Large Metropolitan Areas

Arthur B. Sosslau, *COMSIS Corporation*

The purpose here is to summarize uses of 1990 census data along with user experiences and recommendations for the year 2000 as reported in the case studies for large metropolitan areas presented at this conference. These case studies tend to be different in terms of emphasis, which makes them all the more interesting. (Case studies may be found in Volume 2 of these proceedings.)

Brooks and Bandy of the Baltimore Metropolitan Council (BMC) describe the serious inconsistencies between the originally released Census Transportation Planning Package (CTPP) and BMC's own place-of-work employee tabulations. On the positive side, BMC was able to work successfully with the Census Bureau to produce an acceptable revised version of the CTPP Urban Element. Zakaria of the Delaware Valley Regional Planning Commission (DVRPC) provides a review and evaluation of the CTPP and describes its accuracy and uses. Christopher of the Chicago Area Transportation Study (CATS), Sööt of the University of Illinois, and Stuart of the Chicago Transit Authority prepared a discussion that includes information about CATS, small metropolitan planning organizations (MPOs), the transit community, and the research and university community. This summary will concentrate on reporting the uses of census data at CATS. Limoges of the Southeast Michigan Council of Governments in Detroit reports on an innovative method to assign land use classes to census small-area employment data. Purvis of the Metropolitan Transportation Commission reports on the use of census data in the San Francisco Bay Area. His case study serves as an update on the resource paper he prepared for the 1994 conference.

Endorsements of Census Data

First, some of the endorsements of the census data presented in the case studies will be summarized here. From the Baltimore case study, many of the data tables contained in the CTPP are "one-of-a kind" tabulations that are nearly indispensable in updating trip tables and other components of travel demand modeling. Unless a massive household and travel diary survey were conducted that would be statistically valid at a very small unit of analysis (which

would be astronomically costly), the CTPP alone provides critical data that could not be easily obtained elsewhere.

Zakaria cites the census data for air-quality and transportation planning, travel forecasting, economic base and employment location studies, urban development analysis, and planning and evaluation of transit services. The CTPP minimizes the need for large-scale data collection and decreases the rising costs of surveys required. Under current budget conditions, it is almost impossible to conduct a home interview survey that would provide results similar to those included in the CTPP.

Christopher et al. find that CATS has a substantial history in the use of the planning packages: when CATS first received the CTPP, the data were examined and checked against other local data. CATS was satisfied with the results of the validity check.

Limoges sees the decennial census as by far the single most important source of information on employed persons and jobs and their interrelationships.

Finally, Purvis commends the decennial census data as an independent, observed estimate of various demographic characteristics and travel behavior for many applications. The census long form could be replaced by national or local surveys, but probably at a higher unit cost with lower sampling rates and higher statistical variance and standard errors.

Putting these accolades aside, it would be well to discuss some of the problems with the 1990 census data.

Problems with Census Data

Zakaria found problems in his review of Parts 1, 2, and 3 of the CTPP. All trips were not allocated to transportation analysis zones (TAZs) because the Topological Integrated Geographic Encoding and Referencing (TIGER) file does not contain address ranges for some suburban and rural areas. The DVRPC found the format of the tapes complex and confusing. There was no labeling, and table names were puzzling. There was no documentation of certain record types. As for the data, worker trips by mode included some walk and railroad trips that were unrealistic in terms of travel time or distance. The evaluation of employment by industry showed that some respondents misunderstood the question that used the Standard Industrial Classification (SIC) codes.

The Southeast Michigan Council of Governments (SEMCOG) also found geocoding errors and allocation inaccuracies, which they corrected before use of the data.

The lack of reliable commuting characteristics could have forced the BMC to conduct a costly travel survey. Instead, working with FHWA, BTS, and AASHTO, the Census Bureau produced a revised Urban Element. The BMC's Geographic Base File/Dual Independent Map Encoding (GBF/DIME) file was licensed by the Census Bureau in 1984 as the basis for the TIGER file. However, the Census Bureau never obtained updates after the initial purchase. New streets added after 1984 were never incorporated into the 1990 TIGER file. The Census Bureau's efforts were hampered despite the fact that the BMC created and transmitted to the Census Bureau an Employer Workplace Coding File in 1988 to assist in identifying the location of major employers. Before releasing the data, the Census Bureau informed the BMC that address range problems, primarily in Carroll and Harford counties, had affected small-area coding and that default TAZs had to be created to capture this missed information. The lack of coding for these counties created small-area undercounts that affected 44 and 30 percent of their employment bases. Further comparisons revealed 23 percent small-area undercounts in Baltimore County and pervasive small-area employment differences throughout the region. The BMC staff embarked on an investigation of small-area employment in activity centers in each suburban jurisdiction, which gave an indication as to whether misallocations might be attributed to boundary discrepancy problems or whether they were symptoms of a much larger problem.

Once the Census Bureau decided that it would revise the apparent misallocations, the BMC provided reference materials and tabulations to assist the process. The Census Bureau stated that it could not "re-geocode" census records based on a later and more accurate BMC

JURISDICTION	ORIGINAL CTPP URBAN ELEMENT RELEASE (1994)	REVISED CTPP URBAN ELEMENT RELEASE (1995)
ANNE ARUNDEL COUNTY	214,599	216,860
BALTIMORE CITY	396,360	395,483
BALTIMORE COUNTY	318,654	318,597
CARROLL COUNTY	26,451	40,024
HARFORD COUNTY	50,191	65,724
HOWARD COUNTY	88,995	88,512
BALTIMORE REGION	1,095,250	1,125,200

FIGURE 1 CTPP revision comparisons by jurisdiction, Baltimore region.

BaseMap. Census was informed that serious street coverage and address range problems continued to plague the 1990 and post-1990 TIGER files. The Census Bureau was urged to "regeocode" 1990 records using a more accurate street and address reference file. The Census Bureau could not do so citing a lack of proper computer software and hardware. Because of confidentiality statutes, the Census Bureau could not allow the BMC staff to geocode these records. Thus, the Census Bureau made the adjustments using a combination of automated and manual allocation techniques. The reallocations of small-area employment resulted in dramatic changes in total employment as shown in Figure 1.

Uses of Census Data

Figure 2 summarizes the uses of census data reported in the five case studies. They are the ones that have been mentioned in the past. Most probably occur in all five metropolitan areas, but Figure 2 indicates which ones are specifically mentioned in each case study.

Some of the special studies shown in Figure 3 are interesting, as described next. The first application of interest is the use of 1990 census Public Use Microdata Sample (PUMS) data as part of the Bay Bridge Congestion Pricing Demonstration Project. MTC consultants were able to extract Bay Bridge commuters on the basis of PUMS area of residence and county of work, which allowed MTC to understand their income and modal use characteristics to determine, for example, who would be affected by a toll increase during peak travel times.

An example of a transit application is recent MTC work with the Central Contra Costa Transit Authority on a geographic information system (GIS)–based analysis of transit-

	MTC	DVRPC	BMC	CATS	SEMCOG
Descriptive Analysis - Trends, Database	X	X		X	
Model Estimation	X	X	X	X	
Model Market Segmentation	X				
Model Validation	X	X			
Highway/Transit Corridor Studies		X			
Outreach - Public/Private	X	X			
Input to Land Use Models					X
Socioeconomic Forecasts			X		
Transportation Research	X				X
Special Studies	X	X		X	

FIGURE 2 Uses of census data.

	MTC	DVRPC	CATS
Bay Bridge Congestion Pricing Demonstration	X		
Profile of Those Working at Home	X		
Analysis of Transit Dependency	X		
Evaluate Change in Location of Industry/Commercial		X	X
Evaluate Declining Urban Centers		X	
Potential Users of Downtown Circulator System			X
Establishing HH Travel Surveys			X
Intercity Jobs Accountability			X

FIGURE 3 Examples of special studies.

dependent population in the service area. One of the layers in MTC's GIS is local bus stops and rail stations. The analyst used the GIS to create a buffer zone around each bus stop to represent areas within a certain walking distance. The GIS program then separates demographic data within and outside the buffer zone.

The 1990 CTPP was used in two different studies related to the CATS 1990 Household Travel Survey, first, to help establish the weights for each survey instrument and, second, to establish a model to estimate the nonresponse rates in a mail-out–mail-back surveying procedure.

DVRPC used the 1990 CTPP information on employment to evaluate the significant changes in the type and location of industries and commercial establishments. This evaluation resulted in recommendations and strategies aimed at attracting new industries and high-technology firms to the Delaware Valley. The employment information was also useful in the redevelopment of declining areas of old urban centers and provision of the required physical improvements for their rehabilitation.

Limoges of SEMCOG presented a method for adding land use classes to decennial census employment data. The current version of the land use assignment procedure classifies workers by small-area place of work into six basic land use classes: office; commercial; institutional; industrial; transportation, communications, and utilities; and residential. Before applying the land use assignment method to the special tabulation data, SEMCOG staff conducted a separate project whose purpose was to make improvements to the census data. The improvements addressed geocoding errors and allocation inaccuracies.

Upon examination of the data, the overall quality of tract and block geocoding appeared to be quite good. There were relatively few recognizable major errors, and these were corrected. In the four most urban counties of Southeast Michigan, an average of over 30 percent of all workers needed to be allocated to tract and block. In Detroit City, nearly two-thirds of the tracts had more than 40 percent of their workers allocated by the Census Bureau. SEMCOG believed that the allocation procedure seriously reduced the overall accuracy of the small-area employment data and developed their own reallocation procedure. That procedure accepted the Census Bureau's geocoding to county and to place, and then used the special tabulation's detailed breakdown of industrial class and occupational class to match workers needing to be reallocated to zone with workers who had been geocoded to tract and block and thereby to zone by the Census Bureau. SEMCOG's reallocation greatly increased the accuracy of the zone-of-work geocoding.

SEMCOG next grouped the Census Bureau's 236 industrial classes into 74 and the 501 occupational classes into 39 to develop a matrix. The next task was to assign a land use class to each cell, wherever possible. First, they addressed each industrial class, for example, finance or hospital, where they believed it was justified in assigning all employment of that industrial class to a single land use class. Next, for each occupational class that was assigned entirely to one land use class, all cells in that occupational class column of the matrix would

be assigned to that land use class except where the cell had already been given a land use class because of its industrial class. For example, the occupation of computer programmer was assigned to office land use except for cells belonging to an industrial class, for instance, hospitals, that already had an overall land use class, in that case, institutional. The third step assigned a land use class to each matrix cell that was yet unassigned but that could be assigned a land use class on the basis of the characteristics of that particular combination of industry and occupation. Motor vehicle mechanics and repairers is an example of this cell-by-cell assignment. The remaining cells of the matrix had no assigned land use class.

The matrix was then used to assign employment to land use classes by TAZ. Within each industrial class, the employment in each cell for which a land use class had been assigned was summed by land use class, and the plurality land use class was identified. All employment in the given industrial class, including that in cells for which a land use class had not been assigned in the general matrix, was then reassigned to that plurality land use class. The special tabulation made it possible to assign the plurality land use class of the given industrial class of the given zone to workers in that industrial class in that zone. Land use class became an additional dimension of the cross-tabulation.

The one quantitative comparison made to date is with data collected in SEMCOG's 1994 household travel survey. The two data sets are compared in Figure 4. Considering the differences in data collection method, date, and coverage area, the two sets of numbers are quite close.

Zakaria describes some of the adjustments made in the census data before their use by DVRPC. First, he found that the data on population, households, car ownership, employment, and other socioeconomic characteristics from Part 1 are quite accurate and do not require any major adjustment. Adjustments were made in the CTPP work destinations to account for absentees (2.16 percent for the region) and multiple job holding (6.2 percent). Adjustments at the Minor Civil Division (MCD) level were made to account for coding discrepancies and respondent errors. Employment estimates at the TAZ level were adjusted because of the census allocation to default zones and water tracts. Total regional trips by mode compared favorably with traffic counts and transit surveys. Within smaller areas differences were much larger, with a difference in subway-elevated and bus trips in the central business district of 35 percent. It appears that many respondents confused the access mode to a station with the principal mode. The 1990 CTPP average regional travel time compared very well with DVRPC survey data (24.6 versus 23.8 min).

Recommendations for Census 2000

In each of the five case studies, recommendations were made for Census 2000. Purvis (MTC) states that the long form is critical to provide the accurate and precise data needed to support demographic analysis and transportation planning and research activities. The likely substitute would be a set of metropolitan travel surveys that would be more costly and less accu-

Employment (%) by Land Use Class

Land Use Class	1990 Census	1994 Travel Survey
Office	36.0	33.4
Commercial	23.8	19.8
Institutional	17.3	18.8
Industrial	15.6	18.8
TCU[a]	3.0	2.4
Residential	4.3	6.9
Totals	100.0	100.0

FIGURE 4 Comparison of employment by land use class, 1990 census and 1994 SEMCOG household travel survey.

rate. Purvis also suggests that workplace geocoding is still a major issue and that legal barriers that limit the involvement of local planning staffs should be liberalized. In order to increase the relevance of transportation planning research, it would be desirable to create a census microdata research program that would allow bona fide researchers the opportunity to "add value" to census microdata and prepare more in-depth research. To collect, analyze, and disseminate decennial census data, rapid changes in information systems and information technology should be dealt with along the lines that the Census Bureau has planned. New information technology should lessen the need for "paper-and-ink" publications in favor of electronic data on demand, and public access to the Internet should be a high priority to facilitate collection and dissemination of census data.

Zakaria (DVRPC) suggests that most of the 1990 problems and errors can be avoided in 2000 by quality control edits and a careful review of the census questionnaire as well as the computer formats and programs required for processing the information. Specifically, the journey-to-work questions should be simplified to prevent any confusion on the part of those responding to questions on mode of travel, destination, and industry classification. The questionnaire should be redesigned to capture multimodal trip information. The format of the 1990 CTPP tapes must be simplified. The funding and development of two packages in 1990 was an excellent idea and should be repeated in 2000. AASHTO should again provide the funding for the 2000 CTPP. Finally, DVRPC has not as yet received all parts; a more timely release of data is obviously important to all census data users.

On the basis of their experience, Brooks and Bandy (BMC) make the following recommendations: the Census Bureau needs to maintain the most up-to-date TIGER files and should continue using regional workplace coding; MPOs should prepare data bases to check and validate census data.

CATS recommendations for simplification and timely release parallel those of DVRPC. SEMCOG's recommendations are related to their case study on land use coding. Census place-of-work data would be improved through the correction of geocoding errors and the use of a new procedure to allocate ungeocodable workers. The Census Bureau would give each worker a workplace land use class and would incorporate this land use attribute into a variety of census files and products.