Uses of Census Data in Transportation Planning: San Francisco Bay Area Case Study

Charles L. Purvis, *Metropolitan Transportation Commission*

This case study is an update of a resource paper prepared for the 1994 Conference on Decennial Census Data and Transportation Planning. It focuses on the uses of census data in transportation planning activities in the nine-county San Francisco Bay Area. Attention is paid to the use of decennial census data in various planning analysis activities, including general descriptive analyses, estimation of disaggregate and aggregate travel demand models, market segmentation in travel demand model forecasting systems, and the validation of demographic and travel model simulations. The discussion covers where the census data are critical for the application and where the census data are desirable but perhaps not required for the application. The various census products that are used in the Bay Area—the standard Summary Tape Files, the Public Use Microdata Sample, the Census Transportation Planning Package, and special tabulations—are discussed within the context of the various planning analysis activities occurring in the Bay Area. Recommendations and expectations for Census 2000 are provided.

In this case study the uses of census data in transportation planning activities in the nine-county San Francisco Bay Area are discussed. From an institutional perspective, this covers the uses of census data in the Metropolitan Transportation Commission (MTC), a metropolitan planning organization (MPO); the Association of Bay Area Governments (ABAG); state agencies such as the California Department of Transportation (CalTrans); various transit operators in the region; and local county and city planning, public works, and congestion management agencies. The primary focus is on census activities at MTC and ABAG, with reference to other creative work under way at the county and transit operator levels.

This study also serves as an update of a resource paper (*1*) for the 1994 Conference on Decennial Census Data for Transportation Planning that examined the use of census data in several major metropolitan areas, including the Bay Area, using published reports covering 1970 through 1994. This paper will focus on Bay Area applications as well as new uses of census data in the region between 1994 and 1996.

In terms of census products used in the Bay Area, the case study will cover the use and application of data from standard Census Bureau products such as the Summary Tape Files (e.g., STF1A, STF3A); the Public Use Microdata Sample (PUMS); special products from the Census Bureau and the U.S. Department of Transportation such as the Census Transportation Planning Package (CTPP), including the Statewide Element (CTPP/SE) and the Urban Element (CTPP/UE); and special Census Bureau data files purchased by the MTC and the Santa Clara County Center for Urban Analysis.

Following this introduction, the application of census data in several categories will be reviewed, including descriptive analysis, model estimation, market segmentation for travel forecasting systems, model validation,

and miscellaneous transportation applications. The paper concludes with a set of recommendations and expectations for Census 2000. This last section is essentially a strategic assessment of the decennial census in terms of the strengths and weaknesses of census data, the opportunities for improvement, and the danger of not getting what is needed to maintain analytical tools and data bases for transportation planning.

Descriptive Analysis

In the context of this paper, the term "descriptive analysis" refers to the reports, working papers, summary data files, spreadsheets, maps, press releases, trend reports, newsletters, and so on, related to the dissemination of information from the decennial census. In the Bay Area these analyses have traditionally been in the form of place-level or county-level profile reports (all the information from certain census files, say, STF1A, STF3A, or CTPP/SE) or working papers (more in-depth discussion and trend analysis of census data).

Information systems and technology is a rapidly evolving field, and it is apparent that a new era of data dissemination is also evolving. Although in the past the most common means of dissemination was the hard copy or "dead tree-and-ink" issuance of census data reports, the future (and current) nature of data dissemination involves online, perhaps even real-time provision of data needed in planning activities.

A good example of the use of new information technology for data dissemination is the World Wide Web (WWW) page developed by ABAG (Figure 1), which allows the user to pick any of some 140 places in the Bay Area and any of a set number of topical reports to pull up predeveloped profile reports. This rather straightforward WWW interface with choices such as "Pick Place," "Pick Topic," and "View Data" may likely be superseded in the years to come with more elaborate data-on-demand query-and-display setups that provide the data analyst with exactly those data needed for the analysis at hand.

The issue of access to computers and the Internet is in part addressed by the development of the Public Access Network (PAN), an example of which is PAN Islands, hosted by government agencies such as MTC and ABAG and sponsored by a consortium of private-sector companies in the Silicon Valley (e.g., Smart Valley, Inc., Pacific Bell, 3COM, Yahoo, Surf-Watch, and Arthur D. Little) (http://www.svi.org/PROJECTS/PAN/ISLND). The principal actors involved in public access to the Internet are local public libraries, many of which receive funds from the federal Library Services and Construction Act. Public access to the Internet via the public

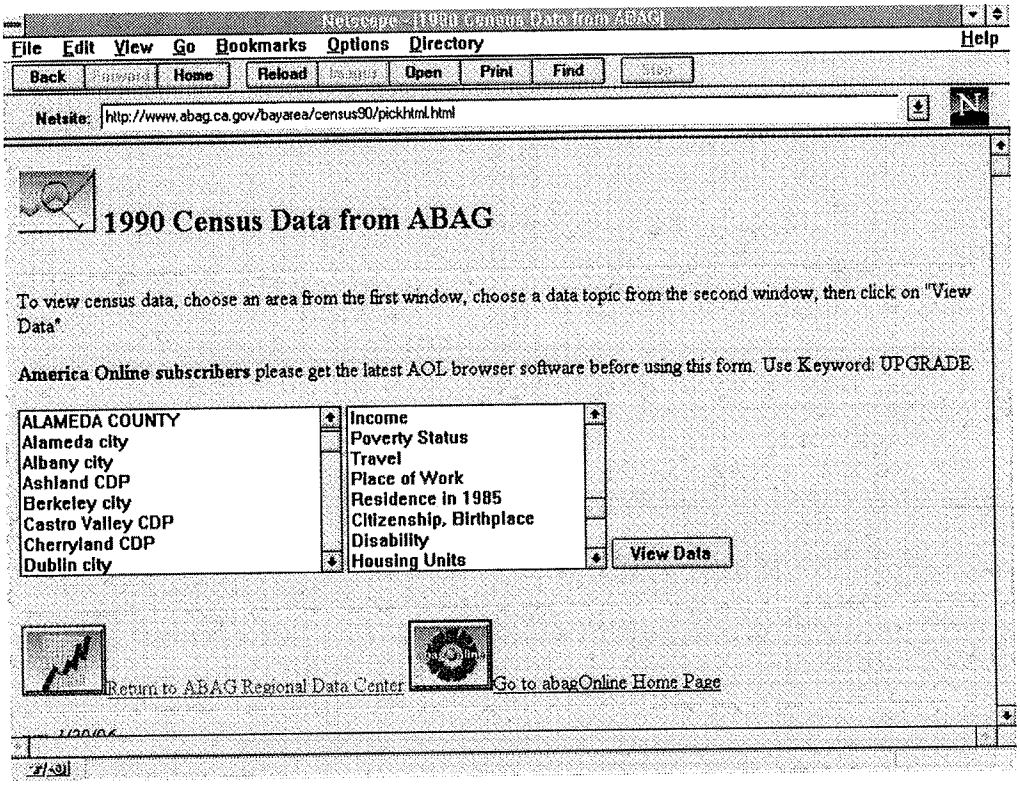

FIGURE 1 Association of Bay Area Governments World Wide Web page: 1990 census data.

library system is critical to ensure universal access to census information.

Many clients and patrons still prefer the paper-and-ink versions of census reports. In the Bay Area, this need is fulfilled by a comprehensive set of profile reports developed by the regional data center for the Bay Area, ABAG, and a set of transportation-related working papers authored by the staff at the MTC. A list of these MTC 1990 census working papers is provided in Table 1 and in the references (*2–12*).

As alternatives to the hard-copy working papers disseminated by MTC, electronic versions of these working papers and spreadsheets as well as special electronic publications (*13–15*) have been developed (Table 2). MTC Electronic Publications 1 and 2 provide comparisons of place-to-place commuters using data from the 1980 census Urban Transportation Planning Package (UTPP) and the 1990 CTPP. Electronic Publication 3 is MTC's version of area-level profile reports using place-level data from the CTPP/SE. These place-level profiles were only issued in electronic format, giving a 13-page profile report for 140 places in the Bay Area, for a grand total of 1,820 pages of information included in 7.9 megabytes of data files.

The need and value of these types of descriptive census reports, electronic or otherwise, should not be underestimated. They serve a vital role in informing the public, the media, policy makers, and planning professionals on many of the most relevant demographic and travel characteristics needed for public policy development. The census working papers have been some of the best-received products issued by MTC and have been of great assistance to librarians and the public information staff at MTC and throughout the Bay Area.

Looking toward the future, it is unlikely that MTC or others will completely abandon the practice of issuing major paper-and-ink reports on census data. It is highly probable, on the other hand, that the rapid changes in information technology will greatly enhance the ability of planners to electronically disseminate and exchange census and other planning data bases with clients and partners.

Model Estimation

The previous resource paper (*1*) provides a fairly detailed discussion on the use, or potential misuse, of census data for estimating demographic and travel models. New applications and insights for the period from 1994 to 1996 will be reported on here.

It bears repeating that the best sources of data for the estimation of travel demand models are household travel surveys. Survey data are essential, disaggregate data that should provide the transportation planner with the necessary input for the estimation and calibration of demographic and travel demand models. Decennial census data can, at best, be used in the estimation of household-level models to predict number of workers in the household or automobile ownership levels (*16*), or perhaps in the estimation of other types of aggregate travel models. For model validation, on the other hand, decennial census data form a critical and invaluable data base for the aggregate validation of various demographic and travel behavior models.

Two new examples of use of census data in model estimation are efforts at MTC in estimating aggregate trip-end mode share models and work in progress

TABLE 1 Metropolitan Transportation Commission 1990 Census Working Papers: 1992–1996

Working Paper No.	Data Source	Date	Title
1	STF1A	April 1992	Bay Area Population Characteristics
2	STF3A	August 1992	Bay Area Travel and Mobility Characteristics
3	STFS-5	December 1992	County-to-County Commute Patterns in the SF Bay Area
4	STFS-5	January 1993	SF Bay Area Interregional County-to-County Commute Patterns
5	CTPP/SE, Part C	April 1993	The Journey-to-Work in the SF Bay Area
6	STF3A	October 1993	Disability, Mobility Limitation and Self-Care Limitation Status
7	CTPP/UE, Part 3	March 1994	Detailed Commute Characteristics in the SF Bay Area
8	CTPP/UE, Part 3	May 1994	Detailed Interregional Commute Characteristics
9	CTPP/SE, Part A,B	September 1994	SF Bay Area: County & Regional Profiles
10	CTPP/UE, Part 4	April 1995	SF Bay Area Detailed Household Characteristics
11	STP-214	January 1996	SF Bay Area Commuters by Household Income Characteristics

TABLE 2 Metropolitan Transportation Commission 1990 Census Electronic Publications: 1992–1996

Electronic Pub. No.	Data Source	Date	Title
1	CTPP/SE; UTPP	April 1993	Bay Area Place to Place Journey-to-Work Characteristics
2	CTPP/SE; UTPP	March 1994	Bay Area Place to Place Journey-to-Work Spreadsheets
3	CTPP/SE	September 1994	SF Bay Area: Place-Level Profiles

at the Santa Clara County Center for Urban Analysis (SCCCUA) in estimating aggregate logit work-destination-choice models stratified by household income quartile.

Trip-End Mode Share Models for Predicting Bicycle and Walk Commuters

One set of new trip-end mode share models developed at MTC are aggregate regression models estimated on zone-level shares of bicycle and walk commuters at the zone of residence and the zone of work (*17*). The models predict the percentage of workers who commute via bicycle or walk modes based on aggregate zonal characteristics such as employment density, share of multifamily dwelling units of total units, local workers-job balance within the traffic analysis zone, and dummy variables to reflect proximity to the university campuses in Stanford and Berkeley. Typical mode share models are estimated using zone-to-zone network levels of service and related demographic and land use characteristics. Typical mode share models are either logit in form or of the "diversion curve" style that was popular in the 1960s. The reason behind the development of these atypical, aggregate zone-level trip-end-based models is concern that there are too few sample walk and bicycle trips in the 1990 MTC household travel survey. Of the 18,300 sample total home-based-work trips in that survey, only 478 (2.6 percent) were by the walk mode and just 222 (1.2 percent) were by bicycle. Such a small and sparse data set on walk and bicycle commuters may prove a challenge in the model estimation process. The final decision on whether to use aggregate trip-end mode share models or a disaggregate trip interchange mode choice model depends on current work in progress at the MTC to estimate best-practice nested work-trip mode choice models.

Aggregate Work-Destination-Choice Models

Another example of travel model estimation using census data is the ongoing effort at SCCCUA to estimate aggregate logit work-destination-choice models. These logit destination-choice models are estimated separately by household income quartile using data from a special census file purchased from the Census Bureau by MTC and SCCCUA. This file, denoted STP-214 (Special Tabulation Product) by the Census Bureau, provides one cross-tabulation of block-group-to-block-group workers (within the nine-county Bay Area) stratified by 12 categories of means of transportation to work by four household income levels (annual household income less than $25,000; $25,000 to $45,000; $45,000 to $75,000; and greater than $75,000). MTC's analysis of this special tabulation product is included in Census Working Paper 11 (*12*). Work at SCCCUA on these aggregate work-destination-choice models will be completed in 1996.

Comparisons of regional average and median trip length (in miles) and average and median commute duration (in minutes) is shown in Table 3; these data show a notable increase in work commute duration and length with increasing household income levels, an indicator that supports the notion for an income-stratified work trip distribution (destination) choice model.

Travel Model Market Segmentation

One of the challenges in travel demand forecasting is the use of disaggregate travel demand models in the aggre-

TABLE 3 Mean and Median Average Commute Length and Commute Duration by Household Income Quartile, 1990 Census, Special Tabulation Product 214: Regional Totals for San Francisco Bay Area

Income Quartile	Median Distance (mi)	Mean Distance (mi)	Median Time (min)	Mean Time (min)
< $25,000	5.87	9.50	14.8	18.0
$25,000–$45,000	7.71	11.57	16.8	20.2
$45,000–$75,000	9.33	13.13	18.2	21.8
> $75,000	9.94	13.46	18.9	22.5
TOTAL	8.58	12.37	17.6	21.1

gate prediction of travel behavior. Methods of aggregation fall into three principal categories: the "naive" method, market segmentation, and sample enumeration (*18*). In a nutshell, the naive method assumes that everyone in a traffic analysis zone has the same characteristics: for example, each household has the same number of workers, the same income, and the same number of persons. Zonal mean values are used exclusively in this method.

The market segmentation method assumes that there are distinct subgroups within each traffic analysis zone, for example, households by automobile ownership level, by workers in the household, or by household income level. In the market segmentation method, the analyst assumes that the average household size, the average workers per household, and so forth, is the same within each subgroup by each zone.

The third aggregation method, sample enumeration, does not use any group or subgroup mean of any input variable. Instead, the disaggregate model is applied at the disaggregate (i.e., household, person, or trip) level, and then the predictions are aggregated for reporting purposes. The sample enumeration technique is also known as microsimulation, where the forecasting of travel or other activity behavior is made at the discrete individual level rather than at the zone level.

At the MTC, use of market segmentation is a key feature of the aggregate forecasting system in place and under redevelopment (*19*). Also, sample enumeration is used at MTC for special analyses, such as evaluating the effectiveness of transportation control measures (*20*).

To apply disaggregate models in a market segmentation framework, analysts at MTC have used the 1980 and 1990 census Public Use Microdata Sample (PUMS) data bases as supplementary inputs to the model application process. Census PUMS data are critical because they provide information the analyst needs to adjust the input model parameters (e.g., household size, household income, percent multifamily) by the desired market segmentation. PUMS data are used because standard census products such as the STF3A or the CTPP do not provide the necessary data at the traffic analysis zone level or any other geographic level.

MTC has a nested workers-in-household–automobile-ownership (WHHAO) choice model. This model splits the households residing in a traffic analysis zone into households by three levels of workers in the household (0, 1, 2+ workers/HH) and by three automobile ownership levels (0, 1, 2+ vehicles/HH). The input market segmentation to the WHHAO model is households by household income quartile. This means that the outputs of the WHHAO model application are the number of households in each traffic analysis zone stratified by household income (four), by workers in the household (three), by automobile ownership level (three), or 4 market segmentations into the model and 36 market segmentations coming out of the model application.

One of the input variables to the WHHAO model is average household size. Rather than zonal average household size, the census PUMS data are used to adjust zonal average household size to zonal average household size stratified by household income level because low-income households (less than $25,000 per year) are smaller in size than higher-income households. PUMS data are used to develop county-level (or PUMA-level) adjustment factors, which are then multiplied by the zonal average household size to yield zonal average household size by household income quartile. County-level household size by income quartile and adjustment factors are summarized in Tables 4 and 5.

An example calculation follows: the North Beach traffic analysis zone in San Francisco has an average household size of 1.719 persons. Using the San Francisco County adjustment factors, it is estimated that the average low-income (Quartile 1) household in North Beach has an average household size of 1.361 persons (1.719 * 0.792); the average medium low-income (Quartile 2) household has an average household size of 1.709 (1.719 * 0.994); the average medium high-income household has an average household size of 2.020 (1.719 * 1.175); and the average high-income household in North Beach has an average household size of 2.228

TABLE 4 Average Household Size by Household Income Quartile by Bay Area County, 1990 Census Public Use Microdata Sample, 5 Percent Sample

County	Income Q 1	Income Q 2	Income Q 3	Income Q 4	Total
San Francisco	1.808	2.269	2.682	2.958	2.283
San Mateo	1.905	2.452	2.965	3.216	2.644
Santa Clara	2.185	2.593	3.067	3.341	2.816
Alameda	2.034	2.487	3.008	3.239	2.582
Contra Costa	2.043	2.428	2.950	3.170	2.638
Solano	2.261	2.927	3.231	3.417	2.867
Napa	1.906	2.547	2.974	2.992	2.495
Sonoma	1.874	2.616	3.050	3.098	2.543
Marin	1.713	2.171	2.526	2.796	2.333
REGION	1.997	2.492	2.968	3.196	2.610

TABLE 5 Household Size Adjustment Factors by Household Income Quartile by Bay Area County, 1990 Census Public Use Microdata Sample, 5 Percent Sample

County	Income Q 1	Income Q 2	Income Q 3	Income Q 4	Total
San Francisco	0.792	0.994	1.175	1.296	1.000
San Mateo	0.721	0.927	1.121	1.216	1.000
Santa Clara	0.776	0.921	1.089	1.186	1.000
Alameda	0.787	0.963	1.165	1.254	1.000
Contra Costa	0.774	0.920	1.118	1.202	1.000
Solano	0.789	1.021	1.127	1.191	1.000
Napa	0.764	1.021	1.192	1.199	1.000
Sonoma	0.737	1.029	1.199	1.218	1.000
Marin	0.734	0.931	1.083	1.199	1.000
REGION	0.765	0.955	1.137	1.225	1.000

persons (1.719 * 1.296). Thus, the subgroup mean household size in North Beach ranges from 1.361 to 2.228 persons. The naive method would just use the zonal mean household size of 1.719.

The adjustment factors used in the market segmentation process could also be developed using data from local household travel surveys. The problem will be the reliability of these factors based on typically too few sample observations in the small-scale household travel survey. For example, the 1990 Bay Area household travel survey provides sample data on 10,800 households. The 1990 census PUMS 5 percent sample includes disaggregate data on 108,500 Bay Area households. A valuable research project would be to calculate the standard errors of these adjustment factors comparing regional travel surveys, such as the 1990 Bay Area travel survey, with the 1990 census PUMS-based adjustment factors.

Demographic and Travel Model Validation

One of the basic uses of census data is for the aggregate validation of demographic and certain travel demand models. *Validation* is the process of comparing predicted values with "observed" values and making the necessary adjustments (calibrations) to each of the component models to produce a valid model simulation. The decennial census data serve a most valuable purpose as an independent, observed estimate of various demographic and travel behavior characteristics. The following is a list of various uses of census data in demographic and travel model validation:

1. Workers in household models can be validated against CTPP/UE data at the regional, county, superdistrict, district, and traffic analysis zone levels.
2. Automobile ownership level models can be validated against CTPP/UE data at the regional, county, superdistrict, district, and traffic analysis zone levels.
3. Analysts can convert the zone-to-zone commuter matrixes, derived from Part 3 of the CTPP/UE, into observed home-based-work person trips using trips-per-worker conversion factors. Home-based-work trip generation (production) and trip attraction models can be validated against these CTPP-derived observed home-based-work trip tables at the regional, county, superdistrict, district, and traffic analysis zone levels.
4. Home-based-work trip distribution models can be validated against CTPP-derived observed home-based-work person trips. In the Bay Area, the census STP-214 file can be used to calibrate and validate work trip distribution models stratified by household income quartile. The validation process for work trip distribution models should include a careful evaluation of model versus observed average trip lengths (and average trip duration) at the regional, county, superdistrict, and perhaps zonal levels; evaluation of regional trip length and trip duration frequency distributions; and evaluation of county-to-county, superdistrict-to-superdistrict, and district-to-district predicted versus observed home-based-work person trips.
5. Home-based-work mode choice models can also be validated against CTPP-derived observed home-based-work person trip tables by means of transportation. Ideally, a multiday household travel survey or employer survey is used to adjust the "usual means of transportation" typology into an "average means of transportation."

The decennial census data cannot be used for the validation of nonwork trip frequency, nonwork trip destination, or nonwork mode choice travel demand models. Given that nonwork trips may typically encompass 75 percent of a large region's travel, it is imperative that a suitably sized household travel survey be on hand for the aggregate validation of nonwork travel demand models. Also, it is very useful to have two competing sets of "observed" home-based-work data. This gives the analyst flexibility in what he or she should be validating against. If the two observed data bases (census and survey) are in

agreement over a certain statistic, that gives the analyst a general indication of confidence, or lack of confidence, in a particular census-based or survey-based estimate of work travel behavior.

Miscellaneous Transportation Applications

Several new applications of census data for miscellaneous transportation planning and transportation research activities in the Bay Area are worth reporting on. The first application of interest is the use of 1990 census PUMS data as part of the Bay Bridge Congestion Pricing Demonstration Project. Funded by FHWA, this demonstration project included an analysis of the demographic characteristics of Bay Bridge commuters. MTC consultants were able to extract these commuters on the basis of Public Use Microdata Area of residence and county of work. The analysis enabled MTC to understand the income and modal usage characteristics of Bay Bridge commuters in 1990 (*21*), or, in other words, who would be affected by a toll increase during peak travel times.

Census PUMS data have also been used by MTC staff to produce a demographic profile of persons working at home in Marin County (*22*). Analysis focused on the industry, occupation, earnings, sex, age, years of schooling, and hours worked "last week." The typical work-at-home person in Marin County is female and highly educated, works part-time, is self-employed, is older, and earns less than workers who commute outside the home. Male work-at-home worker earnings are 140 percent higher than female work-at-home worker earnings ($42,500 male versus $17,500 female). The predominant industry for work at home in Marin County is real estate, management and public relations, and professional and business services. The predominant occupation of work at home in Marin are writers-artists-entertainers and managers-administrators. The work-at-home commute share in Marin ranges from a low of 1.4 percent for government workers to a high of 25.1 percent of self-employed Marinites.

The above examples show but two of the many applications that the creative transportation planner and analyst can produce using the census PUMS data sets, which constitute one of the best transportation research data bases offered by the Bureau of the Census.

An example of transit applications using census data is recent MTC work with the Central Contra Costa Transit Authority (CCCTA) on an analysis based on a geographic information system (GIS) of the transit-dependent population in the CCCTA service area. One of the layers in MTC's GIS is local bus stops and rail stations. The analyst then used the GIS to create a buffer zone around each bus stop to represent areas within a certain walking distance of the bus stop. The GIS program then uses a "cookie cutter" technique to split out demographic data within and outside the buffer zone. Demographic variables such as zero automobile households, population aged 62 and over, and nonworking households were used as measures of transit dependency. This technique can provide the transit market analyst with the demographic characteristics of all persons residing within walking distance of each of the CCCTA routes as well as the characteristics of those not within walking distance of any of the system routes. Similar GIS efforts are under way at other Bay Area transit agencies, including the Bay Area Rapid Transit, Golden Gate Transit, and SamTrans.

Academic researchers at the University of California at Berkeley have made significant contributions to the research literature on urban structure, commuting, residential choice, and job location choice. The most recent research on the San Francisco Bay Area is included in a working paper by Cervero and Wu (*23*), who provide an analysis of the polycentric commuting patterns in the Bay Area using commuter flow data from the CTPP/UE and housing price data from the STF3A data files. They find an emerging hierarchy of employment centers ranging from the San Francisco CBD to outlying suburban business parks, as well as shorter commute times to suburban employment centers.

Summary

As a summary of the San Francisco Bay Area case study, an evaluation is given of the degree to which census data were crucial or essential to MTC planning and research applications and of the possibility that these applications could have been produced using other data bases. The paper concludes with a discussion of recommendations and expectations for Census 2000.

Evaluation of Importance of Census Data and Possibility of Substitution

Descriptive Analysis

Descriptive analyses such as the census working papers and profile reports discussed earlier are best served by a national census including a sample with data from the long-form questionnaire as well as the 100 percent count short-form data. If higher unit costs and lower accuracy are acceptable, a national survey could be conducted that would replace the Census Bureau long-form sample and perhaps could be tailored to each metropolitan area's needs. The national survey would probably cost

more than the decennial census (which costs approximately $25 per long-form questionnaire), the sampling rates would probably be substantially lower, and the statistical variance and standard errors would be substantially higher. Given that one component of descriptive analysis is trend analysis of demographic characteristics at the small-area level (e.g., census tract or block group), a smaller national survey could not be used for trend analysis for finer grains of geography, say, below place level. This would be a critical loss for city, county, and metropolitan area planners who depend on decennial census long-form data for neighborhood-level demographic characteristics.

Demographic Benchmarking

Though not discussed in this paper, one of the primary uses of census data at the metropolitan and local levels is for benchmarking of most demographic variables, including housing units, households, population, household income, and automobile ownership. Much of the most important benchmarking data is included in the decennial census long form. The census long form could be replaced by national or local demographic surveys but, as stated earlier, probably at a higher unit cost, lower sampling rates, and higher statistical variance and standard errors. In addition, total overall costs of a national survey or sets of metropolitan-area surveys could conceivably exceed the cost of conducting the decennial census long-form survey.

Model Estimation

MTC and others have demonstrated the use of decennial census data in estimating demographic and travel demand models, including data on workers in the household, automobile ownership, aggregate work trip mode choice models, and aggregate work trip destination-choice models. The best data sets for demographic and travel demand models, however, are still the locally conducted household travel surveys, often conducted concurrently with the decennial census (e.g., MTC's household travel surveys in 1981 and 1990).

The utility of future decennial censuses for model estimation could be enhanced by developing a "contextual PUMS" program in which the Census Bureau would hire researchers as "special sworn employees" in order to conduct research using raw, disaggregate census microdata records within the confines of a Census Bureau research station. An example of this sort of program is a research data center opened in January 1994 in Boston to examine topics relevant to current economic issues (*24*, pp. 9,12).

Market Segmentation in Travel Models

The decennial census is the largest, most accurate data base available to transportation planners for use in determining the demographic characteristics of subgroups of the population. Census PUMS data could be replaced by national or metropolitan-area survey data, but at higher unit costs, lower sampling rates, and higher variance and standard errors. Research should be conducted to determine just how much less accurate it is to use metropolitan-area survey data as opposed to census PUMS data for market segment adjustments.

Demographic and Travel Model Validation

It may sound redundant, but the decennial census is the best data base that can be used for the aggregate validation of several demographic and travel behavior models, including those of workers in the household, residential and job location choice (land use allocation models), automobile ownership level, and work trip generation, distribution, and mode choice. The decennial census cannot be used for the aggregate validation of nonwork travel behavior models, the best sources for which are metropolitan household travel surveys.

The census is used to represent observed conditions. Household travel surveys can also be used to represent observed conditions, but problems with the lumpiness and sparseness of typical metropolitan travel surveys render them quite difficult to use as aggregate validation data bases at any fine level of geography, say, district or superdistrict. With the decennial census, aggregate validation can be performed at almost any geographical level, perhaps even down to the traffic analysis zone (neighborhood) level.

It is conceivable that metropolitan areas in the United States could return to the 5 percent sample surveys that were more typical in the 1950s and 1960s. A 5 percent sample survey of the 2.465 million households expected in the Bay Area in 2000 would be approximately 123,000 sampled households. At a current dollar cost of approximately $125 per household, this expanded metropolitan household travel survey could cost on the order of $15.4 million in current U.S. dollars. (This is substantially more than the $1.0 million that the Bay Area spent on the 1990 household travel survey.) A national set of metropolitan travel surveys of this size could very well cost more than any conceivable decennial census.

Transportation Research Applications

The transportation planning and research community has made considerable progress in using various census products in research efforts. PUMS data have been used in analyzing markets for congestion pricing and work-at-home commuters and in analyzing the demographics of transit users in the San Francisco Bay Area. The CTPP and other standard census products are used in studies on urban structure and economic development. Certainly the microdata records from household travel surveys provide

the researcher with the most flexibility in any intended analysis. However, the decennial census data, especially the microdata files, afford the researcher the opportunity to delve into the demographic, household, and commuter characteristics of rare or hard-to-reach populations.

Recommendations and Expectations for Census 2000

Maintain Census Long Form

The census 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, in absence of a census long form, would be a set of metropolitan travel surveys that would be more costly and less accurate than a properly conducted national census.

Increase Involvement of Metropolitan and Local Planners To Improve Quality of Workplace Geocoding

Accuracy of workplace geocoding is still an issue to be reckoned with. Improvements in GIS technology will certainly help, as will a cooperative program between the Census Bureau and local persons knowledgeable about local conditions. Legal barriers that limit the involvement of local planning staffs to assist in census data processing should be liberalized to allow greater involvement of local census partners and stakeholders.

Develop Contextual PUMS Program To Facilitate Improved Planning Research

In order to increase the relevance of transportation planning research, it is desirable to create a census microdata research program. This program would allow bona fide researchers the opportunity to "add value" to census microdata and prepare more in-depth research than would otherwise be possible.

Embrace Changes in Information Technology To Provide Better, More Relevant Data

Rapid changes in information systems and information technology should be dealt with in terms of Census Bureau plans to collect, analyze, and disseminate decennial census data. New information technology should lessen the need for paper-and-ink publications in favor of electronic data-on-demand systems. Public access to the Internet should be a high priority to facilitate collection and dissemination of decennial census data.

Given these changes in information technology, it may not make sense to talk about improvements to the CTPP for 2000. Things may be changing to the degree that future analysts may get what they want when they need it and how they asked for it. This type of chaotic flexibility will likely be a challenge in terms of data consistency and comparability, so the Census Bureau may need to prepare "standard" census products in order to facilitate a transition from the highly structured products of the past to the chaotic data-on-demand products of the future.

REFERENCES

1. Purvis, C.L. The Decennial Census and Transportation Planning: Planning for Large Metropolitan Areas. In *Conference Proceedings 4: Decennial Census Data for Transportation Planning,* TRB, National Research Council, Washington, D.C., 1995, pp. 55–67.
2. *Bay Area Population Characteristics: 1990 Census.* Working Paper 1. Metropolitan Transportation Commission, Oakland, Calif., April 1992.
3. *Bay Area Travel and Mobility Characteristics: 1990 Census.* Working Paper 2. Metropolitan Transportation Commission, Oakland, Calif., Aug. 1992.
4. *County-to-County Commute Patterns in the San Francisco Bay Area: 1990 Census.* Working Paper 3. Metropolitan Transportation Commission, Oakland, Calif., Dec. 1992.
5. *San Francisco Bay Area Interregional County-to-County Commute Patterns: 1990 Census.* Working Paper 4. Metropolitan Transportation Commission, Oakland, Calif., Jan. 1993.
6. *The Journey-to-Work in the San Francisco Bay Area: 1990 Census: Census Transportation Planning Package (Statewide Element).* Working Paper 5. Metropolitan Transportation Commission, Oakland, Calif., April 1993.
7. *Disability, Mobility Limitation, and Self-Care Limitation Status in the San Francisco Bay Area: 1990 Census: Summary Tape File 3A.* Working Paper 6. Metropolitan Transportation Commission, Oakland, Calif., Oct. 1993.
8. Purvis, C.L. *Detailed Commute Characteristics in the San Francisco Bay Area: 1990 Census: Census Transportation Planning Package (Urban Element).* Working Paper 7. Metropolitan Transportation Commission, Oakland, Calif., March 1994.
9. Purvis, C.L. *Detailed Interregional Commute Characteristics: 1990 Census: Census Transportation Planning Package (Urban Element).* Working Paper 8. Metropolitan Transportation Commission, Oakland, Calif., May 1994.
10. *San Francisco Bay Area: County & Regional Profiles: 1990 Census: Census Transportation Planning Package (Statewide Element).* Working Paper 9. Metropolitan Transportation Commission, Oakland, Calif., Sept. 1994.
11. *San Francisco Bay Area Detailed Household Characteristics: 1990 Census: Census Transportation Planning Package (Urban Element, Part 4).* Working Paper 10. Metropolitan Transportation Commission, Oakland, Calif., April 1995.
12. *San Francisco Bay Area Commuters by Household Income Characteristics: 1990 Census: Special Tabulation Product #214.* Working Paper 11. Metropolitan Transportation Commission, Oakland, Calif., Jan. 1996.

13. *Bay Area Place to Place Journey to Work Characteristics: 1980–1990.* Electronic Publication 1. Metropolitan Transportation Commission, Oakland, Calif., April 1993.
14. *Bay Area Place to Place Journey to Work Spreadsheets: 1980–1990.* Electronic Publication 2. Metropolitan Transportation Commission, Oakland, Calif., March 1994.
15. *San Francisco Bay Area: Place-Level Profiles: 1990 Census: Census Transportation Planning Package (Statewide Element).* Electronic Publication 3. Metropolitan Transportation Commission, Oakland, Calif., Sept. 1994.
16. Purvis, C.L. Using Census Public Use Microdata Sample to Estimate Demographic and Automobile Ownership Models. In *Transportation Research Record 1443*, TRB, National Research Council, Washington, D.C., 1994, pp. 21–29.
17. Purvis, C.L. Estimation and Validation of Aggregate Home-Based Work Trip Walk and Bicycle Share Models: Technical Memorandum. HBWWALK/BIKE 1. In *San Francisco Bay Area 1990 Travel Model Development Project: Compilation of Technical Memorandum*, Vol. 1. Metropolitan Transportation Commission, Oakland, Calif., March 1995.
18. Horowitz, J., F.S. Koppelman, and S.R. Lerman. *A Self-Instructing Course in Disaggregate Mode Choice Modeling.* Report DOT-T-93-18. Federal Transit Administration, U.S. Department of Transportation, Dec. 1986.
19. Kollo, H.P.H., and C.L. Purvis. Regional Travel Forecasting Model System for the San Francisco Bay Area. In *Transportation Research Record 1220,* TRB, National Research Council, Washington, D.C., 1989, pp. 58–65.
20. Harvey, G.W., and E. Deakin. *Description of the STEP Analysis Package.* Metropolitan Transportation Commission, Oakland, Calif., forthcoming.
21. Harvey, G.W., and E. Deakin. *Incomes of East Bay Households and Transbay Commuters.* Memo to Congestion Pricing Management Board, Metropolitan Transportation Commission, Oakland, Calif., April 11, 1994.
22. Purvis, C.L. *Marin County and the San Francisco Bay Area: Demographic and Commuting Patterns.* Metropolitan Transportation Commission, Oakland, Calif., Nov. 1994.
23. Cervero, R., and K.-L. Wu. *Polycentrism, Commuting, and Residential Location in the San Francisco Bay Area.* Working Paper 640. Institute of Urban and Regional Development, University of California, Berkeley, April 1995.
24. New Research Center Opens in Boston. In *Census and You*, Bureau of the Census, U.S. Department of Commerce, March 1994.