

Using The Planning Database To Plan Data Collection For Census 2000

Linda Akers and Steve Alnwick

US Census Bureau, Los Angeles Regional Office, Van Nuys, California 91406

Key Words: Census, Planning Database, Enumeration

Introduction

Managing a Census is a complex undertaking.

The Decennial Census is the largest civilian activity that the U.S. government performs. It is conducted every 10 years, as mandated by the Constitution, to apportion the House of Representatives. Census data are also used for local redistricting, the allocation of Federal and State funds, and are vitally important for public and private decision makers.

The goal of the census is simple, to count every person in the United States and collect data about the population and its housing. However, the process of taking a census is not simple. The census activities schedule of all operations has over 5600 lines. There are numerous operations and procedures that occur before and after census day (April 1, 2000) which are interlinked to provide for a complete enumeration of our diverse and mobile nation.

This paper will focus on two major operations: the Nonresponse Followup (NRFU) Operation; and the Questionnaire Assistance and Be Counted Center (QAC/BC) Program. The paper will describe how the Planning Database was used to plan for those activities. Many activities, including recruiting, involved the use of the Planning Database, but this paper will limit the discussion to these two activities.

First, we present a brief description of the Planning Database, followed by discussion of the above two operations and how the Planning Database was implemented through a series of Operational Plans. Next, we describe the necessary geographic processing and manipulation of the Planning Database, as well as the dissemination of that data to the Local Census Offices for effective implementation. The results of the use of the Planning Database will be presented, with a focus on the Santa Ana Local Census Office. Finally, an evaluation of the Planning Database by the Local Census Offices will be discussed, along with recommendations for the future.

The Planning Database

The Planning Database was created by Census Bureau staff as a resource tool to aid field management in planning activities for Census 2000. Specifically, it

was used in the identification of geographic areas where the Local Census Offices might experience one or more barriers to enumeration and where special procedures could be applied.

The Planning Database assembles a range of housing, demographic and socioeconomic variables from the 1990 Census which are correlated with 1990 Nonresponse data and 1990 Post Enumeration Survey Undercount Rates. The variables include housing indicators (percent renters, multi-units, crowded housing, lack of telephones, vacancy) and person indicators (poverty, not a high school graduate, unemployment, complex household, mobility, language isolation). A "hard-to-count" (HTC) score is determined for each census tract. The HTC score summarizes 12 variables associated with housing, education, income, employment and linguistic isolation. Each variable is assigned a value of 0 to 11, for a maximum HTC score of 132. Generally, census tracts with a HTC score of 70 or greater were considered likely to be hard-to-count (U.S. Bureau of the Census, 1990 Data for Census 2000 Planning, pp 13-15 of User Documentation).

Operational Plans

To assist in the planning and implementation of the Census the Los Angeles Regional Census Center implemented a series of operational plans. In some cases the operational plans provided guidance in carrying out a task, such as the Local Census Office Grand Opening event. In other cases the operational plans were developed to determine where an office would do certain things, such as team enumeration, hiring persons with special language skills, or identifying Questionnaire Assistance Center and Be Counted Sites. For the latter category of operational plans, the Los Angeles Regional Census Center provided each of its 41 Local Census Offices with a set of instructions explaining the process and expected results. Deliverables were also provided to the Local Census Offices in the form of data and maps. The Planning Database was the primary source of data and was also used for the creation of the thematic maps that were sent to each Local Census Office.

Nonresponse Followup

Nonresponse Followup (NRFU) is the largest

operation the Census Bureau executes. Although the majority of the population return their questionnaire by mail, approximately 35% of the households require personal visits during the Nonresponse Followup operation. This involves sending thousands of enumerators to diverse neighborhoods to conduct interviews. In the Los Angeles Region, approximately 40,000 enumerators were assigned to conduct interviews at more than 2.8 million households.

Two Operational Plans were developed by the Los Angeles Region for use in planning the Nonresponse Followup operation:

Identifying Team Enumeration Areas: The purpose of this operational plan was to identify, down to the census tract level, areas which might require team or paired enumeration. Team or paired enumeration is used in areas where it could be potentially dangerous for census enumerators to perform their work individually. Working an area in pairs, or as a team, provides for better enumerator safety and confidence and therefore a more thorough canvassing.

Identifying Areas for Hiring Enumerators with Special Language Skills: The purpose of this operational plan was to identify, at the census tract level, areas where enumerators with special language skills would be needed. Simply put, the Census Bureau must hire enumerators capable of communicating with the persons to be enumerated. The intent of this plan was to increase participation of linguistically isolated households.

Questionnaire Assistance Center/Be Counted Program

The Questionnaire Assistance Center/Be Counted program was designed to increase participation in the census, in particular, the traditionally undercounted populations. Questionnaire Assistance Centers were typically storefront-type establishments which were staffed by volunteers and/or paid employees. They were designed to: 1) assist persons who have questions about filling out the census questionnaire; 2) provide language/literacy assistance; 3) provide information about the census; and 4) if necessary, give a Be Counted form to persons who did not receive a questionnaire at their home.

Be Counted forms were intended to be used by persons who did not receive a questionnaire at home or believed that they were not included on a census form. Be Counted forms could also be used by persons without conventional housing who were not enumerated during one of the shelter/street enumeration operations. Be Counted forms were

provided in a counter-top display and were available at community organizations, schools, churches, convenience stores and other locations. Questionnaire Assistance Centers could also function as Be Counted sites during the later stages of the operation.

There was an operational plan for this operation as follows:

Identifying Questionnaire Assistance and Be Counted Centers: Designed to provide instructions and deliverables to LCOs to determine the optimal number and location of Questionnaire Assistance Centers/Be Counted (QAC/BC) sites. These sites were to be located in areas where traditionally undercounted persons reside or where persons may need language assistance. The Los Angeles Region had 2028 QACs and 2404 Be Counted sites.

Geographic Processing

Thematic maps were prepared by the geographers at the Los Angeles Region for the Local Census Offices to use when developing their operational plans. To complete this task, Geography staff used the TIGER/Line files, a public product produced by the Census Bureau, along with Arcview™ and Tiger2shape™, both off-the-shelf commercial software packages. Tiger2shape™, a GISTools product, is an intermediary instrument used to convert TIGER/Line files into Arcview™ readable files. This is done by linking the records that form the polygons of basic census geographic units and displaying these graphically in Arcview™. Arcview™, an application developed by ESRI (Environmental Systems Research Institute) is a desktop geographic information system (GIS) which incorporates both geographical and textual information into a single system to allow for data analysis. In order to create thematic maps, a collection of geographic features and associated attributes of those features, or theme as defined in Arcview™, would be aggregated.

The first step in creating a theme for use in thematic mapping was to convert TIGER/Line county-based files into Arcview™ through the application of the Tiger2shape software. Once boundary files were created, information contained in the Planning Database was then joined at the census tract level. The final step in theme composition was to produce a Local Census Office boundary file. A previous spreadsheet used to define Local Census Office boundaries by census tract was joined to the theme by census tract and FIPS (Federal Information Processing Standard) state, county and city codes. At this point, all of the data in the Planning Database was joined to a Local Census Office by census tract and could be mapped

and analyzed.

To create the thematic maps for team enumeration, see Figure 1, a graduated color theme was created based on hard-to-count (HTC) scores for each census tract. For every 20 point interval, a different color was used. Included in these maps were the 1990 Census mail non-return rates. Non return rates are the percent of the questionnaires not returned by occupied units (this is different than the non response rates, which are the percent of questionnaires not returned by total housing). Census tracts with a non return rate of 35% or greater were flagged to identify those tracts where a greater-than-normal ratio of respondents were more likely to not return a questionnaire in 2000. Those areas with the highest HTC scores and identified with a high non-return mail rate were to be evaluated by each local census office through meetings with local law enforcement, community organizations, government agencies and, if necessary, field visits. Generally an HTC score of 60 and above would be considered a potential area for team enumeration. However, depending on previous experiences during earlier field operations and local knowledge, additional census tracts could be included for review.

Creating thematic maps for the operational plan “Identifying Areas for Hiring Enumerators with Special Language Skills” (see Figure 3) required many of the same techniques, but presented some unique challenges. Data in the Planning Database for

the linguistically isolated were broken down into 34 separate language groups. Local management decided that there were simply too many language categories to present to the Local Census Offices for proper evaluation. While it is entirely possible that there are census tracts with isolated language groups such as Greek, Latvian or Croatian, to portray all 34 languages in the 3841 census tracts in the Los Angeles Region was not useful. Instead, the Summary of Population and Housing Summary Files (STF3a) was used to approximate the general characteristics of three linguistically isolated groups: Spanish speaking, Asian speaking and “other” language. Census tracts where 20% or more of the housing units contained linguistically isolated households in one of these three groups were demarcated with an S, A or O symbol (S for Spanish speaking, A for Asian speaking and “O” for “other”). The Local Census Offices then did a more detailed analysis of these identified census tracts.

Included in this same map was a theme overlay representing percentage non-citizens. A graduated color legend was incorporated to break down percentages into five equal increments. This additional data was essential in highlighting areas where non-citizens might be reluctant to answer the census owing to fearfulness or misunderstanding of confidentiality. Bilingual enumerators were sent to these areas with the expectation that respondents would have a greater comfort level when asked questions in their native language.

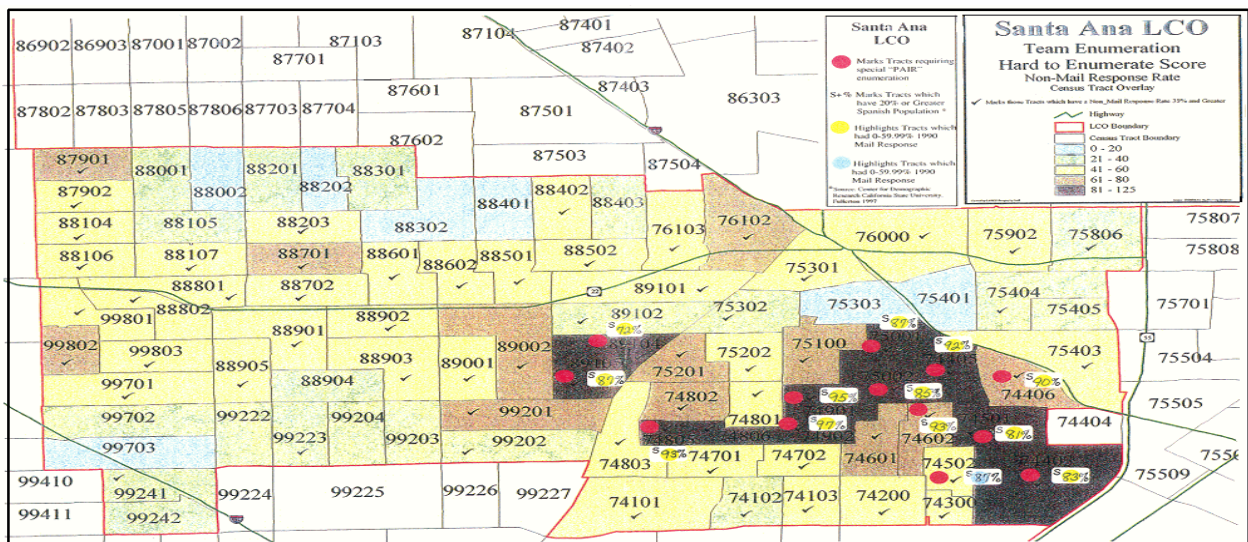


Figure 1. The Santa Ana LCO identified 21 tracts for paired enumeration. 13 tracts were noted as having a mail non response rate of 35% or greater. In addition, 10 out of the possible 11 tracts having an HTC score above 80 were incorporated.

| LOCAL CENSUS OFFICE SANTA ANA 3237, ORANGE COUNTY | | | | | | | | | | |
|---|----------------|--------|-----------|-----------------------|-------------------------|--------------|----------------------|---------------|-----------------|-----------|
| Number of census tracts to be included in team enumeration 21 | | | | | | | | | | |
| CENSUS TRACT | PAIRS ON TEAMS | REASON | TRACT POP | % MINORITY POPULATION | % RENTER OCCUPIED UNITS | INCOME LEVEL | % BELOW POVERTY LINE | GANG ACTIVITY | MAIL RESPONSE | HTE SCORE |
| | | | | | | | | | | |
| 0741.01 | PAIRS | | 8,831 | 92.41% | 35% | Moderate | 15.89% | YES | 60.00 to 69.99% | 30 |
| 0741.02 | | | 6,839 | 83.33% | 26% | Middle | 10.37% | YES | 70.00 to 79.99% | 24 |
| 0741.03 | | | 4,892 | 86.66% | 25% | Middle | 15.65% | YES | 60.00 to 69.99% | 24 |
| 0742.00 | PAIRS | | 8,874 | 91.37% | 36% | Moderate | 19.60% | YES | 60.00 to 69.99% | 30 |
| 0743.00 | | | 4,310 | 86.47% | 35% | Moderate | 16.31% | YES | 60.00 to 69.99% | 26 |
| 0744.03 | PAIRS | | 5,332 | 91.50% | 75% | Low | 28.14% | YES | 0 to 59.99% | 32 |
| 0744.05 | PAIRS | | 6,909 | 94.02% | 84% | Low | 30.50% | YES | 0 to 59.99% | 35 |
| 0744.06 | PAIRS | | 4,299 | 92.46% | 73% | Low | 20.44% | YES | 0 to 59.99% | 29 |
| 0745.01 | PAIRS | | 9,198 | 98.45% | 76% | Low | 34.44% | YES | 0 to 59.99% | 38 |
| 0745.02 | PAIRS | | 6,462 | 96.04% | 41% | Moderate | 17.32% | YES | 60.00 to 69.99% | 29 |
| 0746.01 | PAIRS | | 7,976 | 90.15% | 44% | Moderate | 20.63% | YES | 60.00 to 69.99% | 26 |
| 0746.02 | PAIRS | | 9,341 | 95.04% | 56% | Moderate | 26.13% | YES | 0 to 59.99% | 35 |
| 0747.01 | PAIRS | | 8,956 | 96.46% | 42% | Moderate | 23.53% | YES | 0 to 59.99% | 32 |
| 0747.02 | | | 6,133 | 95.24% | 40% | Moderate | 12.93% | YES | 60.00 to 69.99% | 27 |
| 0748.01 | PAIRS | | 6,101 | 98.31% | 45% | Moderate | 21.81% | YES | 0 to 59.99% | 29 |
| 0748.02 | | | 4,780 | 86.36% | 34% | Moderate | 27.56% | YES | 60.00 to 69.99% | 26 |
| 0748.03 | | | 6,069 | 87.06% | 22% | Middle | 12.16% | YES | 70.00 to 79.99% | 24 |

Figure 2. A table showing demographic data extracted from the PDB, FFIEC and local law enforcement agencies were used to identify paired enumeration areas.

Results and Examples

All forty-one Local Census Offices in the Los Angeles Region used the Planning Database and developed the requested operational plans. The Santa Ana, CA LCO was used as an example for this study. The Santa Ana office put forth an excellent effort on planning census activities and completed the Nonresponse Followup operation ahead of schedule.

The Santa Ana office, like many offices, supplemented the information in the Planning Database with locally obtained data. The Santa Ana office used a 1997 demographic study obtained from

California State University, Fullerton, local knowledge furnished by census staff and community organizations, crime information furnished by local police departments and the Orange County Probation Department and data from the Federal Financial Institutions Examinations Council database.

Figures 1 through 5 are examples of products produced by the Santa Ana Local Census Office from Planning Database information. The Figure 1 map was used to determine where team or paired enumeration would occur. Twenty-one census tracts were eventually identified. Figure 2 is a spreadsheet which shows which census tracts were picked for paired enumeration and the reason for inclusion. Hard-to-count score, mail response, gang activity, poverty and income level and renter occupied units were all used. Figure 3 is a map showing linguistic isolation, Non-citizenship status and mail response rates. This map was used to locate areas to hire bi-lingual enumerators as well as areas for proposed Questionnaire Assistance Centers and Be Counted sites. Figure 4 shows the total number of bi-lingual enumerators needed (746 in seven language groups) and the language needs for each census tract. Finally, Figure 5 is a spreadsheet showing the suggested placement of over 200 Questionnaire Assistance Centers and Be Counted sites in 38 census tracts. Many of the indicators used in Figure 2 were again used in this spreadsheet and it was

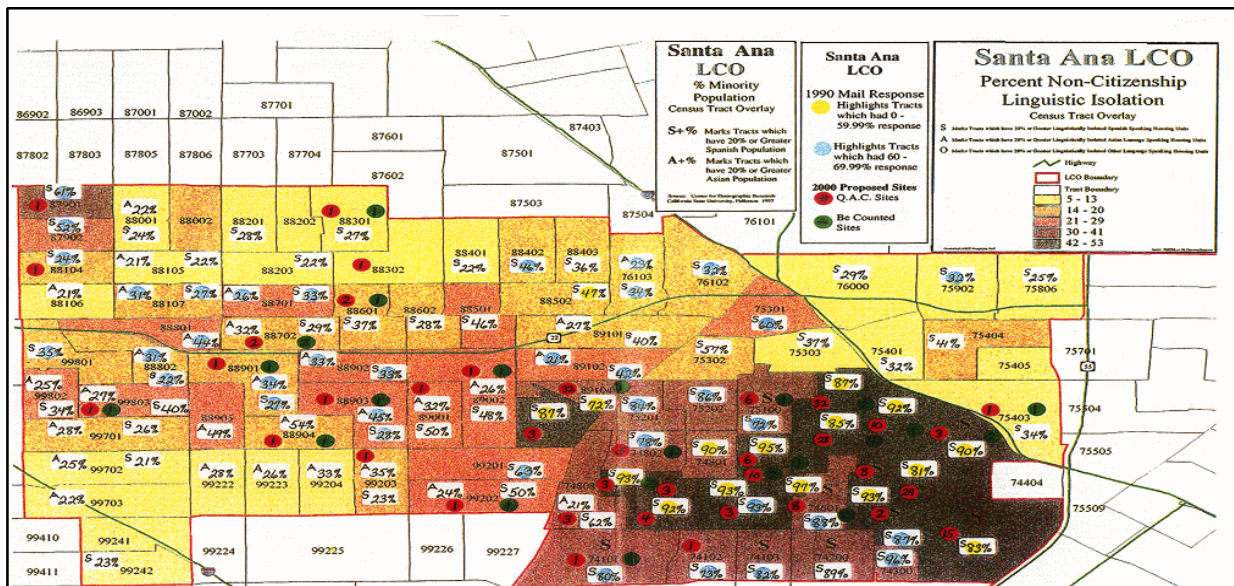


Figure 3 A color gradient scheme was used to identify non-citizenship in each tract within the Santa Ana LCO boundaries, including markers to identify tracts with 20% or more linguistically isolated populations. Minority estimates from a 1997 CSU Fullerton study and 1990 mail response rates were included to identify skilled bilingual speakers to enumerate the population and occupy QAC sites.

| Census Tracts Requiring Enumerators with Special Language Skills | | | | | | | |
|--|------------|--------------------------------------|--------|---------|----------|-----------|---------|
| Santa Ana LCO 3237 | | Number of census tracts included | | | | 87 | |
| Number of | Spanish | speaking enumerators required in LCO | | | | 595 | |
| Number of | Vietnamese | speaking enumerators required in LCO | | | | 94 | |
| Number of | Chinese | speaking enumerators required in LCO | | | | 14 | |
| Number of | Korean | speaking enumerators required in LCO | | | | 18 | |
| Number of | Filipino | speaking enumerators required in LCO | | | | 14 | |
| Number of | Cambodian | speaking enumerators required in LCO | | | | 7 | |
| Number of | Laotian | speaking enumerators required in LCO | | | | 4 | |
| TRACT | SPANISH | VIETNAMESE | KOREAN | CHINESE | FILIPINO | CAMBODIAN | LAOTIAN |
| 0741 01 | X | X | | | | | |
| 0741 02 | X | X | | | | | |
| 0741 03 | X | X | | | | | |
| 0742 00 | X | X | | | | | |
| 0743 00 | X | | | | | | |
| 0744 03 | X | | | | | X | X |
| 0744 05 | X | | | | X | | |
| 0744 06 | X | | | | X | | |
| 0745 01 | X | | | | | X | X |
| 0745 02 | X | | | | | | |
| 0746 01 | X | X | | | | | |
| 0746 02 | X | X | | | | | |
| 0747 01 | X | X | | | | | |
| 0747 02 | X | X | | | | | |
| 0748 01 | X | X | | | | | |
| 0748 02 | X | X | | | | | |
| 0748 03 | X | X | | | | | |
| 0748 05 | X | X | | | | | |

Figure 4 This table describes the number of bilingual enumerators needed for the Santa Ana LCO. 746 bilingual enumerators spanning across 7 language groups in 87 census tracts were identified.

| CENSUS TRACTS TO BE INCLUDED IN Q.A.C. AND BE COUNTED | | | | | | | | | |
|--|-----------|----------------------|--------------|--------------|----------------------|--------------|---------------|----------------------------------|--------------------------------------|
| LOCAL CENSUS OFFICE SANTA ANA 3237, ORANGE COUNTY | | | | | | | | | |
| Number of census tracts to be included in Q.A.C. and Be Counted 38 (1990 Q.A.C. and B.C. 44) | | | | | | | | | |
| CENSUS TRACT | TRACT POP | TRACT % MINORITY POP | % HENNER INC | INCOME LEVEL | % BELOW POVERTY LINE | ACTIVE GAMES | MAIL RESPONSE | Q.A.C. SITES AVAILABLE 1990 2000 | BE Counted SITES AVAILABLE 1990 2000 |
| 0741 01 | 8,831 | 92.41% | 35% | Mod | 15.09% | YES | 60 to 69% | S V | 1 S |
| 0741 02 | 6,839 | 88.93% | 26% | Mid | 10.27% | YES | 70 to 79% | S V | 1 S |
| 0741 03 | 4,892 | 86.66% | 25% | Mid | 15.65% | YES | 60 to 69% | S | 15 S |
| 0742 00 | 3,674 | 91.39% | 38% | Mod | 19.63% | YES | 60 to 69% | E S | E S |
| 0743 00 | 4,310 | 96.47% | 35% | Mod | 16.31% | YES | 60 to 69% | E S | E S |
| 0744 03 | 5,332 | 91.50% | 75% | Low | 28.14% | YES | 0 to 59% | E S | E S |
| 0744 05 | 6,909 | 94.02% | 84% | Low | 30.50% | YES | 0 to 59% | E S | 10 E S |
| 0744 06 | 4,299 | 92.46% | 73% | Low | 20.44% | YES | 0 to 59% | E S | 3 E S |
| 0745 01 | 9,188 | 98.45% | 76% | Low | 34.44% | YES | 0 to 59% | E S | 24 E S |
| 0745 02 | 6,462 | 96.04% | 41% | Mod | 17.32% | YES | 60 to 69% | E S V | 2 E S |
| 0746 01 | 7,976 | 90.15% | 44% | Mod | 20.63% | YES | 60 to 69% | E S | 8 E S |
| 0746 02 | 9,341 | 95.04% | 53% | Mod | 26.13% | YES | 0 to 59% | E S | 8 E S |
| 0747 01 | 8,956 | 96.46% | 42% | Mod | 23.53% | YES | 0 to 59% | E S | 4 E S |
| 0747 02 | 6,133 | 95.24% | 40% | Mod | 12.93% | YES | 60 to 69% | S | 3 S |
| 0748 01 | 6,101 | 98.31% | 45% | Mod | 21.81% | | 0 to 59% | E S | E S |
| 0748 02 | 4,780 | 86.38% | 34% | Mod | 27.58% | YES | 60 to 69% | E S | 5 E S |
| 0748 03 | 5,089 | 97.93% | 22% | Mid | 12.16% | YES | 70 to 79% | S V | 3 S |
| 0748 05 | 7,071 | 97.31% | 67% | Low | 32.26% | YES | 0 to 59% | E S | 3 E S |

Figure 5 Suggested placement of Questionnaire Assistance Centers and Be Counted Sites from the PDB was used in conjunction with data displayed in Figure 3 resulting in the placement of over 200 sites in 38 Census tracts.

cross checked with the map in Figure 3.

Evaluation

In July and August of 2000, the Los Angeles Region conducted extensive debriefings of key Local Census Office management staff. The agenda covered a broad spectrum of topics and included questions about the use of the Planning Database as a tool in the creation of the operational plans and planning in general.

Many managers commented that a major limitation of the Planning Database was that the data was not current, since it primarily came from the 1990 Census. To overcome this, many offices supplemented the Planning Database data with information from local sources. For example, Local Census Offices in

Orange County supplemented the Planning Database data with information from locally organized Complete Count Committees, local law enforcement agencies, the state university, and school districts. Ultimately, managers responded overwhelmingly that the Planning Database was an effective tool in the planning of data collection activities, but should be updated with current information.

Most of the Assistant Managers for Field Operations felt that the PDB was most useful in identifying Questionnaire Assistance Centers and Be Counted sites. On the other hand, the Planning Database was somewhat limited in identifying team/pair enumeration areas because active gang areas constantly shift.

Local Census Offices also felt that they should have been provided more training on how to use and manipulate the data in the Planning Database. Generally it was thought that a detailed training would have lead to a more thorough and complete use of its capabilities.

One suggestion involved the presentation of data on linguistically isolated households. It was suggested that linguistically isolated households of each non-English language in a census tract be compared against the total number of linguistically isolated households in that census tract so that a percentage of linguistically isolated households by specific language could be obtained. These percentages could then be applied to the staffing and recruiting requirements for each census tract and targets could be established.

Another suggestion was that some tallies of special place/group quarters be added at the census tract level. Identification of a number of special place/group quarters from the previous census could serve as an aid during the identification and precanned phase of the special place operations. Taking it one step further it was suggested that an actual list of special places be maintained in the Planning Database. This list, of course, would be confidential under Title 13 of the US Code.

A final suggestion was to provide block group level data in the Planning Database. A block group is the lowest level of geography for which sample data is tabulated. Using block group data would allow managers to focus their research to a smaller level of geography and be able to make more localized decisions. For example, one block group in a census tract might qualify for team enumeration but if the data were examined at the census tract level either the entire census tract would qualify or the census tract would not qualify at all. A related suggestion was to create fixed enumeration districts or assignment areas

prior to the census. These districts would be at the block group level and would be pre-identified in the Planning Database. Census 2000 data, American Community Survey data and other data, could be fed into the Planning Database continually and research about hard-to-count areas and recruiting could begin much earlier.

Conclusion

Even with its above-mentioned limitations, the Planning Database was a useful tool for Regional Office and Local Census Office staff. It should also be mentioned here that it was used successfully by the California State Data Center to analyze data and train Complete Count Committees throughout California.

We would like to see the Planning Database concept continued and expanded. As data is released from Census 2000 and the American Community Survey, it should be incorporated into the Planning Database. We should also be supplementing it with current data from other sources, such as: local population and housing estimates, local crime statistics, employment and unemployment information, BLS Salary Surveys, and information about special places.

As planning for Census 2010 progresses into the new millennium the Planning Database will be an invaluable management tool.