CONCEP 2003

January 2004

Terry Beckhelm

SANDAG

401 B Street, Suite 800 • San Diego, CA 92101-4231 • (619) 595-5300

BOARD OF DIRECTORS

The 18 cities and county government are SANDAG serving as the forum for regional decision-making. The Association builds consensus, makes strategic plans, obtains and allocates resources, and provides information on a broad range of topics pertinent to the region's quality of life.

CHAIR: Hon. Ron Morrison
VICE CHAIR: Hon. Mickey Cafagna
EXECUTIVE DIRECTOR: Gary L. Gallegos

CITY OF CARLSBAD
Hon. Ramona Finnila, Mayor Pro Tem
(A) Hon. Bud Lewis, Mayor
(A) Hon. Matt Hall, Councilmember

CITY OF CHULA VISTA
Hon. Steve Padilla, Mayor
(A) Hon. Patty Davis, Councilmember
(A) Hon. Jerry Rindone, Deputy Mayor

CITY OF CORONADO
Hon. Phil Monroe, Mayor Pro Tem
(A) Hon. Frank Tierney, Councilmember

CITY OF DEL MAR
Hon. Crystal Crawford, Councilmember
(A) Hon. Richard Earnest, Mayor
(A) Hon. David Druker, Councilmember

CITY OF EL CAJON
Hon. Mark Lewis, Mayor
(A) Hon. Gary Kendrick, Mayor Pro Tem

CITY OF ENCINITAS
Hon. Christy Guerin, Councilmember
(A) Hon. Maggie Houlihan, Deputy Mayor

CITY OF ESCONDIDO
Hon. Lori Holt Pfeiler, Mayor
(A) Hon. Ed Gallo, Councilmember

CITY OF IMPERIAL BEACH
Hon. Patricia McCoy, Councilmember
(A) Hon. Diane Rose, Mayor
(A) Hon. Mayda Winter, Councilmember

CITY OF LA MESA
Hon. Barry Jantz, Councilmember
(A) Hon. David Allan, Councilmember
(A) Hon. Ernie Ewen, Vice Mayor

CITY OF LEMON GROVE
Hon. Mary Sessom, Mayor
(A) Hon. Jill Greer, Councilmember
(A) Hon. Jerry Jones, Councilmember

CITY OF NATIONAL CITY
Hon. Ron Morrison, Councilmember
(A) Hon. Frank Parra, Deputy Mayor

CITY OF OCEANSIDE
Hon. Jack Feller, Councilmember
(A) Hon. Terry Johnson, Mayor

CITY OF POWAY
Hon. Mickey Cafagna, Mayor
(A) Hon. Don Higginson, Councilmember
(A) Hon. Robert Emery, Councilmember

CITY OF SAN DIEGO
Hon. Dick Murphy, Mayor
Hon. Jim Madaffer, Councilmember
(A) Hon. Scott Peters, Councilmember

CITY OF SAN MARCOS
Hon. Corky Smith, Mayor
(A) Hon. Pia Harris-Ebert, Councilmember

CITY OF SANTEE
Hon. Hal Ryan, Vice Mayor
(A) Hon. Randy Voepel, Mayor
(A) Hon. Jack Dale, Councilmember

CITY OF SOLANA BEACH
Hon Joe Kellejian, Mayor
(A) Hon. David Powell, Councilmember

CITY OF VISTA
Hon. Morris Vance, Mayor
(A) Hon. Judy Ritter, Councilmember
(A) Hon. Bob Campbell, Councilmember

COUNTY OF SAN DIEGO
Hon. Dianne Jacob, Chairwoman
(A) Hon. Greg Cox, Supervisor

CALIFORNIA DEPARTMENT OF TRANSPORTATION
(Advisory Member)
Jeff Morales, Director
(A) Pedro Orso-Delgado, District 11 Director

METROPOLITAN TRANSIT DEVELOPMENT BOARD
(Advisory Member)
Leon Williams, Chairman
(A) Hon. Jerry Rindone, Vice Chairman
(A) Hon. Bob Emery, Board Member

NORTH SAN DIEGO COUNTY TRANSIT DEVELOPMENT BOARD
(Advisory Member)
Hon. Judy Ritter, Chair
(A) Vacant
(A) Hon. Jack Feller, Board Member

IMPERIAL COUNTY
(Advisory Member)
Hon. Victor Carrillo, Supervisor
(A) Hon. Larry Grogan, Councilmember, City of El Centro

U.S. DEPARTMENT OF DEFENSE
(Advisory Member)
CAPT Christopher Schanze, USN, CEC
Commander, Southwest Division
Naval Facilities Engineering Command
(A) CAPT Richard Gamble, USN, CEC

SAN DIEGO UNIFIED PORT DISTRICT
(Advisory Member)
Jess Van Deventer, Commissioner
(A) Michael Bixler, Commissioner
(A) Peter Q. Davis, Commissioner

SAN DIEGO COUNTY WATER AUTHORITY
(Advisory Member)
Hon. Bud Lewis, Director
(A) Bernie Rhinerson, Director

BAJA CALIFORNIA/MEXICO
(Advisory Member)
Hon. Luis Cabrera Cuaron
Consul General of Mexico

As of February 2, 2004

ABSTRACT

TITLE:	CONCEP 2003
AUTHOR:	San Diego Association of Governments
DATE:	January 2004
SOURCE OF COPIES:	San Diego Association of Governments 401 B Street, Suite 800 San Diego, CA 92101 (619) 699-1900
NUMBER OF PAGES:	34
ABSTRACT:	This report presents a basic description of the Consolidated Estimates Characteristics Program, including a general flow of information and some of the key inputs, assumptions, and computations for each of the components.

TABLE OF CONTENTS

LIST OF TABLES

OVERVIEW

Section 1

OVERVIEW

The Consolidated Characteristics Estimates Program (CONCEP) is a modeling system, rather than a single program. CONCEP was developed to integrate a collection of interrelated but independently operated estimation models. Conceptually the model performs several different functions which fall into the broad categories of estimates of total population and housing units, detailed demographic characteristics estimates (population by age, sex and ethnicity), and household income distribution estimates. Bringing the models under one umbrella streamlines the modeling process and ensures connectivity among the components.

CONCEP includes four major modules or applications and various utilities. The major components are:

- POPEST – housing and population estimates for census tract – city combinations (split tracts)
- PMGRA – housing and population estimates for the SANDAG Master Geographic Reference Areas (MGRAS)
- PASEE – detailed population estimates by ethnicity, sex and age for the region, subregional areas, census tracts and MGRAs
- ESTINC – household income distribution estimates for the region, subregional areas, census tracts and MGRAs.

The major components of CONCEP and two principal utilities are described in detail below.

TECHNICAL DESCRIPTION

Section 2

TECHNICAL DESCRIPTION

POPEST

Population and housing estimates are derived by POPEST. Total housing units are estimated for single and multiple family units and mobile homes and other units. Estimates of occupied units, also known as households, are produced for each type of housing unit. Population estimates are then derived from the number of occupied housing units and the average number persons per household, plus persons not living in households (such as military personnel living in barracks, college students living in dormitories, and prisoners living in correctional facilities).

The model uses a bottom-up approach. Estimates are first completed for each city-census tract combination (referred to as a "split tract"), and then are controlled to city and regional totals published by the California Department of Finance.

Housing Unit Method

Theoretically, the population of any given area can be considered to be the sum of household population (persons living in a housing unit) and non-household population (persons living in group quarters, such as military barracks, college dorms, or convalescent homes). Thus, total population can be represented by the following:

POPULATION = HOUSEHOLD POPULATION + NON-HOUSEHOLD POPULATION

where the household population is represented by:

HOUSEHOLD POPULATION = HOUSING UNITS x OCCUPANCY RATE x AVG. HOUSEHOLD SIZE

In theory, if the number of housing units, occupancy rate, average household size, and non-household populations were known precisely, then the true total population would be known. However, in practice these inputs are rarely known with complete accuracy. Therefore, using local building completions data, occupancy rates and household size parameters derived from the 2000 Census, and trends in occupancy rates and household size, the individual components are estimated and are used to estimate the population.

Split Census Tract Population Estimates

For each split census tract, the following procedure is performed:

1. Update total housing units

TOTALSF = SFBASE + SFCHNG
TOTALMF = MFBASE + MFCHNG

TOTALMH = MHBASE + MHCHNG
TOTALHS = TOTALSF + TOTALMF + TOTALMH

where

TOTALSF is total single family housing units;
SFBASE is base year single family housing units;
SFCHGN is single family housing unit change since the base year;
TOTALMF is total multi family housing units;
MFBASE is base year multi family housing units;
MFCHGN is multi family housing unit change since the base year;
TOTALMH is total mobile home and other housing units;
MHBASE is base year mobile home and other housing units;
MHCHGN is mobile home and other housing unit change since the base year, and
TOTALHS is total housing units.

Housing stock changes are calculated by:

CHNG = CP – D ± A ± M ± C

where

CP is housing units completed;
D is housing units demolished;
A is housing units annexed or de-annexed from an area;
M is housing units moved into or out of an area, and
C is housing units converted from other uses.

Theoretically, if each jurisdiction reports the same unit changes to both SANDAG and the California Department of Finance (DOF), the results from step 1 should match DOF's official housing estimates for each jurisdiction. This is seldom the case, however, as DOF may not receive all inputs in time for the year's estimates release, and DOF can adjust the jurisdiction-level data sent to them. In addition, changes by split tract reported to SANDAG do not always match the jurisdiction totals reported to DOF. Consequently, minor adjustments to the housing stock change inputs are often needed to get the initial POPEST housing estimates by jurisdiction to match those produced by DOF. These adjustments are made to control jurisdiction housing stock totals to DOF total before proceeding to the next step.

2. Estimate occupied housing units

Occupied housing units, also referred to as households, are estimated from the housing stock by applying occupancy rate factors. The occupancy rate is equivalent to one minus the vacancy rate.

Estimates of total housing stock are more detailed than those for occupied households because housing unit permits and annexation information allow an accurate update by structure type. Vacancy rates vary by structure type and single family vacancy rates are generally lower than multi-family rates. While using aggregate rates could produce large errors in areas where the housing mix has significantly changed, research has shown that vacancy rate estimates for small areas are not reliable. In fact, using separate rates for single and multiple family units may introduce more error than using an aggregate rate. Separate estimates of single family and multiple family vacancy rates are not usually attempted for areas with less than 10,000 people.

HH = TOTALHS * OCC

where

HH is occupied housing units, and
OCC is occupancy rate (the occupancy rate = 1 – vacancy rate).

Occupied housing unit (household) estimates are controlled to DOF annual estimates for each city and the unincorporated area using a factoring algorithm and a +1/-1 adjustment to control for rounding. Changes in households on military bases are an external input and are not adjusted during controlling.

3. Estimate population in households

Household population is estimated from households by applying persons per household factors (household size). Estimates of household size for small areas have the same problems as vacancy rates. Therefore, household size is applied to the households for single and multiple family units combined and to mobile homes households.

$$HHP = HH * PPH$$

where

HHP is household population, and
PPH is average persons per household.

Household population estimates are controlled to DOF annual estimates for each city and the unincorporated area using a factoring algorithm and a +1/-1 adjustment to control for rounding. Changes in household population on military bases are an external input and are not adjusted during controlling.

4. Estimate total population

As noted above, total population is the sum of household population plus non-household population. Non-household population is further defined as including military non-household population (military group quarters, such as persons living in barracks) and civilian non-household population (civilian group quarters, such as persons living in dormitories or convalescent homes). Thus, by adding the estimates of household population, military group quarters population, and civilian group quarters population, the total population estimate is obtained.

$$TPOP = HHP + MILGQ + CIVGQ$$

where

TPOP is total population;
MILGQ is military group quarters population (uniformed military personnel living in barracks or on ships), and
CIVGQ is civilian group quarters population (civilian population living in college dorms, prisons, convalescent homes and other non-household living quarters).

5. Completing the estimates

At this point, POPEST has derived final estimates of housing stock, households, and household population. POPEST re-computes the vacancy rates and household sizes from these normalized data for use as starting values for the following year's estimates. Group quarters population estimates are added to household population to determine the total population for each split tract. POPEST estimates both civilian and military group quarters population. These estimates are based on changes added to base year values. If necessary, POPEST controls the civilian group quarters population to DOF estimates for each city and the unincorporated area. Changes in military group quarters population are an external input and are not controlled.

PMGRA

PMGRA distributes split tract estimates to MGRAs. The distribution scheme is essentially a controlling algorithm. The algorithm requires MGRA level input for the change in civilian and military group quarters population and for the changes in single family and multiple family housing units. The group quarters changes are derived from the split tract level changes by assigning each change to an MGRA. Since there are only a few group quarters changes, this is a straightforward, manual process. Housing unit changes, however, may number in the hundreds or thousands and require a different methodology. Estimates of housing unit changes by address are obtained from local jurisdictions. These records are addressed matched to assign a MGRA to each unit change. These MGRA unit and group quarters changes are then added to the previous year's POPEST_MGRA data as an initial estimate of current MGRA totals. The MGRA estimates of housing stock and group quarters population are then controlled to their respective split tract estimates.

Once the MGRA housing stock and group quarters population estimates are completed, households are computed using vacancy rates from each split tract and controlled to the split tract households. Similarly, MGRA household population is estimated using household size factors from each split tract and controlled to the split tract household populations. Total population is the sum of the household and group quarters populations.

PASEE

PASEE produces annual population estimates by single year of age, cross-classified by sex and race/ethnicity for eight groups. Estimates are produced for special and non-special population groups. Special populations include persons living on military bases, in large off-base military housing, and around major universities.

For non-special population groups, the model uses the cohort-component method, described below, to develop the population estimates. For reasons described below, different procedures are applied to accurately estimate the special population characteristics.

The model uses a top-down approach. Estimates are completed first for the region, then for each subregional area (SRA), and finally for census tracts (CT). All estimates are controlled to the total population estimates from POPEST.

Cohort-Component Method

PASEE uses a cohort-component method to estimate population characteristics annually from base year census data. The cohort-component method models components of population change (natural increase and net migration) for age cohorts grouped by sex and ethnicity. Natural increase is the number of births minus the number of deaths occurring to a population. Net migration is the difference between the number of people moving into an area and the number who leave. The relationship of population change to its components can be expressed as:

$$P_t - P_0 = (B - D) + NM$$

where

P_t is the population at the end of the time period;
P_0 is the population at the beginning of the time period;
B is births;
D is deaths, and
NM is net migration, which may be positive or negative.

Special Population Estimates

While most populations in an area experience the typical life cycle of births, aging and deaths, some populations experience a continual in and out rotation of persons sharing similar demographic characteristics. Consider a typical neighborhood – most residents will remain in the same home, each resident aging as time passes. There may be a few births in the area. There may be a few deaths over the course of a year. A few residents may move to other neighborhoods or leave the region. Most areas in the region experience this trend over time. However, certain special populations do not. Consider a college campus and nearby student housing. Each year, the oldest students graduate and leave the area, and a new group of younger students move in. Over time, the population in this area retains a fairly consistent age range in the late teens and early twenties. We do not expect that a university campus will be filled with thirty year olds ten years from now. Thus, a special estimation procedure must be used for these "special population" areas.

In the San Diego region, special population census tracts include the following:
Military: 38.00, 63.00, 64.00, 94.00, 99.01, 99.02, 106.02, 113.00, 114.00, 187.00
Off-base military: 66.00, 87.01, 92.01, 95.08
Universities and other: 28.01, 29.04, 55.00, 62.00, 83.05, 91.06

A bottom up approach is used to estimate the special population characteristics. Estimates are done first for each special population census tract. Special population estimates for the region and SRAs are created by summing the census tract estimates. For most of these tracts (sct), a raking procedure is used to adjust the base year age, sex and ethnic distribution to the total population estimate from POPEST:

$$POPE_{s,e,a,sct} = POPB_{s,e,a,sct} * (POPE_{sct} / POPB_{sct})$$

where

$POPE_{s,e,a,sct}$ is the POPEST control, and
$POPB_{s,e,a,sct}$ * is the POPEST base population.

Three military bases, the Amphibious base, Miramar and Camp Pendleton, use changes in group quarters and household populations to update the base year characteristic estimates. These areas have large group quarters and household populations that have unique demographic characteristics. Estimates are generated by:

$$POPE_{s,e,a,sct} = POPB_{s,e,a,sct} * + GQCHG_{s,e,a,sct} + HPOPCHG_{s,e,a,sct}$$
$$GQCHG_{s,e,a,sct} = (GQE_{sct} - GQB_{sct}) * UMILDIST_{s,e,a}$$
$$HPOPCHG_{s,e,a,sct} = (HPOPE_{sct} - HPOPB_{s,e,a,sct} *) * (.75 * MILDEPDIST_{s,e,a} + .25 * UMILDIST_{s,e,a})$$

where

GQCHG and HPOPCHG are the change in military group quarters and on-base household populations, respectively;

GQE is the military group quarters population estimate from POPEST;
GQB is the base year military group quarters population from POPEST;
UMILDIST is the 2000 regional proportionate distribution of uniformed military in each age, sex and ethnic group from PUMS; (see Data Sources and Storage for details on PUMS)
HPOPE is the on-base household population estimate from POPEST;
HPOPB is the base year on-base household population from POPEST; and
MILDEPDIST is the 2000 regional proportionate distribution of military dependents in each age, sex and ethnic group from PUMS. This method assumes that an average of four persons live in on-base housing and that one person is in the military; therefore, .75 and .25 are used in the weighted average of the two regional distributions.

Because of rounding error, these estimates for special population census tracts do not exactly match the total population estimates from POPEST. A routine (called vector rounding) adds or subtracts one from randomly selected age, sex and ethnicity categories to correct the rounding problem. This routine starts by calculating the difference between the POPEST census tract control and the sum of the estimates over the sex, ethnicity and age categories:

$$DIFFPOP_{sct} = POPE_{sct} - \Sigma POPE_{s,e,a,sct}$$

where

Σ is the summation over the sex, ethnic and age categories; and
DIFFPOP is the difference between the control total and the POPEST estimate.

A random number from 1 to 1,616 is selected (there are 1,616 age, sex and ethnic categories) and one is added to or subtracted from the corresponding $POPE_{s,e,a,sct}$. One is added if $DIFFPOP_{sct}$ is positive and subtracted if $DIFFPOP_{sct}$ is negative. Categories with zero values are not adjusted and no estimate can have a negative value. The $DIFFPOP_{sct}$ value also is adjusted by plus or minus one. This operation is repeated until $DIFFPOP_{sct}$ reaches zero.

Regional Population Age, Sex and Ethnicity Estimates

1. Components of change

First determine total population change from POPEST. Then determine which portion of the change is attributable to migration, rather than natural increase. Migration is then separated into special and non-special components.

$$POPCHNG = \Sigma POPE_{ct} - \Sigma POPB_{ct}$$
$$NMIG = POPCHNG - BIRTHS + DEATHS$$
$$SPECMIG = \Sigma POPE_{sct} - \Sigma POPB_{sct}$$
$$NSPECMIG_{reg} = NMIG - SPECMIG$$

where

Σ is the summation over all census tracts (or over special population census tracts when denoted by “sct”);
POPE is POPEST estimate;
POPB is POPEST base population;
NMIG is total net migration;
BIRTHS is total region births;
DEATHS is total region deaths;
SPECMIG is net migration of special populations, and

NPECMIG is net migration of non-special populations.

2. Adjusted and survived non-special population

To determine the non-special population by sex, ethnicity and age, start with the base population and subtract the special population deaths. The remaining population is “aged” into the next year of age group. Births become the new 0-year group.

$APOP_{s,e,a} = POPB_{s,e,a} – SPECPOPB_{s,e,a} – DEATHS_{s,e,a}$
$SPOP_{s,e,0} = BIRTHS_{s,e}$
$SPOP_{s,e,a} = APOP_{s,e,a-1}$ (for populations aged 1 through 99)
$SPOP_{s,e,a,100} = APOP_{s,e,a,99} + APOP_{s,e,a,100}$

where

APOP is adjusted base year population;
s is sex;
e is ethnicity/race;
a is age;
SPOP is survived population one year from the base year, and
SPECPOPB is base year special population aggregated from census tracts.

3. Non-special net migration

$NSPECMIG_{s,e,a} = SPOP_{s,e,a} * (CIVMIGRATE_{s,e,a} / 1,000)$
$NSPECMIGSUM_{s,e} = \Sigma NSPECMIG_{s,e,a}$
$NSPECMIGFAC_{s,e} = NSPECMIG_{reg} / NSPECMIGSUM_{s,e}$
$NSPECMIG_{s,e,a} = NSPECMIG_{s,e,a} * NSPECMIGFAC_{s,e}$

where

CIVMIGRATE is the civilian net migration rate, and
NSPECMIGSUM is the sum of migration of non-special population over all age groups,
NSPECMIGFAC is the sex and ethnic specific adjustment factors used to control the initial age-specific migration estimates.

4. Non-special and total population estimates

To get the region total population by sex, ethnicity and age, complete the non-special estimates by adding the survived and migration populations. Then add the special and non-special estimates by sex, ethnicity and age.

$NSPECPOPE_{s,e,a} = SPOP_{s,e,a} + NSPECMIG_{s,e,a}$
$POPE_{s,e,a} = SPECPOPE_{s,e,a} + NSPECPOP_{s,e,a}$

where

NSPECPOPE is the non-special population estimate, and
SPECPOPE is the special population estimate aggregated from census tracts.

Subregional Area Population by Sex, Ethnicity and Age Estimates

The main objective is to develop non-special population estimates for each SRA. Once this is done, they are added to the special population estimates to get the total population estimates by age, sex and ethnic group.

Adjusted and Survived Non-Special Population

The same procedure described for the regional estimates is used to adjust and survive the base year population in each SRA. Births, deaths and base year special population are aggregated from census tracts. There are two modifications of the adjusted population before calculating the survival equations. It is possible in age, sex and ethnic groups with small population sizes that the subtraction of deaths could cause a negative adjusted population. If this occurs, the adjusted population is set to zero. Also, two SRAs (16- Miramar and 43- Pendleton) are entirely military so their non-special adjusted populations are set to zero. There are a few births and deaths in these SRAs, but they are ignored.

If these adjustments were not made, the sum over the SRAs of the survived non-special population would match the regional survived non-special population in each age, sex and ethnic group within rounding error. Because of this rounding error and the modifications to the adjusted population, a routine called Pachinko is used to control the SRA survived population to the regional totals. This controlling allows for the proper calculation of SRA overall net migration level without the use of births and deaths (discussed below in the section on Net Migration Estimates by Age, Sex and Ethnicity). The first step is to create a regional survived population total by adding across the SRAs for each sex, age and ethnic group and then calculating the difference between the sum and the regionwide control:

$$SPOPSUM_{s,e,a} = \Sigma SPOP_{s,e,a,sra}$$
$$DIFFSPOP_{s,e,a} = SPOP_{s,e,a} - SPOPSUM_{s,e,a}$$

where

Σ is the summation over the SRAs;
SPOPSUM is the resulting survived population regional sum, and
DIFFSPOP is the difference between the regional control and the summation over the SRAs.

The second step is to create probability and cumulative probability distributions across the SRAs for each sex, age and ethnic group:

$$PRSPOP_{s,e,a,sra} = SPOP_{s,e,a,sra} / SPOPSUM_{s,e,a} * 100$$
$$CPRSPOP_{s,e,a,1} = PRSPOP_{s,e,a,1}$$
$$CPRSPOP_{s,e,a,2} = CPRSPOP_{s,e,a,1} + PRSPOP_{s,e,a,2}, \text{ etc.}$$
$$CPRSPOP_{s,e,a,41} = CPRSPOP_{s,e,a,40} + PRSPOP_{s,e,a,41}$$

where

PRSPOP is the probability distribution across SRAs (i.e., the percent ranging from 0 to 100 of the region's survived population in each SRA);
1, 2, 40 and 41 represent a sequential count of the SRAs, and
CPRSPOP is the cumulative probability distribution across SRAs.

Starting with the first age, sex and ethnic group, a random number from 1 to 100 is selected. This number identifies the SRA to be adjusted based on its cumulative probability value. For example, a random number of 20 is selected and the third SRA has a cumulative probability of 15 percent and the fourth SRA has a cumulative probability of 25 percent, the fourth SRA would be adjusted. One is subtracted from or added to $SPOP_{s,e,a,sra}$ depending on the value of $DIFFSPOP_{s,e,a}$. One is added if $DIFFSPOP_{s,e,a}$ is positive and subtracted if $DIFFSPOP_{s,e,a}$ is negative. No adjustment is made if there are fewer than five survived persons. $DIFFSPOP_{s,e,a}$ is then adjusted by plus or minus one. This process is repeated until $DIFFSPOP_{s,e,a}$ reaches zero. Pachinko is then applied to the next age, sex and ethnic group and continues until all groups are processed. This method causes the size of the controlling adjustment to be directly related to the size of the survived population in an SRA.

Completing the Non-Special Population Estimates

The survived population is controlled directly to the regional non-special population estimates by age, sex and ethnic group. Controlling the survived population to regional non-special population estimates is done in three steps.

1. Sex and Ethnicity Estimates

Estimates are first developed by sex and ethnicity. The initial step is to aggregate the survived population by age, sex and ethnicity into sex and ethnic categories for each SRA. The $SPOP_{s,e,sra}$ matrix is adjusted to match the SRA non-special population estimate ($NSPECPOPE_{sra}$) as well as the regional estimate of non-special population by sex and ethnic group ($NSPECPOPE_{s,e}$). The SRA control total is computed by:

$$NSPECPOPE_{sra} = POPE_{sra} - SPECPOPE_{sra}$$

where

$POPE_{sra}$ is the POPEST total population estimate based on the aggregation of all census tracts, and
SPECPOPE is the POPEST special population estimate based on the aggregation of the special population tracts.

An iterative proportionate adjustment method is used to factor the $SPOP_{s,e,sra}$ matrix to both row totals (SRA non-special population estimate) and column totals (regional sex and ethnicity non-special population estimate). It is important that the grand totals of the $NSPECPOPE_{sra}$ and $NSPECPOPE_{s,e}$ match; that is:

$$TROW = \Sigma NSPECPOP_{sra}$$
$$TCOL = \Sigma NSPECPOP_{s,e}$$
$$TROW = TCOL$$

where

TROW is the sum over the SRA non-special population estimates, and
TCOL is the sum over the regional sex and ethnicity non-special population estimates.

The adjustment occurs by an iterative sequential raking of the SPOP matrix, alternating with rows and columns. Starting with a proportional adjustment to the SRA non-special population estimate (rows), the matrix is changed by:

$$SUMROW_{sra} = \Sigma SPOP_{s,e,sra}$$

$NSPECPOPE_{s,e,sra}^{(1)} = SPOP_{s,e,sra} * (NSPECPOPE_{sra} / SUMROW_{sra})$

where

SUMROW is the summation across sex and ethnicity categories for each SRA, and
$^{(1)}$ represents the first adjustment to the matrix.

This is followed by a proportionate adjustment to the regional sex and ethnicity estimates (columns) by:

$SUMCOL_{s,e} = \Sigma NSPECPOPE_{s,e,sra}^{(1)}$
$NSPECPOPE_{s,e,sra}^{(2)} = NSPECPOPE_{s,e,sra}^{(1)} * (NSPECPOPE_{s,e} / SUMCOL_{s,e})$

where

SUMCOL is the summation across SRAs for each sex and ethnic category, and
$^{(2)}$ represents the second adjustment to the matrix.

The second adjustment to the regional estimate throws the figures out of line with the SRA non-special population estimate; that is, the sum across sex and ethnicity categories no longer matches the SRA population figure. The next step is to readjust the figures with respect to the row totals and obtain $NSPECPOPE_{s,e,sra}^{(3)}$. This readjustment to the row totals throws the figures out of line with the regional non-special population estimate by sex and ethnicity; that is, the sum across SRAs no longer matches the regional estimates, and so the figures are again adjusted to conform to the column totals to obtain $NSPECPOPE_{s,e,sra}^{(4)}$. This adjustment sequence is continues until convergence with a final unchanging matrix, which adds to the required SRA non-special population estimate and regional non-special population estimate by sex and ethnicity. The program allows for 15 iterations that reaches convergence within rounding error. The rounding error can be significant if many cells of the matrix have fewer than 10 persons.

Because rounding error still remains after the application of the iterative proportionate adjustment method, a routine (called matrix rounding) was developed. This routine adds or subtracts one from each cell of the matrix until the rounding error for both the rows and columns is eliminated.

The first step is to compute the difference between the SRA control and the sum across sex and ethnic categories and the difference between the regional sex and ethnicity controls and the sum across SRAs. Because of the mathematics of the iterative proportionate adjustment routine, the totals of the row and column differences match. These differences along with their totals are computed by:

$DIFFROW_{sra} = NSPECPOPE_{sra} - SUMROW_{sra}$;
$DIFFCOL_{s,e} = NSPECPOPE_{s,e} - SUMCOL_{s,e}$;
$TDIFFROW = \Sigma DIFFROW_{sra;}$
$TDIFFCOL = \Sigma DIFFCOL_{s,e}$; and
$TDIFFROW = TDIFFCOL$

where

DIFFROW is the rounding error of the rows;
DIFFCOL is the rounding error of the columns;
TDIFFROW is the sum over the row (SRA) differences, and
TDIFFCOL is the sum over the regional sex and ethnicity differences.

The matrix rounding procedure starts with the first row with a non-zero difference. A random number for 1 to 16 is selected that identifies the sex and ethnicity category to be adjusted. A one is added to or subtracted from $NSPECPOPE_{s,e,sra}$, if that category has at least five persons. One is added if $DIFFROW_{sra}$ is positive and subtracted if $DIFFROW_{sra}$ is negative. $DIFFROW_{sra}$ is then adjusted by plus or minus one and $DIFFCOL_{s,e}$ for the appropriate sex and ethnic group is also adjusted like $DIFFROW_{sra}$. This process is repeated until $DIFFROW_{sra}$ reaches zero for a particular SRA. Matrix rounding is applied to the next SRA with a non-zero $DIFFROW_{sra}$ and continues until all SRAs are processed.

2. Age within Sex and Ethnic Group Estimates

The last step is to estimate the non-special population by age within each sex and ethnic group. This is done by re-adjusting the $SPOP_{s,e,a,sra}$ matrix to match the SRA non-special population estimate by sex and ethnic group described in the preceding section ($NSPECPOPE_{s,e,sra}$) as well as the regional estimate of non-special population by age within each sex and ethnic group ($NSPECPOPE_{s,e,a}$). For these estimates, the SPOP array is essentially split into sixteen matrices representing age by SRA within each sex and ethnic category. The iterative proportionate and matrix rounding methods convert the survived population into the non-special population by age and SRA. They are applied sixteen times to complete the non-special population estimate by age, sex and ethnicity ($NSPECPOPE_{s,e,a,sra}$). Table 2-1 illustrates the matrix organization using Hispanic males as an example.

Table 2-1
Matrix Organization to Convert Survived Population by Age to SRA Non-Special Population Estimate
Hispanic Males

	Survived Non-Special Population									
SRA Code	0	1	2	.	.	98	99	100+		Hispanic Male Non-Special Population Estimate
1										HM1
2										HM2
3										HM3
4										HM4
5										HM5
.										.
.										.
.										.
63										HM63
	R0	R1	R3	.	.	R98	R99	R100+		
	Regional Hispanic Male Non-Special Population Estimate									

2. Total population estimates

The total population estimate is the sum of the special and non-special population estimates:

$$POPE_{s,e,a,sra} = SPECPOPE_{s,e,a,sra} + NSPECPOPE_{s,e,a,sra}$$

where

SPECPOPE is the special population estimate aggregated from census tracts (see SPECIAL POPULATION ESTIMATES section).

Census Tract Estimates

There are two groups of census tracts, special population (i.e., military and college areas) and non-special population. Estimates for the special population census tract are described in the SPECIAL POPULATION ESTIMATES section. A modified cohort-component method that does not estimate migration is applied to the non-special population census tracts. The survived population estimates for these census tracts are controlled to the SRA non-special population estimates.

Adjusted and Survived Population

A procedure similar to that for the region and SRAs is used to adjust and survive the base year population in each non-special population census tract. The difference is that the base year population is only adjusted for deaths, because by definition there are no special populations in these census tracts. If the subtraction of deaths causes a negative adjusted population, the population in the tract is set to zero. Also, the adjusted population in all special population census tracts is set to zero. Setting the adjusted population to zero lets the matrices, which contain all census tracts, represent only non-special population tracts for computational purposes.

Controlling The Survived Population to SRA Estimates

The survived population by census tract is controlled to the non-special population estimates by age, sex and ethnic group for its SRA and the total population estimate for the census tract. A two step estimation process is used. Non-special population census tract estimates are developed by sex and ethnicity. These estimates are then disaggregated in age categories within each sex and ethnic group.

Sex and Ethnicity Estimates

The initial step is to aggregate the survived population by age, sex and ethnicity to sex and ethnicity categories for each non-special population census tract (nsct). The $SPOP_{s,e,nsct}$ matrix is adjusted to match the census tract population estimate ($POPE_{nsct}$) and the SRA estimate of non-special population by sex and ethnic group ($NSPECPOPE_{s,e,sra}$). The $POPE_{nsct}$ control represents the POPEST estimate. The iterative proportionate adjustment method matches the $SPOP_{s,e,nsct}$ matrix to both row totals (non-special population census tract estimate) and column totals (SRA sex and ethnicity non-special population estimates). The matrix rounding routine completes the sex and ethnic group estimates in the non-special population tracts ($POPE_{s,e,nsct}$).

Estimates of Age within Sex and Ethnic Group

The last step is to estimate population by age within each sex and ethnic group for non-special population census tracts. This is done by adjusting the $SPOP_{s,e,nsct}$ matrix to match the sex and ethnic group estimates described in the preceding section ($POPE_{s,e,nsct}$) and the SRA estimate of non-special population by age within each sex and ethnic group ($NSPECPOPE_{s,e,a,sra}$). For these estimates, the SPOP array is essentially split into sixteen matrices representing age by non-special population census tract within each sex and ethnic category. The iterative proportionate and matrix rounding methods convert the survived population estimate into the population by age and non-special population census tracts. These methods applied sixteen times yield estimates by age, sex and ethnicity for non-special population census tracts ($POPE_{s,e,a,nsct}$). Table 2-2 illustrates the matrix organization using Hispanic males.

Table 2-2
Matrix Organization to Convert Survived Population by Age to Census Tract Non-Special Hispanic Males

	Survived Non-Special Population								Hispanic Male Non-Special Population Estimate
CT Code	0	1	2	.	.	98	99	100+	
1.00									HM1
2.00									HM2
3.00									HM3
4.00									HM4
5.00									HM5
.									.
.									.
.									.
213.04									HM63
	R0	R1	R2	.	.	R98	R99	R100+	
	Regional Hispanic Male Non-Special Population Estimate								

Final Adjustment

A top down approach is used to prepare the census tract estimates by age, sex and ethnicity. SRA estimates are controlled to regional estimates and census tract estimates are controlled to their parent SRA. Several routines are used to match both the POPEST totals for each geographic area and the age, sex and ethnicity estimates for the larger geographic area (e.g., region for the SRAs). Of primary concern is that the sum over all age, sex and ethnicity categories for any geographic area matches the POPEST total population estimates. A final adjustment is made to the age, sex and ethnicity estimates for the non-special population census tracts. This is the same method used to adjust the special population census tract detail to the POPEST control.

The vector rounding routine adds or subtracts one from randomly selected age, sex and ethnicity categories to correct the rounding problem. This routine starts by calculating the difference between the POPEST census tract control and the sum over the sex, ethnicity and age estimates:

$DIFFPOP_{nsct} = POPE_{nsct} - \sum POPE_{s,e,a,nsct}$, where $\sum$ is the summation over the sex, ethnicity and age categories; $POPE_{nsct}$ is the population estimate from POPEST; and DIFFPOP is the difference between the control total and the estimate total.

A random number from 1 to 144 is selected (there are 144 age, sex and ethnic categories) and one is added to or subtracted from $POPE_{s,e,a,nsct}$. One is added if $DIFFPOP_{nsct}$ is positive and subtracted if $DIFFPOP_{nsct}$ is negative. Categories with zero values are not adjusted and no estimate can have negative value. The $DIFFPOP_{nsct}$ value also is adjusted by plus or minus one. This operation is repeated until $DIFFPOP_{nsct}$ reaches zero.

ESTINC

ESTINC uses a mathematical model that represents the shape of the household income distribution curve. This model has undergone extensive evaluation and accurately replicates known income distributions for the region and its geographic areas. Three factors are needed to specify this model. The first, median household income represents the central point of the household income distribution. The second factor represents the distribution's variability. The third factor is a non-linear adjustment that provides a more accurate estimate of the income distribution.

Estimates of households are produced for eight income ranges: less than $10,000, $10,000 to $14,999, $15,000 to $24,999, $25,000 to $34,999, $35,000 to $49,999, $50,000 to $74,999, $75,000 to $99,999, and $100,000 and above. These income distribution estimates are expressed in constant (real) 1999 dollars. This follows the convention used in the 2000 census, which reports income for 1999. Median household income is calculated in both constant 1999 dollars and current dollars.

The estimates are produced using a top-down approach. Estimates are done first for the region, then for each SRA, and finally for census tracts.

Income Model

Income distributions are characterized by a single mode and are positively skewed. ESTINC uses a mathematical model, known as the modified lognormal curve, to estimate this distribution. The classical lognormal curve fits the middle of the income distribution well, but it underestimates the proportion of low-income units and overestimates the proportion of high-income units. A modification was made to the lognormal curve to remedy this problem. To better understand the development of the modified lognormal curve, the lognormal model is first presented.

A variate X, int this case, income, is lognormally distributed if the natural logarithm of X follows a normal curve. The lognormal distribution, Y, is minimally specified by its mean and standard deviation:

$$MED(Y) = \ln[MED(X)]$$
$$SD(Y) = [2 * (\ln[MEAN(X)] - \ln[MED(X)])]1/2$$

where

ln is the natural logarithm;
MED is the median;
SD is the standard deviation; and
MEAN is the average.

Cumulative probabilities for units with incomes less than X are computed by:

$$Y_i = \ln(X_i);$$
$$Z_i = (1 / [SD(Y) * \sqrt{(2)}]) * [Y_i - MED(Y)];\text{ and}$$
$$CUMP_i = (1 + erf[Z_i / \sqrt{(2)}]) / 2$$

where

i is income class;
Y is the upper limit of the income class;
Z is the lognormal standard score;

CUMP is the cumulative probability for income less than class X_i, and
erf is the error function of the normal curve.

The modification of the lognormal curve is made to the Z_i equation by:

$PARM = 1 / [SD(Y) * \sqrt{2}]$
$Z_i = PARM * Y_iEXP * [Y_i - MED(Y)]$

where

i is the income class;
PARM reflects the constant terms, and
Y_iEXP is a non-linear adjustment parameter.

The addition of Y_iEXP to the lognormal curve helps to correct its estimation problems. PARM and EXP are determined empirically by iteration to minimize the index of dissimilarity between the estimated curve and the observed income distribution (INCOMECALIB).

As the Z_i equation shows, three parameters are needed to estimate the household income distribution–median income, PARM and EXP. Their derivation for the postcensal time period is described in Volume III, Data Definition. The cumulative probabilities (CUMP) are translated into a probability distribution, P, of a household being in an individual income class by:

$P_1 = CUMP_1$, where 1 is the lowest income category
$P_i = CUMP_{j+1} - CUMP_j$, where i ranges from 2 to n-1; j ranges from 1 to n; and n is the number of income categories
$P_n = 1 – CUMP_{n-1}$
$\sum P_i = 1$

where

$\sum$ is the summation over the income groups.

To provide estimates that are consistent with the 2000 census, this probability distribution is adjusted by the difference between the 2000 census proportionate income distribution and the 2000 proportionate distribution estimated from the modified lognormal. An additive adjustment insures that the resulting probability distribution will sum to one. This process is represented by:

$ADJ2000_i = CEN2000_i – EST2000_i$
$\sum ADJ2000_i = 0$

where

$\sum$ is the summation over the income groups;

$ADJP_i = P_i + ADJ2000_i$

where

i is the income class;
ADJ2000 is the census calibration factor;
CEN2000 is the proportionate distribution from the census;
EST2000 is the 2000 estimated proportionate distribution from the modified lognormal curve; and

ADJP is the adjusted proportionate distribution for the estimate year, and

$\sum ADJP_i = 1$

where

$\sum$ is the summation over the income groups.

Households in each income class are determined by multiplying the adjusted probability distribution by the total households. This method insures that the sum over the households in each income class will equal the total households for a geographic area, within rounding error. The calculation of the household income distribution is done by:

$ESTINC_i = HH * ADJP_i$

where

i is the income class;
ESTINC is the estimated household income distribution; and
HH is the estimated households from POPEST.

Household Income Distribution Estimates

A household income distribution is first developed for the region based on the modified lognormal curve. Next distributions are developed for SRAs, which are controlled to the regional distribution. Finally, census tract household income distributions are controlled to their parent SRA. This multi-stage controlling process insures consistent information across all geographic levels.

Initial Estimates

The initial or uncontrolled estimates of the household income distributions apply to SRAs and census tracts but not to the region, since the region's income distribution estimate is not affected by controlling. Also, the modified lognormal curve is not used for every SRA or census tract to generate the uncontrolled distribution, because the calibration showed too large a discrepancy between the observed and estimated 2000 income distributions. The initial income distributions for these areas are generated in one of two ways. For the SRAs and for census tracts with under 500 households, the base year (e.g., 2001 for the 2002 estimate) income distribution is used. The initial distribution for the other census tracts is the base year distribution adjusted by the change in the SRA income distribution between the base and estimate years. This is done as follows:

$INCHG_{i,sra} = ESTINC_{i,sra} / BYINC_{i,sra}$
$ESTINC_{i,ct} = BYINC_{i,ct} * INCHG_{i,sra}$

where

i is the income class;
INCHG is the change in SRA income;
ESTINC is the estimated income distribution, and
BYINC is the base year income distribution.

Controlling

An iterative proportionate adjustment method is used to factor the initial SRA income distribution matrix ($ESTINC_{i,sra}$) to both row totals (SRA household estimate from POPEST (HH_{sra})) and column totals (regional income distribution estimate ($ESTINC_i$)). It is important that the grand totals of HH_{sra} and $ESTINC_i$ match; that is:

$TROW = \sum HH_{sra}$;
$TCOL = \sum ESTINC_i$, and
$TROW = TCOL$

where

TROW is the sum over the SRA household estimates; and
TCOL is the sum across the regional household income distribution estimates.

The adjustment occurs by an iterative sequential raking of the $ESTINC_{i,sra}$ matrix, alternating with rows and columns. Starting with a proportional adjustment to the SRA household estimate (rows), the matrix is changed by:

$SUMROW_{sra} = \sum ESTINC_{i,sra}$
$ESTINC_{i,sra(1)} = ESTINC_{i,sra} * (HH_{sra} / SUMROW_{sra})$

where

SUMROW is the summation across income categories for each SRA; and
(1) represents the first adjustment to the matrix.

This is followed by a proportionate adjustment to the regional income distribution estimate (columns) by:

$SUMCOL_i = \sum ESTINC_{i,sra(1)}$
$ESTINC_{i,sra(2)} = ESTINC_{i,sra(1)} * (ESTINC_i / SUMCOL_i)$

where

SUMCOL is the summation across SRAs for each income category; and
(2) represents the second adjustment to the matrix.

The second adjustment to the regional estimate throws the figures out of line with the SRA household estimate; that is, the sum across income categories no longer matches the SRA household figure. The next step is to readjust the figures with respect to the row totals and obtain $ESTINC_{i,sra(3)}$. This readjustment to the row totals throws the figures out of line with respect to the regional income distribution estimate; that is, the sum across SRAs no longer matches the regional estimates, and so the figures are again adjusted to conform to the column totals to obtain $ESTINC_{i,sra(4)}$. This adjustment sequence is continued until convergence with a final unchanging matrix, which adds to the SRA household estimate and regional income distribution estimate. The program allows for 15 iterations that reaches convergence within rounding error. The iterative proportionate adjustment routine is then used to control the initial census tract income distribution matrix to both row totals (census tract household estimate from POPEST) and column totals (SRA income distribution estimate).

DATA DEFINITION

Section 3

DATA DEFINITION

DATA SOURCES AND STORAGE

Data for the CONCEP model are gathered from a wide variety of sources. Data are then stored in two places, depending on data type. MS Excel files are stored in the directory **M:\RES\EstimatesData** and database tables are stored in the CONCEP database on the SQL server.

Base Year Population and Housing

2000 population counts from the U.S. Census Bureau comprise the base data for the population by age, sex, and ethnicity. The Census Bureau data file that contains 100% population counts for the region and its census tracts and sufficient detail about age, sex and ethnicity is Summary File 1 (SF1). The Summary File 3 (SF3) contains detailed information on housing stock by structure type, as well as household income distributions.

The U.S. Census Bureau did not control its Summary File 3 (sample survey) variables to the 100% count Summary File 1. To ensure proper model function, SANDAG performed controlling procedures to ensure, for example, that housing stock by structure type counts in Summary File 3 summed to total housing stock counts from Summary File 1. In addition, some minor adjustments were made (such as correcting the location of housing units that were erroneously allocated by the Census Bureau to roads and open space) to more accurately reflect the region's true population and housing distribution.

The SQL server table **census_mgra_sf1_sr10** contains basic population characteristics, and **census_mgra_sf1_5ya_sr10** contains detailed demographic (five year age group) characteristics from the Census 2000 Summary File 1. Summary File 3 data from Census 2000 is found in the SQL server tables **census_mgra_sf3_1_sr10**; **census_mgra_sf3_2_sr10**, and **census_mgra_sf3_3_sr10**.

Population and Housing Control Totals

Each year the California Department of Finance (DOF) produces estimates of population and housing stock for each county and city in California. Known as the "E-5 City/County Population and Housing Estimates," these county and jurisdiction estimates serve as the control totals for the POPEST model for the San Diego region.

These control totals are stored in **M:\RES\EstimatesData\DOF**.

Housing Stock Changes

Base year housing stock information is derived from the decennial census, conducted by the U.S. Census Bureau (see above). However, annual housing stock change information is derived from a variety of sources.

Address level housing stock permit data for completions, conversions, and demolitions are collected from local jurisdictions. Housing changes received from the jurisdictions are entered into a spreadsheet that is stored in **M:\RES\EstimatesData\Housing\.** Data are stored by calendar year. Thus, completions in the directory **M:\RES\EstimatesData\Housing\2002HousingCompletions** show completions from January 1, 2002 through December 31, 2002, and are used to update the region's housing stock for the January 1 estimate of the next year. (For example the CY2002 housing change data are used to update base year housing stock data for the January 1, 2003 estimate).

Data are entered in the following format.

BC code	addr	zip	apn	apn_8	SF ANNX	MF ANNX	SF COMP	MF COMP	SF DEMO	MF DEMO	SF MOVE	MF MOVE	SF CONV	MF CONV	MH ADD	DATE	city no
1010			5801710800	58017108			1									0103	19
1010			5031730900	50317309			1									0103	19
2141			5990200900	59902009			1									0103	19
6450			4003303000	40033030					1							0103	19
6450			2810820500	28108205					1							0103	19
6450			6581400400	65814004					1							0103	19
1010			5793770500	57937705			1									0203	19
1010			3754000100	37540001			1									0203	19
1019			2225501800	22255018			1									0203	19
1050			4831013200	48310132				18								0203	19
1140			2441306400	24413064											1	0203	19
2141			2691722600	26917226			1									0203	19
2141			5202100100	52021001			1									0203	19
6450			2663401200	26634012					1							0203	19
6450			3020601200	30206012					1							0203	19
6450			1240112100	12401121					1							0203	19
6450			1055133300	10551333					1							0203	19
6470			1831703900	18317039						3						0203	19
7020			5844900500	58449005							2					0203	19

BC code is the building code assigned to particular types of permits (e.g. new single family construction is BC code 1010). APN is the county Assessor's Parcel Number, and APN_8 is the left-most eight digits for any 10-digit APN.

The data are collected by Assessor's Parcel Number (APN) and are matched to SANDAG's mgra layer using an address matching technique. Street address data are recorded if an APN cannot be matched. The mgra-level housing changes are then controlled to the California Department of Finance (DOF) annual housing estimates for jurisdictions and the region. For modeling purposes, the mgra data are then aggregated to city census tract combinations (split tracts) and utilized as an input to the POPEST model.

Occupancy Rates and Household Size

Occupancy rates are estimated from the 2000 Census. Average household size is based on trends between the 1990 and 2000 Censuses and on data including school enrollment and vital statistics counts.

Group Quarters Population

The California Department of Finance collects military population information and SANDAG is authorized to receive that information from DOF. The Department of Finance provides an annual tabulation of on-base military population for each military installation in the region. In addition, DOF provides a report of group quarters populations for local universities and correctional facilities. These estimates form the basis of the annual non-household population estimate used in the POPEST model. Allocations of group quarters population to MGRAs are completed in a spreadsheet that is stored in **M:\RES\EstimatesData\GroupQuarters**.

Detailed demographic information about the characteristics of the uniformed military population and military dependent population is not available at the census tract level, but the Public Use Microdata Sample (PUMS) from the Census 2000 provides characteristics at the PUMA level. Census data from 1990 were used to develop some of the parameters of the estimates model.

Vital Events

For all vital events (births and deaths) a certificate is issued. Historically a summary table of these certificates was used to determine the region's vital events. Over time, it has become clear that these records provide the most detail regarding vital events. However, reporting restrictions and data errors have reduced their effectiveness in tracking vital events for modeling purposes. Therefore SANDAG has shifted to a two-pronged method. The state's vital records are used to determine region controls for births by ethnicity/race and sex, and for deaths by age, ethnicity and sex. The region control totals are then used along with along with fertility rate and survival rate data to produce subregional estimates of births and deaths. As a result of reporting differences (birth records show a high degree of divergence from Census reporting in terms of race/ethnicity) births control totals for small populations (American Indian, Asian, Hawaiian, Other and 2 or More Races) are smoothed to more closely reflect the Census distribution of 0-4 year olds. Then the births are distributed to census tracts based on the distribution of women of "child bearing age" (20-29). The controlling of births and deaths is done with the CONCEP utility program CONTROLVITALS.

Vital events data are stored in **M:\RES\EstimatesData\VitalEvents**.

Net Migration

Census Bureau data are used to calculate net migration. The files used in the age, sex and ethnicity estimates model include the 1990 Modified Age, Race, Sex (MARS) file, and the 1990 and 2000 Public Use Microdata Sample (PUMS) files. The survival rates used in the migration calculations were based upon vital events records. (See above.)

Migration rates by sex, age and ethnicity were calculated between the 1990 and 2000 Census. These migration rates are stored in **M:\RES\EstimatesData\Migration**.

Net migration estimates from 1990 to 2000 are calculated for the civilian population. Civilian population excludes the uniformed military. For the region, this is done in two stages. Civilian net migration is first calculated by sex and ethnic group using the residual method. The life table survival rate procedure is then used to calculate net migration by age within each sex and ethnic group. The regional net migration levels are converted to annual rates. Civilian net migrants by sex and ethnic group also are estimated for each SRA using the residual method.

Income

Base year household income distribution information is derived from the decennial census, conducted by the U.S. Census Bureau (see above).

Trends in median income are determined from a range of data sources, including the U.S. Census Bureau's Current Population Survey (CPS) which reports data for the state of California and American Community Survey (ACS) which tracks data for the San Diego Region. Other sources used to estimate the regional trend for median income include the California Department of Finance, and the California Franchise Tax Board. During model testing, the Census Bureau growth rates (from the CPS and ACS) tracked most closely with historical trends, and thus were used for estimation purposes.

Income trends are calculated and stored in **M:\RES\EstimatesData\Income**. Income trends are estimated for the region and then SRAs and tracts are estimated based upon the regional trend.

Income Model Calibration

The income modified lognormal curve is calibrated using the CONCEP utility program INCOMECALIB. It calibrates to known income distributions such as the census. Using the median income and income distribution as inputs, this program estimates the other two parameters needed to fully specify the modified lognormal income distribution model (PARM and EXP). An iterative method is used to find the parameters that minimize the error between the observed and the estimated income distributions. This error is measured by the index of dissimilarity.

Index of Dissimilarity

The index of dissimilarity (IOD) is often used to measure the inequality of two distributions. It is computed by comparing two percentage distributions as follows:

$$IOD = \Sigma \left| ACT_i - EST_i \right| * .5$$

where

Σ is the summation across income groups;
$| \, |$ is the absolute value;
ACT is the observed percent of households in an income group; and
EST is the estimated percent based on the modified lognormal curve

IOD is interpreted as the percentage that the estimated distribution would have to change in order to match the actual distribution. Given that the objective is to provide reliable estimates, and based on the analysis of the calibration to 2000 census data for the region, SRAs, census tracts, the following guidelines were used to evaluate the IOD. An IOD less than 7 percent indicated that the modified lognormal curve generated an accurate income distribution estimate, while an IOD greater than 9 indicated that the curve's estimate was not precise enough. Areas with IODs between 7 and 9 percent were examined on a case-by-case basis. The modified lognormal curve produced an acceptable estimate for most of them. For those areas where the modified lognormal curve could not be used to estimate the income distribution, other procedures were developed (see Initial Estimates section in the previous chapter).

Table 3-1 illustrates the correspondence between the magnitude of the IOD and the precision of the income distribution estimates. These data show a range of IODs based on the 2000 census calibration. For IODs around 1 percent, the greatest divergence in any income class is 0.8 of a percentage point, but most classes are within 0.0 - 0.4 of a percentage point. An IOD around 5 or 6 percent shows a reasonable match to the census distribution. The largest divergence is around 3 percentage points. For IODs above 9, the fit is poor.

Table 3-1
Illustrative IOD Distributions by Value of IOD

Income Class	(IOD = 1.20)		(IOD = 5.73)		(IOD = 9.46)	
	2000 Census	Estimate	2000 Census	Estimate	2000 Census	Estimate
< $10,000	7.8	7.9	2.8	2.9	1.6	2.8
$10,000-14,999	5.8	5.5	4	6.2	6.7	3.3
$15,000-24,999	12.7	12.9	24.1	21.9	10	9.6
$25,000-34,999	14.3	13.5	25	24.4	14.6	11.7
$35,000-49,999	17.2	17.9	22.3	25.6	13.1	17.5
$50,000-74,999	20.6	20.6	15.7	15.5	22	23
$75,000-99,999	10.9	11.1	4.3	2.9	17.1	14.3
$100,000+	10.5	10.5	1.9	0.5	14.9	17.7
	100	**100**	**100**	**100**	**100**	**100**

* Based on the modified lognormal curve

Algorithm

For each geographic area, INCOMECALIB produces an estimated income distribution based on the modified lognormal curve for 9,250 combinations of PARM and EXP in conjunction with the observed median and income distribution. The PARM and EXP values are controlled by user defined calibration limits, which are 125 to 310 for EXP and 1 to 50 for PARM. It determines which of the 9,250 (185 values for EXP X 50 values for PARM) estimated income distributions are the most accurate using the following logic:

```
DISSL = 1,000,000,000
FOR EXP 125 to 310
        FOR PARM 1 to 50
                EXP = EXP / 100                 /* Convert EXP to appropriate scale
                PARM = PARM / 1000   /* Convert PARM to appropriate scale
                Solve Curve
                Compute IOD
                IF (IOD < DISSL)
                        X1 = EXP
                        X2 = PARM
                        DISSL = IOD
                        CONTINUE LOOPS
                ELSE
CONTINUE LOOPS
```

The first time through the loop (EXP = 125 and PARM = 1) the IOD will be less than the initialized value of DISSL and it will replace the DISSL value, and EXP and PARM will be stored in X1 and X2, respectively. The next time through the loop (EXP = 125 and PARM = 2), the IOD from the newly created estimate is compared to the prior value. If it exceeds the prior value, the loop continues and X1, X2 and DISSL are unchanged. If the IOD is less, X1, X2 and DISSL are replaced before continuing the loop. This process is repeated until all values for EXP and PARM, represented in the loops, are examined; therefore, the final values of X1, X2 and DISSL represent the EXP, PARM and IOD of the income distribution estimate that best matches the observed income distribution. INCOMECALIB then re-generates the income distribution estimate based on the optimal values for EXP and PARM.

The accuracy of the estimated income distribution is evaluated by examining the IOD and the actual and estimated percentages. The other information in the table relates to mechanics of the modified lognormal curve and is not that pertinent to the end user. The last income class contains no information for either Z-score or the adjustment, because its estimated percentage is derived as a residual.

The actual Z-score is computed from the observed (e.g., census) cumulative proportionate income distribution by:

```
FOR I = 1, 1000
        ZOBS = (I - 500) / 100    /* Convert to Z-score value scale (-4.99 to 5.00)
        PROB = (1 + erf((Z / √2)) /2
        IF (PROB ≥ CUMP) END LOOP
CONTINUE LOOP
```

This routine computes the cumulative probability under the normal curve for Z scores that can range from -4.99 to 5.00 (PROB). It checks PROB for a given Z-score against the observed cumulative proportion for an income group. The observed Z-score for a given income class is the one that yields a PROB that matches the observed cumulative portion (CUMP). This routine is repeated for each income class.

The estimated Z and adjustment come from the derivation of the modified lognormal curve as follows:

$Z_i = PARM * Y_i^{EXP} * [Y_i - MED(Y)]$; and
$ADJ_i = PARM * Y_i^{EXP}$

Where

i is the income class;
Z is the estimated Z-score;
Y is the natural logarithm of the upper limit of the income class; and
MED(Y) is the natural logarithm of the median income.

PROGRAMMER'S REFERENCE

Section 4

PROGRAMMER'S REFERENCE

This document provides general programming information for CONCEP, including a description of source files, data, and utilities. The discussion covers the current program organization and operation, and provides general detail to assist in making changes to source code to prepare executable programs. Additional technical information describing the various relationships and equations is available in Section 2 of this document. Detailed program documentation is provided in each source file as described below. Furthermore, some of what is discussed here may be described in more detail in Section 3 – Data Design and Database Documentation, which contains the detailed database formats and organization for the input and output data discussed in each section.

CONCEP encompasses several computer programs and a variety of data sources. In the following discussions, the environment is described in terms of the primary programs and ancillary programs, utilities, and procedures.

ENVIRONMENT

The public version of CONCEP is stored on SANDAG's shared PC network, in the common directory **m:\res\concep**. Most of the CONCEP data is maintained in the SQL Server 2000 relational database management system on the SANDAGNET server **Kahuna**. The primary database for CONCEP is **concep**. (In practice, the CONCEP project is copied to a local workstation for execution. CONCEP is not a program for general staff use; therefore, the necessary steps to make it executable from the network have not been added.) Table 4-1 summarizes the organization of the principal CONCEP development and operational software.

Table 4-1
CONCEP Software and Data Organization

Label	Description
M:\res\concep	CONCEP main directory
M:\res\concep\code	CONCEP code
M:\res\concep\code\ccode	CONCEP main program code directory
M:\res\concep\code\controlVitals	CONCEP utility program to control vital statistics
M:\res\concep\code\IncomeCalib	CONCEP utility program to calibrate income estimates model
M:\res\concep\annual stuff	Annual data checks and utilities
M:\res\concep\misc	Miscellaneous scripts and utilities
M:\res\concep\temp	Temporary working directory
M:\res\concep\doc	Formal CONCEP document
Kahuna.concep	CONCEP SQL Server 2000 database

PROGRAMMING

CONCEP is written in Microsoft's C# using the Visual Studio.Net development environment. Data originating in or being sent to database tables are managed in a variety of ways. Programmatically, access to database tables is provided through Active-X Data Objects (ADO).NET. Some utilities for constructing or analyzing data tables are written as SQL scripts that are executed with SQL Server query tools.

NOMENCLATURE

In general, a file or table name is descriptive and its extension, if applicable, identifies its purpose. The Visual Studio.Net environment is case sensitive. Most program names are in mixed (upper and lower) case, may be up to 64 characters, and may use the underscore character ("_") or the dot character (".") in the name or extension. Database tables have no extension, but may make use of the underscore character. The following nomenclature is used to help identify the assorted programs, tables, files and scripts used in CONCEP (the <> symbols represent a name or parameter and are not typed). Only the principal source and data files are listed here; some source files used for cosmetic effects are not listed. Consult the documentation for the Visual Studio.Net for a complete description of the available file types.

- **<file_name>.csproj**: identifies C# project files.
- **<file_name>.cs*:*** identifies C# source files.

Unless stated otherwise, files are assumed to reside in the **code** directory. Path and file names are listed in **bold face**. Files with various year designators are listed as "_YYYY" for YYYY = 2000, 2001, and so on. Files with year designators lagged one year are listed as "_L" for _L = YYYY -1.

COMPONENTS

Table 4-2 lists the major source code components of CONCEP. There are other source files in the CONCEP project that are used to display help windows; they are not described here. Table 4-3 lists the source file components for the two CONCEP utilities, CONTROLVITALS and INCOMECALIB.

Table 4-2
CONCEP Code Components

Component	Description
ccode.csproj	CONCEP project file
cmain.cs	CONCEP C# main menu source file
estinc.cs	ESTINC C# source file
pasee.cs	PASEE C# main source file
pasee_compute.cs	PASEE C# utilities source file
pasee_ct.cs	PASEE C# census tract classes and methods source file
pasee_mgra.cs	PASEE C# MGRA classes and methods source file
pasee_sra.cs	PASEE C# SRA classes and methods source file
pmgra.cs	POPEST_MGRA C# source file.
popest.cs	POPEST C# source file

Table 4-3
CONCEP Utilities

Component	Description
controlVitals.csproj	CONTROLVITALS C# project file
controlVitals.cs	CONTROLVITALS C# main menu source file
IncomeCalib.csproj	INCOMECALIB C# source file
IncomeCalib.cs	INCOMECALIB C# main source file

INPUT AND OUTPUT

CONCEP database tables are listed in Table 4-4. Some database tables serve as input for one component and output from another.

Table 4-4
CONCEP Database Tables

Table	Module	I/O	Description
births_ct_L	PASEE; CONTROL VITALS	I/O	Base year births by sex, ethnicity; census tracts; (calendar year prior to estimates year)
deaths_ct_L	PASEE; CONTROL VITALS	I/O	Base year deaths by age, sex, ethnicity; census tracts; (calendar year prior to estimates year)
detailed_pop_ct_L	PASEE	I	Base year population by age, sex, and ethnicity; census tracts; L = base year -1
detailed_pop_ct_tab_YYYY	PASEE	O	Tabular output; Estimates year population by age, sex and ethnicity; census tracts
detailed_pop_ct_YYYY	PASEE	O	Estimates year population by age, sex, and ethnicity; census tracts; normalized
detailed_pop_mgra_tab_YYYY	PASEE	O	Tabular output; Estimates year population by age, sex and ethnicity; MGRAs
gq_civ_chg	POPEST	I	Civilian group quarters changes by year; POPEST split tracts
gq_civ_chg_mgra	PMGRA	I	Civilian group quarters changes by year; MGRAs
gq_mil_chg	POPEST	I	Military group quarters changes by year; POPEST split tracts
gq_mil_chg_mgra	PMGRA	I	Military group quarters changes by year; MGRAs
hhovr	POPEST	I	Household size and vacancy rate overrides for POPEST split tracts
hs_chg	POPEST	I	Housing stock changes by year; POPEST split tracts
hs_chg_mgra	PMGRA	I	Housing stock changes by year; MGRAs
inc_estimates_YYYY :	ESTINC	O	Household income distribution; CT, SRA, Region
inc_estimates_YYYY_mgra	ESTINC	O	Household income distribution; MGRAs

income_model_adj	ESTINC	I	Income adjustment factors computed from 2000 base year and applied to each estimate year
income_model_parms_YYYY	ESTINC	I	Income model asd parm, median and exponent
mig_distribution	PASEE	I	Regional migration distribution factors.
migration_rates	PASEE	I	Age, sex, ethnic migration rates.
pop_mil_dep_pct_2000	PASEE	I	Military dependent population distribution factors (decade).
pop_mil_pct_2000	PASEE	I	Military population distribution factors (decade).
popest_L	PASEE; POPEST	I	Base year POPEST data; Census Tracts; L = base year -1
popest_L_mgra	PMGRA	I/O	Base year POPEST data; MGRAs
popest_YYYY	ALL	I/O	Estimates year POPEST data; Census Tracts
popest_YYYY_mgra	PASEE; PMGRA; ESTINC	I/O	Estimates year POPEST data; MGRAs
sandag_dof_YYYY	POPEST	I	SANDAG's version of DOF hs and pop controls
data_cafe.census_income_8groups	INCOME CALIB	I	Census 2000 income distribution for 8 groups
income_calibration_parms	INCOME CALIB	0	Income curve calibration parameters
births_deaths_controls	CONTROL VITALS	I	Birth and death control totals by year, sex and ethnicity
concep_survival_rates_2000	CONTROL VITALS	I	Census 2000 survival rates
concep_survival_rates_roc	CONROL VITALS	I	Rates of change in survival rates 2000 - 2020
concep_fertility_rates_2000	CONTROL VITALS	I	2000 fertility rates
concep_fertility_rates_roc	CONTROL VITALS	I	Rates of change in fertility rates 2000 - 2020
xref_ct00tosra	PASEE	I	Cross-reference census tracts to SRAs.

VERSION DESCRIPTION

Section 5

VERSION DESCRIPTION

CONCEP Version 1.0 changes:

1. Coded in Visual Studio.Net with C#, combining ESTINC, PASEE, POPEST and PMGRA with additions for 2003 Current Estimates and Final 2030 Forecast.