Expansion of CATS 1990 Household Travel Survey for Transit

Siim Sööt and Ashish Sen
with
Paul Metaxatos and Lise Dirks

Urban Transportation Center
University of Illinois at Chicago
1033 West Van Buren Street
Suite 700 South
Chicago, IL 60607-2919
e-mail: siim@uic.edu
Tel: (312) 996-4820
Fax: (312) 413-0006

Final Report

submitted to the
Regional Transportation Authority

January 21, 1997

1 Background

While survey data used for modeling purposes should generally use raw unfactored data [5], this is not true when the data are used for descriptive purposes. The initial factoring of the CATS 1990 Household Travel Survey (HHTS) [3] showed some discrepancies between expanded trip totals and observed passenger counts for CTA, Pace and Metra. It is the purpose of this study to analyze in more detail the HHTS in order to assess where corrections in survey expansion factors should be made to better replicate observed transit passenger counts. These corrections will then be implemented, resulting in revised expansion factors and new transit trip totals for the HHTS.

In a similar previous effort [2] a two-step adjustment to the original survey was proposed. In a first step, the expansion factors of commuters in the survey were adjusted by matching the distribution of survey commuters (persons who participated in work trips) by travel modes and residence-workplace locations to the Census journey-to-work data. In a second step, the expansion factors for survey non-commuters were adjusted iteratively by matching the expanded transit trips by submodes to the regional transit ridership estimates. This corrective action was suggested under the assumption of nonresponse bias at the household and person level in the original survey.

In the HHTS steps were taken during the planning stages of the survey to eliminate anticipated problems with total nonresponse bias at the household and person level and item nonresponse bias at the trip level [1]. The discrepancies, however, between the expanded survey transit trips and the observed transit passenger counts raise the suspicion that item nonresponse bias problems remain. Evidently, transit trips are recorded at a higher rate than trips by other modes, and this results in inflated expansion factors for those trips in the HHTS. The adjustment method devised to alleviate the problem suggests a stepwise procedure, which is described in the next section.

2 Suggested Methodology and Preliminary Results

The method of re-adjusting the survey weights acknowledges a hierarchy of mode use. This hierarchy stems, in part, from the trips which are linked as access–egress modes coupled with rail modes. This hierarchy is:

- Metra,
- CTA Rail,
- CTA Bus and
- Pace.

Transit trips (trips taken by either Metra, CTA bus or rail, or PACE) from the HHTS were expanded using the initial factors, summarized at the township geographic level, and mapped together with observed passenger counts at the same township level. On inspection, it became clear to the investigators of the study that transit trips heading to the Chicago CBD were overpredicted. Those trips are mainly taken by Metra and CTA rail, and, to a lesser extent, by PACE. Consequently, the focus of the investigation fell exactly on those trips.

In the HHTS the access and egress trips from and to the Metra or CTA rail modes are treated as separate trips by design. The access and egress trips from and to the PACE and CTA bus modes, however, were not identified separately from the transit trip itself, but only as distances walked from and to the bus stop. As a result, if a survey participant failed to report on an access or egress trip to the Metra or CTA rail modes she would cause an item nonresponse problem, and consequently, would raise the value of the factor for that particular trip unduly. The survey participant, however, could not miss reporting on the access or egress trip from or to the PACE or CTA bus modes because, simply, the participant was not asked to. Those transit trips, therefore, would have been factored up correctly had the initial expansion method been error free.

If the above conjecture for the Metra and CTA rail trips is correct, then a suitable correction for the item nonresponse bias would be to expand the access and egress trips from and to those two transit rail modes in the same way as the (Metra or CTA rail) trip itself. If, for example, the Metra or CTA rail trip is to be adjusted by only ninety percent of its original expansion, the access and egress trips should be factored up in the same proportion. The investigators decided to adjust first the trips taken by Metra because in those cases the access and egress modes would frequently include other transit modes. Therefore, an adjustment to the expansion of the trips taken by Metra would adjust, at the same time, other transit trips by CTA bus and Pace, which are linked to Metra trips as access or egress modes. This is very desirable.

An adjustment to the expansion of trips taken by Metra (or any other mode) requires determining the adjustment factor, f_{adjust}, a number normally between 0 and 1, which then would have to be applied to the original expansion factor using the formula,

$$T^{new}_{expanded} = T_{sample} \times w_{initial} \times f_{adjust}$$

where,

$T^{new}_{expanded}$ – a new expanded number of trips correctly adjusted,
T_{sample} – a survey trip (not expanded) and
$w_{initial}$ – the trip factor computed in the initial survey expansion.

Note the above formula is generic, because there are many instances during which the access and egress trips are taken by automobiles or non-motorized modes.

2.1 Correcting the Survey Expansion Factors for Metra trips

To calculate the adjustment factor, f_{adjust}, for Metra trips the HHTS Metra trips were summarized at the township level. Boarding counts were obtained and summarized also at the township level. Every effort was made to ensure that the boarding counts had been taken on a date as close as possible to the date of the survey. The ratio by township of the survey data divided by the boarding counts determined the desired factor. In cases where a township had not been sampled or if the original factor is correct, $f_{adjust} = 1$. The adjustment factor was then appended to the access (egress) sample trips (in the HHTS) from (to) the Metra rail mode.

Table 1. Adjustments for METRA Trips

Transit Mode	Original Survey	Adjusted Survey	Actual Counts	Percent Diff.[a]
Pace	147.136	143.479	131.430	9.2
Metra	340.572	257.563	256.846	0.2
CTA Bus	1.480.154	1.473.455	1.258.000	17.1
CTA Rail	575.964	575.271	459.350	25.2

[a]Percent Diff. is the difference between adjusted survey and actual counts.

The system-wide results of such an adjustment appear in Table 1. The regional totals for Metra trips now appear to be in agreement with the observed system-wide passenger counts.

2.2 Correcting the Survey Expansion Factors for CTA Rail Trips

The correction of the survey expansion factors for Metra trips also adjusted other transit modes but not completely. For example, the Pace total dropped by almost 4,000, but it is still 12,000 above Pace's passenger counts. CTA bus declined by almost 7000 but is still 17% too high.

Since CTA rail changed very little and for reasons explained earlier, corrective action was also taken for these rips. The adjustment factors for CTA rail trips were obtained in a similar manner as those for Metra trips, and were subsequently applied to the access (egress) sample trips in the HHTS. The effect of the compound adjustment of both the Metra and CTA rail trips can be seen in Table 2. It seems that the adjustment of the CTA rail trips had a dramatic effect on adjusting also the CTA bus trips as well as the PACE trips. CTA Bus dropped by over 57.000 and Pace decreased by approximately 5,700.

Table 2. Adjustments for trips made by METRA and CTA Rail

Transit Mode	Original Survey	Adjusted Survey	Actual Counts	Percent Diff.[a]
Pace	147.136	137.702	131.430	4.8
Metra	340.572	257.563	256.846	0.2
CTA Bus	1.480.154	1.416.174	1.258.000	12.6
CTA Rail	575.964	461.204	459.350	0.4

[a]Percent Diff. is the difference between adjusted survey and actual counts.

2.3 Correcting the Survey Expansion Factors for CTA Bus Trips

Despite the fact that the adjustment of the Metra and CTA rail trips had a considerable impact in adjustment of the CTA bus trips, the system wide results for trips taken by the latter mode were still at some distance from the desired target value. The adjustment of the CTA bus trips followed a different logic from the previous two situations with the Metra and CTA rail trips, and is illustrated in Figure 1. The different method is attributed to the lack of boarding counts. Since only route-level data are available for CTA Bus, it was not possible to assign these data to specific township, the bases for the previous two methods. The CTA Bus system is too dense and each route is too long to be able to extract township level ridership estimates.

The adjustment factors calculated are shown in Figure 1 for the CTA rail trips (top number) and the Metra trips (bottom number). The number in the middle represents the adjustment factor for the CTA bus trips calculated from the formula,

$$f_{adjust}^{CTA-BUS} = 1 - \{\frac{1}{2}[1 - \frac{(f_{adjust}^{METRA} + f_{adjust}^{CTA-RAIL})}{2}]\}.$$

Note that this adjustment factor applies only to CTA bus trips and not to access or egress trips as in the previous two cases for reasons explained earlier.

The aforementioned method for CTA bus trips was used with the expectation that the previously performed Metra and CTA rail adjustment provides information about how to adjust the CTA bus trips. Moreover, since the difference between the CTA bus survey data and the CTA ridership data are relatively small, the adjustments only need to be half as large as the average for both rail modes.

The system-wide summary of the transit trips after the combined adjustment of the Metra, CTA rail and bus trips has been performed is presented in Table 3. It can be seen that the last adjustment has brought the expanded CTA bus trips very close to the target value.

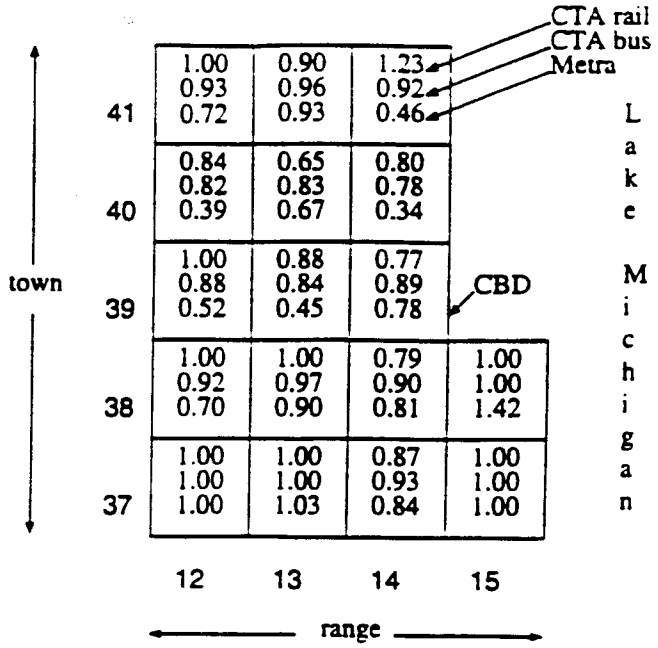

Figure 1: Adjustment factors for Metra, and CTA rail and bus trips within the City of Chicago

2.4 Correcting the Survey Expansion Factors for Pace Tips

The final adjustment was made for trips by Pace. The target number is reported as being 131,430 (Table 3) but there are several reasons why this number is difficult to establish. First, Pace carries many school children and the Household Travel Survey only records trips for persons of age fourteen and over. Second, Pace reports approximately 142,000 for the 1990-1991 period when the survey was conducted. The adjusted number of 136,860 is therefore offered as the target number as a compromise among the various factors contributing to the uncertainty of the target number.

Table 3. Adjustments for trips made by METRA, CTA Rail and CTA Bus

Transit Mode	Original Survey	Adjusted Survey	Actual Counts	Percent Diff.[a]
Pace	147,136	137,702	131,430	4.8
Metra	340,572	257,563	256,846	0.2
CTA Bus	1,480,154	1,269,491	1,258,000	0.9
CTA Rail	575,964	461,204	459,350	0.4

[a]Percent Diff. is the difference between adjusted survey and actual counts.

This number was derived by making adjustments to the survey using route-level ridership data for April, 1991 reported in *Bus Route Descriptions* [4]. This was possible in only limited areas, namely in the outer suburbs where bus routes tend to fall within one or a small collection of townships (the basic factoring zone). This is true for each of the following cities and surrounding areas: Waukegan, Elgin, Aurora, Joliet, Naperville and Crystal Lake. There were a few other townships in the periphery of the metropolitan area where adjustments were made; none were made in the inner ring of suburbs where the tight network of bus routes cross numerous township boundaries.

The net effect of these adjustments, however, was to decrease the ridership too much and therefore all unadjusted townships received a factor of 1.06, as seen in Figure 2, resulting in the 136,860 ridership total. Again, this is the adjusted Pace ridership total in Table 4 and is a compromise between alternative data sources.

Table 4. Adjustments for trips made by METRA, CTA Rail, CTA Bus and Pace

Transit Mode	Original Survey	Adjusted Survey	Actual Counts	Percent Diff.[a]
Pace	147,136	136,860	131,430 [b]	4.8
Metra	340,572	257,563	256,846	0.2
CTA Bus	1,480,154	1,269,491	1,258,000	0.9
CTA Rail	575,964	461,204	459,350	0.4

[a] Percent Diff. is the difference between adjusted survey and actual counts.
[b] Alternative source reports 142,000

3 Data Limitations

While the survey encompasses 19,314 households some subsamples of these data represent rather small numbers. A good example is the Pace ridership by township (Figure 3). In the southeastern corner of the study area in Range/Township numbers 13/34 and 14/34 there were three and one respondents respectively who reported using Pace. These responses were factored up to 220 and 157 riders respectively. Despite the large factored ridership data these two townships only had a total of four total trips in the raw data and therefore care must be used in employing these data in small areas especially if the variable in question (Pace ridership) is a small portion of the entire data set. Only in the inner suburbs are there consistently more than ten trips in the raw data; a number which itself is a small sample.

The case for Metra riders (Figure 4) is analogous but the service area is much better defined and there appear to be several rail lines which have moderately high samples along

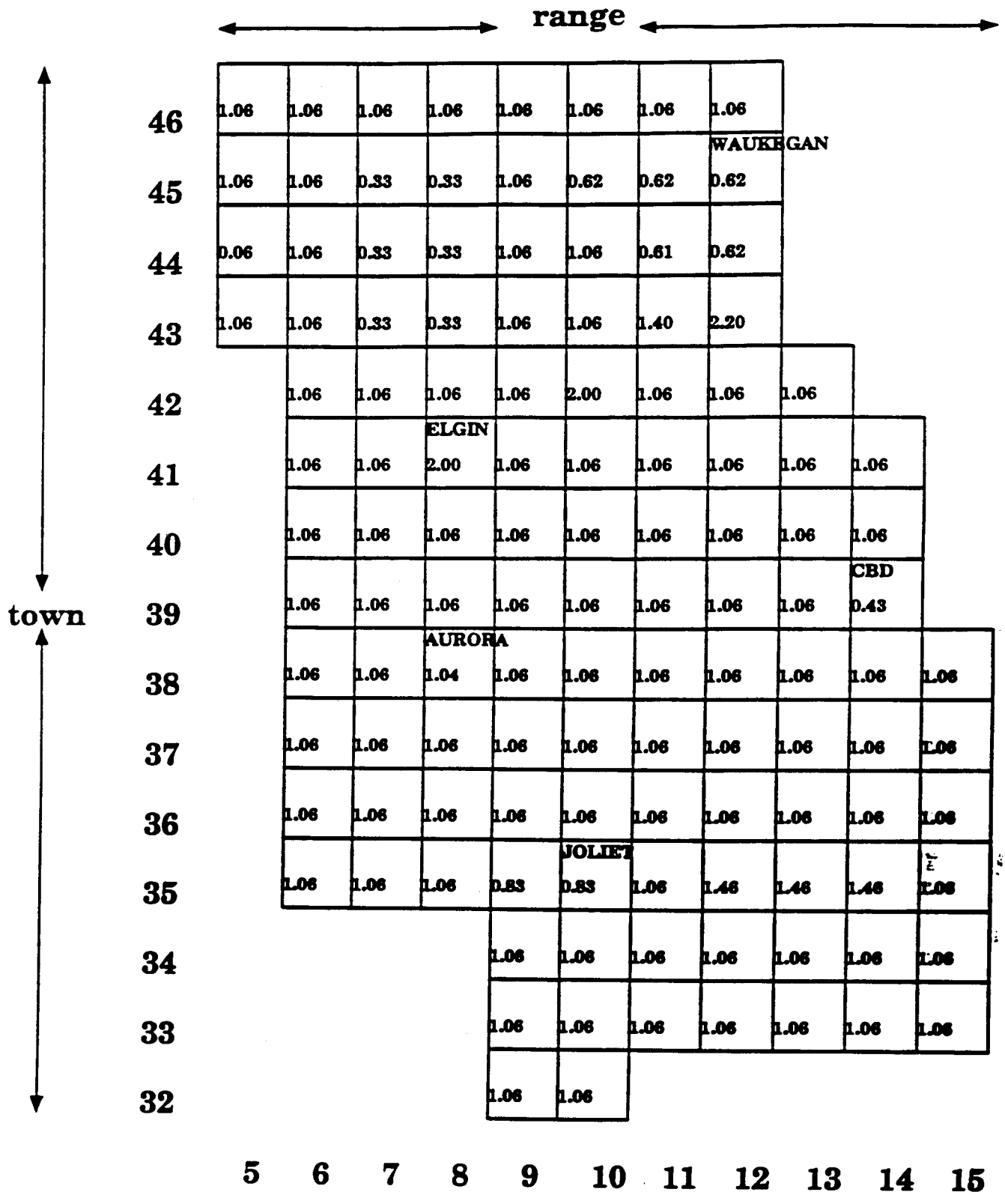

Figure 2: Adjustment factors for Pace trips

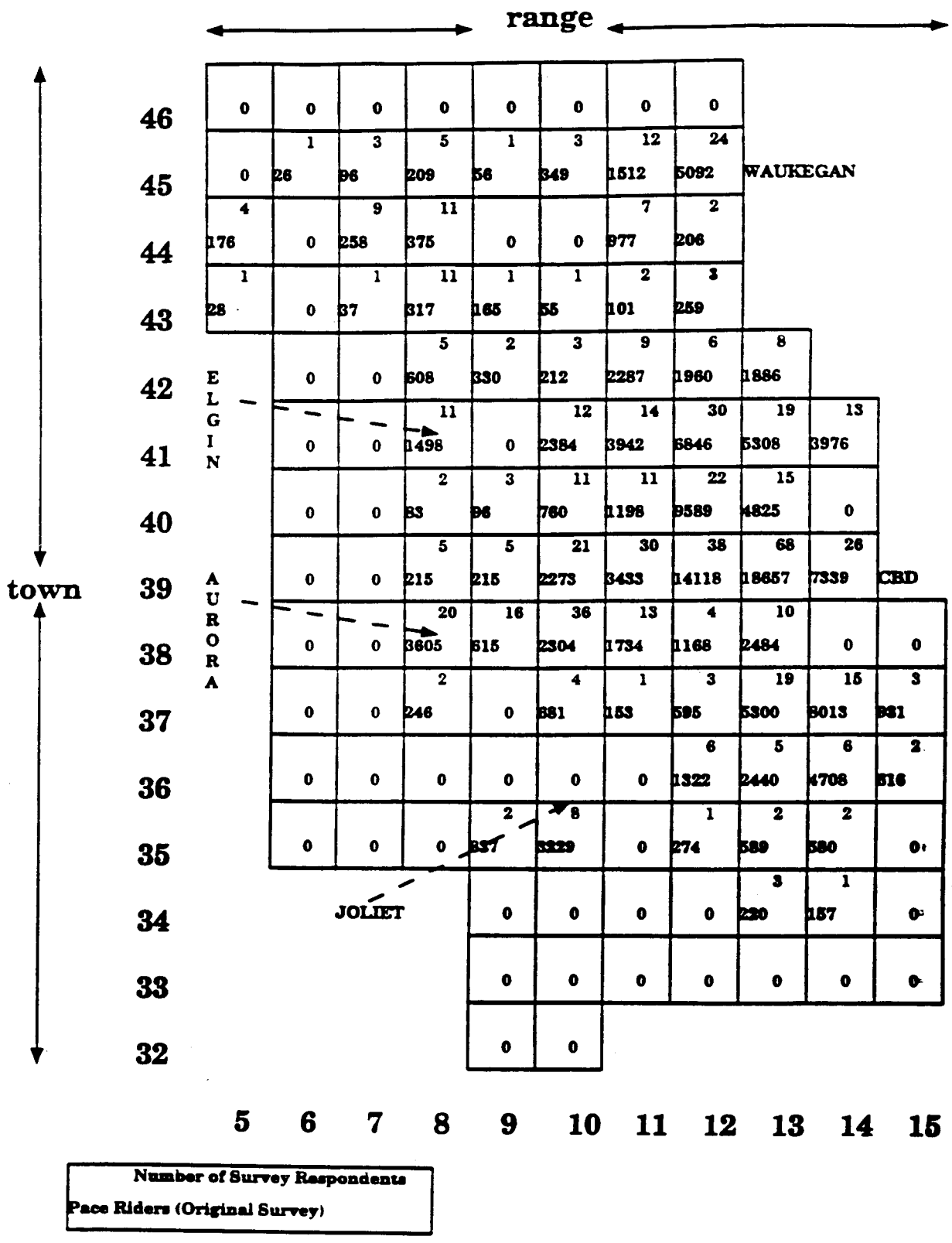

Figure 3: Number of survey respondents and Pace riders in the original survey

the entire line. The Burlington Northern service to Aurora has over 30 riders per township in the raw data all the way to Naperville and even 23 at the terminal township at Aurora. The lowest numbers are found in the two Metra lines servicing the Joliet area. Even the Joliet township, two Metra lines has only three records in the original file.

Since the CTA Rail system is much more confined and the usage levels are high, the survey numbers are much higher (Figure 5). Still there are only ten townships which have more than ten records in the raw data. With these numbers it is possible to perform some township-level comparisons in and near the city of Chicago. For example, it would be possible to determine the proportion of the trips destined to the Chicago CBD or perhaps to suburban Chicago.

Finally, ridership data have been used to adjust the CATS Household Travel Survey data to account for over estimation of transit ridership. While none of the adjusted totals are exactly the targets reported here, given the degree of fluctuations in weekday ridership data the adjusted totals are typically within the range of month-to-month variations.

4 How to Use the Data

The new 1990 CATS HHTS trip file has all of the previous data plus an additional weight for the final adjustment. The proper weighting of the file for use with transit data now includes multiplying two weights, the old and the new. The weights are multiplicative factors, that is, the estimated total is obtained by multiplying each data value by the appropriate weight (the product of the old and the new weight) and summing the results.

Since the data files now have two weights it is important to use the weights properly. As such the files more resemble the Nationwide Personal Transportation Survey which has three different weights and is therefore more difficult to use. For example, if one sums the number of workers in the NPTS by using the household file the total is approximately 103 million.

If the person file is used, then the total is 118 million workers. The reason for the difference is that the person file is factored for missing persons and the household file is not. Similarly, we have now adjusted the trip file for over and under reporting of transit trips. When statements are to be made about transit users then the ***new weights*** are to be used but when the subject is household demographics, then only the ***original weights*** should be used. Similarly characteristics from the person file should be summarized using only the weights from the household file.

5 Data Format

The new data set contains all of the information in the original release (Table 5) plus the new weight and new variables which facilitate processing inquiries on trip chaining. Therefore,

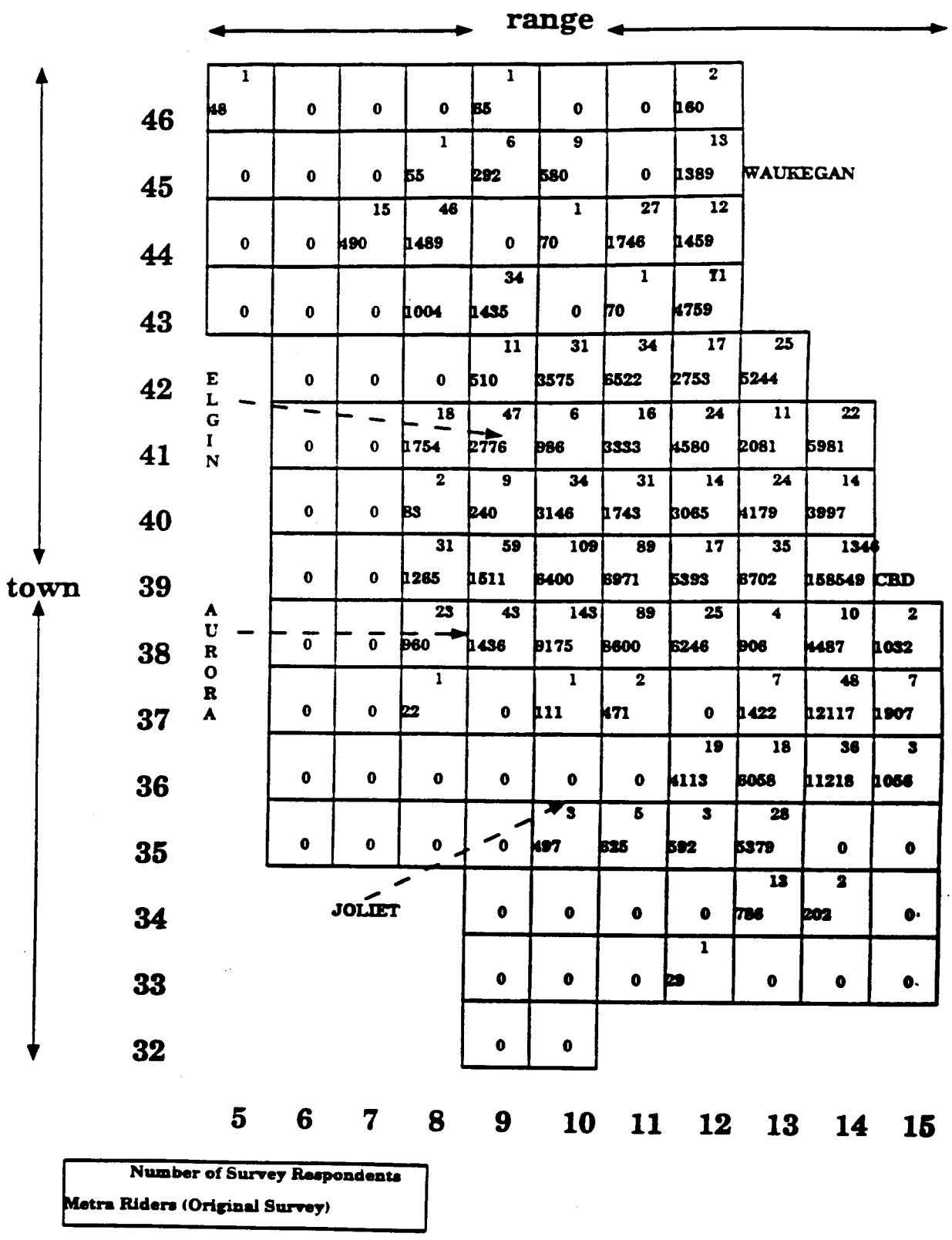

Figure 4: Number of survey respondents and Metra riders in the original survey

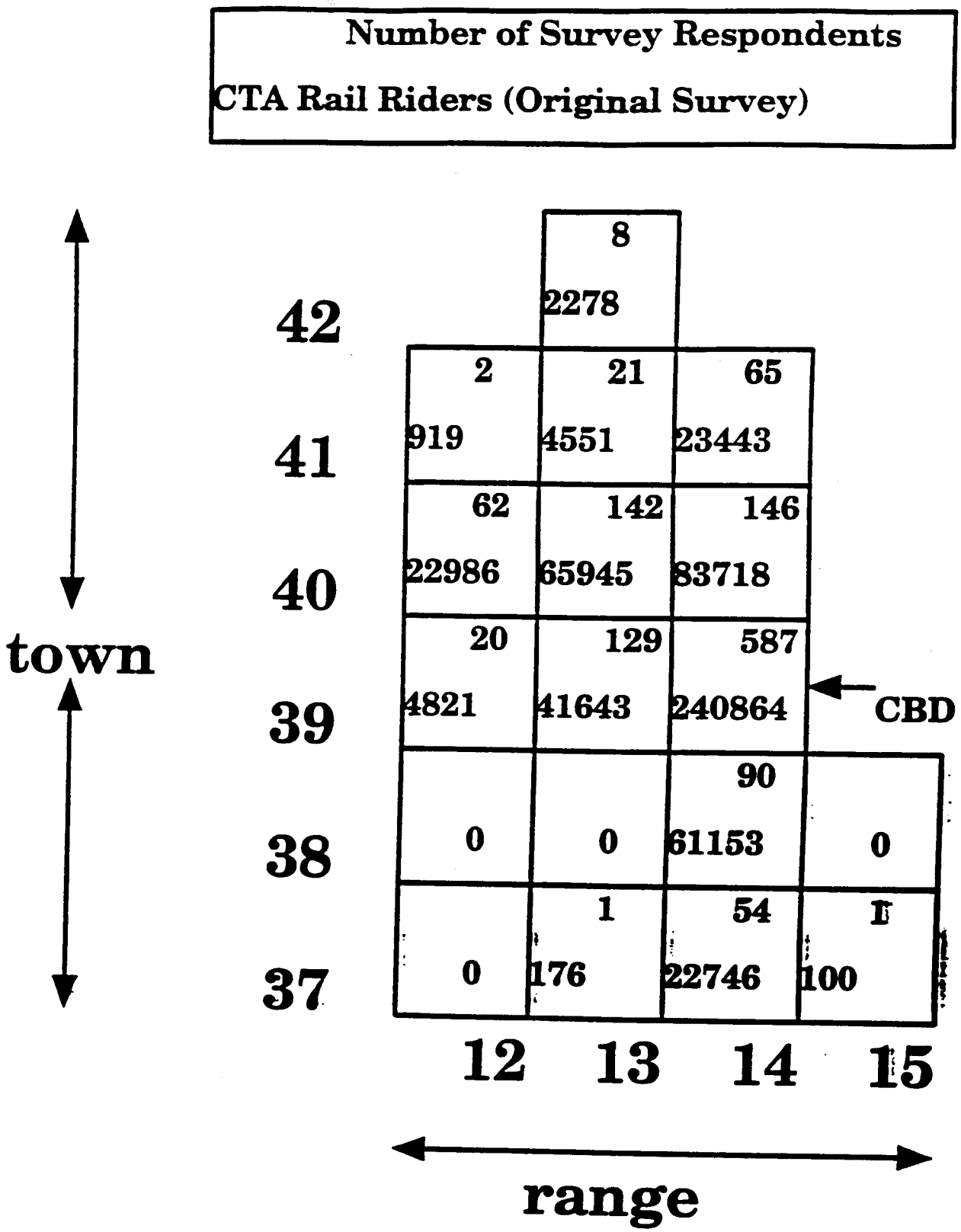

Figure 5: Number of survey respondents and CTA rail riders in the original survey

this discussion only pertains to the trip file and the household and person files remain unaltered.

Five new variables were added to assist studies of trip chaining. A trip chain is defined as a sequence of trips starting and ending at home. The simplest chain consists of a trip to a destination and back home but more complex chains may include over a dozen stops. Increasing trip chaining is an important aspect of changing travel behavior since it affects the propensity to use transit.

The first new variable (CHAINUM) is the sequence number in the chain. The trip file already includes a sequence number for each person in the file but since a person can complete several chains in a day this number is different. The second variable, WCHAIN, is either a one if there is a work destination in the chain or a zero if there is not. It is then possible to distinguish what happens in conjunction with travel to and from work. Also it provides important information about the reason why the peak period is increasing in length as travelers conduct other business on the way to and from work.

SCHAIN is an analogous variable except it reports the presence of shopping in the chain. A chain may have both work and shopping, one and not the other or neither trip purpose.

The variables TOFROMW and TOFROMS identify whether the trip, as part of a work or shopping chain respectively, is conducted on the way to or from work or shop. For work chains TOFROMW it is a one if the trip is on the way to work and a three if it is on the way from work to home. A two designates a subchain starting and ending at work. For shopping chains the variable is similar except the value two is not necessarily a subchain but rather there is a shopping destination before and after this trip in the chain. These two shopping stops do not need to be at the same location, as is common for work destinations.

6 Potential Future Uses of the Adjusted CATS 1990 HHTS

The main reason for employing the CATS 1990 HHTS was to use the data to update the regional transportation models. However, there are numerous other applications. The authors would like to offer a brief list of potential future uses of the data:

- Make comparisons with other surveys which are more specialized in scope and coverage to test for reasonableness.
- Obtain an intimate knowledge of trip and mode-chaining situations either by means of descriptive analysis (spatial or non-spatial) or by developing and estimating mathematical models (we have already started this effort).
- Evaluate the barriers to transit use caused by extensive trip chaining.

Table 5. Augmented CATS 1990 HHTS Trip File Documentation

Variable	Type	Position	Length	Description
CNTYID	NUM	1	3	County Code Number
HHID	NUM	4	5	Household Identification Number
PERID	NUM	9	2	Person Number
TRIPID	NUM	11	2	Trip Number
STARTIME	NUM	13	4	Trip Start Time
ENDTIME	NUM	17	4	Trip End Time
TRTIME	NUM	21	4	Elapsed Travel Time (in minutes)
MODE	NUM	25	1	Type of Transportation
VEHOCC	NUM	26	2	Auto Occupancy for Auto Trips
BLOCTO	NUM	28	2	Blocks Walked to Bus and Rail Modes
BLOCFROM	NUM	30	2	Blocks Walked from Bus and Rail Modes
PURFROM	NUM	32	2	Trip Origin Activity
PURTO	NUM	34	2	Trip Destination Activity
ORGCODE	NUM	36	8	CATS Geog. Reference Code for the Trip Orig.
ORGCBD	NUM	44	3	CBD CAAS/CUTD Zone for Trips Orig. in the CBD
ORGLAT	NUM	47	7	Latitude of Trip Origin
ORGLONG	NUM	54	8	Longitude of Trip Origin
DESTCODE	NUM	62	8	CATS Geog. Reference Code for the Trip Dest.
DESTCBD	NUM	70	3	CBD CAAS/CUTD Zone for Trips Dest. to the CBD
DESTLAT	NUM	73	7	Latitude of Trip Destination
DESTLONG	NUM	80	8	Longitude of Trip Destination
AIRMILE	NUM	88	8	Trip Length in Miles
SPEED	NUM	96	7	Trip Speed in Miles/Hour
MORE	NUM	103	1	Another Trip Indicator
WGT	NUM	104	8	Original Weight
NEW VARIABLES				
CHAINUM	NUM	112	2	Sequence number in the chain
WCHAIN	NUM	114	1	1 if work destination in the chain, else=0.
SCHAIN	NUM	115	1	1 if shopping destination in the chain, else=0.
TOFROMW	NUM	116	1	1 is to work, 2 is work subchain, 3 is work to home
TOFROMS	NUM	117	1	1 is to shop., 2 is shop. subchain, 3 is shop. to home
NWGT	NUM	118	5	New Weight

- Develop models for access to public transit.
- Estimate catchment areas along major transit routes.
- Contrast sociodemographic profiles of transit users with non-transit users, e.g., automobile ownership and household income.
- Determine the use of public transit for purposes other than work.
- Study cold starts in automobiles (automobiles compete effectively in many markets in energy use per passenger mile).

References

[1] Ghislandi, A.C., A.R. Fijal, and E.J. Christopher, *CATS 1990 Household Travel Survey: A Methodological Overview*. Working Paper 94-05, Chicago Area Transportation Study, Chicago, Illinois, April 1994.

[2] Harrington, I., and C-Y. Wang, Modifying Transit Mode Share in Household Survey Expansion, (1995) *Transportation Research Record, 1496*, TRB, National Research Council, Washington, D.C., 25-34.

[3] Kim, H., J. Li, S. Roodman, A. Sen, S. Sööt, and E. Christopher (1993), Factoring Household Travel Surveys, *Transportation Research Record, 1412*, TRB, National Research Council, Washington, D.C., 17-22.

[4] PACE, *Bus Route Descriptions*, April 1991.

[5] Sen, A., S. Sööt, L. Yang, and E. Christopher (1995), Household Travel Survey Nonresponse Estimates: The Chicago Experience, *Transportation Research Record, 1493*, TRB, National Research Council, Washington, D.C., 170-177.