Adequacy of Clustered Home Interview Sampling For Calibrating a Gravity Model Trip Distribution Formula

KEVIN E. HEANUE, LAMELLE B. HAMNER, and ROSE M. HALL
Respectively, Highway Engineer and Statisticians, Urban Planning Division, U. S. Bureau of Public Roads

•INCREASED USE of the gravity model formulation in transportation planning has brought about a wide variation in study procedures, many of which, developed without the benefit of research, are characterized by a reduction in travel data accumulated for model calibration. In several instances, urban planning engineers have advocated the use of a reduced sample home interview survey, clustered in selected zones, to supply the basic data for model calibration (1, 2, 3).

A clustered sample may best be described by contrasting it with the typical systematic sample normally used in home interview origin-destination surveys. In a systematic sample every nth dwelling unit in each zone in the entire study area is surveyed. In clustered home interview sampling, a systematic sample is conducted, but only in selected zones. In other words, every nth dwelling unit is interviewed in the selected zones, but none are interviewed in the remaining zones. Each cluster, therefore, consists of a systematic sample in a selected zone.

The selected zones are generally chosen to reflect a range in the urban characteristics known to be correlated with travel habits. For example, residential density, car ownership, income level, and distance from the central business district (CBD) are among the variables which might be utilized as criteria for determining zones to select for the clustered sample.

The objective of the clustered sampling technique is to reduce study time and costs by reducing the total number of home interviews required. This objective is successfully achieved only if the clustered sample data are sufficient to develop a travel model which can then be used to estimate trip distribution patterns for the entire study area.

This paper reports the results of a research study designed to evaluate the use of clustered home interview samples as the basic source of data for developing a gravity model that will accurately synthesize areawide travel patterns. The study was initiated by the Urban Planning Division of the U. S. Bureau of Public Roads with the financial assistance of the Pennsylvania Department of Highways and in cooperation with the Pittsburgh Area Transportation Study (PATS).

THE GRAVITY MODEL

The gravity model theory may be simply described as follows. The trips produced in any zone will distribute themselves to other zones in the study area in direct proportion to the trip opportunities or attractions in the other zones and in inverse proportion to some function of the spatial separation between the zones.

The gravity model equation used in this research is stated as follows:

$$T_{(i-j)} = \frac{P_i A_j F_{(t_{i-j})} K_{(i-j)}}{\sum_{x=1}^{n} A_x F_{(t_{i-x})} K_{(i-x)}} \tag{1}$$

Paper sponsored by Committee on Origin and Destination.

where

$T(i-j)$ = trips produced in zone i and attracted to zone j;
P_i = trips produced in zone i;
A_j = trips attracted to zone j;
(t_{i-j}) = travel time in minutes between zone i and zone j;
$F(t_{i-j})$ = empirically derived travel time factor expressing average areawide effect of spatial separation on trip interchange between zones that are (t_{i-j}) apart; and
$K_{(i-j)}$ = specific zone-to-zone adjustment factor to allow for incorporation of effect on travel patterns of defined social or economic linkages not otherwise accounted for in the gravity model formulation.

The travel time factor $F(t_{i-j})$ is approximately equivalent to the more traditional $1/t^n$. The use of a set of travel time factors to express the effect of spatial separation on zonal trip interchange, rather than the traditional inverse exponential function of time, simplifies the computational requirements of the model. It also allows more complex mathematical functions (for example, with n varying by t) to be conveniently represented.

Trip production (P_i) and trip attraction (A_j) take on specific definitions when considered in the gravity model formulation. First, consider home-based trips. Home-based trip productions per zone are all those trips made by residents of the zone originating at or destined to their homes. Home-based trip attractions per zone are the nonhome trip ends of home-based trips arriving at or departing from the nonhome zone. For example, a person living in zone 1 who travels from his home to zone 2 and home again makes two trips. Zone 1 is credited with having produced two home-based trips; zone 2 is credited with having attracted two home-based trips.

Nonhome-based trips are those trips having neither end at the residence of the trip maker. Nonhome-based trip productions are the origins of nonhome-based trips and, accordingly, nonhome-based attractions are the destinations.

Data Sources

Basic trip data sources for this research were the 1958 standard origin-destination (O-D) home interview survey conducted in the 226 internal zones of the Pittsburgh study area and a special high sample rate clustered home interview survey conducted in 1960 in 13 selected zones of the same study area. Both of these home interview surveys were conducted by the staff of PATS.

The 1958 survey consisted of a systematic 4 percent sample of all the dwelling units in the study area, providing over 16,000 completed interviews (4, 5, 6). The 1960 clustered survey was designed with a sample rate varying by zone from 10 to 33 percent, yielding a total of 4,250 interviews. The variable sample rate was designed to provide trip data of approximately the same degree of accuracy from each of the selected zones. Previous research by Sosslau and Brokke on the statistical accuracy of home interview O-D data was used as the basis for the selection of the sample rates (7).

The 13 zones chosen for the 1960 survey were selected from among the 226 internal zones in the study area as a result of an examination of certain land-use and socio-economic characteristics tabulated from the 1958 survey data. Principal factors examined were net residential density, car ownership, and distance from the CBD, which in Pittsburgh is the Golden Triangle. The 13 zones were selected to yield a range in these characteristics and hopefully to reflect the full range of income levels. Income data were not available from the 1958 survey; however, some preliminary studies had shown income level to be highly correlated with trip production and a question on this subject was included in the 1960 survey.

Figure 1 illustrates the geographical location of the 13 selected zones and the area encompassed by the 1958 study. North, south, and east zones noted later are referenced to the three rivers.

The basic data sources for land-use and socio-economic data were the land-use survey conducted during the PATS study which summarized land area measurements

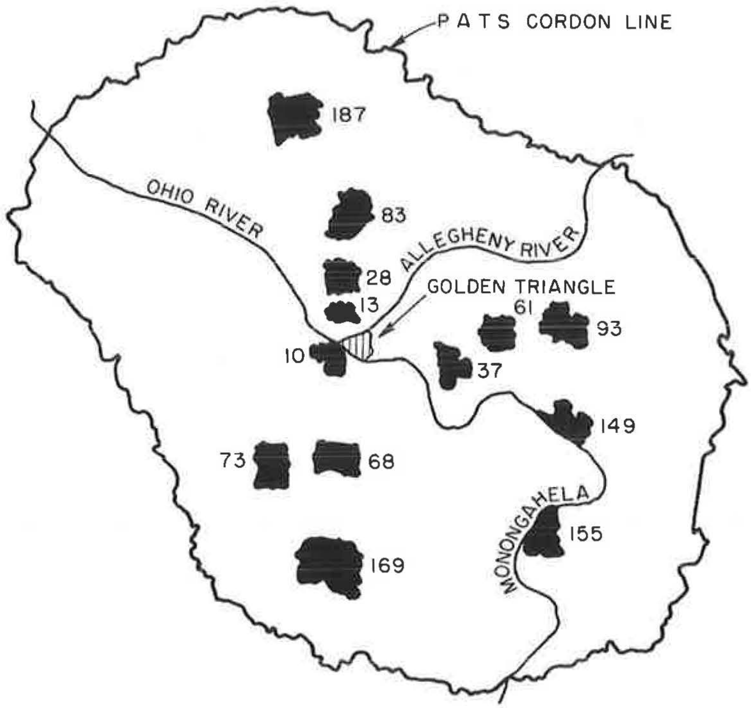


Figure 1. Location of 13 selected analysis zones used in Pittsburgh research project.

by two digit classification and the 1958 home interview survey which obtained population and car ownership statistics by zone.

Objectives

The study was designed to answer four basic questions.

1. Were the trip data from the 13-zone clustered survey sufficient to develop a set of travel time factors, $F(t_{i-j})$, representative of the effect of zonal separation on trip making for the entire study area?
2. Did the limited survey contain enough trip information so that trip productions (P_i) and trip attractions (A_j) could be related to specific land-use and socio-economic data through the use of regression analyses and thereby expanded to every zone in the study area?
3. How accurate, when compared to the 1958 O-D survey, was a gravity model developed solely through the use of the 13-zone survey trip data and all available land-use and socio-economic data when applied to the total study area?
4. How accurate, when compared to the 1958 O-D survey, was a gravity model developed by making use of the full 1958 O-D survey and the same land-use and socio-economic data?

Procedures

To provide answers to these questions the work of the study was organized into three major phases: (a) gravity model calibration, (b) trip production and attraction estimates, and (c) combined analysis—trip production, attraction and distribution. Each of these major phases was in turn divided into two subphases, one relating to the 13-zone clustered survey and the other to the total study area survey.

Gravity Model Calibration.—Travel time factors were developed using only the 13-zone survey trip data. These were representative of the effect of spatial separation on trip making as expressed by the clustered survey data. Travel time factors were then developed using the 1958 total study area trip data and were representative of the effect of spatial separation on trip making for the entire study area. Deviations between the clustered survey data factors and the total area data factors were analyzed for significance.

Trip Production and Attraction Estimates.—Regression analysis was used to develop trip production estimating equations from the 13-zone survey trip productions and the available land-use and socio-economic data. These equations were then solved for every zone in the study area to develop trip production estimates for each of the 226 internal zones. Regression analysis was also used in an attempt to develop trip attraction estimating equations from the 13-zone data.

Trip productions and trip attractions from the 1958 total study area survey were related to the available land-use and socio-economic data through the use of regression analysis.

Combined Analysis.—The trip production and attraction estimates, the calibrated travel time factor estimates developed using only 13-zone survey trip data, and the available land-use and socio-economic data were used as input to a total study area gravity model and a trip distribution was calculated. The accuracy of this distribution when compared to the 1958 O-D survey trip data was then determined.

Trip production and trip attraction estimates and the travel time factors developed using the 1958 O-D total study area trip data and the available land-use and socio-economic data were used to calculate another gravity model trip distribution. The accuracy of this distribution when compared to the 1958 O-D survey trip data was also determined.

A detailed analysis was then made of the accuracy of the alternate gravity model distributions. Each of the gravity model trip distributions was compared statistically to the 1958 O-D trip distribution. Selected movements were isolated and compared to measure geographical bias.

GRAVITY MODEL CALIBRATION

The basic calibration procedure and descriptions of the IBM 1401 and 7090 computer programs used for both the 13-zone and total study area phases of this study have been fully documented (8).

Data Processing

The initial data processing work was similar for both the 13-zone and the total study area survey analyses. The data necessary for use in the calibration process consisted of the following: (a) trip productions (P_i) and trip attractions (A_j) by purpose and zone; (b) minimum path travel times (t_{i-j}) between all zones; and (c) trip length frequency distributions by purpose for the O-D trip data.

Tables of expanded zone-to-zone person trip movements were first developed by purpose from the home interview survey data. A byproduct of the trip table building program was a zonal summary of trip production and trip attraction.

The trip tables and trip production and attraction summaries were tabulated for six trip purpose categories. The five home-based trip purposes were work, shop, school, social-recreational, and miscellaneous. The home-based trip purpose definitions were taken directly from the standard home interview questionnaire classifications, with the exception of miscellaneous which is a grouping of personal business, medical-dental, and eat meal trips. The sixth purpose category, nonhome-based trips, included all those trips where neither the purpose to nor purpose from was home.

Change mode and serve passenger trip purpose classifications had previously been eliminated when these trips were linked. In trip linking, certain segmented trips in the previous classifications are connected or linked to assign more meaningful origins, destinations, and purposes.

The trip tables were then used with the minimum path travel times between all zones to develop trip length frequency distributions by purpose. The minimum path travel time between any two zones is made up of the minimum path driving time between the zones and the terminal times in the zones of production and attraction. The minimum path driving times, more commonly called trees, were calculated by the traffic assignment process. To develop the driving times for use in this study, the PATS base network of coded link distances and times was recoded to fit the BPR tree building program format. The driving times on the PATS coded network were developed from a limited number of operating speed studies.

Terminal times were estimated by the research staff to allow for the effect of differences in cruising, parking, and walking times as caused by differences in congestion and availability of parking facilities. Terminal times, ranging from 1 min in suburban residential areas to 6 min in downtown areas, were added to the driving times to develop a more realistic measure of the actual spatial separation between zones. The same travel times were used to develop the trip length frequency distributions and as input to the gravity models.

For the gravity model travel time factor calibration runs, both trip productions and trip attractions were taken directly from the home interview survey results. The actual P_i and A_j from the survey were used rather than estimates to prevent error from this source from affecting the travel time factor calibrations. This is standard practice when calibrating gravity models.

13-Zone Gravity Model Calibration

Due to the limitations of the clustered survey data, an abbreviated trip distribution was utilized in the 13-zone gravity model calibration. For both the clustered survey data and the gravity model, home-based trip productions occurred in only 13 of the 226 internal zones, whereas trip attractions occurred in all 226 internal zones, 46 external analysis zones, and at the 8 external stations. Nonhome-based trip productions and attractions occurred in all 280 analysis units. For each of the trip purposes, the total of the trip productions equaled the total of the trip attractions.

For a first approximation of travel time factors, a set of values previously developed for Washington, D. C., was used (9, 10). Using these travel time factors and the clustered survey trip productions and attractions, the gravity model program was used to calculate a first trip distribution.

The trip length frequencies of the resultant distribution were plotted by purpose along with the O-D trip length frequencies. The trip length frequency distributions were not in agreement and the initial travel time factors were therefore modified by using the relationship between the percent O-D trips per 1 min of time interval and the percent gravity model trips per time interval (8). The gravity model program was then rerun using the adjusted travel time factors, and the plotting and adjustment procedures were repeated. This iterative process was continued until a satisfactory agreement was reached between the O-D and gravity model trip length frequency distributions for each purpose. In addition to the basic agreement in the closeness of fit of the curves, the total person hours of travel and mean trip length were continually checked by purpose. When all three parameters were in close agreement, the model was said to be calibrated. Figure 2 shows the plot of the trip length frequency distributions for all trip purposes of the 13-zone O-D vs the 13-zone calibrated gravity model along with the relationship between the total person hours and mean trip lengths. The set of travel time factors that resulted in an acceptable trip length distribution comparison was, therefore, that set which best approximated the effect of spatial separation on trip interchange as exhibited by the 13-zone survey data.

Trips from each of the 13 zones to the Golden Triangle were isolated for both the O-D selected zone survey trips and the 13-zone calibrated gravity model distribution to determine if there was any geographical bias inherent in the gravity model distribution. Table 1 presents these data. Several attempts were made to relate the gravity model's deviation from the O-D to selected socio-economic data. Figure 3 shows the results obtained when the ratios of gravity model to O-D Golden Triangle-oriented movements were plotted vs zonal income. Although the plot shows a slight correlation between the ratios of the movements and zonal incomes, the correlation ($r = -0.71$) was not considered significant enough to form the basis for the development of adjustment factors.

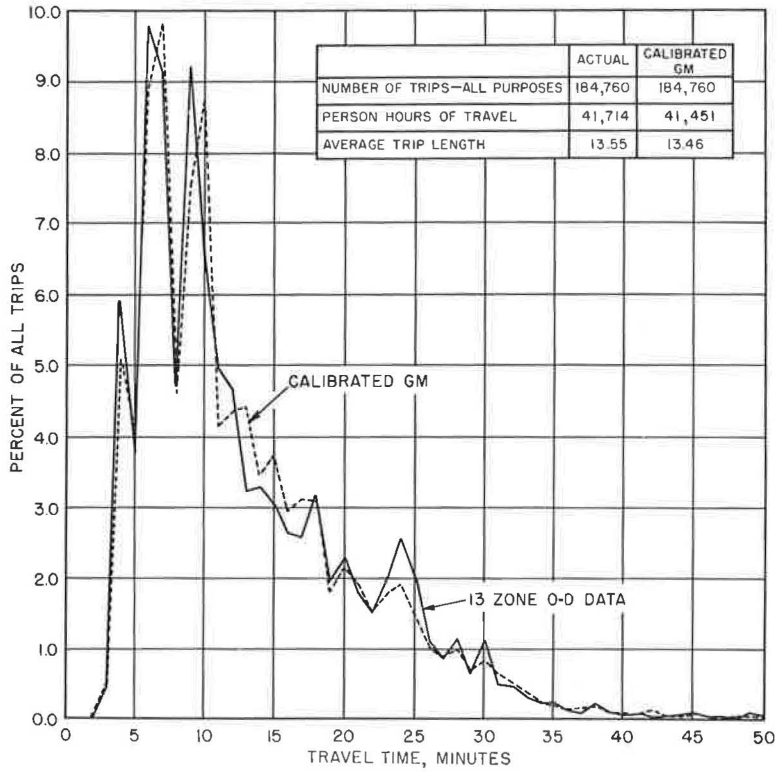


Figure 2. Trip length frequency for all trips, 13-zone data.

TABLE 1

SELECTED MOVEMENTS TO GOLDEN TRIANGLE; 13-ZONE GRAVITY MODEL DATA VS CLUSTERED SURVEY DATA

Zone of Origin	Home-Based Trips (No.) Clustered Survey	Home-Based Trips (No.) Model	Difference No.	Difference %
10	2,409	3,375	966	40.1
13	2,450	3,595	1,145	46.7
28	1,324	2,136	812	61.3
37	2,086	1,983	-103	-4.9
61	2,471	1,319	-1,152	-46.6
68	2,161	1,364	-797	-36.9
73	2,353	932	-1,421	-60.4
83	593	418	-175	-29.5
93	1,092	711	-381	-34.9
149	764	642	-122	-16.0
155	282	453	171	60.6
169	835	433	-402	-48.1
187	320	261	-59	-18.4
Total	19,140	17,622	-1,518	-7.9

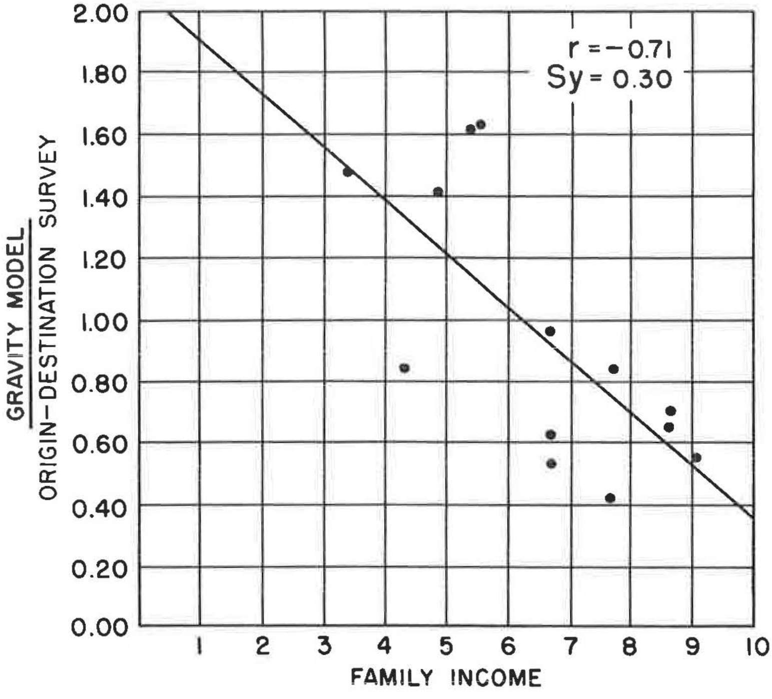


Figure 3. Differences in observed and estimated home-based trips from 13 selected zones to Golden Triangle related to mean family income per zone.

TABLE 2

SELECTED MOVEMENTS, CALIBRATED 13-ZONE GRAVITY MODEL DATA VS CLUSTERED SURVEY DATA

From	To	Home-Based Trips (No.)		Difference	
		Clustered Survey	Model	No.	%
South zones	North zones	1,962	1,981	19	1.0
North zones	South zones	2,650	3,135	485	18.3
South zones	East zones	13,114	11,217	-1,897	-14.5
East zones	South zones	6,017	6,298	281	4.7
East zones	North zones	2,138	2,476	338	15.8
North zones	East zones	8,907	11,269	2,362	26.5
North zones	North zones	21,281	18,434	-2,847	-13.4
South zones	South zones	33,169	35,047	1,878	5.7
East zones	East zones	66,239	65,620	-619	-0.9
North zones	Golden Triangle	4,687	6,410	1,723	36.8
South zones	Golden Triangle	7,758	6,104	-1,654	-21.3
East zones	Golden Triangle	6,695	5,108	-1,587	-23.7
Internal zones	External zones	4,501	4,337	-164	-3.6

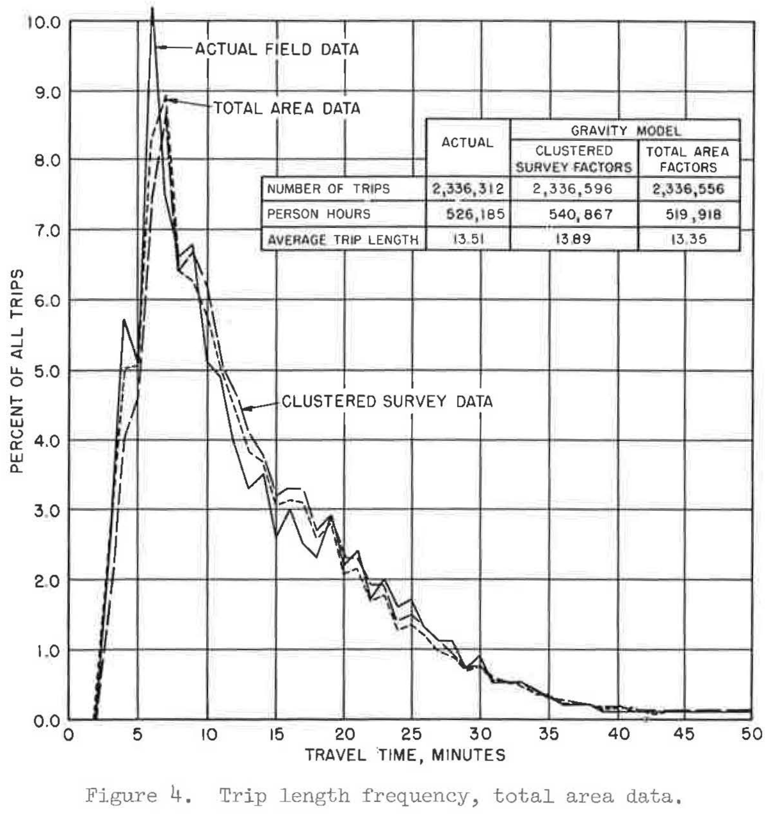
If geographic bias, which showed a more direct numeric relationship to specific land-use or socio-economic characteristics, had been found, it would have been eliminated through the use of adjustment factors. These factors are represented by $K_{(i-j)}$ in the gravity model formulation (Eq. 1).

A comparison of selected movements was made to determine the accuracy of the model and to see if a time barrier was needed to bring river crossings into agreement with O-D data. Table 2 gives these data. The gravity model river crossing trips were not systematically high or low when compared with the O-D data and no time barrier was used. Several previous studies in other cities have shown the need for such a time barrier over rivers (10, 11).

Total Study Area Model Calibration

The total study area trip distribution for both the O-D and the gravity model had trip productions in all of the 226 internal zones and trip attractions in all of the 280 analysis units. The travel time factors developed using the 13-zone data were applied to total study area data. To eliminate error from other sources, the trip productions and attractions for all zones were taken directly from the 1958 O-D reportings. A new trip distribution was calculated and the trip length characteristics of this distribution were then compared with the 1958 O-D trip length characteristics. Figure 4 depicts the trip length distribution comparison for all trip purposes combined. The trip length curves were similar but the analysis indicated a definite need for further travel time factor calibration. This meant that the travel time factors developed from the clustered survey data were not sufficiently accurate when applied to the total study area.

A statistical comparison of district-to-district movements, given in Column (d) of Table 3, indicated significant differences between the 1958 O-D data and the clustered



	ACTUAL	GRAVITY MODEL CLUSTERED SURVEY FACTORS	GRAVITY MODEL TOTAL AREA FACTORS
NUMBER OF TRIPS	2,336,312	2,336,596	2,336,556
PERSON HOURS	526,185	540,867	519,918
AVERAGE TRIP LENGTH	13.51	13.89	13.35

Figure 4. Trip length frequency, total area data.

TABLE 3

DISTRICT-TO-DISTRICT MOVEMENTS, PITTSBURGH, PA.

1958 Survey Data All Purposes			Percent RMS Error[a]				
			Trip Ends Based on				
			1958 Survey Data	1958 Survey Data	13-Zone Regression Est.	226-Zone Regression Est.	226-Zone Adj. Regression Est.
			Travel Time Factors Based on				
Volume Group	Mean of Volume Group	Frequency	13-Zone Study	226-Zone Study	13-Zone Study	226-Zone Study	226-Zone Study
(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
0- 499	141	2,121	107.09	99.29	118.44	134.04	104.96
500- 999	698	261	53.72	50.72	65.90	62.46	53.72
1,000- 1,999	1,476	143	52.10	48.78	54.47	59.89	49.19
2,000- 2,999	2,451	81	42.80	42.15	40.43	47.00	41.33
3,000- 3,999	3,437	37	34.27	34.45	34.97	39.80	33.28
4,000- 4,999	4,373	21	38.74	40.04	47.82	39.42	42.35
5,000- 5,999	5,539	17	35.28	33.16	37.39	49.86	36.18
6,000- 7,999	6,797	20	28.83	29.56	29.66	50.62	30.74
8,000- 9,999	8,843	11	27.98	26.19	29.05	39.30	27.62
10,000-14,999	11,882	18	23.74	23.62	27.18	39.51	24.66
≥15,000-	31,159	22	21.07	16.08	21.83	26.08	17.89

[a]Percent RMS error = $100 \left(\frac{\sqrt{\frac{\Sigma (\text{diff.})^2}{n}}}{\bar{x}} \right)$

where

diff. = difference between surveyed and estimated movements;
n = number of district-to-district movements; and
$\bar{x}$ = mean of survey volume.

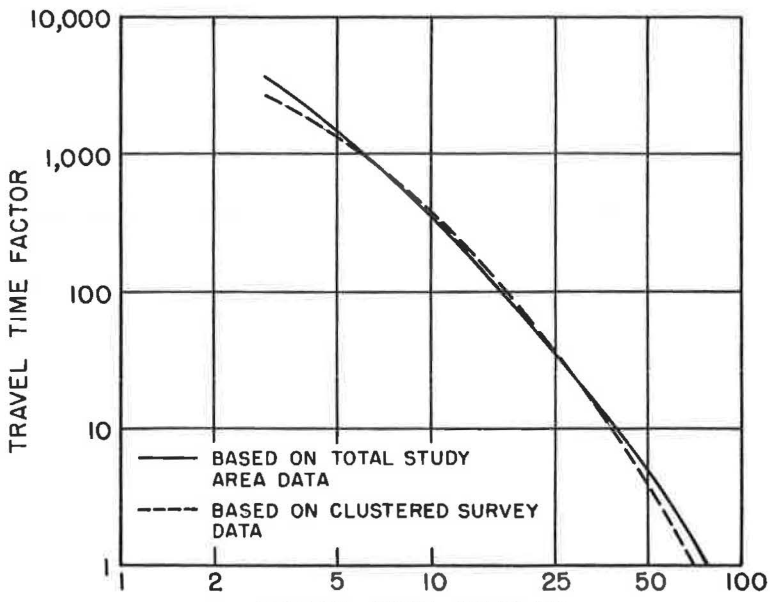


Figure 5. Home-based work travel time factors.

survey model results. The travel time factors were, therefore, adjusted using the trip data from the comprehensive 1958 O-D survey. The model calibration process was repeated until an acceptable relationship existed between the O-D and the gravity model trip length curves, mean trip lengths, and person hours of travel.

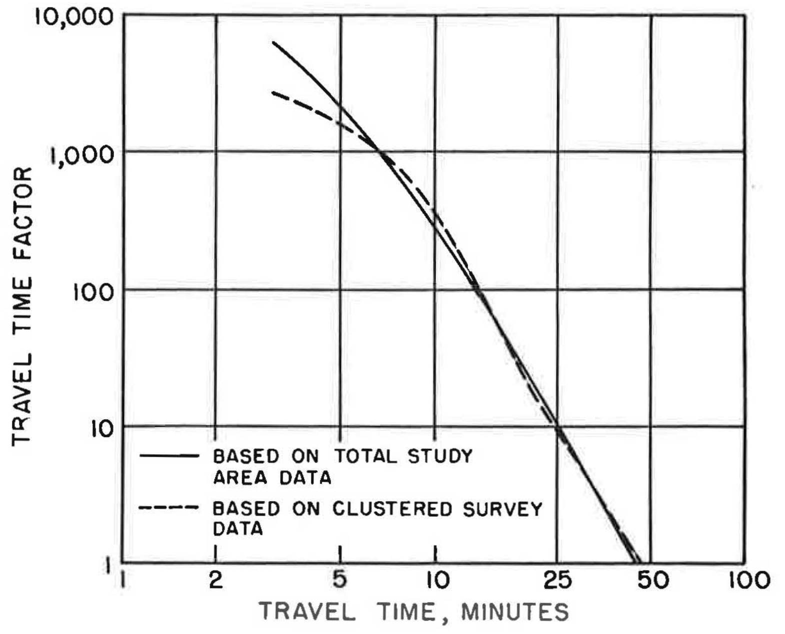


Figure 6. Home-based other travel time factors.

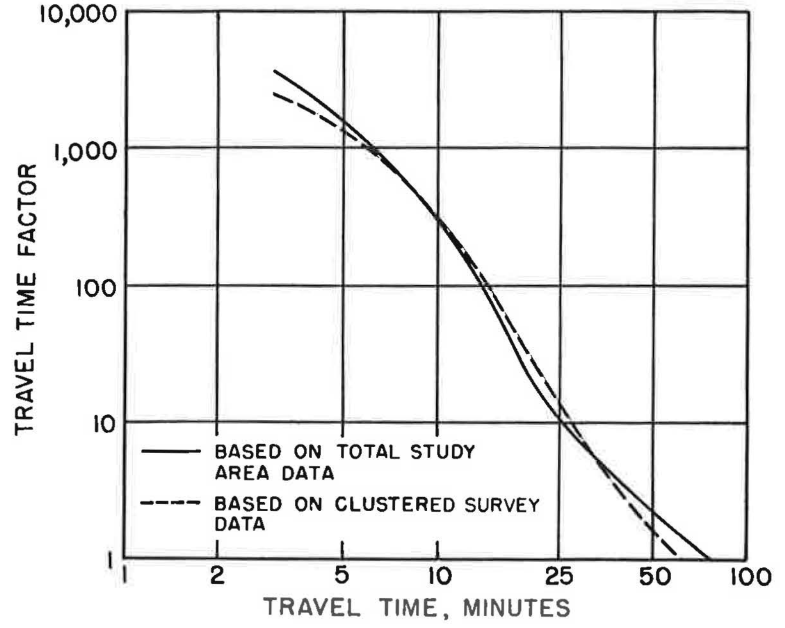


Figure 7. Home-based social and recreation travel time factors.

The travel time factors developed using 13-zone data and the revised set developed using total study area data are reflected in the curves in Figures 5 through 10. The significance of the differences between the two curves in each figure will be discussed later in this paper.

Figure 4 also illustrates the final agreement in trip length characteristics between the calibrated total area gravity model and the 1958 origin-destination survey data.

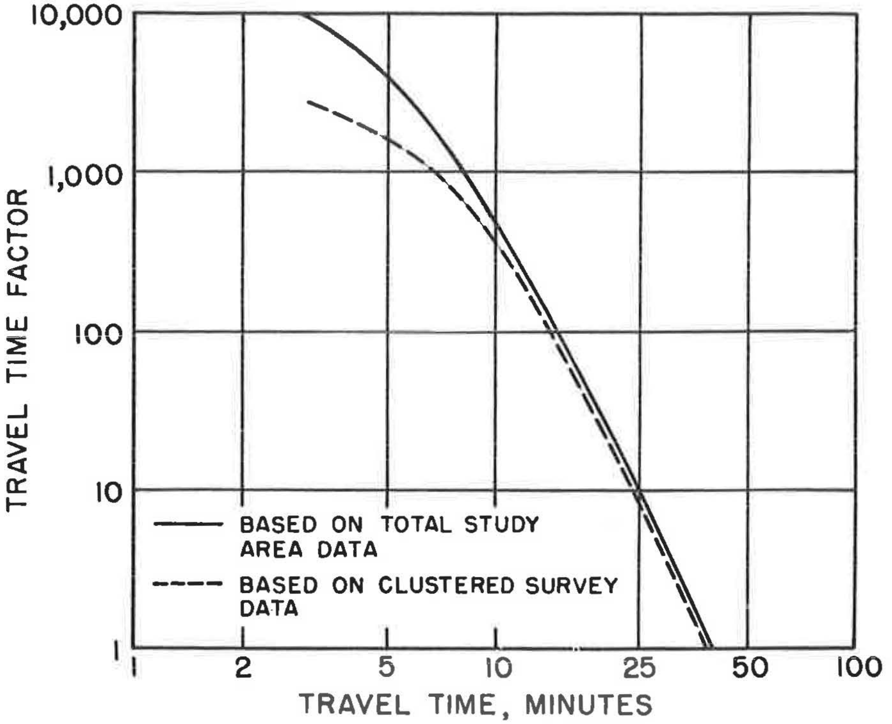


Figure 8. Home-based shop travel time factors.

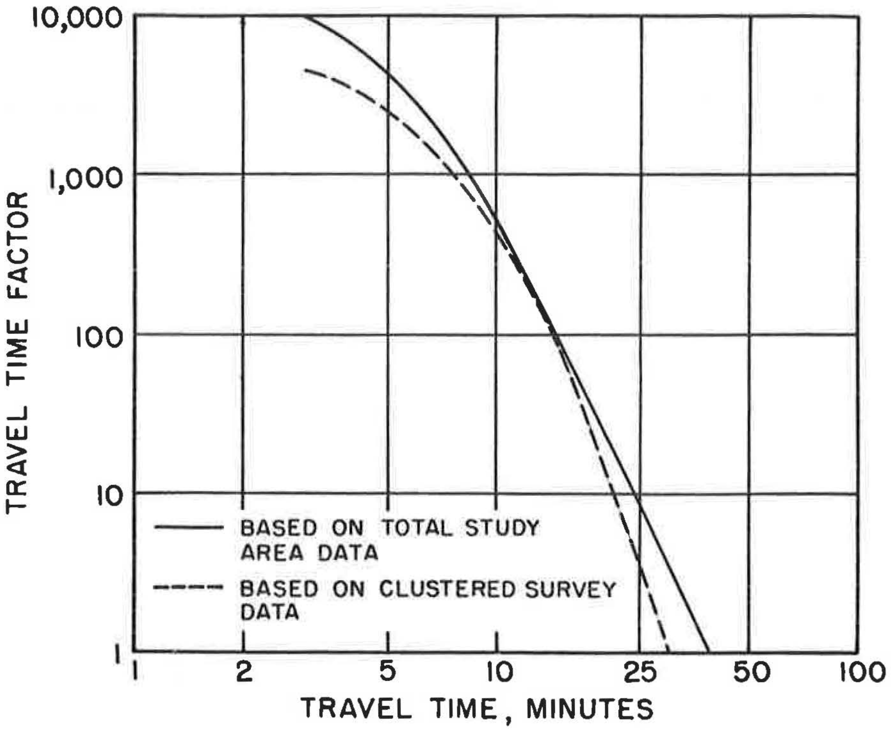


Figure 9. Home-based school travel time factors.

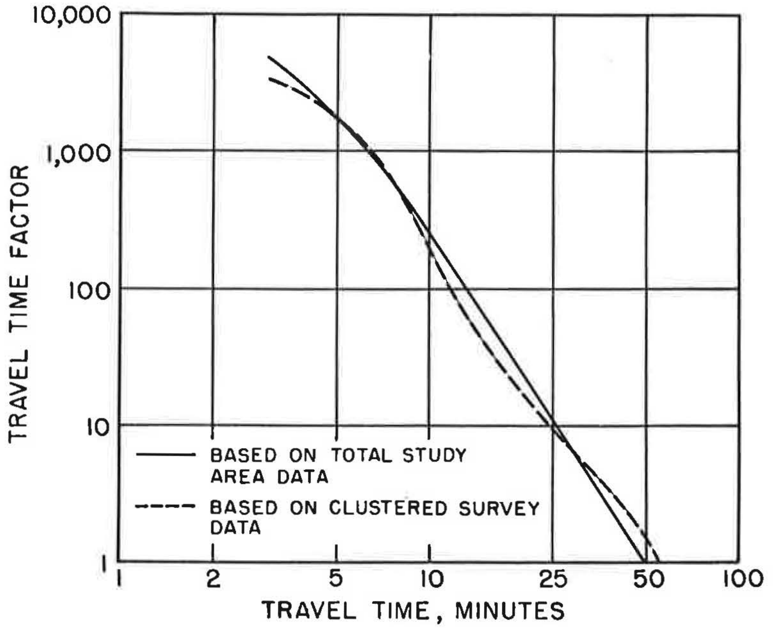


Figure 10. Nonhome-based travel time factors.

TABLE 4

SELECTED MOVEMENTS, CALIBRATED GRAVITY MODEL DATA VS 1958 HOME INTERVIEW DATA[a]

From	To	Total Trips (No.)		Difference	
		1958 Survey	Model	No.	%
South zones	North zones	44,651	53,962	9,311	20.9
North zones	South zones	35,558	46,344	10,786	30.3
South zones	East zones	228,330	230,975	2,645	1.2
East zones	South zones	77,409	99,172	21,763	28.1
East zones	North zones	46,389	61,513	15,124	32.6
North zones	East zones	113,762	133,869	20,107	17.7
North zones	North zones	310,937	280,272	-30,665	-9.9
South zones	South zones	631,193	619,037	-12,156	-1.9
East zones	East zones	848,183	811,452	-36,731	-4.3
North zones	Golden Triangle	52,755	57,754	4,999	9.5
South zones	Golden Triangle	107,660	94,999	-12,661	-11.8
East zones	Golden Triangle	95,291	91,476	-3,815	-4.0
Golden Triangle	North zones	4,919	4,709	-210	-4.3
Golden Triangle	South zones	5,887	4,290	-1,597	-27.1
Golden Triangle	East zones	13,100	14,853	1,753	13.4
External zones	Internal zones	10,254	7,002	-3,252	-31.7
Internal zones	External zones	90,451	92,288	1,837	2.0
All zones	All zones	2,336,412	2,336,596	184	0.0

[a]Based on total study area data.

A statistical comparison of the differences in the district-to-district movements between the 1958 O-D and the calibrated gravity model is given in Column (e) of Table 3.

An analysis of selected movements for the O-D and the calibrated gravity model is given in Table 4. These values do not represent the maximum accuracy the total area model could attain. No attempt was made at this stage to perform a full river crossing or $K_{(i-j)}$ factor analysis. The calibration process was stopped because of the previously noted differences in the clustered survey and total area travel time factors.

TRIP PRODUCTION AND TRIP ATTRACTION ESTIMATES

Because trip productions and trip attractions are two of the key gravity model inputs, considerable time was spent in preparing these estimates and in subjecting them to statistical analyses. Indicators of zonal trip production such as residential density, car ownership, family size, and distance from the CBD were used for estimating trip production.

The trip attraction portion of the research study, however, was hampered by the fact that no data were available on zonal employment, school enrollment, or retail sales from either the 1958 or 1960 studies. These factors provide a basis for specifying nonresidential land-use intensity. They allow for a differentiation between, for example, a downtown office building, shopping area, or school and their suburban counterparts. Land area measurements alone do not reflect intensity of use. For example, a downtown office building on a 10,000-sq ft site may be 10 stories high and fill the entire site, whereas a two-story suburban office building may be located on a site of the same size and fill only one-half the site. Land area measurements would rate these sites equally.

One method by which trip attractions may be related to land use without a direct measure of intensity is through the use of a trip attraction rate analysis. For example, ratios of work trips attracted per acre of industrial land would be developed by individual zone or group of zones. These ratios are, in effect, measures of intensity. By making assumptions with respect to the stability of the ratios over time, these and similar ratios for other land-use activities could then be used to determine future trip attraction. The attraction estimate, therefore, would be based on a future land-use plan which would specify the projected areas of land-use types.

The PATS trip generation analysis did not require detailed measures of land-use intensity as it consisted of a study of land-use trip generation rates as the basis for trip end estimates. A study of trip generation rates, as opposed to a regression study, does not lend itself to a statistical analysis of its accuracy. It also requires an extensive knowledge of the study area. For example, large or specialized employment centers may be isolated for detailed analysis.

For these reasons, regression analysis was selected as the tool best-suited to the development of trip estimates for this research. In some of the trip attraction analyses, however, it was necessary to use modified ratio procedures because meaningful correlations could not be developed by regression analysis on the available data. Certain of these ratio trip attraction estimates, which exhibited high error when compared with O-D survey data, were arbitrarily adjusted to within preset limits of accuracy to make the final trip distribution analysis more meaningful. These preset limits were designed to yield trip attraction estimates with a level of accuracy that could be obtained had the land-use data been more complete and the research staff's knowledge of the study area more extensive. An indication of the type of accuracy to be expected was gained from analysis of data from other cities.

An alternative to either regression or ratio analysis is a procedure sometimes referred to as the synthetic procedure. This is particularly applicable when the trip attraction data are weak but the land-use data are reliable and extensive. In these instances, certain land-use or socio-economic indicators may be used directly as trip attractions. For example, total employment can be used to indicate work trip attraction, school enrollment for school trip attraction, and retail sales for shopping trip attraction. However, procedures such as these would require a detailed zonal analysis to account for such items as walk-to-work trips, school bus policy, and type of shopping center. Although the accuracy of this procedure is difficult to evaluate, it is more reliable than a regression analysis when data on trip attractions are insufficient.

The adjustment of the unsatisfactory attraction estimates was, in part, an attempt to match the type of accuracy that could be obtained by substituting selected land-use statistics (had they been available) directly for estimated trip attractions.

This portion of the research study made extensive use of the Bureau of Public Roads 1401 Regression Analysis Programs, consisting of a basic regression program, an equation solver, and a root-mean-square (RMS) error program. The latter program was used to calculate the accuracy of the 13-zone trip end estimates when applied to the total study area and of total study area estimates that were adjusted to eliminate negative trip production or trip attraction values.

Total Area Estimates Based on Study of 13 Zones

Trip Production.—Due to the high sampling rates, stable values of actual trip production by home-based purposes were available from the 1960 selected zone survey for each of the 13 zones. Regression analysis was used to relate these trip production values to land-use and socio-economic data for the zones. Six basic land-use and socio-economic variables, available for each of the 13 zones as well as for all remaining zones, were selected in a preliminary correlation analysis for use in this phase of the study. These were population, automobiles owned, residential acreage, nonwhite population, airline distance in miles from the CBD, and dwelling units. The number of variables was increased to 10 by using combinations and transformations of the basic variables.

Five equations were developed for each purpose by using varying combinations of independent variables. Two of these equations were then solved for each trip purpose using land-use and socio-economic data for all 226 internal zones. One set of equations had two independent variables and the other five. The five-variable equations had the lower standard errors of estimate based on the 13-zone dependent variable trip data but they showed higher RMS errors when applied to all 226 zones. The following two-variable equations were therefore selected for use in estimating trip production for each of the 226 zones:

$$\begin{aligned} Y_1 &= 2.6817 - 0.5045\,X_1 - 0.3812\,X_2 \\ Y_2 &= 1.5843 - 0.6584\,X_1 + 0.1205\,X_2 \\ Y_3 &= 0.7119 - 0.3503\,X_1 + 0.5374\,X_2 \\ Y_4 &= 1.7773 - 0.8095\,X_1 - 0.2639\,X_2 \\ Y_5 &= 0.7186 - 0.4638\,X_1 + 0.2154\,X_2 \end{aligned} \tag{2}$$

where

Y_1 = home-based work trips produced per dwelling unit;
Y_2 = home-based other trips produced per dwelling unit;
Y_3 = home-based social-recreational trips produced per dwelling unit;
Y_4 = home-based shop trips per dwelling unit;
Y_5 = home-based school trips per dwelling unit;
X_1 = log of residential density; and
X_2 = car ownership.

The accuracy with which the selected equations estimate total trip production by purpose when compared to the O-D totals is given in Table 5. Although the total areawide home-based trip production estimate was 3.4 percent higher than the O-D value, the percent error by purpose varied from -20.5 to +21.6. The precision of the zonal estimates by trip purpose both before and after adjusting is given in Table 6.

It was not possible to develop nonhome-based trip production estimates by zone using the 13-zone data. The total nonhome-based productions from the 13-zone study amounted to 29,283 or slightly more than 100 trips per zone since they were produced in all

TABLE 5

PRECISION OF TOTAL HOME-BASED TRIP PRODUCTION ESTIMATES DEVELOPED FROM 13-ZONE DATA AND APPLIED TO TOTAL STUDY AREA

Trip Purpose	Trip Productions		Difference	
	1958 Survey	Clustered Survey Est.	No.	%
Work	796,646	813,761	17,115	2.1
Other	423,226	439,911	16,685	3.9
Social-rec.	292,195	355,226	63,031	21.6
Shopping	283,555	303,976	20,421	7.2
School	231,092	183,674	-47,418	-20.7
Total	2,026,714	2,096,548	69,834	3.4

TABLE 6

PRECISION OF ZONAL HOME-BASED TRIP PRODUCTION ESTIMATES DEVELOPED FROM 13-ZONE DATA AND APPLIED TO TOTAL STUDY AREA

Trip Purpose	1958 Survey Data		RMS Error[a]			
			Unadjusted		Adjusted	
	Mean	Std. Dev.	Abs.	%	Abs.	%
Work	3,525	2,698	735	20.85	723	20.51
Other	1,873	1,360	578	30.86	572	30.54
Social-rec.	1,293	997	661	51.12	543	42.00
Shopping	1,255	901	430	34.26	419	33.39
School	1,023	726	530	51.81	489	47.80

[a]Percent RMS error = $100 \left(\frac{\sqrt{\frac{\Sigma (\text{diff.})^2}{n}}}{\bar{x}} \right)$

where

diff. = difference between surveyed and estimate productions;
n = number of zones; and
$\bar{x}$ = mean of survey volume.

zones in the study area. A regression analysis showed that the 13-zone O-D nonhome-based trips were correlated with the 1958 O-D nonhome-based trips, as indicated by the correlation coefficient of +0.83. An examination of the data, however, showed that much of the variation in the two sets of data was due to bias in the vicinity of the 13 zones. Each of the 13 zones and the zones immediately surrounding them had a higher percentage of nonhome-based trip productions than shown by the total study area O-D data. Conversely, zones located at some distance from any of the 13 zones had a lower percentage of the nonhome-based trip productions. An alternate procedure was therefore developed which made use of the nonhome-based trip data from the 1958 O-D survey. Total areawide nonhome-based trips can be estimated from the 13-zone O-D data by relating these trips to the characteristics of the trip maker's zone of residence. A separate analysis must then be used to determine the zones of origin of these trips.

Trip Attractions.—Due to the lack of land-use or socio-economic data on zonal employment, school enrollment, or retail sales, etc., it was impossible to develop reliable trip attraction estimates using regression analysis. This was true even when the trip attraction data from the 1958 O-D survey were substituted for the 13-zone survey trip attraction data.

A further discussion of the 13-zone survey attraction data is warranted, however, since even if the land-use data had been sufficient the trip data would still present problems. The 13-zone trip attractions occur in all study area zones and are of low statistical significance numerically. More importantly, the distribution of trip attractions among the zones is highly biased by the location of the 13 zones. Therefore, a set of trip attraction estimates developed from the total area survey was substituted for the 13-zone trip attraction estimates.

Estimates Based on Study of All Zones

Trip Production.—This phase of the study made use of all available data, including 1958 O-D survey trips and all of the various land-use and socio-economic information. Regression analysis was used to estimate trip productions for all trip purposes. Several equations were developed and analyzed for each trip purpose. The selected equations and their standard errors of estimate (designated as ± values) are as follows:

$$Y_1 = -489.4913 + 0.2753\,X_1 + 0.3773\,X_2 + 0.6729\,X_3 \pm 521.0435$$

$$Y_2 = 61.1273 + 0.0351\,X_1 + 0.0214\,X_2 + 0.8878\,X_3 \pm 551.5152$$

$$Y_3 = 64.9557 + 0.0281\,X_1 - 0.0465\,X_2 + 0.6561\,X_3 \pm 542.6655$$

$$Y_4 = 32.6220 + 0.0946\,X_1 - 0.3636\,X_2 + 0.7680\,X_3 \pm 379.1923$$

$$Y_5 = 307.6413 + 0.1083\,X_1 - 0.6711\,X_2 + 0.7782\,X_3 \pm 491.8209$$

$$Y_6 = 352.8010 - 0.4573\,X_1 + 2.0154\,X_2 + 0.0432\,X_3 \pm 1010.1590 \qquad (3)$$

where

Y_1 = home-based work trips produced per O-D zone;
Y_2 = home-based other trips produced per O-D zone;
Y_3 = home-based social-recreational trips produced per O-D zone;
Y_4 = home-based shop trips produced per O-D zone;
Y_5 = home-based school trips produced per O-D zone;
Y_6 = nonhome-based trips per O-D zone;
X_1 = population per O-D zone;
X_2 = dwelling units per O-D zone; and
X_3 = cars owned per O-D zone.

The precison of the estimates developed from these equations both before and after adjusting for the elimination of negative estimates is given in Table 7.

Trip Attraction.—A regression analysis was made using the available land-use data yielding estimates with standard errors of estimate given in Table 7. Because these standard errors of the regression estimates were unacceptably high, even after the estimates were adjusted for negative values, rates of trip attraction per unit of service acreage were developed by trip purpose for each district in the study area. The correlation analysis had shown that service acreage was the best single variable for estimating trip attraction for each of the six trip purposes. These attraction rates developed for each district were then applied to each zone in the district. These estimates were also very poor and were arbitrarily adjusted to within ±30 percent of the O-D values. The accuracy of these estimates both before and after adjustment is shown in Table 8.

TABLE 7

PRECISION OF TOTAL STUDY AREA REGRESSION ESTIMATES

Trip Purpose	1958 Survey Data		RMS error[a]			
			Unadjusted		Adjusted	
	Mean	Std. Dev.	Abs.	%	Abs.	%
(a) Productions						
Home-based						
Work	3,525	2,698	521	14.78	509	14.44
Other	1,873	1,360	552	29.47	546	29.15
Social-rec.	1,293	997	543	42.00	537	41.53
Shopping	1,255	901	379	30.20	371	29.56
School	1,023	726	492	48.09	476	46.53
Nonhome-based	1,326	1,534	639	48.19	385	29.03
(b) Attractions						
Home-based						
Work	3,355	6,721	3,745	111.62	3,200	95.38
Other	1,817	2,257	1,112	61.20	883	48.60
Social-rec.	1,191	1,254	690	57.93	587	49.29
Shopping	1,245	2,381	1,471	118.15	723	58.07
School	958	1,094	863	90.08	888	92.69
Nonhome-based	1,321	1,440	600	45.42	513	38.83

[a]RMS error of unadjusted regression estimates is equal to standard error of estimate.

TABLE 8

PRECISION OF TOTAL STUDY ATTRACTION RATE ESTIMATES[a]

Trip Purpose	1958 Survey Data		RMS Error			
			Unadjusted		Adjusted	
	Mean	Std. Dev.	Abs.	%	Abs.	%
Home-based						
Work	3,355	6,721	3,900	116.24	583	17.38
Other	1,817	2,257	1,252	68.90	497	27.35
Social-rec.	1,191	1,254	740	62.13	250	20.99
Shopping	1,245	2,381	1,520	122.09	246	19.76
School	958	1,094	974	101.67	212	22.13
Nonhome-based	1,321	1,440	650	49.21	269	20.36

[a]Estimates per origin-destination district based on 1958 trips per acre of land used for nongoods-handling activities serving both individuals and business.

COMBINED ANALYSIS—TRIP PRODUCTION, ATTRACTION AND DISTRIBUTION

The objectives of this phase of the study were (a) to evaluate the amount of error introduced into a gravity model trip distribution by the use of trip production and trip attraction estimates in the gravity model formulation, rather than trip production and trip attraction values taken directly from O-D data; and (b) to provide statistical measures of the accuracy of gravity model trip distributions calculated with parameters developed from 13-zone clustered survey trip data and from total study area trip data.

13-Zone Data

Two levels of areawide gravity model trip distribution accuracy for travel time factors were developed with 13-zone survey trip data.

1. The calibrated gravity model previously described used O-D trip ends (i.e., trip productions and trip attractions). The accuracy of this trip distribution when compared to the 1958 O-D on a district-to-district movement basis is given in Column (d) of Table 3.
2. A new gravity model trip distribution was calculated using the same travel time factors but with modified trip end input. The five home-based trip production purpose estimates were developed as previously described, using only 13-zone trip data and related land-use data for the entire study area. To develop nonhome-based trip productions and all trip attractions, it was necessary to use trip data from the 1958 survey. The accuracy of the trip production and trip attraction estimates is given in Tables 6 and 8, respectively.

The results of the gravity model trip distribution calculated with the travel time factors developed from 13-zone data and the trip end estimates developed partially from 13-zone data are given in Column (f) of Table 3. The increased error of the second distribution—Column (d) vs Column (f)—may be related entirely to the decreases in the accuracy of the trip production and trip attraction estimates.

Total Study Area Data

Three levels of trip distribution accuracy using the travel time factors were developed with total study area trip data and alternate trip production and trip attraction estimates.

1. The calibrated gravity model, which used O-D trip ends, was described previously. The accuracy of this trip distribution when compared to the 1958 O-D district-to-district movements is given in Column (e) of Table 3.
2. A gravity model was next run which used the same total study area travel time factors but with unadjusted regression estimates of trip production and trip attraction. Column (g) of Table 3 gives the accuracy of these estimates. This run was made to analyze the effect on trip distribution of trip attraction estimates that were significantly in error. The trip production estimates were satisfactory, but the trip attractions exhibited a high standard error. In addition, the error was biased with contiguous zones such as the CBD all under- or overestimated. The error in this trip distribution was the highest of any of the gravity models.
3. The trip attraction estimates used as input to the last run were adjusted by zone for the reasons previously discussed. The accuracy of these adjusted trip attraction estimates is given in Table 8.

The gravity model trip distribution was recalculated using as input parameters the total study area travel time factors, the regression trip productions, and the adjusted trip attractions. The accuracy of this distribution with respect to the O-D is given in Column (h) of Table 3.

SUMMARY

Trip Production and Trip Attraction Estimates

Home-based trip productions were estimated by zone for the entire study area using the clustered survey trip data. These estimates showed relatively minor losses in accuracy when compared to the regression estimates based on the 1958 O-D survey data. The percent RMS error of the zonal estimates of home-based trips developed from the clustered survey data was, on the average, approximately 5 percent greater than the percent RMS error of the zonal estimates based on trip data from the 1958 O-D survey. Areawide, total home-based trips were overestimated by 3.4 percent based on the clustered survey analysis. A statistical analysis of the total study area trip data showed that a random sample of the same size as the cluster sample would have estimated total study area trips within +1.7 percent (12).

Trip attractions for any purpose or the distribution of nonhome-based trip productions cannot be estimated from the clustered survey data. It must be assumed that if a clustered survey is used as the sole source of trip calibration data for a gravity model, synthetic trip end measures must be utilized. That is, specific land-use and socio-economic variables must be used as indices for the zonal distribution of trip attractions and for nonhome-based trip productions.

With a comprehensive survey, it might possibly be shown that the accuracy of these synthetic procedures is very high. However, the main disadvantage of these procedures is that there is no accurate means of checking the estimates without a complete home interview survey.

Travel Time Factor Calibration

Travel time factors developed with the 13-zone clustered survey trip data show significant differences when plotted and compared with the factors developed from the total study area data (Figs. 5 through 10). The differences are particularly significant for travel time values of less than 10 min. This portion of the travel time factor curve is used primarily to determine the number of intrazonal trips. A special analysis was made on intrazonal trips which showed that these trips were underestimated by 8.0 percent using only the 13-zone trips, but when the 13-zone factors were applied to the total study area, intrazonal trips were underestimated by 32.2 percent. In addition to variation in the upper portions of the curves, there was slope variation that was difficult to evaluate, but, in general, the total study area factors had steeper slopes.

Trip Distribution Accuracy

When the travel time factors developed using the 13-zone clustered survey data were applied to the total study area there was a loss in accuracy of the trip distribution as indicated in Table 3, Column (d) vs Column (e). The loss in accuracy was not significant enough to draw negative conclusions on the adequacy of clustered surveys, but it was significant enough to point up the value of the more accurate calibration data. The percent RMS error was 5 percent greater (21.07 vs 16.08) for trip volumes of 15,000 and over when the 13-zone factors were used as opposed to the total study area factors.

Socio-Economic Adjustments

The gravity model calibrated with the 13-zone data did not show a specific need for the use of socio-economic adjustment factors, $K_{(i-j)}$. However, if these factors had been necessary, there were not sufficient trip data from the clustered survey to establish them. The 1958 O-D trip data were extensive enough for a full $K_{(i-j)}$ factor analysis and for a meaningful river crossing analysis.

Full Gravity Model Trip Estimation

The combined analysis demonstrated that over the range of the larger trip volumes there was very little difference in the accuracy of the gravity model trip distributions when either O-D trip ends or trip end estimates were used as trip production and trip attraction. The analysis did demonstrate that when poor trip end estimates, such as the unadjusted regression attractions, were used as model input, a very significant decrease in accuracy resulted.

Adequacy of Clustered Samples

The clustered sample provided very stable trip volumes but did not provide good calibration data. An evaluation of the many study analyses would seem to indicate that one reason for the inadequacy of the clustered survey data is inherent in clustered sampling itself. The clustered survey trip attractions were biased by the location of the 13 zones. This bias had a direct effect on the travel time factors. The travel time factors developed from the total study area when compared directly with the 13-zone factors (Figs. 5 through 10) show higher travel time factors for the low time increments. If other parameters of the model remain fixed, these higher time factors mean that the total study area data place a higher weight on short trips than do the 13-zone data. If it is considered that the total study area data present a greater range of potential attractions, then it may be hypothesized that the travel time factors at the lower time intervals would have to be greater to keep the correct proportion of short trips.

The 13 zones selected for clustered sampling were chosen with the utmost care. During the course of the research work there was no reason to criticize the selection or to feel that a different set of 13 zones would have been more representative of the entire study area. The basic problem seemed to be the bias created by the fact that the clustered survey zones amounted to such a small portion of the total zones in the study area.

Further Research

The authors recommend that further research be undertaken on the reduction of data requirements for travel model calibration. They suggest that the most promising approach to this research would be to examine the adequacy of small systematic or random samples. This suggestion is not derived from any specific study finding, but rather from the general impression that any clustered survey data will be biased by the location of the selected zones. To eliminate this bias, the number of zones selected for interviewing would have to be increased to a point where the clustered sample would take on the characteristics of a systematic sample.

CONCLUSIONS

1. The clustered survey did provide sufficient trip data for the development of zonal estimates of trip production.
2. The clustered survey did not provide enough trip data for the development of trip attraction estimates.
3. The travel time factors developed from the clustered survey were significantly different from those developed with the total study area data.

REFERENCES

1. Kudlick, Walter P., and Fisher, Ewen S. Use of Pre-Interview Trip Cards in Developing a Traffic Model for the Hamilton Area Transportation Study. Highway Research Board Bull. 347, pp. 186-202, 1962.
2. Davidson, Robert G. Developing a Traffic Model With a Small Sample. Highway Research Board Bull. 297, pp. 106-108, 1961.
3. Smith, Bob L. Gravity Model Theory Applied to a Small City Using a Small Sample of Origin-Destination Data. Highway Research Record 88, 1965.
4. Sullivan, S. W., and Pyers, C. E. Results of Use of Pre-Interview Contacts in Pittsburgh. Highway Research Board Bull. 298, pp. 42-51, 1961.
5. Sullivan, Sheldon W. Comparisons of Selected Home Interview Data, 1958 and 1960. PATS Research Letter, Vol. 3, No. 3, pp. 1-10, May-June 1961.
6. Pittsburgh Area Transportation Study Final Report, Vol. 1, Nov. 1961; Vol. 2, Feb. 1963.
7. Sosslau, Arthur B., and Brokke, Glenn E. Appraisal of Sample Size Based on Phoenix O-D Survey Data. Highway Research Board Bull. 253, pp. 114-127, 1960.
8. Calibrating and Testing a Gravity Model for Any Size Urban Area. U. S. Bureau of Public Roads, Office of Planning, July 1963.

9. Hansen, Walter G. Evaluation of Gravity Model Trip Distribution Procedures. Highway Research Board Bull. 347, pp. 67-76, 1962.
10. Bouchard, Richard J., and Pyers, Clyde E. Use of Gravity Model for Describing Urban Travel—An Analysis and Critique. Highway Research Record No. 88, 1965.
11. Hartford Area Traffic Study Report, Vol. 1, July 1961.
12. Ben, Constantine, Bouchard, Richard J., and Sweet, Clyde E. An Evaluation of Simplified Procedures for Determining Travel Patterns in a Small Urban Area. Highway Research Record No. 88, 1965.