods. Therefore, the reporting of vehicle reliability data stops at whatever the lowest replaceable unit might be for a vehicle subsystem. The structure of some maintenance information systems is such that a link is provided between primary and secondary maintenance data. In general, secondary maintenance (maintenance performed in the shop on components that have been removed from the vehicle) statistics are kept by the properties for the purpose of production control and material and time-cost accounting.

Transit properties maintain permanent files of the hard-copy forms. One property also enters the data into a computerized data system via interactive terminals. Other properties keypunch their data for batch entry into their respective computer systems. Several properties use their computers for all processing and manipulation of data and for production of their various reports.

The investigation of existing sources of reliability information on transit vehicle equipment has indicated that

1. The total extent of data that would normally be collected to support the classical, detailed reliability analysis of transit equipment is not available for all potential data sources;
2. Only maintenance data that pertain to what was done to and on the vehicle itself (primary maintenance) can generally be correlated to a vehicle failure;
3. It is difficult to separate primary failures from secondary failures based on existing bases; and
4. Failure data are recorded to different levels of equipment detail at different properties (e.g., only to subsystem level versus the component level).

Based on this characterization of failure data collected in the transit industry, it is important to establish a clear definition of revenue service reliability. The most common measure of hardware reliability is mean time (or miles) between failures. However, an alternative statistical measure of reliability that more closely corresponds to the data that are actually available is mean miles between replacement.

The dynamic data that describe unscheduled maintenance activities do not, in many cases, include information in sufficient detail to accurately determine failure cause and effect. From the transit property viewpoint, it is only important to verify and record the fact that, whatever the apparent problem, a correction was made that returned the vehicle to revenue service availability. Data that are generated, therefore, most often present the unscheduled maintenance action that took place, and a description of components that were replaced.

Based on this characterization of data for vehicle repair, the most meaningful information that can be obtained includes unscheduled maintenance activities and related equipment replacement. Although these outputs do not specifically describe vehicle or equipment reliability, they will be proportional to hardware reliability values and, more important, describe major contributing factors to the cost of maintenance operations.

In most cases, the primary measure available for transit vehicle and equipment reliability-related analysis is the replacement of components and equipment on vehicles. Although it may not be possible to draw a direct correlation between equipment failures and replacements, the data that are available from transit properties record these change-outs with the greatest degree of accuracy and completeness. When this replacement information is combined with the generally available data on vehicle utilization (i.e., miles per reporting period by vehicle type), a measure of mean miles between replacements can be determined that combines both equipment reliability and replacements relating to apparent failures.

DATA AVAILABILITY AT THE CTA

In view of the preceding and other recent work in this area, the general statement can be made that the data items needed to carry out traditional reliability or maintainability analyses (or both) are not routinely collected at a majority of U.S. rail transit authorities. Where they are maintained at all, the format, frequency, and scope of these data items vary significantly between transit authorities so that substantial data reduction would be required to make an existing data bank meaningful. Because of the existing dearth in reliability and maintainability information, the task of rail transit performance evaluation is, to say the least, difficult. Many transit authorities maintain that the benefits from extensive and consistent data collection do not justify the expense needed to maintain such a data bank. The CTA has recently implemented a new maintenance management information system (MMIS) that is currently being "debugged." This program should be a valuable source for future maintenance and reliability information and data.

The CTA is no exception to the problems with reliability and maintainability information previously discussed. The only routinely collected maintenance information available is a chronological log of car failures or maintenance actions, or both. This log is accumulated on a car-by-car basis in a computer data file called the Railcar History. Railcar his- tories are usually available on-line for a period of 6 months after which they are transferred into a related data base (not on-line) called the VMT. (Note that the acronym originated with the CTA's bus MMIS and actually stands for Vehicle Maintenance--Terminal in that system.)

CTA car histories include unscheduled and scheduled maintenance actions for individual cars in the fleet. The logs also contain brief descriptions of the types of failures experienced as well as times that indicate when the reports were logged in and out of the computer. Except for information about the cars that experienced particular failures, none of the data contained in this data file are helpful in establishing the reliability of the equipment. In addition, because times recorded in this data file are not indicative of actual repair times for failures, equipment maintainability could not be established on the basis of this file alone. Another problem with using the car history data relates to the problem of "wrong calls." Normally, when a car fails in service, the identity of the car that is logged in the railcar history is that of the lead car in a multiple-car train, although this car may not be the one that experienced the failure. This introduces significant problems with any attempt to study the performance of individual car series because the appropriate number of maintenance actions for a specific car cannot be accurately determined. The other problem involves the description of failures in the car histories. The computer operator only identifies maintenance actions on the basis of information received from line reports. Moreover, maintenance actions are logged by job numbers that correspond to major work categories established at the CTA. Hence, car failures are only related to major car subsystems. For example, the work category Propulsion (Code = 1000) includes work on traction motors or traction motor controls, or both; the category Car Body (Code = 2000) includes work on the car structure, windows, seats, destination signs, lighting systems, coupler and drawbar, and such activities as battery-charging, converter/motor generator/motor alternator repairs, and maintenance of safety equipment.

The VMT report, which is routinely compiled from car histories, also does not contain data on repair times. However, the reports go one step further than railcar histories in that they attempt to reconcile problems of wrong calls and incorrect failure descriptions by establishing two distinct segments of the report--one for "problem reported" and the other for "problem found."

The CTA keeps extensive maintenance cost information for each series of cars in the fleet. Maintenance costs are summarized monthly in the Vehicle Series Report under two major categories--maintenance performed at the rail terminals and maintenance performed at the shop areas (e.g., parts rebuilding). Monthly records for total rail maintenance for both labor and material functions can be generated from these data. Costs are generally broken down by car series when possible. When costs cannot be identified by vehicle series, they are listed as unassigned in the Vehicle Series Summary.

RESULTS OF DATA REDUCTION EFFORTS

Considerable data reduction was performed in order to use available data to establish current performance levels for each of the three car series under study. The data reduction effort was supplemented by information generated from extensive discussions with CTA maintenance personnel as well as estimates and assumptions made by the study team. In general, data items were broken into three major categories--maintenance data, system operational data, and maintenance cost information.

Maintenance Data

Because VMTs were only available for the first half of 1984, railcar histories and VMTs were combined to reflect the performance of the 2200, 2400, and 2600 Series for a 1-year period. These data included all maintenance actions during the year for a randomly selected sample consisting of ten 2200-, ten 2400-, and eight 2600-Series cars. VMT data covered the period January-May 1984 and railcar histories covered the period June-December 1984.

Because neither of these records includes any information on repair times, the study team obtained three independent estimates of repair times for typical types of failure for each car series. CTA railcars are maintained at 11 inspection and maintenance shops. Repair times at these shops vary as a result of the differences in available manpower and facilities. Hence, any estimates of times must recognize these variations. Based on the independent estimates obtained from CTA maintenance personnel, repair times were generated for all failures experienced by each car in the sample. These times were then accumulated for the period under investigation. Information on all maintenance actions for the data sample is given in Table 1. The number of maintenance actions includes both unscheduled and scheduled maintenance (inspections). Repair times are based on an average of two men per repair action. Times for routine inspections are based on an estimate of 4 hr per inspection with six people performing each inspection activity.

TABLE 1 Summary of Equipment Maintenance Actions for Railcar Sample

Data Item	2200 Series	2400 Series	2600 Series
Sample size	10	10	8
Number of failures	485	476	275
Number of inspections	71	64	48
Number of maintenance actions	556	540	323
Repair time (hr)	430	330	182
Inspection time (hr)	284	256	192
Maintenance time (hr)	714	586	374

Note: The fleet size was comprised of 144 railcars in the 2200 Series, 199 railcars in the 2400 Series, and 250 railcars in the 2600 Series at the time of the data collection.

In addition to car maintenance actions, subsystem failure distribution and corresponding repair times were also generated. These are given in Table 2. Major subsystems covered include propulsion, car body, brakes, doors/communications, heating, ventilating, and air conditioning (HVAC), truck, and automatic traffic control (ATC) in accordance with work categories established at CTA. In view of the extent of necessary data reduction effort, the subsystem failure and repair data are based on a sample size consisting of five 2200-, five 2400-, and four 2600-Series cars.

System Operational Data

System operational data obtained from the CTA include a breakdown by lines of the total number of cars required for service, the number scheduled for main-

TABLE 2 Summary of Subsystem Failure Distribution and Repair Times for Railcar Sample

	2200 Series		2400 Series		2600 Series	
Subsystems	Number of Failures	Repair Time (hr)	Number of Failures	Repair Time (hr)	Number of Failures	Repair Time (hr)
Propulsion	44	81.5	61	42.0	17	9.5
Car body	56	28.0	40	22.5	27	19.0
Brakes	34	28.5	41	32.0	21	20.5
Doors and communications	59	41.5	48	22.5	53	26.5
HVAC	25	14.5	9	6.0	12	6.5
Trucks	6	7.0	3	6.5	4	2.5
ATC	26	19.5	26	19.0	13	10.5

tenance, and the number in reserve. This breakdown does not disaggregate the number of cars required for service by car series. Hence, estimates of railcar requirements by car series were made and then a calculation of car-hours scheduled was made on the basis of operational schedules. Once the car-hours scheduled for service were calculated for each series, they formed the basis for estimating both the mean time between failure and the mean time between maintenance actions.

The number of 2200-, 2400-, and 2600-Series cars required for service was estimated by considering only those lines on which these cars are assigned and by assuming that all cars on a particular line have an equal chance of being scheduled for service. The results are given in Table 3. Railcar requirements by period of day were furnished by the CTA.

TABLE 3 Estimated Number of Cars Needed for Service

Car Series	Route				Total Required for Service[a]	Current Fleet Size
	W-NW	W-S	N-S	Ravenswood		
2200	106	–	–	–	106	144
2400	–	140	16	–	156	194
2600	100	58	–	34	192	250

Note: W = west, S = south, N = north, and NW = northwest.

[a] Adjusted to account for assignment by married pairs.

Original data show two requirement levels--rush and base. The "owl" requirement was estimated on the basis of one-third the base requirement. Based on total fleet requirement by period of day and the calculated number of cars scheduled for service, car-hours scheduled per day were calculated for each of the three series for weekday and weekend and holi- day schedules. These calculations translate to the following annual car-hours requirement by series.

Series	Estimated Annual Car-Hours
2200	349,000
2400	512,000
2600	629,000

It is important to reiterate that these calculations have been based on the assumption that each car has an equal chance of being scheduled for service. There may be other considerations that affect this assumption but the scope of this study did not permit more detailed analyses of this aspect. Moreover, iterative analyses based on estimated availabilities for each car series showed that the final results of the analyses are not significantly affected by the original assumption.

Maintenance Cost Data

Monthly records of maintenance costs summarized by car series were compiled by the Financial Services Department of the CTA. These records also disaggregate maintenance costs by function (i.e., labor, material, and other). Costs that were not readily allocable to car series are identified as unassigned in these records. The breakdowns in the Vehicle Series Summaries, as they are called, were used to generate adjusted maintenance costs for each of the three railcar series over 1 year. The adjustments reflect amounts that were unassigned to car series in the original data source.

CURRENT RELIABILITY AND MAINTAINABILITY STATUS

Fleet Availability

The availability of each car series is determined by its reliability and maintainability status. To establish current reliability and maintainability, failure and maintenance time data generated for the fleet sample were extrapolated to reflect all cars in each series. The extrapolation procedure assumes that the entire fleet is operable and that all cars in each series are circulated as required. Details of the extrapolation procedure are presented in Muotoh and Elms (3). Using these data, reliability and maintainability factors for each series have been calculated. These are given in Table 4. Because of the scarcity of data, the mean time to restore (R_e) has been estimated from results of an earlier study (3). Based on the estimated values of mean time to restore and the mean time between maintenance, fleet availabilities for the 2200, 2400, and 2600 Series have been calculated as 80 percent, 86 percent, and 89 percent, respectively. These translate to the following estimated fleet size requirements:

Series	Estimated Fleet Size Requirement (cars)
2200	134 (including 28 spares)
2400	182 (including 26 spares)
2600	216 (including 24 spares)

It is important to recognize some implications of the factors derived in Table 4. The reliability

TABLE 4 Fleet Reliability and Maintainability Factors

	Series		
Factor	2200	2400	2600
MTBF (car-hr/failure)	50	56	73
MTBM (car-hr/maintenance action)	44	49	62
MTBI (car-hr/inspection)	341	412	419
MTTR (hr)	0.9	0.7	0.7
MTTM (hr)	1.3	1.1	1.2
R_e (car-hr)	11.0	7.8	7.8

(mean time between failures, or MTBF) of the 2600-Series cars is considerably higher than for the older cars despite the fact that the 2600-Series cars appear to be less frequently inspected. The mean time to repair the 2200-Series cars is approximately 30 percent higher than for the other series.

Examination of Subsystem Reliability and Maintainability

The evaluation of railcar performance has also been conducted by investigating the contribution of major subsystems to overall car reliability and maintainability (1). The three worst offenders, in order of severity, are as follows for each of the three car series under study:

Series	Subsystem
2200	Doors/communications, propulsion, brakes
2400	Propulsion, doors/communications, brakes
2600	Doors/communications, brakes, propulsion

Car body is not considered in this ranking because this category includes miscellaneous items not directly allocable to a single subsystem. Notice that, generally, the three worst offenders are the same for each series and also reflect the results of the UMTA Transit Reliability Information Program (TRIP). The percent distribution of repair time and the mean time to repair each subsystem are given in Table 5. Notice that the mean time to repair the 2200-Series propulsion system is considerably higher than for the 2400 and 2600 Series. Repair times for all other subsystems are similar for all the car series. Expansion of the sample size, however, might provide more insight into differences among the series' subsystems.

TABLE 5 Distribution of Repair Time and Subsystem Maintainability

	Percent of Total Repair Time (and MTTR)		
Subsystem	2200 Series (hr)	2400 Series (hr)	2600 Series (hr)
Propulsion	37 (1.9)	28 (0.7)	10 (0.6)
Car body	13 (0.5)	15 (0.6)	20 (0.7)
Brakes	13 (0.8)	21 (0.8)	22 (1.0)
Doors/communications	19 (0.7)	15 (0.5)	28 (0.5)
HVAC	7 (0.6)	4 (0.7)	7 (0.5)
Trucks	3 (1.2)	4 (1.0)	2 (0.6)
ATC	8 (0.8)	13 (0.7)	11 (0.8)

Note: Numbers in parentheses are in percent.

FLEET PERFORMANCE AND COST MODELS

Potential benefits from proposed improvements in railcar performance must be quantifiable so that these benefits can be compared with benefits from competing alternatives. This section presents a procedure that is used to estimate potential economic benefits that can be derived from improvements in CTA railcar reliability and maintainability. The procedure uses mathematical models that estimate both potential savings in maintenance cost as well as fleet capital cost savings. Operating cost savings that may result from reduction in service delays have not been addressed because these have been found to be minimal. Detailed development of these models is presented elsewhere (1,3).

The following paragraphs give a more detailed explanation of the basis for both the maintenance cost savings model and the fleet capital cost savings model. Each of these models has been formulated for the 2200-, 2400-, and 2600-Series railcars based on CTA data; for the sake of brevity, only the results of the 2200-Series models are included herein.

BASIC RATIONALE FOR MODELS

Equipment breakdowns can result in system downtime, lost car-hours, higher levels of maintenance, and, consequently, increased operating and maintenance costs. Equipment breakdowns also result in higher capital costs because transit authorities must make allowances for car unavailability in new car acquisitions. This results not only in higher capital commitment for increased fleet size, but also in the increased cost needed to provide larger maintenance and storage facilities. By reducing railcar failure rates or car downtime, or both, operating costs will be reduced through reduced service delays, maintenance costs will drop as a result of lower labor and parts requirements, and fleet capital cost will also be lower because the need for spare cars would have been minimized. Two major areas that are relevant to this study include maintenance and fleet capital cost savings from improved railcar performance.

Maintenance Cost Savings Model

The maintenance cost model estimates both potential labor and spare parts cost savings from improved reliability and maintainability. Although labor cost savings are derived from reductions in failure rate or mean time to repair failed cars, or both, the spare parts cost savings result only from reduced failure rates. Because maintenance costs are incurred from all maintenance actions, the maintenance cost savings model takes account of both service- and nonservice-related failures.

The relationship for estimating potential maintenance cost savings based on unscheduled maintenance actions alone has been developed in a recent study (3); the relationship is as follows:

$$\Delta C_{um} = N_f\{K_s R_s[(p_f + p_r)/(1 + p_f)] + K_{up}[p_f/(1 + p_f)]\} \qquad (1)$$

where

ΔC_{um} = unscheduled maintenance cost savings;
N_f = total number of unscheduled maintenance actions for fleet during the period under investigation;
K_s = unscheduled maintenance labor cost factor. This represents the cost per shop car-hour and is expressed in dollars per car-hour. It is given by C_{ul}/D_s where C_{ul} = total labor cost for all unscheduled maintenance actions and D_s = shop time for unscheduled maintenance actions;
R_s = MTTR = mean time to repair (car-hours) = D_s/N_f;
K_{up} = unscheduled maintenance parts cost factor. It relates spare parts cost due to unscheduled maintenance to number of unscheduled maintenance actions and is given by C_{up}/N_f, where C_{up} = cost of spare parts consumed in unscheduled maintenance and N_f is as previously defined;
p_f = improvement in MTBF = change in MTBF/initial MTBF; and
p_r = improvement in MTTR = change in MTTR/initial MTTR.

By reformatting Equation 1 to reflect both scheduled and unscheduled maintenance actions, the total maintenance cost saving is given by

$$\Delta C_m = N_m\{K_m\ R_m[(P_m + P_t)/(1 + p_m)] + K_p[p_m \div (1 + p_m)]\} = C_{m1}[(P_m + P_t)/(1 + p_m)] + C_{mp}[(p_m/(1 + p_m)] \quad (2)$$

where

ΔC_m = total maintenance cost savings (unscheduled and scheduled),
C_{ml} = total labor cost for unscheduled and scheduled maintenance,
C_{mp} = total spare parts cost for unscheduled and scheduled maintenance,
N_m = total number of maintenance actions (scheduled),
K_m = total maintenance labor cost factor = C_{ml}/D_m (where D_m = shop time for both unscheduled and scheduled maintenance actions),
R_m = MTTM = mean time to maintain = D_m/N_m,
K_p = total spare parts cost factor for scheduled and unscheduled maintenance = C_{mp}/N_m,
p_m = improvement in mean time before maintenance (MTBM) = change in MTBM/initial MTBM, and
p_t = improvement in MTTM = change in MTTM/initial MTTM.

Fleet Capital Cost Savings Model

Savings in fleet capital cost are reflected in the reduction of the spare car requirement realized as a result of improved car reliability and maintainability. Car-hours are lost because of failures occurring in service as well as failures detected when the car is in the shop for other maintenance. Hence, the fleet cost model also considers both service- and nonservice-related incidents.

The reduction in the spare car requirement, and consequently fleet cost, is directly related to the reduction in car downtime realized through improvements in car performance. The car-hours of downtime saved can be translated into the number of cars saved through the following relationship, the details of which have been explained elsewhere (2).

$$\Delta N_c = N_0[(P_m + P_t)/(1 + p)]R_e/L \quad (3)$$

where

ΔN_c = number of cars that can be saved,
N_0 = number of cars required for service,
L = mean time between maintenance actions, and
R_e = mean time to restore.

CALIBRATING AND USING THE MODELS

The relationships for maintenance cost savings and fleet capital cost reduction (number of cars saved) must be calibrated for each transit authority before they can be used to investigate benefits from performance improvements. Results of the data reduction efforts previously discussed have been used to calibrate these models for the CTA (1, Appendix E). While expressing cost directly as a function of car reliability and maintainability factors, it should be pointed out that the results of this analysis are only indicative of the general level of savings that is achievable through improved reliability and maintainability.

Potential Maintenance Cost Reduction

Maintenance cost savings for the three series can be obtained by calibrated equations for each series; the equation for the 2200 Series is as follows:

$$\Delta C_m = 1{,}100{,}000[(P_m + P_t)/(1 + p_m)] + 1{,}700{,}000[P_m/(1 + p_m)] \quad (4)$$

Recognizing that p_m is improvement in MTBM and p_t is improvement in MTTM, potential maintenance cost savings at any level of reliability or maintainability (or both) improvements can be examined by varying p_m or p_t (or both) in each of the preceding equations. A sensitivity analysis has been conducted by calculating potential cost savings for each series for varying levels of p_m and p_t. For each car series, cost savings are calculated first by keeping mean time between maintenance constant at a known level and varying the mean time to maintain. A second analysis is then conducted by holding mean time to maintain constant and varying the mean time between maintenance.

To facilitate the use of the models, the calculated results can also be presented as a set of easy-to-use graphs that can be employed without reference to the mathematical formulation once the defining parameters have been established. Figures 1 and 2 show two sets of graphs relating improvements in MTBM and MTTM with associated potential savings in maintenance cost for the 2200-Series cars. Figure 1 shows variations in the MTTM at fixed levels of MTBM and Figure 2 shows variations in MTBM for fixed levels of MTTM. The vertical axes represent the potential annual maintenance cost savings (dollars), while the horizontal axes represent the improvements in MTTM (Figure 1) and MTBM (Figure 2) expressed as percentages of their respective values before improvements. Either of these two sets of graphs can be used to analyze potential maintenance cost savings for improvements in reliability or maintainability of the 2200 Series of cars.

Consider Figure 1, which plots a set of linear relationships between cost savings and changes in MTTM for various levels of MTBM. The set of lines is

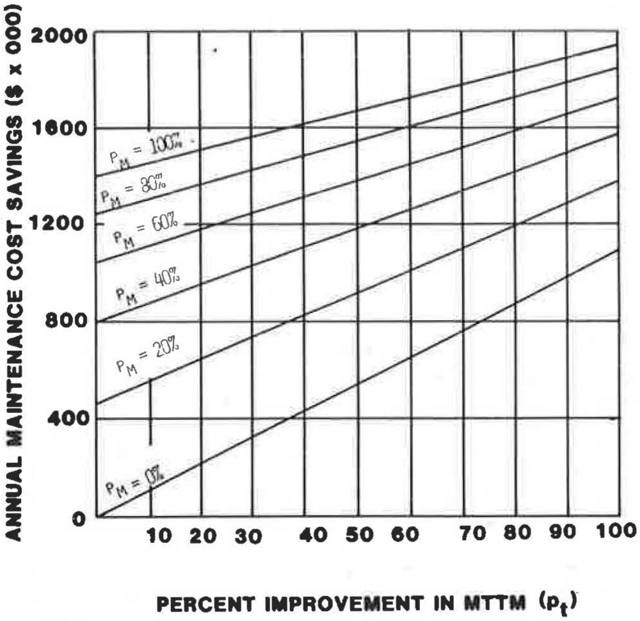

FIGURE 1 Maintenance cost savings for the 2200 Series—percent improvement in MTTM (p_t).

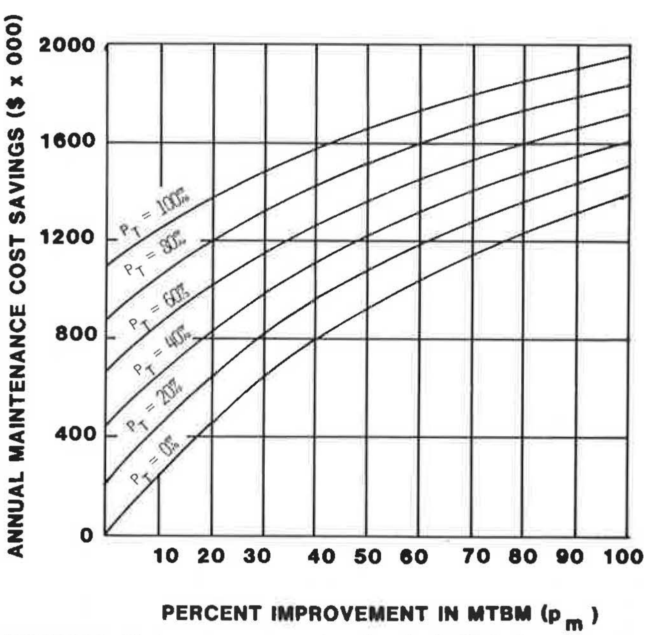

FIGURE 2 Maintenance cost savings for the 2200 Series—percent improvement in MTBM (p_m).

plotted only in the positive quadrant representing improvements in MTTM (reduction in MTTM) and improvements in MTBM (increase in MTBM). Potential cost savings can be realized by either or both of these improvements by entering this chart with the percent improvement in MTBM or MTTM, or both, and reading the corresponding cost savings from the vertical axis. Notice that the maximum improvement in MTTM will correspond to the hypothetical case where it takes zero time to maintain a car, that is p_t = 100 percent. On the other hand, maximum improvement in MTBM will occur when the car operates perpetually without the need for maintenance, that is, $p_m = \infty$. For purposes of this analysis, the set of graphs is, however, bounded between p_m = 100 percent (that is, doubling the MTBM) and p_t = 100 percent (zero time to maintain). Maximum annual maintenance cost savings for this hypothetical case is about $2 million and is given by the topmost line in Figure 1. For zero MTTM reduction (p_t = 0 percent) and a 100 percent increase in MTBM (that is, p_m = 100 percent), an annual savings of about $1.4 million (or half of estimated annual maintenance cost for the 2200 Series) could be obtained. Note that a 100 percent increase in MTBM is equivalent to a 50 percent reduction in number of maintenance actions.

It is important to recognize that the horizontal axis (change in MTTM) can be extended to the left (negative values) to represent increases in MTTM. Notice also that improving the MTBM can allow the MTTM to increase substantially without any maintenance cost penalty. For example, a 40 percent improvement in MTBM (p_m = 40 percent), without a change in MTTM, has the potential to save $800,000 in annual maintenance cost in the 2200 Series. Extending the p_m = 40 percent line to intersect the horizontal axis indicates that a 40 percent improvement in MTBM would permit the MTTM to increase by up to twice the original value before costs are increased. In other words, by doubling the MTTM (to allow for more thorough inspection and repair of the 2200-Series cars) and thereby realizing a 40 percent improvement in MTBM, the overall annual maintenance cost will remain unchanged. A net maintenance cost savings can be realized if 40 percent improvement in MTBM is achieved at less than double the current MTTM. Detailed implications of these results relative to the CTA maintenance program are discussed in a later section.

Potential Fleet Capital Cost Reduction

Potential fleet capital cost savings have been expressed in terms of the number of cars that can be saved in Equation 3. The calibrated equation for the 2200-Series is

$$\Delta N_c = 28 (p_m + p_c)/(1 + p_m) \qquad (5)$$

Using the same approach of separately varying one parameter while keeping the other constant, the potential fleet size reduction at various levels of reliability and maintainability improvements can be determined for each car series (1). (It should be noted that the number of cars that can be saved is given as a fraction of the required number of spares estimated for each series.) The calculated results are shown graphically in Figures 3 and 4 for the 2200 Series. It should be noted that the maximum number of cars that can be saved converges to the number of spares estimated for each series. For a hypothetical 100 percent improvement in MTTM (that is, zero maintenance time), there will, theoretically, be no need for spares.

If the percent improvements in MTTM and MTBM are known for the 2200 Series, these values can be used to enter Figure 3 or 4 in order to determine the number of 2200-Series cars that can be saved as a result of the improvements. These cars will be in addition to the number of cars (10) calculated as excess on the basis of current spare allowance. For example, consider a change in maintenance practice or subsystem modification (or both) that results in a modest 20 percent improvement in MTTM only. Figures 3 or 4 can be entered with p_t = 20 percent or p_m = 0 percent to obtain a savings of six 2200-Series cars. This means that the spare requirement can be reduced to 20 cars, down from 26 cars initially estimated for the series before this improvement. Hence, the total excess cars of the 2200 Series will be 18 cars

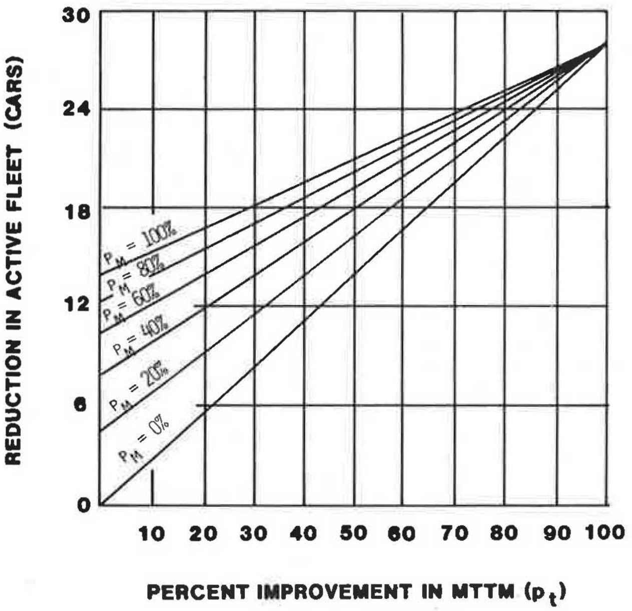

FIGURE 3 Fleet cost savings for the 2200 Series—percent improvement in MTTM (p_t).

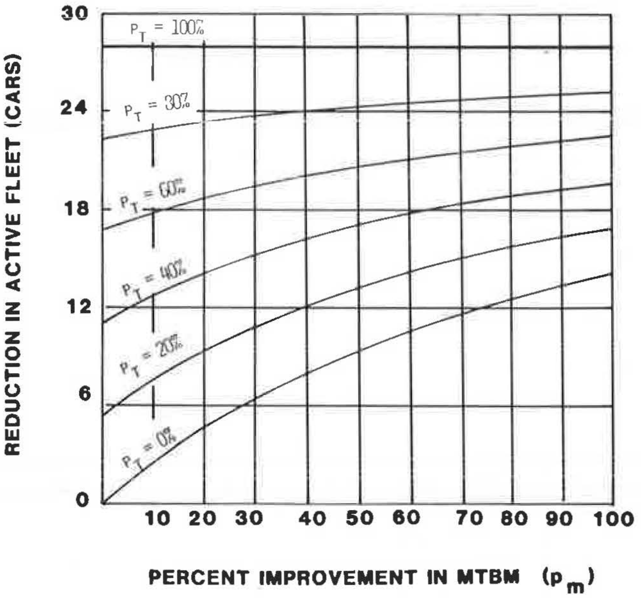

FIGURE 4 Fleet cost savings for the 2200 Series—percent improvement in MTBM (p_m).

based on the estimate of fleet requirement for this series.

IMPLICATIONS RELATIVE TO CTA MAINTENANCE PROGRAM

The following paragraphs demonstrate the implications of results from the models relative to the CTA maintenance program. It should be kept in mind that the data sample used to develop the model was limited and that it would be desirable to expand on this data base before performing any extensive analyses with the models. However, the results that have been obtained can be considered indicative of the existing situation.

Unit Maintenance Cost Factors

Unit maintenance cost factors for each car series have been estimated. These factors, which are direct derivatives of the development of the models, provide an indication of the cost-effectiveness of current CTA maintenance practice. The maintenance labor cost factor (k_m) relates labor cost for scheduled and unscheduled maintenance to total maintenance time for each car series. It represents the maintenance labor cost per car-hour of maintenance. The spare parts cost factor (k_p) relates the parts cost for scheduled and unscheduled maintenance to the number of maintenance actions experienced by each car series. It represents the average parts cost per maintenance action.

Table 6 gives a comparison of the maintenance experience for each car series with related unit maintenance cost factors for the data and period examined. Notice that labor cost per maintenance car-hour is highest for the 2600-Series cars, followed by the 2400 Series, and then the 2200 Series. This is not completely surprising in view of the following: the newer cars have more sophisticated electronic components that (a) must operate in a harsh transit environment (i.e., electrical disturbances, vibrations, temperature variations, and dust and dirt), and (b) require an additional level of troubleshooting time. In addition, the new cars have been experiencing excessive burn-in problems.

From the point of view of parts costs, however, the older 2200-Series cars have the highest average parts cost per maintenance action. This can be partially explained by the fact that the parts on the older cars are worn and that there are likely to be more worn parts, many of which are large or expensive or unavailable except through a special order; on the other hand, the parts on the newer cars are still in their early life, should be readily available, and might be less expensive.

TABLE 6 Comparative Maintenance Data and Maintenance Costs

	Series		
Data Item	2200	2400	2600
Fleet size (cars)	144	194	250
Number of maintenance actions	8,006	10,476	10,094
Average number of maintenance actions per car	56	54	40
Number of failures	6,984	9,234	8,594
Average number of failures per car	49	48	34
Mean time to maintain (hr)	1.3	1.1	1.2
Maintenance labor cost factor ($ per car-hr)	107	167	188
Spare parts cost factor ($ per maintenance action)	212	191	109

Rebuild Versus Replace

As with most transit authorities, one of the primary considerations of the CTA in adopting a maintenance strategy concerns the choice between rehabilitating and replacing cars. The primary impetus for considering the rebuilding or replacing of an aging fleet is deteriorating reliability accompanied by increasing maintenance costs and by worsening fleet availability. This issue has been raised at the CTA in connection with some of the fleet.

A preliminary evaluation of the rebuild-or-replace decision must be made to determine if rebuilding is an option for remedying acute reliability and maintainability problems. Rebuilding existing cars may not be considered an option for a number of reasons. First, new cars may be preferred if the current fleet, even at 100 percent availability, cannot meet peak service demand. Second, existing cars may require such extensive work that rebuilding cannot be considered. This was found to be the case with the 2000 Series. Third, rebuilding may not be attractive because of limitations in available facilities and manpower. Space for rebuilding and the storage of replacement or rebuilt parts may not be available on the transit authority property. Also, available manpower may be insufficient for the requirements of a rebuild program. The fourth major consideration in deciding if rebuilding is an option is the availability of funds. Budget restrictions may even rule out both replacing and rebuilding and the transit authority may be compelled to rely on existing cars.

If an examination of the foregoing considerations indicates that rebuilding is a viable option, then a more comprehensive financial analysis should be made to estimate the value to the property of each alternative--rebuild or replace (buy new). Facility, rebuilding, logistics supply, and all overhead costs related to a rebuild program must be estimated. Because new railcars are regularly being ordered and delivered somewhere in this country, the cost of new cars can be easily estimated.

Estimates of rebuild costs for specific railcar series are included in the CTA report on rapid transit car rehabilitation and purchase plan. To complete the rebuild-or-replace analysis, it is necessary to estimate the expected life of a rebuilt car and that of the new car in order to amortize the costs discussed previously. This economic analysis should also include the required maintenance costs for both alternatives. The cost-performance model discussed in this study can be used as part of this life cycle cost analysis to assist in making decisions between rebuilding or replacing CTA cars. The process involves the comparison of potential net benefits that can be realized by improving the performance and extending the life of an existing car against the net benefits from buying new and possibly more reliable equipment. To do this, each car series would have to be analyzed on the basis of its current reliability and maintainability.

The value of the models in this type of decision is the ability to perform "what if" analyses quickly and inexpensively. For example, estimates of the expected improvement in reliability and maintainability for a rebuilt fleet can be made in conjunction with the maintenance staff. The models for that fleet can then be used to generate estimates of maintenance cost savings. These cost savings can be compared to the rebuild costs to determine the value of the rebuild. Additional analyses can be made of the value of a new car purchase over time as compared to the value of a rebuild.

Equipment Retrofit

On the basis of the data that were examined, the doors/communications, propulsion, and brakes subsystem areas were identified as the worst offenders for each of the three series studied. For example, the data show that the propulsion subsystem was the least reliable for the 2400 Series. It was estimated to account for about 27 percent of all unscheduled maintenance actions, a significantly high cause of failure for this particular fleet. Two things can be done to improve the reliability of this subsystem and, consequently, the overall performance of the 2400-Series cars. First, CTA may opt for a retrofit program that will reflect changes in the subsystem design; if only a few propulsion system components are causing much of the problem, it may be helpful if these are replaced. A second option, which may be more feasible, could involve changes in the existing maintenance practice for this subsystem.

If it is believed that a retrofit of a subsystem or a component would be helpful but the available information is not sufficient for making a commitment to retrofit the entire series or fleet, then it should be possible to perform a sample retrofit on a small number of railcars. The performance of the sample railcars can be monitored before and after the retrofit. Using the performance measures, the cost models can then be exercised to determine estimates of maintenance cost savings. These savings can be compared with the retrofit cost to determine whether or not the retrofit will pay off.

Changes in Maintenance Practice

With regard to changes in existing maintenance practice, two courses of action can be taken, as discussed previously. The first alternative is to change the maintenance procedure (increase MTTM) and the second is to change the maintenance interval (decrease the mean time between inspections, or MTBI). The three worst offenders for each of three series were indicated previously. In addition, a summary of the subsystem reliability (MTBF) estimates is of interest in view of the related estimates of MTBI detailed in Diewald and Muotoh (1). This summary is given in Table 7. The data indicate that for some of the worst offenders, the MTBF is less than the MTBI. For example, for the 2200 Series, three subsystems, doors/communications, car body, and propulsion, can be expected to fail between inspections. For the 2400 Series, there are four such subsystems and, for the newer 2600 Series, there are two.

TABLE 7 Summary of Subsystem Reliability Calculations for Sample Set by Railcar Series

	Mean Time Between Failure (hr)		
Subsystem	2200 Series	2400 Series	2600 Series
Doors/communication	205	275	190
Car body	216	330	373 (419)[a]
Propulsion	275 (348)[a]	216	529
Brakes	356	322 (412)[a]	479
ATC	446	507	774
HVAC	485	1,467	839
Trucks	2,017	4,414	2,516

[a] Indicates the calculated value of mean time between inspection for the car series.

In view of the foregoing, it would be prudent to establish the nature of the problem with the identified subsystem so that appropriate adjustments can be made in either the inspection or repair procedures. This can involve documenting (through discussions with repairmen and maintenance supervision) the kinds and extent of repairs that are being required. Then, for some period, say 1 month, additional repair documentation on maintenance of the subsystem can be required. In addition, intensive investigations of failures and repairs for the subject subsystem can be conducted during the period. This documentation should form a sufficient basis for further action regarding inspection or repair procedures or schedules. Changes in maintenance actions include changes in the frequency of maintenance or changes in maintenance procedures or both.

As an example, consider the 2400-Series propulsion subsystem. At present, scheduled maintenance is performed as part of the routine 6,000-mi inspection for each car. Changes in existing maintenance practice can be accomplished by either changing the scheduled maintenance interval for this particular subsystem or changing the maintenance procedure. Recognizing that it may be impractical to do this on a fleet-wide basis, CTA could conduct a test on a test sample of the 2400 Series (say 20 cars) for a period of about 3 months. The scheduled maintenance frequency for the propulsion system of these sample cars could be increased to, say, 4,000-mi intervals. This may help in spotting more incipient failures before they cause service disruption. Then the impact on railcar performance and cost can be estimated through the cost model. Alternatively, the maintenance frequency could be decreased to 8,000 mi and a performance-cost analysis performed. (Inquiries by the authors indicated that the evolution of maintenance intervals at the Port Authority Transit Corporation, involving a purely trial-and-error approach, took about 4 years to find the optimum interval.)

Alternatively, more thorough inspection of the propulsion subsystem for these sample cars can be conducted within the current 6,000-mi interval. These changes in maintenance schedules or procedures should be monitored to determine if the performance of the subsystem has been affected. Also, the additional cost (labor or material, or both) needed to accomplish such changes in schedules or procedures for maintaining the sample cars should be determined. Suppose it is found that these changes result in a 10 percent reduction in total number of maintenance actions experienced by the sample. If so, approximately $400,000 in annual maintenance cost can be saved (see Table 4). In addition, this could result in a lower fleet requirement. To justify the change in maintenance practice, these potential cost savings can then be compared with the estimated additional cost to effect such changes in maintenance for the entire fleet.

REFERENCES

1. W. Diewald and D. Muotoh. Rapid Transit Car Maintenance and Overhaul Analysis, Final Report. Chicago Transit Authority, Chicago, Illinois, May 1985.
2. Rapid Rail Transit Vehicle Guidelines for the Operation and Use of the TRIP Data Bank. Contract DOT-TSC-1559. U.S. Department of Transportation, April 1979.
3. D. Muotoh and C. Elms. Cost Savings Potential From Improvement in Railcar Reliability and Maintainability. Report UMTA-It-06-0273-84-1. UMTA, U.S. Department of Transportation, April 1984.

Publication of this paper sponsored by Committee on Rail Transit Systems.