Trip Generation in the Transportation Planning Process

CHRISTOPHER R. FLEET and SYDNEY R. ROBERTSON, Urban Planning Division, U.S. Bureau of Public Roads

This paper attempts to serve two purposes with respect to trip generation analysis. Research findings concerning data variations and aggregative effects are outlined. Also, the implications of the continuing phase of transportation planning for trip generation methodology are discussed.

Trip generation estimating procedures have generally relied heavily on statistical methodology. Such procedures require a sound knowledge of the structural relationships contained in the basic data set. Several suggestions are offered for the operational transportation planner concerning trip generation methods. By utilizing standard statistical techniques, it is possible to identify the sampling variations associated with zonal aggregate data and to illustrate the realistic level of accuracy to be expected from zonal estimating procedures. Other simple statistical procedures make explicit the effects of aggregating trip-making relationships which are, in actuality, most meaningful at the individual household level. It is concluded that the typical "fine tuning" of a multiple regression trip generation analysis with aggregate data may be of marginal value. The compelling logic of a disaggregate analysis of trip generation relationships is indicated.

Trip generation estimating relationships are derived from cross-sectional data and are subject to change with time. These relationships must be evaluated for stability periodically. Little has been written concerning the trip generation phase of the continuing urban transportation planning process. Several of the existing and proposed procedures are particularly valuable as a basis for an efficient monitoring and reevaluation program. Three trip generation analysis procedures are suggested for use in the continuing planning phase. Rather than implementing an extensive resurvey, well-designed small sample cross-sections and on-site surveys appear to hold promise for periodic reevaluation of basic trip generation estimating procedures.

•A NUMBER of papers have appeared concerning the methodology of what has been generally called trip generation analysis [see summary of literature (1) and discussion (2)]. The bulk of this work has been directed toward the improvement of estimating procedures relating observed travel to household characteristics and land use activities. Major interest and effort have been in the direction of statistical efficiency and selection of optimum relationships. Little effort has been devoted to understanding the inferences concerning travel behavior that are implicit in these procedures.

Paper sponsored by Committee on Origin and Destination and presented at the 47th Annual Meeting.

The first section of this paper explores the structure of several simple travel relationships and household characteristics, using standard statistical procedures, in order to make explicit the consequences of typical trip generation study assumptions. Two general aspects are treated: data variations and the effects of areal aggregation. The sampling variations of zonal aggregate data are identified and related to a level of accuracy that can be realistically expected from zonal estimating procedures. Partitions of variance are utilized to illustrate the effects of areal aggregations (zones, districts, etc.) of what are, in actuality, trip-making relationships meaningful at the individual household level.

These notions are by no means new. For example, the problems attendant to areal aggregations of behavioral data have been discussed in the literature of sociology (3) and of geography (4). An excellent discussion of the important aggregation considerations in the construction of a behavioral system simulation model can be found in Orcutt, et al (5). A recent project suggests the importance of this problem in the trip generation study (1). The purpose here is to illustrate, with a typical transportation study data set, the magnitude of the problem and to offer some suggestion as to the implications of these findings for the typical trip generation analysis.

A second, and somewhat different, problem is the role of trip generation in the continuing study. Trip generation estimating relationships are derived from cross-sectional data and are subject to change with time. These relationships must be periodically evaluated for stability. Though other transportation study models require equal consideration, trip generation supplies the most direct link to the vital changes in the land use pattern and deserves particular attention.

Little has been written concerning the methodology for continuing reevaluation of the derived estimating procedures. The second portion of this paper is intended to evaluate some of the current methods, plus some newer and less widely known procedures, to determine if they can be adapted to provide an efficient monitoring and reevaluation program in the continuing phase.

The trip generation phase is of interest due to its relative sensitivity to the quality and adequacy of the data used in the estimating procedures. Previous studies in which comprehensive travel models have been developed have noted that results are particularly dependent upon the adequacy of the data available for the trip generation estimates (6, 7, 8, 9, 10, 11). The justification for the collection of several of the detailed household travel data items is primarily the trip generation study. Other technical phases of the transportation study require significantly fewer data and less detail in those data. With this premise in mind, this discussion will delineate some of the problems that the analyst faces in using the standard home interview survey data (12). Suggestions will be made, both for traditional studies and continuing analysis, concerning both the use of these data and the efficacy of alternative methods of using these data.

The findings reported here are from research conducted by the Urban Planning Division involving a number of aspects of the typical trip generation study. This research has also formed the basis for a publication which documents more fully the state of the art of trip generation methodology (2).

TRIP GENERATION IN THE TRANSPORTATION STUDY

Decisions concerning expenditures for transportation facilities and services in urban areas have become particularly important. Each choice has complex implications for the entire urban community. To aid in making these decisions, effective and accurate forecasts of travel demand are necessary. Given the impetus of the 1962 Federal-Aid Highway Act, these forecasts are generally derived within the framework of the urban transportation study.

These studies are characterized by a systematized process which, when carried through, serves as a basis on which to plan, design, and evaluate transportation systems. This process is generally considered to be comprised of the following key technical phases: population and economic studies, land use allocation, trip generation, trip distribution, modal split, and traffic assignment. Though necessarily oversimplified, the essential elements are illustrated. Of these, only trip generation will be treated in this paper.

The trip generation phase is intended to prepare forecasts of travel demand by small geographic area. Travel demand is used here in the restrictive sense of trip-making frequency. The result is, in essence, a spatial distribution of travel demand, or simply a frequency of trip-making, defined at one end of the trip and stratified by the type of trip being made. This is the stage of the travel forecasting process where the traditional linkage between land use and travel is introduced.

This paper deals primarily with a restricted subpopulation of generated travel—those trips beginning and ending at the household. These trips are of primary interest because of their direct relationship to the characteristics of the household. More information concerning travel motivation is available in this situation. The nonhome end of the trip presents other problems which would cloud the issues—a discussion of some of these problems can be found in two recent reports (13, 14).

Various approaches to trip generation estimation have evolved within the last two decades. In almost all cases, however, the analyses are similar. Typically, the total number of trips generated in a given areal unit is related to either average measures of the characteristics of the households in that unit or aggregated characteristics of the unit itself. The geographic unit is generally characterized by land use activity or intensity measures, while households are identified by certain "socioeconomic" data. Using this information, estimating procedures are normally derived by statistical methodology. The derived relationships are then presumed to remain stable over time, and trip forecasts are obtained by introducing changes in the socioeconomic and land use characteristics. This process uses vast amounts of descriptive data. The data set concerned with household travel habits is both extensive and costly. In many cases, the travel inventories comprise over one-half of the urban transportation study budget. It is important, then, that efficient use be made of these data.

The transportation study typically works with traffic analysis zones. This need for relatively fine geographic detail pushes the typical household travel survey data set to the limits of its usefulness. An examination of the sampling variations that are associated with this small area use of data will offer insight into data efficiency. Relatively little has been done to actually examine the nature and effect of these variations and to indicate what level of accuracy one should expect from these data. The areal aggregation introduced by using traffic zones requires relatively restrictive assumptions concerning the spatial distribution of variation in the derived travel parameters. Rarely are these assumptions actually evaluated. This situation also has major implications for efficient data utilization.

AGGREGATION EFFECTS AND DATA VARIATIONS

Most current methods of trip generation analysis rely on areally aggregated data. The level of aggregation is generally the traffic zone and occasionally the traffic district. This level of analysis has primarily evolved out of the procedures used in the traffic assignment process. The extent of aggregation is usually limited, on the upper side, by the point-loading concept of traffic assignment (that all trips emanating from an areal unit are loaded on the network at a common point, usually the center of activity within the unit). The lower limit of aggregation is a function of statistical estimating efficiency.

Aggregation Effects

The underlying assumption of areal aggregations is that contiguous households exhibit some similarity in family and travel characteristics, thereby allowing them to be grouped or aggregated with mean parameter values used for each group. Statistically this implies that the mean value of any particular parameter is reasonably representative of all households within the specific areal unit. It also implies that differences between areal unit mean values are more expressive of the spatial distribution of the parameter than are the extremes to be found in a single areal unit. Obviously, if there are major differences within single areal units, this assumption of homogeneity is inappropriate.

Strict homogeneity is practically impossible to obtain, particularly with the complex combinations of urban land uses. If all areal units are fully homogeneous, complete use is made of the disaggregate dwelling unit information. The degree to which these units are not homogeneous results in a loss of disaggregate detail. If sufficient detail is lost, then full value is not obtained for the survey dollar and changes in sample design or use should be considered. As an example of information loss, consider the commonly used car ownership variable. The number of cars owned by a particular family has direct implications for the family's frequency of trip-making. When this variable is aggregated by areal unit and any degree of heterogeneity exists, the relationship between travel and vehicle ownership at the family level is clouded. The net result is that the significance of a vital estimating relationship is reduced.

Aggregation effects are dramatically illustrated by looking at changes in the simple correlation coefficient during the process of aggregation. The simple correlation coefficient (r) is a measure of the degree of association between two variables (15). In Table 1, the lower correlation at the dwelling unit level implies that a differentiation between family size and car ownership exists at this level. The correlation between equivalent zonal aggregates is significantly higher. This implies that most of the distinction between car ownership and family size is lost in the process of aggregation. This illustrates the inherent danger in making statistical inferences from aggregate data concerning disaggregate relationships (3).

To make these problems of aggregate explicit, it is necessary to examine them in terms of measures of variation. A one-way analysis of variance (ANOVA) computer program (16) was used to partition the total variation into that attributable to the differences between groups and that associated within each group. Unexpanded dwelling unit data were used. Data variation was expressed in terms of summed squared deviations about a mean value. Mathematically, this is the total sum of squares (TSS) and is expressed by:

$$\sum_{j=1}^{k} \sum_{i=1}^{n_j} (X_{ij} - \overline{X})^2 \qquad (1)$$

where

X_{ij} = the i th observation on some variable (X) taken in the j th group (areal unit);

$\overline{X}$ = the grand mean of variable X $= \dfrac{\sum_{j=1}^{k} \sum_{i=1}^{n_j} X_{ij}}{N}$;

N = the grand total number of observations $= \sum_{j=1}^{k} n_j$;

n_j = the total number of observations in the j th group; and

k = the number of groups.

If the data are grouped, then the TSS can be partitioned into two components. The first is the sum of the squared deviations between the group means and the grand mean, weighted by the number of observations in each group. This is the between sum of squares (BSS). The within sum of squares (WSS) is computed by summing the squared

TABLE 1

SIMPLE CORRELATIONS BETWEEN FAMILY SIZE AND CAR OWNERSHIP AT THE ZONAL AND DWELLING UNIT LEVELS

Relationship	Observation Unit	r
Family size vs cars owned	Households	0.400
Total persons vs cars owned	Households summed by zone	0.955

deviations of each observation within the group from the group mean. Mathematically, the two components of the TSS may be stated as:

$$\mathrm{TSS} = \sum_{j=1}^{k} n_j \, (\overline{X}_j - \overline{X})^2 + \sum_{j=1}^{k} \sum_{i=1}^{n_j} (X_{ij} - \overline{X}_j)^2 \qquad (2)$$

where $\overline{X}_j$ is the j th group mean of variable X.

The two components in the equation (BSS and WSS) can be interpreted as measures of the variation between the groups and within the groups, respectively. Typical multiple regression trip generation estimates are developed utilizing the variations between groups. Variations between the groups are attributed to the differences between them. The amount of this component (BSS) that is actually "explained," however, is a function of the degree of homogeneity within the groups. Homogeneity, in this case, implies a small (or zero) within-group sum of squares. By computing the WSS for various typical areal units, the homogeneity assumption can be evaluated. At the same time, effects of increasing levels of aggregation on the BSS and WSS can be seen. This information is summarized in Table 2. The results indicate that, by far, the greatest portion of the total variation is within areal units and is lost, insofar as its usefulness to trip generation analysis is concerned. In particular, a zonal level regression analysis applied to the data of Table 2 would deal with only a little over 20 percent of the total variation of trip-making by dwelling units. At the district and sector level, the proportions are progressively less.

As the degree of aggregation increases, the amount of variation between groups decreases. This is of particular importance in a regression analysis (1, 2, 13). As the variation between groups decreases, it becomes less difficult to obtain a good data fit, as there is less variation to deal with. Much of the meaningful variation has been eliminated. It is this deceptive "increase" in statistical reliability that has led some analysts to work at higher levels of aggregation (districts, for example). "Increased" statistical efficiency, however, is not meaningful unless the comparison is being made on a common base. Measures of accuracy derived at different levels of variation are not comparable. An example will illustrate both the problems involved and the manner in

TABLE 2

BSS AND WSS OF TOTAL TRIPS PER DWELLING UNIT, GROUPED BY THREE AREAL UNITS

Areal Unit (Number)	WSS	% of TSS	BSS	% of TSS
Dwelling units (5,255)	0	0.0	213,936[a]	100.0
Zones (247)	170,270	79.6	43,666	20.4
Districts (57)	184,864	86.4	29,072	13.6
Sectors (10)	192,895	90.2	21,041	9.8

[a] This is also the total sum of squares in all three groupings.

which estimating procedures derived at differing areal levels can be compared as to efficiency.

To determine the effect of this spurious accuracy, multiple regression trip generation analyses were conducted on the same trip and socioeconomic data at two levels of aggregation, zones and districts. The final estimating equations are given in Table 3. Typically, estimates are required at the zone level and the district equations are adjusted to this level. The statistical tests used to evaluate the district equations say nothing about the adequacy of these equations at the zonal level of application. Moreover, it is statistically incorrect to attempt to "explain" the between-zone variation of the dependent variable with an equation that was developed using data aggregated by districts.

If the adjusted district level equations are utilized to derive zonal trip end estimates, and then these estimates are compared, by zone, to the actual values from the expanded survey, a comparison with the zonal regression results at a comparable level is possible. This was accomplished by deriving two squared correlation coefficients (15). One expresses the degree of fit between the zonal regression estimates and the base data (actually the multiple R^2 yielded by the regression program). The other expresses the fit between the zonal estimates from the district equation and the base data. The latter is the square of the simple correlation (r^2) between the zonal estimates from the district equation and the base data. This value (0.77) is given in Table 3. The comparison is enlightening. The security of the high r^2 at the district level (0.95) has vanished when the equation is applied at the zone level.

The aggregation problem has many facets, and perhaps the most important point of this discussion relates to the utilization of the basic household travel survey. The travel survey is costly and obviously should be used efficiently. In practice, this does not seem to be the case. Detailed data concerning household travel and characteristics are obtained, but the richness of this detail is washed out by aggregation. This becomes a difficulty when the analyst attempts to make inferences concerning the disaggregate behavior of households from these aggregate data. Many of the real reasons for trip-making are hopelessly concealed. If disaggregate inferences are to be made, Stowers and Kanwit conclude that it is logical for analysis to precede, rather than succeed, aggregation (17). Several other authors have noted the problems of utilizing the travel survey for trip generation analysis. Of particular interest are the comments by Shuldiner (18, 19). This discussion is not intended to suggest that the O-D survey be abandoned or in large part modified. Rather, it is hoped that more efficient use can be made of the data. Stowers and Kanwit suggest a possible alternate approach (17). They argue for the household as the basic unit of analysis, citing the compelling logic of such an analysis. Meaningful relationships are first developed at the household level and then aggregated to give zonal estimates. The important point is that full use can then be made of the household level information prior to aggregation. Details are left to a later section of the paper. If this method is to be useful, it must offer reliability at the same level (at least) as existing methods. A comparison was made by deriving zonal estimates using both aggregate and disaggregate data and determining the efficiency of both estimating procedures at this level.

TABLE 3

RESULTS OF REGRESSION ANALYSES CONDUCTED AT TWO LEVELS OF AGGREGATION

Level	Equation	Number of Observations	r^2	$\bar{Y}$
Zones	WP = 18.2 + 1.68 CO	283	0.80	315
Districts	WP = 143.0 + 2.12 CO	57	0.95	1,708
Zonal estimates with district equation		–	0.77	–

WP = home based work trip productions; CO = total cars owned; $\bar{Y}$ = mean of dependent variable.

Table 4 gives two regression equations for estimating total home-based trips. The first equation was developed using trip and socioeconomic data expanded and aggregated to zones prior to analysis. These data were dwelling unit samples from a home interview survey. Trips were expressed as total home-based trips per zone. The second equation was derived from the raw dwelling unit data. The data were taken as a group and were left unexpanded. Here, the data represented a cross-section of all households in the study area. Trips were expressed as total home-based trips by each dwelling unit.

The immediate conclusion (Table 4) based on the standard errors ($S_{y.x}$) and coefficients of determination (R^2), is that the zonal equation is superior (15). These results are misleading. As noted before, differing definitions of variation are associated with different levels of aggregation. These two statistics should be recomputed after the dwelling unit equation is applied at the zone level.

Using the number of households in each zone, zonal trip end estimates were derived from the dwelling unit equations. In the same manner as before, a squared correlation coefficient (r^2) was calculated comparing these zonal estimates to the base data. The percent standard error was also calculated. These adjusted values are shown in Table 4. The adjusted proportion of the variance explained (r^2) was 0.94, and the adjusted percent standard error was 19.4. This reveals that the dwelling unit equation, applied at the zonal level, can "explain" 94 percent of the zonal variation in trip-making. This, taken with the percent standard error, shows that either method yields practically the same accuracy at the zonal level.

This discussion implies that the typical "fine tuning" of zonal regression equations may be of questionable value. In comparison with the information losses due to aggregation, whether the derived zonal equation has a coefficient of determination of 0.75 or 0.90 is relatively unimportant. Considering that much of the meaningful variation has been washed out, whether 75 percent or 90 percent is explained may not be worth the effort commonly devoted to improvement.

The more logical behavioral base of dwelling unit analysis, given that it is comparable to zonal analysis, supplies a compelling argument for the use of this type of analysis. The dwelling unit relationships, because they have not lost their meaning to aggregation, are more likely to be stable over forecast intervals. These relationships are most accurately measured at the level of the greatest detail, the household. In transportation planning, the interest lies in household travel patterns, not in aggregate zonal changes. Thus, logic would suggest that the analysis should be conducted at the household level. The main point of this entire section on aggregation is that these problems are important and that their recognition is of importance to any transportation study. Such an analysis will permit a more efficient trip generation study.

Household Variations and Homogeneity

One of the major considerations in the efficiency of aggregation is the degree of homogeneity in the areal units. Ideally, aggregated households should exhibit almost ex-

TABLE 4

RESULTS OF REGRESSION ANALYSES CONDUCTED AT THE ZONE AND DWELLING UNIT LEVEL

Level	Equation	Number of Observations	R^2	$S_{y.x}$	$\bar{Y}$	% $S_{y.x}$
Zones	TT = -36.03 + 5.09 CO	143	0.95	296.07	1,679	17.6
Dwelling units	TT = -0.69 + 1.39 NP +1.94 CO	5,255	0.36	3.89	5.20	74.9
Dwelling units at the zone level		—	0.94	—	—	19.4

TT = Total home based trips
CO = Total cars owned
NP = Number of persons 5 years of age or older
R^2 = Coefficient of determination
$\bar{Y}$ = Mean of the dependent variable
$S_{y.x}$ = Standard error of estimate
$\%S_{y.x}$ = Standard error divided by the mean of the dependent variable

TABLE 5

STATISTICS ON CARS OWNED PER DWELLING UNIT
FOR EIGHT ARBITRARILY SELECTED ZONES

Zone No.	Mean	Standard Deviation	Standard Error of the Mean	Percent Standard Error	Maximum	Minimum	Number of Observations
20	0.7143	0.6437	0.1405	19.7	2	0	21
35	0.9861	0.3137	0.0370	3.8	2	0	72
52	0.9583	0.6829	0.0986	10.3	3	0	48
65	1.1000	0.6074	0.1109	10.1	3	0	30
136	1.2373	0.5177	0.0477	3.9	3	0	118
140	0.8889	0.5774	0.1111	12.5	2	0	27
168	1.3881	0.7579	0.0926	6.7	3	0	67
216	1.3667	0.5813	0.0750	5.5	2	0	60
Total area	1.0558	0.7264	0.0100	0.9	5	0	5,255

actly the same characteristics. Differences should primarily occur between areal units. The results reported above point to the conclusion that zones are not homogeneous. This can be tested directly by studying the distribution of characteristics within a single zone. The standard deviation (s) can be used to represent the extent that a range of values of a characteristic can be found in a single zone (15). Comparing this statistic to the standard deviation computed for the entire study area yields a relative index of homogeneity.

Table 5 presents several statistics for dwelling unit car ownership calculated by zone. The zones were arbitrarily selected. Note that in most of the zones s is nearly as large as the value of s is for all of the sampled dwelling units in the area. The hypothesis of zonal homogeneity, in the case of car ownership, is not supported. The range of dwelling unit car ownership is nearly as extensive within each zone as within the entire area. If all zones are considered, this conclusion is reinforced. A crude index of homogeneity was calculated for each zone by dividing the value of s for the zone by the value of s for all dwelling units in the study area (Table 6). Again, the characteristic is household car ownership. In general, there are large variations in dwelling unit car ownership within each individual zone. Of the 207 zones, 18 percent have as large or larger variations than the total area, and 72 percent have at least three-fourths of that for the total area. Similar calculations were made for typical household characteristics (family size, for example) and for various trip types, yield-

TABLE 6

FREQUENCY DISTRIBUTION OF THE HOMOGENEITY INDEX
FOR CAR OWNERSHIP OF ALL ZONES

Homogeneity Index $= \frac{s_{\text{d.u.'s in zone i}}}{s_{\text{all d.u.'s}}}$	Number of Zones	Percent of Total Zones	Cumulative Percent
1.55 & over	5	2	2
1.28 to 1.54	5	3	5
1.14 to 1.27	11	5	10
1.00 to 1.13	16	8	18
0.86 to 0.99	52	25	43
0.72 to 0.85	60	29	72
0.59 to 0.71	49	24	96
0.31 to 0.58	9	4	100
Totals	207[a]	100	

[a] Excludes zones with no dwelling units.

ing similar results. For the sake of brevity, these results will not be repeated here. It seems safe to conclude that, for the purposes of trip generation analysis, zones are heterogeneous.

Sampling Variations

In a zonal regression trip generation analysis, aggregate trip and socioeconomic characteristics are utilized to develop estimating equations to forecast future trips. For each zone, either average or total values of each characteristic are taken to be representative. Both average and total values rely upon the adequacy of the zonal mean as an estimate. In the first case, the average is used directly. Zonal total values can be thought of as being the zonal average multiplied by the number of dwelling units in the zone.

Depending on the number of samples (dwelling units) in each zone, the mean value of each zonal characteristic will not be known exactly, but as within some range. As the number of samples within a zone decreases, the range of likely values for the mean expands, and we become less confident of the mean as being representative. The variance of the characteristic also affects the magnitude of this range. These consequences are well known from elementary sampling theory.

Rarely is any consideration given to the magnitude of the variations introduced, because of sampling, into the inputs to zonal regression trip generation analysis. As the estimating equations are directly developed from these inputs, some consideration is essential. The degree to which the regression analyses can be reasonably expected to fit these data is dependent upon the amount of variation in the data. A degree of "fit" to closer tolerance limits than those associated with the input data due to sampling is spurious.

The effects of sampling variation on the sample mean of each zone can be represented by the standard error of the mean (15):

$$S_{\bar{x}} = \frac{s_x}{\sqrt{n}} \tag{3}$$

where

$S_{\bar{x}}$ = standard error of the mean;
s_x = standard deviation of X; and
n = sample size.

The standard error of the estimate ($S_{y.x}$) from the zonal regression equation indicates the variation to be expected in the estimates of the dependent variable derived from Eq. 3. For each of these statistics, it is more meaningful to standarize them by expressing each as a percent of the mean value to be estimated. They are expressed as the percent standard error of the mean and the percent standard error of the estimate, respectively.

An analysis of the distribution of the zonal percent standard errors of the mean ($S_{\bar{x}}$) compared with the regression standard error of the estimate ($S_{y.x}$) can provide an indication of the extent to which the analyst should attempt to improve $S_{y.x}$. Regression standard errors that are pushed to greater accuracy than the majority of the zonal standard errors result in false precision. Estimates from such an equation have less variation than that known to exist due to sampling. The accuracy of the estimating procedure can only be expected to approach or equal the accuracy of the data being "fitted" and further "fine tuning" is meaningless. Thus, zonal percent standard errors of the mean can be used to indicate the point at which, in terms of $S_{y.x}$, the regression analysis can be terminated. A suggested graphic representation is shown in Figure 1. Because a value of the percent $S_{\bar{x}}$ is derived for each zone, a frequency distribution of these values is convenient. The distribution illustrates the pattern of zonal sampling variation.

In Figure 1, the regression dependent variable is home-based work trip productions per zone. The mean of the distribution of the work trip zonal percent standard errors ($S_{\bar{x}}$) is represented by the vertical dashed line. The solid line represents the regression percent standard error for a zonal work trip estimating equation developed using these same data. If the mean value (23.3) is selected as a cutoff criterion, then a similar value of $S_{y.x}$ should be an indication that the equation needs little further improvement. The equation in Figure 1 has, perhaps, been carried far enough. In this case, regression is the limiting factor. If the regression analysis had been carried further, such that the percent standard error became 15.0 (vertical dotted line, Fig. 1), this would be a good example of unfounded "fine tuning."

If such an analysis is included as an integral part of the trip generation study, the analyst is provided with considerable information regarding the nature and quality of the available data. In addition, it is possible to decide the extent to which the accuracy of regression trip generation equations can be improved without overextending the accuracy of the input data.

TRIP GENERATION IN THE CONTINUING TRANSPORTATION PLANNING PROCESS

When the necessary steps have been taken to prepare the first target year forecasts, the transportation planning process is by no means complete. These forecasts must be responsive to changes occurring in the urban area, and particularly to changes in the basic forecasting relationships. The continuing transportation planning phase supplies the framework for this monitoring and reevaluation function. This continuing process necessarily requires that current information on the basic parameters of travel be available in order to make comprehensive evaluations and revisions of the original forecasts. Two types of information are required, current estimates of the inputs to the forecasting procedures and data describing the changes (or stability) in the estimating relationships.

In the continuing phase, trip generation relationships are of particular interest. They deserve attention because of the direct link that is provided to the vital changes in the land use pattern. Any major changes in household travel behavior are reflected here. Trip generation estimating relationships are derived from cross-sectional data and are subject to change with time (20). These relationships must be evaluated periodi-

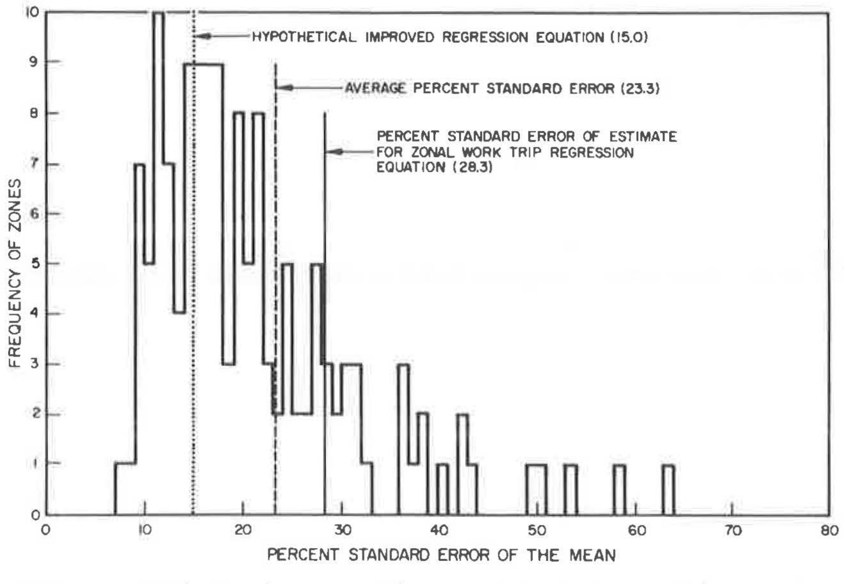

Figure 1. Frequency distribution, by zone, of the percent standard error of the mean for work trips.

cally for stability. This continuing reevaluation of trip generation estimating procedures is a vital element of the continuing process. Little has been written concerning this element.

The second major consideration in providing for adequate continuing trip generation is the availability of pertinent recent data to allow the preparation of current estimates. One way to assure this availability would be periodic full-scale surveys. Cost rules this alternative out except at relatively long time intervals. Between surveys, other methods must be used.

The extent of and procedures for continuing trip generation analysis will be different in each study. It is not possible to offer any detailed procedures for any part of the continuing phase. Much of what will be done in the continuing program will depend on what was done in the initial phases. For this reason, the procedures outlined should only be taken as illustrative of what might be accomplished. Each study must adapt the procedures as appropriate.

Several new approaches to trip generation analysis have recently emerged, although their applicaton has been somewhat limited. Of these, three are presented that have particular potential as efficient tools in the continuing process.

Dwelling Unit Analysis

Travel behavior is most efficiently studied at the household level. Vital trip-making relationships are most explicitly identified at this level of detail. Continuing monitoring and evaluation of the trip-making estimating procedures will require information on these relationships at this level of detail. Dwelling unit analyses can efficiently summarize these relationships for the entire area, allowing important changes to be quickly identified. This information can then be used to revise the basic trip generation procedures.

Previous discussion has emphasized the mechanical and statistical aspects of "data fitting." There is a more important argument for the use of a disaggregate, dwelling unit analysis. It is generally felt that a family's travel behavior is, in part, a product of the unique characteristics of that family. These relationships lose their meaning if they are aggregated. Monitoring of the aggregate relationships, however, gives little information on the components of change. If the subtle changes in household travel behavior are to be understood, it will be at the disaggregate level and not at the aggregate level. Aggregate measures of change are only the external manifestation of numerous possible combinations of travel behavior shifts.

Several more practical reasons suggest the use of a dwelling unit level analysis in the continuing phase. The costs of monitoring changes are particularly important. Small sample surveys are easily and effectively used with dwelling unit analyses, particularly where the intent is to determine the changes which have taken place in the basic relationships. Though some effort will be necessary to obtain a well-designed sample, the appropriate monitoring information can be obtained at a small fraction of the cost of a full survey. The linearity assumptions of aggregate regression analyses also present some difficulty. If, for example, there is a nonlinear relationship at the dwelling unit level, bias is introduced by an aggregate regression analysis. It is possible to avoid this problem by working at the disaggregate level and employing a different method of variable definition.

Often, the areal units of an aggregate trip generation analysis change, making the base year work incompatible with subsequent work. Most aggregate analyses use variables which directly reflect the size of the areal units. If these units change, the analysis must often be redone. The application of a dwelling unit equation is not limited to established analysis units. Past experience indicates that the flexibility to adapt to changes in areal units is a valuable asset in the continuing process.

A further consideration is the expectation that there will be growing awareness of the need to study human behavior at a disaggregate level. This paper has argued for the logic of disaggregate travel analysis, as have others (1, 17). In the area of land use development models, these same arguments are appearing, e.g., Harris, Garrison, and Schlager (21). On the empirical level, there is the summary of the work done by the University of Michigan in studying household location choice and travel behavior (22).

Though these ideas have affected the analysis procedures of only a few transportation studies, it seems logical to expect that this will be a growing trend. If, as Schlager suggests, the household location decision is to be stratified by particular household types, then this same identification will be available for trip generation. This supplies one of the major inputs to a disaggregate trip generation analysis. It would appear that the state of the art of land use models is turning toward more disaggregate model development. It follows that trip generation studies, because of the intimate relationships of travel and land use, should begin moving in this direction.

The actual application of dwelling unit estimating procedures to the development of areal estimates is relatively efficient. While a single areawide small sample cross-section can be used to evaluate the magnitude and direction of changes, small samples for each area having identified changes will be required for updating. Means of the independent variables for each areal unit and the total number of households contained in the unit are also required. The sample would never actually be expanded, however. The following general equation could then be used to obtain trip estimates:

$$\hat{Y}_w = Na + Nb_1\overline{X}_1 + \ldots + Nb_i\overline{X}_i + \ldots + Nb_n\overline{X}_n \qquad (4)$$

where

$\hat{Y}_w$ = the estimate of the dependent variable for area w;

N = total number of dwelling units in w;

$\overline{X}_i$ = the i th independent variable mean for the analysis area, based on the sample;

b_i = the regression coefficient of the i th independent variable; and

a = the regression constant.

Equation 4 presumes linear relationships. Many of the significant household characteristics are not easily scaled in this manner. Stage in the family life cycle, occupation of the head of the household, or structure type are typical examples (17). "Dummy variable" regression analysis allows such qualitative variables to be used and circumvents the restrictions usually associated with linearity.

Using this technique, household characteristics are stratified into meaningful categories. Household variation is then associated with the differences between the several categories of household types rather than by absolute scale ratios, as in the linear regression case. Thus, qualitative household groupings are used to explain the behavior of the households.

The application of the dummy variable technique is identical to the usual dwelling unit regression methods, except that now the number of households associated with each household type is required. The mean values of the linear regression parameters, by area, are also required and can be obtained by small sample surveys. Estimates can be used to obtain the count of households in each dummy variable category. If a residential location model with household type stratifications has been developed, household counts by type are readily available. The general utility of this method and its use of behavioral data make it an extremely useful tool.

Cross-Classification Analysis

Another recently introduced technique offers many of the same capabilities as dwelling unit analysis. This method, cross-classification analysis, has been largely limited to research applications (10, 11, and 18). It is becoming more widely used by operational transportation studies (23, 24).

Essentially, a multidimensional matrix is constructed, each dimension representing an independent variable (household characteristic). These characteristics are then stratified into meaningful categories. Each dwelling unit observation is allocated to a cell of the matrix based on the values (categories) of the independent variables. The

TABLE 7

RELATIONSHIP OF FAMILY SIZE AND AUTO OWNERSHIP TO AVERAGE TOTAL PERSON TRIPS PER DWELLING UNIT

Number of Persons per Dwelling Unit	Average total person trips per dwelling unit				
	Number of autos owned per dwelling unit				
	0	1	2	3 & Over	Weighted Average
1	1.03	2.68	4.37[a]	–	1.72
2	1.52	5.13	7.04	2.00[a]	4.38
3	3.08	7.16	9.26	10.47	7.46
4	3.16	7.98	11.56	12.75	9.10
5	3.46[a]	8.54	12.36	17.73[a]	10.16
6-7	7.11[a]	9.82	9.61	16.77[a]	11.00
8 & over	7.00[a]	9.65	6.18	11.00[a]	12.24
Weighted average	1.60	6.62	10.53	13.68	6.58

[a] Average based on fewer than 25 samples.

dependent variable (trips of some type) is accumulated by cell, and the average determined for each cell. A typical application (two-dimensional) is given in Table 7, and a graphic representation is shown in Figure 2.

Changes in a dependent variable are studied when two or more independent variables are varied. In this respect, the methodology is the same as regression analysis using only dummy variables. By holding one or more of the independent variables constant, the effect of varying a particular variable can be studied.

This technique makes particularly full use of the household travel survey data and, by its nature, offers the analyst the opportunity to work closely with the data. It is also not bound by the usual assumptions of linearity. For these reasons, the same positive comments made about the data efficiency associated with dwelling unit analysis apply here. Perhaps the greatest limitation is imposed by the amount of data required for adequate representation and statistical stability. Despite this, the technique is straightforward and efficient and offers none of the problems often encountered with curvilinearity and the treatment of qualitative variables.

Much of the early work with this technique as a tool in trip generation analysis was undertaken by the Puget Sound Regional

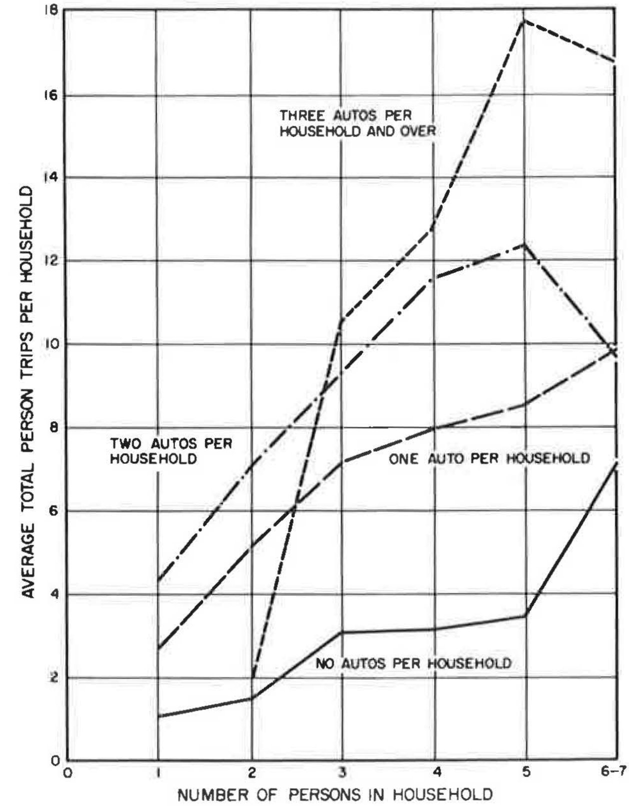

Figure 2. Average total person trips per household related to family size and car ownership.

Transportation Study (23, 24). The matrix of values was initially developed using the dwelling unit data and then applied at the zonal level. Household and density characteristics were used in a method that employed weighted zonal rankings based on zone averages of five independent variables: household size, cars per household, median income of the head of the household, population per acre, and population per net residential acre.

To evaluate the changes in basic estimating relationships, the use of the cross-classification and dwelling unit techniques are identical. In practice, this matrix method may be less complex and more efficient in the monitoring of change than a full dwelling unit analysis. The data, as before, need only be from an areawide small sample survey.

The preparation of new forecasting procedure inputs is also similar to that discussed in the previous section. Again, small sample estimates of zonal characteristics are required. In the case of the Puget Sound method, the new ranking of each zone must be determined. The more general case requires an estimate of the number of dwelling units, by analysis area, that are expected to fall into each matrix cell.

Site Analysis

Most current methods of analyzing nonresidential trip generation rates rely on trip information collected in the home interview survey, while the independent variables are usually obtained from other sources. Home-based work trip attractions, for example, are usually based on employment estimates at the nonhome end of the trip. This approach assumes, of course, that the characteristics of trips attracted to nonresidential land, obtained in the sample at the home end, are representative of similar trips by all other households that the sample represents.

A recent study has indicated the utility of using O-D data for estimating nonresidential trip generation (25). Another study has noted that O-D data are not adequate for this purpose (13). Both views are within the context of nonresidential trip generation analysis as a tool in project and facility planning, rather than in traditional transportation systems planning. In the development of base year relationships, the interest is in total systems planning rather than individual sites. This changes in the continuing phase when the impact of new major single site traffic generators is of particular importance. While steady growth will often be found throughout the region, evaluating the transportation consequences of new major generators will be a large task in the continuing phase. Examples of such generators would be shopping centers, airports, and hospitals.

In these cases, better information on the generation of travel can be obtained by collecting data at the site rather than relying on home interview data. This is not as prodigious a task as it seems. These major generators are relatively few in number, even in an entire area. One study found, for example, that a large percentage of the nonhome trip ends (70 percent) were actually attracted to a relatively small proportion (15 percent) of the parcels in the study area (14). Concentrating on these relatively few important attractors would provide more accurate estimates than spreading the same effort thinly over all areas. This does not suggest that home interview trip end data should be replaced. Rather, it is intended that the basic data be supplemented with more information for the few sites which contribute large amounts of traffic. In combination, on-site data and home interview data can place nonresidential (and some residential) trip generation analyses on a much more stable base.

In the base year procedural development, site analyses will be useful to improve the accuracy of nonresidential trip generation estimates. Here the use of this technique is supplementary. In the continuing phase, on-site data collection at new major generators can supply much of the necessary update information. These data, in combination with small sample updates of the nonresidential trip generation rates for the entire area, can provide the framework for a continuing trip generation program.

Site analyses have been little used by transportation studies except for one-shot single site studies. A good example of a comprehensive program is that reported by the California Division of Highways (26, 27). Traffic entering and leaving selected sites was related to characteristics of the particular facility to obtain trip rates. Such rates as trips per employee, trips per square foot of floor area, and trips per hospital bed were obtained. Counts were obtained on an hourly basis over a period of from two or three days to a week.

The Chicago Area Transportation Study has recently employed these methods to develop trip end estimates for a high rise apartment, for O'Hare International Airport, and for walking trips in the Loop (28). Here, both interview techniques and traffic counts were used.

It seems reasonalbe to conclude that on-site surveys for major generators should play an increasing role in both the initial and continuing phases of transportation planning. Such methods should make trip end estimates more reliable and allow the home interview data to be used more appropriately.

SUMMARY

The intent of this paper has been to bring trip generation analysis into perspective in light of recent practice and developments. Major emphasis has been placed on illustrating the consequences of the typical assumptions inherent in the derivation of generation forecasting procedures. Much of this has been based on standard statistical methodology. Though it is hoped that transportation studies will find these procedures useful, the primary motive has been one of understanding. A major failing of most trip generation forecasting procedures has been the lack of effort devoted to understanding the inferences concerning travel behavior that are implicit in these procedures.

The second purpose of the paper has been to offer a beginning framework for continuing trip generation analysis. Two major elements, monitoring and updating, are associated with a continuing trip generation program. Trip generation forecasting procedures are subject to change with time. The monitoring function supplies the information necessary to indicate significant change in the derived forecasting relationships. In addition, inputs to these forecasting procedures are periodically required to develop new estimates. This is the updating function. In the discussion of continuing procedures, the intent has been to outline efficient procedures to accomplish the two continuing generation elements. Adaptations of three procedures are suggested. Major data collection costs are reduced by utilizing small sample cross-sections and on-site surveys.

Perhaps the primary point of the first section of this paper is that trip generation analyses are too often conducted by rote. An inordinate amount of faith is typically placed in the traditional approaches to analysis. This particular phase of transportation planning is frequently too product oriented. Each of the points made in the first section refer to a method of dissecting these procedures to make explicit the assumptions made. It is this disassembly which is vital to understanding.

Much of what has been discussed relates to the basic travel data. Here two points are important. The high cost of these data requires efficient use. Several of the comments made above indicate that more efficient use could be made of the data by employing different procedures. In particular, disaggregate dwelling unit level analyses are strongly supported, based both on data utilization and compelling logic. The second point is that existing methodologies tend to impute more validity to statistically derived forecasting procedures than is warranted. The false security of superficial accuracy tends to hide the real value of each procedure and to allow the selection of the wrong one.

The problems of aggregation are paramount in trip generation analysis. It is not realistic to infer disaggregate (household) travel relationships from aggregated data. It is vital that the trip generation analyst understand the problems attendant to aggregated data, as almost all studies are performed at an aggregate level. The procedures suggested in the paper can be used by any analyst for this purpose. Perhaps the problems of aggregation are best summarized by the general comment that aggregation should follow analysis, rather than the reverse.

The arguments for disaggregate analysis can be carried over into the continuing phase. The same advantages, plus the benefit of an effective and efficient method for periodic evaluations and updates of trip end estimates apply. The large costs of data collection are circumvented because of the ready adaptability of dwelling unit and cross-classification analysis to small sample survey methods. In addition, the analyses are not limited to established methodologies. Innovations, such as dummy variable regression techniques, allow more flexible analysis.

Nonresidential trip generation estimates have generally been a problem. In both the initial and continuing phases, on-site surveys can supply the information necessary to develop estimates at major sites. This is of particular value in the continuing phase where major efforts must be devoted to evaluating the impacts of new major generators. Site analyses also are a valuable supplement to the basic household travel surveys.

As trip generation estimates provide a basic ingredient for transportation planning, it is important that these estimates have as great a degree of reliability and validity as possible. Additional data collection or improvements in statistical precision do not necessarily result in increased forecasting capability. The real objective should be the design and utilization of techniques that both recognize data limitations and are structured in a sound and logical manner. This type of care, rather than statistical complexity, will yield usable forecasting procedures.

ACKNOWLEDGMENT

The data set used in this research was obtained from the Madison Area Transportation Study. The authors appreciate the cooperation of the Wisconsin State Highway Commission in making these data available.

REFERENCES

1. Garrison, W. L., and Worrall, R. D. Monitoring Urban Travel (draft copy of Final Report, Project 2-8, NCHRP). Transportation Center at Northwestern University, 1966.
2. U.S. Bureau of Public Roads. Guidelines for Trip Generation Analysis. U.S. Govt. Print. Off., June 1967.
3. Robinson, W. S. Ecological Correlations and the Behavior of Individuals. Amer. Sociol. Rev., Vol. 15, p. 351-7, June 1950.
4. Duncan, O. D., Cuzzort, R. P., and Duncan, B. Statistical Geography. The Free Press, Glencoe, Ill., 1961.
5. Orcutt, G. H., et al. Microanalysis of Socioeconomic Systems: A Simulation Study. Harper and Brothers, New York, 1961.
6. Heanue, K. E., Hamner, L. B., and Hall, R. M. Adequacy of Clustered Home Interview Sampling for Calibrating a Gravity Model Trip Distribution Formula. Highway Research Record 88, p. 116-36, 1965.
7. Ben, C., Bouchard, R. J., and Sweet, C. E., Jr. An Evaluation of Simplified Procedures for Determining Travel Patterns in a Small Urban Area. Highway Research Record 88, p. 137-70, 1965.
8. Horn, J. W., et al. An Investigation To Correlate Synthetic Land Use Origin and Destination Techniques to Field Conducted Origin and Destination Surveys. North Carolina State of the University of North Carolina at Raleigh, June 1964.
9. Horn, J. W., et al. Examination and Comparison of ADT Gravity Models. North Carolina State of the University of North Carolina at Raleigh, June 1966.
10. Parsonson, P. S. Comparison of Techniques for Estimating Zonal Trip Productions and Attractions. North Carolina State of the University of North Carolina at Raleigh, June 1966.
11. Parsonson, P. S., and Cribbins, P. D. Estimating Urban Trip Production and Attraction. Jour. Highway Div., Vol. 93, No. HW2, ASCE, p. 115-34, 1967.
12. U.S. Bureau of Public Roads. Manual of Procedures for Home Interview Traffic Study. U.S. Govt. Print. Off., Oct. 1954 (out of print).
13. Shuldiner, P. W. Non-Residential Trip Generation Analysis. Transportation Center at Northwestern University, Nov. 1965.

14. Thomas, E. N., Horton, F. E., and Dickey, J. W. Further Comments on the Analysis of Non-Residential Trip Generation. Transportation Center at Northwestern University, Nov. 1966.
15. Ezekiel M., and Fox, K. A. Methods of Correlation and Regression Analysis. Wiley and Sons, New York, 1959.
16. Dixon, W. T. (Ed.). BMD-Biomedical Computer Programs. School of Medicine, University of California, Los Angeles, Jan. 1964 (rev. Sept. 1965).
17. Stowers, J. R., and Kanwit, E. L. The Use of Behavioral Surveys in Forecasting Transportation Requirements. Highway Research Record 106, p. 44-51, 1966.
18. Oi, W. Y., and Shuldiner, P. W. An Analysis of Urban Travel Demands. Northwestern University Press, 1962.
19. Shuldiner, P. W. Trip Generation and the Home. HRB Bull. 347, p. 40-59, 1962.
20. Worrall, R. D. The Urban Panel as a Longitudinal Data Source. Highway Research Record 194, p. 62-77, 1967.
21. Economic Forecasting, 6 Reports. Highway Research Record 126, p. 1-37, 1966.
22. Lansing, J. B., and Hendricks, G. Automobile Ownership and Residential Density. Survey Research Center, University of Michigan, 1967.
23. Cowan, G. R., and Walker, J. R. Rationale for Trip Production Generation Analysis. Staff Report No. 6, Puget Sound Regional Transportation Study, Seattle, March 1964.
24. Cowan, G. R., and Walker, J. R. Forecasting Future Trip Ends. Staff Report No. 16, Puget Sound Regional Transportation Study, Seattle, Oct. 1964.
25. Keefer, L. E. Urban Travel Patterns for Airports, Shopping Centers, and Industrial Plants. NCHRP Rept. 24, 1966.
26. Goodrich, D. K., and Smith, C. L. Progress Report on Trip Ends Generation Research Counts. Calif. Dept. Public Works, Division of Highways, Sacramento, Dec. 1965.
27. Drosendahl, R. G., and Smith, C. L. Second Progress Report on Trip Ends Generation Research Counts. Calif. Dept. Public Works, Division of Highways, Sacramento, Dec. 1966.
28. Orzeske, J. D. How To Collect Data on Special Generators. Paper presented at the Origin and Destination Committee Meeting (Sept. 14-15, 1967), HRB.
29. Johnston, J. Econometric Methods. McGraw-Hill, New York, 1963.
30. Melichar, E. Least-Squares Analysis of Economic Survey Data. Proc. Business and Economic Statistics Section, Amer. Statistical Assoc., Washington, 1965, p. 373-85.
31. Suits, D. B. The Use of Dummy Variables in Regression Equations. Jour. Amer. Statistical Assoc., Vol. 52, 1957, p. 548-51.
32. Fleet, C. R., Stowers, J. R., and Swerdloff, C. N. Household Trip Production, Results of a Nationwide Survey. Highway Planning Technical Report No. 2, U.S. Bureau of Public Roads, p. 16-24, Aug. 1965.
33. Morgan, J. N. A Note on the Interpretation of Multiple Regression Using Dummy Variables. Unpublished notes, Survey Research Center, University of Michigan, April 1964.

Discussion

R. D. WORRALL, Northwestern University—Fleet and Robertson are to be complimented on an interesting and useful paper. Their comments on the statistical hazards associated with the conventional trip-generation study are very well taken; they should be examined carefully by anyone likely to engage in such analysis.

My remarks here are directed to four points raised by the authors, two relating to statistical analysis and two to the more general theme of trip generation and its place within the total transportation planning process.

First, however, a minor quibble. The authors suggest that the study of trip generation and the analysis of travel demand may, in the sense that both deal with the frequency of trip-making, be considered as one and the same thing. The two however are very definitely not the same. They are very different, with different objectives, different output, and substantially different analytic rationale. In the one case the analyst is concerned simply with developing an efficient estimate of zonal trip-ends for subsequent distribution and assignment. In the other he is concerned with the much more complex question of the relationship between travel demand, travel "price," and the urban resident's basic demand for goods and services. There are obvious similarities between the two—one may argue that the first is an inelastic subset of the second—but they are certainly not the same thing and it is somewhat misleading to suggest, even in passing, that they are equivalent. I will return to this point later.

I like the authors' discussion of the problems of spatial aggregation, particularly the sections dealing with variance partitioning and the use of disaggregate models. The former illustrates clearly the seductive nature of aggregation bias, while the latter provides the analyst with a convenient mechanism, not only for circumventing problems of spatial aggregation, but also for extending the scope of a conventional linear regression model to include categorical and nonlinear relationships. Recent empirical work at Northwestern University supports the authors' comment that the disaggregate household model used in their example yields zonal trip-end estimates as accurate as those obtained from a standard aggregate model, with the added advantage of a sharper, more coherent model structure. Obviously the method is equally applicable to the study of nonresidential trip generation, or to any situation in which the analyst wishes to avoid aggregation of a heterogeneous data set. It should be noted, however, that its use does involve a number of statistical problems which the authors do not discuss—most notably those involving the limitations on the use of dummy variables. Several excellent discussions of the topic are available (2, 29).

Disaggregation for its own sake is of little value—it merely increases the variance of the basic data set. Its virtue lies mainly in the flexibility which it affords the analyst to develop alternate aggregative structures, each appropriate to a specific phase of his analysis and each yielding a sharper, more meaningful model format. The criterion for the aggregation in each case is, of course, the clustering of trip-making units which are homogeneous with respect to their trip-making characteristics. In the case of trip generation analysis it has the particular virtue of releasing him from an initial, arbitrary aggregative structure, frequently based upon a fortuitous spatial proximity rather than any logical relationship.

I am somewhat less happy with the authors' discussion of sampling errors. I am also a little unhappy with some of their empirical examples. In Table 1, for example, the variables "family size" and "total persons" are not disaggregate and aggregate equivalents, and the difference between the two values of r is indicative not of a reduction in the "significance of a vital estimating relationship," but simply of a reduction in variance due to the process of aggregation. Similarly, in Table 4, the differential composition of the two equations (the zonal equation contains an extra variable) suggests that specification bias may have influenced the result. The points are minor, but they suggest that a different selection of examples may perhaps make the authors' points more clearly.

Certainly one should avoid "overtuning" a regression equation (presumably the authors are thinking here of overindulgence in variable transformations or interactance terms). Certainly, also, one should strive for the simplest possible model structure. Their suggested cut-off criterion based on the comparison of percent standard errors, however, is somewhat arbitrary. Their point concerning the danger of overtuning could be made much more clearly, and with greater validity, in terms of the partitioning of variance discussed earlier in the paper. The whole question of sampling error and sampling design in travel surveys is extremely complex. It deserves more attention than it has received in the past, particularly from the cost-conscious.

Relatively little research has been directed toward the important question of the temporal stability of travel forecasts. Again there are compelling arguments in favor of disaggregate rather than aggregate modeling. Predictions of future trip-rates based purely on zonal aggregates, for example, will inevitably be in error if a significant change occurs in the composition of the zone over time. Equally, a prediction of household trip-making based upon estimates of residential density and car ownership is likely to yield erroneous forecasts if the relative importance of these parameters, i.e., the values of the β coefficients in the regression equations, changes over time. Such a change is not only possible but almost inevitable. For example, as car ownership levels tend towards a general saturation point it may be expected that differences in car ownership rates are likely to account for less and less of the variance in zonal trip production. Other factors—personal travel preferences, transportation systems improvements, substitution effects, etc.—will and do exert a more significant influence on the trip-making process, consequently the old proxy relationships will no longer hold.

All this is to argue in effect not only for a continuing monitoring of trip generation data, but rather for a continuous, microscopic analysis of travel behavior, set at the level of the individual decision unit (e.g., the household or the firm) and aimed at isolating the fundamental relationships between the decision unit's daily activity pattern on the one hand (i.e., the set of activities in which the members of unit regularly engage and which generate a demand for travel) and the characteristics of its travel demand on the other. Equally, it is to argue for a basic change in the analyst's orientation, away from the description of aggregate travel patterns and toward the specification of disaggregate travel demand relationships, sensitive to system and activity variables. Several interesting starts have been made in this direction (1, 34, 35).

Unfortunately, all this is likely to be rather expensive. The cost of a possible longitudinal monitoring scheme for Chicago, for example, based upon the surveillance of a panel of 2000 households has been estimated at approximately $200,000 per year. Alternate sample designs, based on a combination of randomized cross-sectional samples and partial overlap designs reduce the cost somewhat but not significantly (1).

The question of cost brings me to my final point. Any statement concerning the efficacy of current expenditures for the collection and analysis of travel data should, logically, be set in the context of the objectives of the total transportation planning process. More specifically, it should take account of the importance of these expenditures to the achievement of these objectives. To make such a statement we require at least five pieces of information: a statement of the desired accuracy of the final travel forecast, a statement of the cost (or penalty) of failing to meet this accuracy (e.g., a misinvestment of construction funds), a statement of the sensitivity of the forecast to errors in specific phases of the planning process (e.g., errors incurred in the process of generation, distribution or assignment, etc.), a statement of the sampling distribution for each phase, and a statement of the cost of achieving a particular degree of accuracy in each phase. This, in effect, is to ask for a combined sensitivity/cost-effectiveness analysis of the transportation planning process, in both its initial and continuing phase—a pretty tall order, but probably the most critical single requirement in current urban transportation planning. A start has been made on such a project by CONSAD Research Corporation under a contract sponsored by the Federal Department of Transportation.

Finally let me once again compliment the authors on an interesting and extremely useful paper. Coupled with the recent "Guidelines for Trip Generation Analysis" published by the Federal Department of Transportation, their discussion should be of considerable assistance to any agency involved in the analysis of trip generation data.

References

34. Kraft, G., and Wohl, M. New Directions for Passenger Demand Analysis and Forecasting. Transportation Research 1, 3, p. 205-30, Nov. 1967.

35. Baumol, W., and Quandt, R. Abstract Multi-Mode Travel Model: Theory and Measurement. Technical Report 4, N.E. Corridor Project, U.S. Dept. of Commerce, 1966.

HAROLD D. DEUTSCHMAN, *Tri-State Transportation Commission*—Multiple linear regression equations, utilizing an arbitrary zonal scheme for observational units of independent and dependent variables, have become a common method for trip generation analyses. The authors are to be commended for indicating some of the frequently overlooked pitfalls inherent in the trip generation process. The primary message of the paper concerns the statistics of regression and cautions the analyst on the statistical paradoxes in measuring the accuracy of a regression equation when zonal data are used. It is emphasized that the conventional coefficient of correlation only explains the variation between the zonal units and not between the original data set. If the zones are in fact not chosen specifically as homogeneous units, then the actual variation in the trip-making characteristics may be as little as 20 percent as explained by the aggregated zones, even when the variance in the regression analyses indicates a figure of 95 percent. The authors point out that "aggregation will tend to cloud much of the variation and also many of the relationships that may explain this variation." They address themselves to the problem of how fine the regression equations should be tuned to—what are the best results that may be obtained, considering the relative homogeneity of the zones? By analyzing the standard error of the mean for each zone, an excellent measure of reliability may be inferred from the data set being utilized by regression. The illustration that the percent standard error of the estimating (regression) equation may not be smaller than the average standard error of the mean of the dependent variable (by zone) provides a significant measure of the limits of efficiency for the regression equation. There is much merit in this portion of the paper, in relating the accuracy inherent in the trip generation equations with that of the data set.

The authors then turn to alternate approaches of trip generation, including dwelling unit analysis and cross-classifications. These approaches are broached, but an in-depth analysis of the applicability and effectiveness of the processes is not undertaken. It is indicated that a dwelling unit analysis could use the dwelling unit as the basic observational unit instead of zones, thus eliminating the aggregation effect. In lieu of average zonal figures such as median income and average auto ownership, dummy variables may be employed to describe the socioeconomic and density characteristics. While data aggregation is avoided, basic limitations arise such as (a) there is a significant increase in the number of independent variables, using the distribution of variables instead of means; (b) some variables are only amenable to a zonal description, such as gross residential density, and could not be effectively utilized in a dwelling unit analysis; (c) the variables used in the basic regression equation must be projected to a future year. Since the number of variables is greater, this task is more formidable. In addition, since the number of zones effectively equals the number of households, a complex methodology must evolve to estimate a number of characteristics for each household for the forecast year.

Much research is needed in the trip generation phase of travel analyses. I take a pragmatic view of the analytical techniques that may be used. The one to be chosen should be most efficient from the standpoint of not only reproducing the survey data, but must also include the efficiency of (a) estimating the independent variables and (b) relative stability of the estimating equation over time (forecasting capability).

It was pointed out that the regression technique of trip generation using aggregated data has basic limitations; it must also be noted that the alternate methods posed have their limitations as well. Since the number of variables, equations, and techniques used in trip generation analyses are many in number, it is strongly suggested that once an enriched data set becomes available over two periods in time (detailing the full spectrum of socioeconomic data with the travel data), all of the suggested methods of trip generation be analyzed on a common basis. It is only through this analysis that the various procedures may be systematically evaluated.

CHRISTOPHER R. FLEET and SYDNEY R. ROBERTSON, Closure—In preparing this paper, the authors chose to narrow the perspective somewhat by dealing primarily with the existing state of the art. The discussants have ably contributed the appropriate and, perhaps, essential extension of the authors' comments into a broader context.

Worrall's concern over our equation of travel demand and trip-making frequency is understandable. Though it would perhaps be easier to pass over this point as an argument in semantics, we are sympathetic. Too few of the ideas concerning the demand of persons for transportation services have penetrated the methodology of transportation planning. Particularly lacking is the interaction between travel demand and the transportation system variables. At this point in time, only the most subtle interaction between trip generation and the transportation system is present.

Worrall suggests that the argument for disaggregate analysis must be carried further, to allow a "microscopic analysis of travel behavior," relating daily activity to travel demand. Though this is certainly theoretically attractive, we do not think that the required tools are yet available. Also, as he notes, the costs are prohibitive. And, even if it were to be possible to develop this type of analysis, is the transportation planning process capable of using the resulting information effectively?

This brings us to Worrall's final point, with which the authors completely agree. It is difficult to seek the efficient use of the travel survey data when we have no real notion of how effectively the data are used in the transportation planning process. At present, we have little or no knowledge concerning the balance of the technical phases in terms of accuracy. And, even if we could identify where the imbalances occur, no rational basis exists for allocating resources to remove them.

Regarding our discussion of sampling errors, it appears that Worrall was expecting more than the authors intended. Certainly a comprehensive sampling error analysis was not done. The intent was to illustrate the kinds of variation associated with the zonal regression input data due to sampling. Our interest was in the regression dependent variable—trip-making. It seemed reasonable to conclude that the degree of precision in estimating the dependent variables using regression should be no greater than the precision associated with that dependent variable due to sampling. Carefully fitting an optimum equation to data where the numerical values are known to vary over a wide range does not seem appropriate. It was this that was considered to be "overtuning."

Deutschman suggests that disaggregate trip generation analysis is a two-way street and that there are basic limitations, some of which he enumerates. Perhaps we did not devote sufficient time to this. However, the paper was intended to be suggestive in this area, rather than definitive. Of the limitations noted, those concerning forecasting bother us most. Though we disagree that "the number of zones effectively equals the number of households," the identification of stable (particularly over time) household types is difficult. The estimation of the future number of households of each type also presents a significant problem. As noted in the paper, however, many of the disaggregate household location models can supply this information.

The authors completely agree that the next step must be the evaluation, with time series data, of the notion that disaggregate trip generation models are more likely to remain stable over forecast intervals. We also can find no argument with Deutschman's pragmatic criteria for an efficient trip generation procedure.