

will have to evaluate their specific cases, and the Bureau of the Census provides preliminary data that are helpful in that evaluation. We think the package is good; it just needs refining.

## The Tri-State Region

Haden Boswell, Tri-State Regional Planning Commission

In order to show how the Tri-State Regional Planning Commission has used data from the 1970 census, I want first to define our region briefly and describe our role as a census processing center. I shall also describe in more detail our experience so far with the journey-to-work and the Urban Transportation Planning Package (UTPP) because I believe it is unique and will be of interest to you.

### THE REGION

The Tri-State region comprises 8,456 square miles, including parts of 3 states, 27 counties and planning regions, and more than 600 communities. It contains 11 (1970) SMSAs, about 140 Universal Areas, and 4,521 census tracts.

According to the 1970 census, the population of the region was 18.7 million, a gain of 11 percent since 1960. More than 15 percent of this population is nonwhite. The population total includes 75 percent of the population of New Jersey, 67 percent of New York, and 52 percent of Connecticut. On a typical workday in 1970 there were 7.4 million journeys to work; 25 percent of those ended in the Manhattan central business district. Sixty percent of all the work trips were by automobile, an increase of 39 percent since 1960. Transit trips dropped from 48 percent in 1960 to 37 percent in 1970. There are 7.5 million registered vehicles in the region, and 85 percent of all their travel is on highways.

### DATA BASE

To establish its planning data base, Tri-State in 1963 conducted field surveys in the intensively developed urbanized area of the region. A home interview survey of 1 percent of the households (57,000 interviews) gathered detailed demographic and travel data. A land use survey field-listed all blocks in the urbanized area portion (except those in New York City, which were compiled from tax assessor records). Truck, goods movement, taxi, and external (cordon crossing) surveys were also accomplished at that time.

To cope with the processing and analysis of data for a region this size, Tri-State developed the square-mile data cell, defined by X-Y coordinates. Data from the surveys, originally coded to the block, were aggregated to these square miles, which are represented visually in many forms in our planning process. The origins of these are the data map, computer produced in strips and photographically reduced. Another form in which square-mile data are represented is the "stick model," a 3-dimensional approach that also begins with the data map.

The square-mile data cell is an example of our regional approach to this large land area, an approach that is often at variance with census geography.

## PROCESSING CENTER

As a census processing center, Tri-State has obtained tapes for the region of the first, second, third, fourth, and sixth counts; the public use sample; address coding guides; DAULIST 1, 2, 3, and 4; the admatch program; grids mapping system; and Medlist. We were deeply impressed with the efficient advance preparation that the census bureau provided, by means of which we were able to prepare programs in advance of the receipt of all counts except the fourth. We simply put on the tapes and ran the tables.

A total of 841 tables have been formatted and printed at varying levels of geography. Copies were sent to the state and county planning offices in the region, and a copy of each table was reduced to  $8\frac{1}{2} \times 11$  in. and placed in a master file containing approximately 35,000 such pages. These are copied on order for any agency, private company, or person for the cost of copying. The Metropolitan Map Series on mylar made possible the reproduction of any part of the urbanized area and was sold to accompany the tabulations. Thus, the detailed data were made available in the region far ahead of census bureau publication. Special tabulations at cost are prepared for public agencies only because Tri-State is not in competition with commercial processing centers.

For our mailing lists and for public consumption, we have prepared a group of regional profiles that contain maps and tables and show distributions, densities, growth, and changes in the region. So far, 6 have been released on population, housing units, automobile availability, work places and work travel, education, and labor force.

Users of our data include a varied group: family agencies, those interested in local redistricting, Madison Square Garden, real estate offices, Ford Motor Company, agencies for the aged, mayors' offices, the High-Impact Anti-Crime Program, Manufacturers Hanover Trust Bank, Worldwide Volkswagen, the NAACP, the Rand Corporation, the Chase Manhattan Bank, Hospital and Health Plan for Northern New Jersey, the Federal Reserve Bank, Hoboken Model Cities Program, and the U.S. Corps of Engineers.

Tri-State is actively supporting the census bureau's geographic base file efforts by encouraging the adoption of the CUE program by the various state, county, and local agencies. So far, the only DIME file in the region is in Nassau and Suffolk Counties, although New York City is developing a related program, and other counties and cities have indicated an interest.

## AS A USER OF CENSUS DATA

First, of course, the census data permit us to update parts of the 1963 surveys. The third-count population and housing unit block-level totals were summarized to the square mile by using our geographic converter file. We have devised a program to convert tract level summaries to square mile, but have not yet used it.

Our 1963 land use survey was the base for estimating major categories of land use for 1970 at the square mile. Comparison of 1963 and 1970 aerial photography and 1970 census housing unit counts were the new inputs. These estimates were then used in a land development allocation system to forecast housing, jobs, and population under conditions of maximum development. Housing data were also used to obtain replacement-development ratios for 10-year periods to the planned capacity of the region, to test the correlation of Tri-State's measures of accessibility against actual development between 1963 and 1970, and to obtain data-base consistency with state and subregional planning agencies.

Many data for updating the household part of the home interview survey (at the home site) were available, especially from the fourth count. For the travel portion, county-level demographic characteristics closely related to travel demands were studied and analyzed, and Table 35 from the fourth count has been converted to a county-to-county matrix (planning region in Connecticut). A computer program, which was developed from the 1963 survey, synthesized non-work-trip interchanges among counties based on 1970 demographic data.

The second travel analysis approach involves small-area (square-mile) data, which will be obtained from the Urban Transportation Planning Package (or our version of it), and I will return to that later.

An understanding of how Tri-State has used census data can be conveyed, I believe, by reports that have been completed: Housing Losses and Demolitions; 1970 Census Housing Statistics and Analysis of Housing in the Tri-State Region; Detailed Characteristics of the 1970 Population for the Tri-State Region; 1960-1970 County Migration Patterns; Regional Net Migration Patterns; Regional Population Forecasts, 1985 and 2000; Balancing Jobs and Housing in the Tri-State Region; Occupational Employment Projections for the Tri-State Region, 1985 and 2000; Income and Gross Rent Distributions among Non-Working Household Heads; Income Distribution of Households, 1985 and 2000; Trends in Regional Income Inequality; Floor Space per Employee Change, 1963 to 1970, Including Nonresidential and Employment Densities; A Decade of Change in the Journey to Work; Changes in Auto Availability during the 1960s; Fourth-Count Census Journey to Work; Transit-Oriented Age Groups; and The Extent of Public Water and Sewer Systems in the Tri-State Region.

To make travel forecasts, we need trip ends at the square mile for our models. Tri-State's traffic assignment for highway travel is obtained through use of the direct traffic estimation method, which uses a coded highway network, trip ends, and a decay function. Transit forecasts use the Fratar method for distributing central business district trips and the intervening opportunities model for distributing all other commercial transit trips. Transit traffic assignment for both methods is obtained through the assignment portion of the intervening opportunities model. We have designed 2 models, one for highway and one for transit, to estimate the remaining 1970 nonwork travel.

#### URBAN TRANSPORTATION PLANNING PACKAGE

Assuming that we would be able to obtain from the UTPP work trips coded to census tract and block, we requested in 1971 a cost estimate for the UTPP. Our zones were to be square miles in the urbanized area (to be obtained by using our converter file, which contains census block/Tri-State block equivalents), and the Universal Area Codes (UAC) for the entire region. Because the census bureau was not then prepared to provide cost estimates, we had some time to think about this request, and in March of 1972 we asked that our zones be quarter square miles in the Manhattan CBD, square miles for the remainder of the urbanized area, and MCDs for the entire region. At this point our difficulties began. The census bureau responded that

1. The large number of workers in the Tri-State region prohibited the use of the UTPP,
2. The place-of-work coding guide coverage was considerably smaller than we had realized, and
3. The estimated cost would be \$80,000.

Tri-State realized that it had responsibilities to other planning organizations in the region with regard to the form in which the UTPP would be requested. Therefore, Tri-State carefully reconsidered its approach to the UTPP and also discussed the matter with a "regional team" consisting of the Port Authority of New York and New Jersey, the Regional Planning Association, the New York City Planning Commission, the departments of transportation of the 3 states, and later the Nassau-Suffolk Regional Planning Commission and the Westchester County Planning Board, both of whom had indicated an intense interest in the UTPP.

Our discussions raised so many questions that we could not answer that we asked the census bureau to send representatives to New York City to meet with us. We learned a great deal from the representative who came, but also made a discovery that filled us with dismay. We learned that the UTPP was designed for an SMSA, that 16-digit coding (to tract and block) had been done in coding guide areas of all of the 11 SMSAs in the region, but that very little inter-SMSA coding had been done. For ex-

ample, none of the cross-Hudson detail coding was done for trips from New Jersey SMSAs to the New York City SMSA. All that was available was UACs. We were told that the census bureau had begun this coding but did not have sufficient funds to complete the job.

This was a blow to all of us—even more so to some of the other agencies that needed smaller work site zones than Tri-State did. Perhaps we should have realized this before, but none of us did. In the meantime, the census bureau informed us that to recode the work trips from the New Jersey SMSAs to the New York City SMSA would cost \$80,000. We began to feel that \$80,000 specials had been designed just for us!

Tri-State then took a long look at the UTPP and decided to ask the census bureau to provide for us a quite different product. We were spoiled with our home interview survey, which we could slice as and when we wished. We needed something more flexible than the UTPP, and so we developed what we call the Worker File. The census bureau agreed that the approach was possible and helped us work out details.

Described briefly, the Worker File is a series of tapes containing for each worker residing in the region the following record:

1. Primary or secondary earner (primary = highest earner)
2. Household relation (basic relation code)
3. Mode of travel
4. Size of household (number of persons)
5. Number of employed persons in household
6. Automobiles available to household
7. Earnings of individual
8. Household income
9. Occupation
10. Industry
11. Value of owner-occupied unit (1 family, detached, without business)
12. Other owner-occupied unit
13. Monthly gross rent
14. Number of units in structure
15. Age
16. Sex
17. Race [white, Negro, Puerto Rican (Spanish origin in Connecticut)], other
18. Class of worker
19. Hours worked (during week)
20. Years of school completed (highest grade attended + finished grade distribution)

The record should be coded to the finest geography possible, as follows:

1. Residence—state, county, MCD, place, tract, block
2. Work site—state, county, UAC, zip, tract, block

Because the records are confidential, these tapes will, of course, stay with the census bureau and be accessed by Tri-State and others at cost. Any sort of cross stratification of the variables in summary form can be requested, but limited, of course, by the bureau's rules for suppression.

After discussions of the need for recoding the inter-SMSA trips and the costs of doing so, arrangements were made with the census bureau (at a cost of \$5,000) to conduct a test of 3,500 randomly selected work trips from residents in Bergen, Morris, and Monmouth counties who worked in the New York City SMSA and were coded to a UAC. We needed to know the degree of success of coding to tract and block before we proceeded to spend the larger sum. Coding success in the New York City SMSA was, according to the bureau, more than 80 percent in New York City and lower for other areas. For this test, the regional team asked that it be allowed to code the problem cases—trips the bureau could not code to 16 digits. The bureau agreed, as long as the records were treated as confidential, which meant that team members had to be sworn in and to work on the bureau's premises.

The bureau found to its regret (and to ours) that, because of a serial numbering defect in its files, 15 percent of the records were not retrievable and could not be retrieved for any future recoding. This meant that even with perfect coding no better than 85 percent success could be achieved. The system we had set up worked very well. There were 604 problem addresses (75 percent in New York City), and the problem solving group was able to code 91.5 percent to the census tract (with 85 percent to tract and block) and 7.6 percent to UAC only. Only 5 addresses could not be coded at all.

Encouraged by the success of the test, Tri-State investigated possibilities for raising the \$85,000 finally estimated to do the larger part of the recoding. The Port Authority of New York and New Jersey agreed to pay one-third of the cost; Tri-State is paying the other two-thirds. The recoding effort will require 6 to 8 months to complete; the regional team will again code the problem addresses.

To obtain some measure of the cost to access such a file, we designed a county-to-county table of several cross-stratified data items about the traveler and his household. The bureau estimated the cost to be \$5,000 and warned us that a great deal of suppression would doubtless occur. To determine how much suppression, the bureau offered, at its expense, to run a test of this table on Rhode Island records, which were then being used for other forms of testing.

The table showed that all but 13 percent of the summary records was suppressed; however, that 13 percent represented 57 percent of the employed labor force. We then cut the number of data items, and the table was rerun. This time the results showed that 46 percent of the summary records was suppressed, but the remaining 54 percent represented 95 percent of the workers. Staff analysis of the test results provided what we believe to be a good set of measures for suppression that we can use in planning future requests. Copies of the analysis report are available.

The Worker File permits any agency using it to establish its own zones and specify its own data requirements. The regional team helped construct the list of data items and agreed to the scheme. At the request of the 3 state departments of transportation, all the residents of each entire state were included. The file was completed in July, and the estimated cost is \$9,000. Tri-State has already ordered from the Worker File a summary tape of tract-to-tract origin and destination by mode for the entire states of New York and New Jersey and the Tri-State region portion of Connecticut. (Connecticut had already ordered a town-to-town origin and destination table from the census bureau and stated that no smaller zones were needed.) Since mode is not a confidential data item, we will receive the tapes and be able to process them. The estimated cost to prepare the tapes is \$7,500.

The Worker File will, we hope, prove to be a rich source of data for many agencies in the region. It is undoubtedly the only source of detailed, regionwide travel data available to us until the 1980 census. In the use of the file, Tri-State will act in a coordinating capacity. We will assist agencies in making their requests to the census bureau for data from the file; we will coordinate requests and inform other agencies who may wish to obtain tabulations, sharing costs where possible. We will certainly share any data we obtain from the file, and the members of the regional team have indicated that they intend to do the same. Tri-State intends to order a county-to-county table with several variables cross stratified. We intend to process copies of the tract-to-tract origin and destination file by mode and send them to the states and county planning offices.

In spite of the lengthy process, the occasional cries of pain, and the feeling that we would never reach the goal, the end is at last in sight. We hope that what we learned along the way will be useful in planning for the future, and we are grateful to those in the Bureau of the Census who stayed with us until this point was reached.