

# The Claritas Demographic Update Methodology

---

## Contents

- Introduction
- The Claritas Demographic Estimation Program
- The Annual Demographic Update
- 2005 Methodology: Executive Summary
- The Claritas 2005 Methodology
- Demographic Characteristics
- Smoothed Data
- Additional Information

---

## Introduction

Nationwide sets of small area demographic estimates and projections were pioneered by the private sector 30 years ago, and such “updates” are still a unique product of the private suppliers. These suppliers have developed a variety of approaches to annual demographic estimation, and the results for small areas can vary widely. Users are encouraged to familiarize themselves with the methods used to produce such data.

This document explains the methodology used to develop the 2005 Claritas Demographic Update.

---

## The Claritas Demographic Estimation Program

The Claritas Demographic Estimation Program traces its history to the industry’s earliest years, and is completing its third decade in the hands of the industry’s most experienced demographers. The demographers now with the Claritas team did the industry’s groundbreaking work in small area estimation, and continue to make contributions to the profession of applied demography. The team is always looking ahead to new methods and data sources, and is actively contributing to the planning of the 2010 census.

In fact, the Claritas Demographic Estimation Program draws upon the strengths of five of the industry’s pioneering programs, including:

- National Planning Data Corporation
- Donnelley Marketing Information Services
- National Decision Systems
- Claritas
- Market Statistics

## Evaluation and Support Materials

The Claritas Estimation Program is supported by extensive research and evaluation, and the results often are documented in professional papers. In addition to this methodology document, papers describing the following topics are available:

- Evaluation of 2000 estimates against 2000 census results
- Evaluation of consumer database counts against 2000 census results
- Use of Equifax population counts as input to tract estimates
- Comparison of 1996 estimates and 2001 projections from alternative suppliers
- Evaluation of geometric data retrieval methods

## **The Annual Demographic Update**

The Demographic Update is a shorthand term for the massive set of demographic estimates and projections produced each year by Claritas. Estimates are data prepared for current year, and projections (sometimes called forecasts) are prepared for dates five years in the future.

The Claritas Demographic Update is produced each year for many geographic levels including national, state, county, place (city/town), census tract, and block group. Data also are available for commonly used areas such as metropolitan areas, ZIP Codes, and media areas such as DMAs. Because they are produced for small areas, the Updates can be easily aggregated to custom geographic areas specified by the user.

The Update starts with the estimation and projection of “base counts,” such as total population, household population, group quarters population, households, family households, and housing units. Characteristics related to these base counts are then estimated. Population characteristics include age, sex, race, and Hispanic ethnicity. Households are estimated by age of householder and income, family households are estimated by income; and owner-occupied housing units are estimated by value.

The updates are prepared first for large geographic areas, then for progressively smaller areas, with adjustments ensuring consistency from one level to the next.

**NOTE:** In order to take full advantage of methodological refinements and new data resources, each set of updates begins not with the previous year’s estimates, but with data from the most recent decennial census. For this reason, the difference between estimates for consecutive years is not an estimate of change from one year to the next. Change is estimated with reference to the previous census numbers.

### **New for 2005**

As of the 2005 Update, the target date for estimates and projections is January 1 of the relevant year. In previous years, an April 1 target date was used. The change achieves consistency with other Claritas products that use the January 1 date, and is of no substantive significance.

### **Contents**

The Claritas Demographic Update includes the following data items:

#### ***Base Counts***

- Population
- Households (occupied housing units)
- Family households (households with two or more related persons)
- Group quarters population (e.g., dormitories, military barracks, prisons)
- Housing units (house, apartment, or group of rooms intended as separate living quarters)

#### ***Population Characteristics***

- Population by age
- Population by sex
- Population by race
- Population by Hispanic ethnicity
- Population by age by sex by race by Hispanic ethnicity

#### ***Household Characteristics***

- Households by income
- Average household size (persons per household)
- Households by size (number of persons)
- Age of householder

- Income by age of householder
- Households by Effective Buying Income
- Householders by race and Hispanic ethnicity
- Householders by income by race.
- Householders by income by ethnicity
- Households by year householder moved into unit

**Housing Characteristics**

- Total owner-occupied units
- Owner-occupied units by value
- Housing units by year structure built

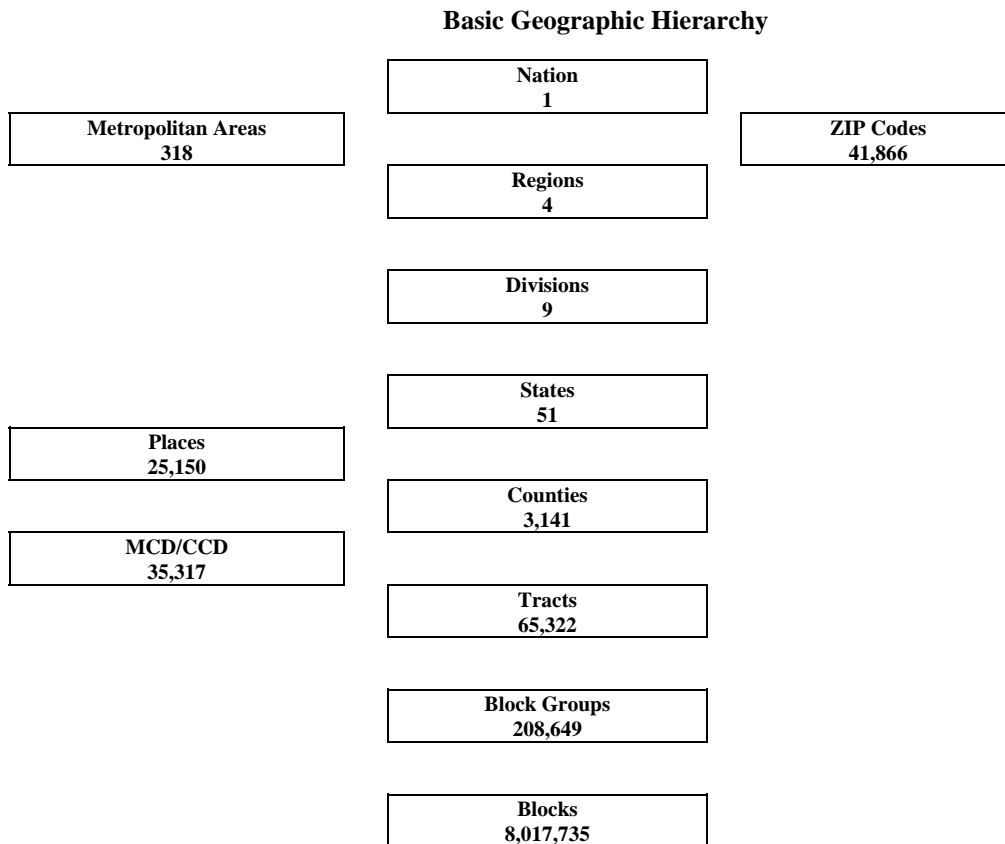
**Means and Medians**

Means and medians are provided for numerous items including:

- Mean and median household income
- Mean and median family household income
- Per capita income
- Median age of population
- Median age of householders
- Median home value

**Geography**

The Claritas Demographic Update is prepared for a wide range of census and other geographic areas. The chart below indicates the basic structure and approximate number of census and other common geographic units. The totals are those for the 2000 census geographies for which the 2005 Update was produced.



In addition to the core geographic levels identified in the chart, the Update is also produced for the areas listed below.

- Designated Market Areas (DMAs)
- Congressional Districts
- Telephone service areas
  - NPA/NXXs
  - Wire Center
- Cable Television Franchise Areas
- Yellow Pages Directory Areas
- Electric service areas
- Natural gas service areas

---

## **2005 Methodology: Executive Summary**

### **Base Counts**

At the national, state, county, and place levels, total population and household estimates are based on estimates produced by the Census Bureau, and in some cases by state demographers. At the census tract and block group levels, change is estimated based on sources including local estimates, trends in USPS deliverable address counts, and trends in consumer counts from the Equifax TotalSource database.

For 2005, national and state population estimates were based on Census Bureau estimates provided at those levels. County population estimates were based on Census Bureau county population estimates, combined with state-produced county estimates in selected states. Census tract and block group estimates were based on local estimates and post-2000 trends in USPS address counts and TotalSource consumer database households.

### **Population by Age/Sex**

Age/sex distribution is estimated with a modified cohort survival method, which ages population based on age/sex specific survival probabilities, and estimates births over the estimation period. Group quarters and other populations that do not “age in place” are not aged. The method is applied first at county level, using the Census Bureau’s most recent estimates of county population by age/sex as a starting point. Tract age/sex estimates are produced next, and controlled to the county estimates, then block group age/sex estimates are produced and controlled to tract level.

### **Population by Race/Ethnicity**

Race by Hispanic ethnicity is estimated for 14 categories reflecting “single classification” race. County estimates are produced first, based on the Census Bureau’s most recent county race/Hispanic estimates. Tract estimates are produced next based on 1990-2000 census trends, and are controlled to county level. Block group race/Hispanic estimates are produced next based on projected 1990-2000 census trends, and are controlled to tract level. The 1990-2000 census trends are identified through Claritas’ bridging of 1990 census race data to the 2000 census race definitions. Estimates of “all-inclusive” race are derived from the “single classification” estimates through the use of Census 2000 ratios of race counts and tallies.

### **Income**

Income estimates and projections reflect the census money income definition, and are produced for current dollar values. Rates of change in mean income are estimated first, and then the 2000 census income distributions are advanced to reflect the estimated rate of change. Income estimates at the county level and above reflect income change estimated by the Bureau of Economic Analysis as well as income change indicated by income statistics from the Internal Revenue Service. Income change at the tract and block group levels is estimated based on a combination of 1) change in consumer financial information from the Equifax Consumer Marketing Database, 2) change in income summarized from the TotalSource consumer household database, and 3) projections of intercensal trends.

Distributions of 2000 census income were advanced to the estimated and projected years through a process that estimates the movement of households from one income category to the next based on the specific area's estimated rate of income growth.

### **Income by Age of Householder**

The “income by age” estimates are produced after those for population by age and households by income. The household by income estimates serve as totals for the income dimension, but persons by age are converted to householders by age through the use of “headship rates” reflecting 2000 census householder by age data. The households by income and householders by age estimates serve as “income” and “age” row and column totals for the estimated income by age table. Cell values (specific income by age categories) are estimated through iterative proportional fitting of Census 2000 income by age data to the estimated income and age totals. This process yields income by age values, which not only sum to the income and age estimates, but preserve the statistical relationship between income and age for each area—as measured by the census.

### **Households by Size**

The distribution of households by size starts with the 2000 census distributions, and advances them to current year based on estimated change in persons per household (average household size). Iterative proportional fitting is then used to ensure consistency with previously estimated household totals and average household size.

### **Housing Value**

Housing value is estimated for all owner occupied housing units. As with income, the method begins with the estimation of a rate of change, which is used to advance the 2000 census distribution to current and then projection year.

At the state national levels, “target” rates of change in value are based on change in value estimated by the 2003 American Community Survey, as well as change in the “House Price Index” from the Office of Federal Housing Enterprise Oversight (OFHEO).

At county level, the OFHEO data are combined with change in median sales price data from the National Association of Realtors to estimate change at the county level. Tract rates of change are based on a combination of projected intercensal trends and post-2000 change in average mortgage amounts from the Equifax Consumer Marketing Database.

As with income, estimated rates of change are used to advance the 2000 census distributions to current year. The national and state rates serve only as targets (not control totals) for the county estimates, while the tract and block group estimates are controlled to the next higher level.

---

## **The Claritas 2005 Methodology**

### **Base Counts**

Base counts include basic totals such as population, household population, group quarters population, households, family households, and housing units.

### **Total U.S.**

Total population is estimated using the Census Bureau’s estimates of total resident population in the United States (i.e., all persons residing in the United States, regardless of citizenship). The 2005 estimate was a short projection beyond the Census Bureau’s most recent post-2000 estimate.

When the 2005 base counts were produced, the Census Bureau had produced, but not released post-2000 estimates of group quarters population. However, the Census Bureau's 2002 group quarters population estimate was inferred as the difference between its 2002 estimate of total population and the 2002 American Community Survey totals – which reflect estimated persons in households (i.e., those not in group quarters).

Total estimated households are derived by subtracting the estimated group quarters population from the estimated total population (to derive total persons in households) and dividing by the estimated average household size, or persons per household (PPH). Average household size was estimated based on change in PPH indicated by the Census Bureau's Current Population Survey, as well as test data from the new American Community Survey.

Five-year projections of the national base counts were produced with similar methods targeted at the five-year projection date. The Census Bureau's national level population projections serve as a guideline for the population projection, but not control totals.

## **State**

State population estimates are projections from the Census Bureau's most recent population estimates (Census 2003 estimates were used for the 2005 Update). Households are estimated indirectly from the completed population estimates. Specifically, group quarters population is estimated forward from the 2000 count, and subtracted from estimated population (to estimate household population), and divided by estimated average household size (itself based on intercensal trends) to produce estimated households. Total family households and housing units are estimated by applying 2000 census ratios to the household estimates.

## **County**

County population estimates are based on the Census Bureau's most recent county population estimates, in combination with county population estimates produced by selected states. The Census Bureau estimates usually lag a year and a half behind the Claritas estimation date, so a series of long and short-term projections is produced for the target date (in this case, January 2005). The mean of these projections serves as the Census-based county population estimate. Where state-produced estimates are available, and evaluations document their contribution to accuracy, these estimates also are projected to current year, and averaged with the Census-based estimates. The resulting estimates are then adjusted to conform to the state population estimates described above. As at the state level, household estimates were derived by subtracting estimated group quarters population (based on ACS inference in large counties and intercensal trends elsewhere) from total population, and dividing by estimated persons per household.

## **Place**

Population estimates for places and county subdivisions (cities and towns) are based on Census Bureau estimates for these geographic units. Population estimates for these units—and unincorporated county balances—are controlled to the county population estimates described above, and serve as control totals for the tract population estimates described below.

## **Census Tract**

### ***Background***

Post-census sources of tract level data are relatively scarce, so demographic data suppliers are on their own to identify, acquire and incorporate small area data for input to estimates. The Claritas method involves the acquisition and review of data from a variety of sources. The objective is to identify sources reflecting the extent of population and household change since the 2000 census, and to adapt standard estimation methods for use with these sources.

### ***Sources***

Among the data sources contributing to the 2005 tract level estimates are:

- Estimates produced by local governments or planning agencies
- Counts of deliverable addresses from the U.S. Postal Service
- Household counts from the Equifax TotalSource consumer database.

- Military employment counts from the Defense Manpower Data Center

Nationwide sets of small area estimates are not produced outside the data industry, but some local governments produce estimates for the census tracts in their jurisdictions. Because such data often are the best information available on small area trends, Claritas contacts a large number of local agencies each year in order to obtain, review, and incorporate the work being done by local demographers and planners.

The local data do not come in a neat package. Methodologies vary, as do estimation dates and the content provided. Some sources estimate population, others households, and still others housing units. The data are reviewed, and prepared for input to programs that account for these differences. In all cases, estimates of tract level average household size (persons per household) are critical to tying the varied input (population, households and housing units) together into a consistent set of current year base count estimates.

Alternative sources supplement locally produced estimates, and are the primary input in areas where local estimates are not available. The most widely used alternative consists of deliverable address counts from the U.S. Postal Service. These counts are summed from carrier routes to census tracts, and a time series from 2000 is maintained to enable the measurement of coverage and rates of change. Using a modified “housing unit method,” the tract-specific rates of change in USPS address counts are applied to 2000 census household counts to establish preliminary tract level household estimates. The conversion of postal address counts from carrier routes to tracts is subject to imprecision in some areas, so separate conversions are made with two methods. The results differ in some areas, and thus provide “second opinions” of the address count for each tract.

Another set of preliminary estimates is based on trends in household counts from the Equifax TotalSource household database. Again, a modified “housing unit method” is used, with trends in TotalSource household counts driving estimated change in total households since 2000.

The sources described above are used to produce alternative household estimates for each census tract, and preliminary (prior to adjustment to control totals) household estimates are established as the mean of the alternative estimates.

***Event Tracts, Military Base Closings, and Rapid Change Review***

The preparation of tract estimates is necessarily an exercise in demographic mass production, but Claritas demographers take the time to identify selected tracts that merit individual attention. For example, it is during the acquisition and review of local demographic estimates that Claritas demographers account for events such as earthquakes, fires, and hurricanes that can have a dramatic impact on the population of specific areas. Local, USPS, and Equifax data all contribute to this effort, as does consultation with local demographers.

During the 1990s, military base closings impacted the population of some communities, and additional base closings and realignments are planned. For this reason, Claritas tracks base closings and realignments, and estimates their impact at the census tract level. This effort is accomplished with military employment data from the Defense Manpower Data Center, which indicate the timing and magnitude of downsizing by installation.

In a process called the Rapid Change Review, tracts where input data indicate dramatic growth or decline are reviewed individually. The review process involves a closer examination of the input data and cross checks among the alternative sources described above. In the midst of a mass production effort, the Rapid Change Review identifies a subset of tracts where estimates are done by hand.

The preliminary household estimates described above are not adjusted directly to larger areas, but are the basis for tract level population estimates which are adjusted to the place and county subdivision estimates described above. Estimated group quarters population is then subtracted, and the remaining household population divided by estimated persons per household to produce the final estimate of total households. In other words, households are adjusted indirectly with the population estimates.



Estimates of group quarters population are rare, but Claritas estimates reflect change in group quarters since 2000. Census Bureau estimates of group quarters population are the primary input at county level and above. A few local areas provide tract group quarters estimates, and change in military quarters is estimated (in selected tracts) through the “base closing” checks described above. In most small areas, however, change in group quarters population is estimated very conservatively, unless there is specific knowledge of the opening or closing of major facilities.

Family households are estimated by applying Census 2000 (family household/household) ratios to total estimated households. Housing units are similarly estimated by applying Census 2000 (housing units/households) ratios to total estimated households.

Five-year projections of tract level base counts are produced as nonlinear projections from 2000 through the current year estimates. Rapid rates of growth and decline are moderated into the future to reflect the assumption that extreme rates of net migration are unlikely to be sustained over long periods of time. Event tracts, such as those described above, are projected separately, in order to reflect the extent of rebuilding or recovery from the relevant event. Initial five-year tract projections are ratio-adjusted to county level control totals.

### **Block Group**

Block group estimates are a challenge because change can be volatile and quality input data is scarce. Although used where available, local block group estimates are rare, so in most areas, USPS deliverable address counts are the primary input. As at tract level, trends in the USPS counts since 2000 are used to estimate 2000 census household counts forward to 2005. However, the block group application focuses on change in the block group-within-tract ratios observed in the USPS counts. For example, a block group that had 15 percent of its tract’s USPS count in 2000 might have 18 percent of the tract’s households by current year. This increase is the basis for estimating the block group-within-tract ratio of total households for current year. Similar estimates are produced based on TotalSource household counts, and are averaged with the USPS results to complete the block group household estimates. Population estimates are derived from the household estimates using methods similar to those described above for tracts.

### **Block**

The 8.2 million census blocks in the United States are the smallest units of census geography, and the census reports only data from the complete count census (or short form) at this level. Claritas produces estimates of population, households and population age 18+ at the block level. However, these estimates are not incorporated into major data products, and were produced by proportioning the block group estimates to blocks based on 2000 census (SF1)

### **ZIP Code Estimates and Projections**

#### **Background**

Estimates and projections for ZIP Codes are aggregations of estimates already prepared for block groups and block group parts. As such, there is not a distinct “ZIP Code methodology.” However, the process used to build ZIP Code estimates is important to understand, as are the complications of ZIP Code data.

ZIP Code demographic data are widely used, but involve complications not encountered with standard geographic areas. ZIP Codes are defined by the U.S. Postal Service for the delivery of mail—not for the presentation of data. They lack definitive boundaries, and change frequently at the discretion of postal officials. In addition, ZIP Codes do not conform to the boundaries of standard geographies such as counties, cities, census tracts, or census blocks.

Further complicating the specification of ZIP Code demographics is the imperfect correspondence between where people live and where they get their mail. Some people live in rural areas where there is no mail delivery, and pick up their mail at a post office—perhaps in a nearby town. The boundaries of such P.O. box ZIP Codes (there are about 5,000 of them) are not formally defined. Also, some urban residents elect to pick up some or all of their mail at a P.O. box—perhaps near their place of work. They reside in one ZIP Code, but receive mail at another. Such ZIP Codes often consist exclusively of P.O. boxes at a post

office in a nonresidential area. They have no definable boundaries, as the people receiving mail there may reside in neighborhoods scattered across a wide area.

### **Different Definitions, Different Applications**

Such disparities reveal that there are two ways to define ZIP Code demographics:

1. **Spatial Definition:** ZIP Code demographics relate to the persons and households living within the land area approximated for the ZIP Code--no matter where they get their mail.
2. **List Definition:** ZIP Code demographics relate to the persons and households receiving their mail at addresses with a common ZIP Code--no matter where they live.

The two definitions do not always produce consistent demographic data. Four thousand households might live within the boundaries of a ZIP Code, but one might be able to send mail to 5,000 households at addresses with that ZIP Code. In another ZIP Code, one might be able to send mail to only 3,000 out of the 4,000 households living within a ZIP Code's boundaries. One definition is no more correct than the other. They are different, and preferred for different applications. For example, retailers often prefer spatially defined ZIP Codes because of their correspondence with trade areas around store locations. However, direct marketers and others dealing with customers by mail may prefer list definition ZIP Code data.

### **Census Data for ZIP Codes**

Contrary to common belief, ZIP Codes have not been a standard geography for the reporting of census data. The Census Bureau did release 1980 and 1990 census ZIP Code products, but these products were non-standard, and had a short shelf life. The 1980 product reflected the "list definition," and ZIP Code definitions current as of about 1979. In contrast, the 1990 census ZIP Code product reflected "spatial definition," based on estimated ZIP Code boundaries current as of about 1992.

With Census 2000, the Census Bureau has released data for what it calls ZIP Code Tabulation areas (ZCTAs). ZCTAs approximate ZIP Code areas based on the allocation of whole census blocks. Although a significant development, the Census Bureau points out that ZCTAs are not ZIP Codes, and users need to understand that ZCTA data do not constitute "official" ZIP Code estimates. Also, even if the Census Bureau updates ZCTA definitions, Census 2000 data would not be updated from the ZCTA definitions that are now five years old.

### **ZIP Code Data from the Claritas Demographic Update**

Claritas ZIP Code estimates and projections are aggregations of Claritas estimates for block groups and block group parts. The process is similar to that used to retrieve data for circles and polygons. Census data, estimates, and projections already exist for block groups, and are aggregated to the current roster of ZIP Codes reflecting current definitions. Data for all years (including 1990 and 2000 census) are aggregated the same way to maintain a consistent reference to current ZIP Code definitions.

All Claritas products provide spatial definition ZIP Code data. Spatial definition ZIP Codes are based on a block group-to-ZIP Code correspondence, which is updated once (and for some products twice) each year. This correspondence is based on the location of block centroids (latitude/longitude points) within current ZIP Code boundaries estimated by Geographic Data Technology (GDT). If a block's centroid falls within a ZIP Code boundary, it is allocated to that ZIP Code. These block-to-ZIP allocations define which block groups (or partial block groups) are included in a given ZIP Code. For block groups allocated to more than one ZIP Code, percent inclusion factors are based on 2000 census block population counts. For all ZIP Codes with a GDT boundary, the resulting block group-to-ZIP correspondence establishes a geographic definition that is used to aggregate block group data to current ZIP Codes.

Claritas products do not provide demographic data for rural P.O. box ZIP Codes. These ZIP Codes serve residents in rural areas where there is no mail delivery, and who pick up mail at post office boxes. Although included in the roster, these ZIP Codes have no clearly defined spatial dimension, and therefore have no demographic data associated with them. Instead, the data for these ZIP Codes are included in the spatially defined ZIP Code (or multiple ZIP Codes) covering the area near the post office. These are sometimes known as "parent ZIP Codes."

It is not unusual to find spatial definition ZIP Code data that appear to be discrepant with deliverable address counts. For example, spatial definition data might indicate no data for a rural P.O. box ZIP Code for which the post office reports 600 residential deliveries, and spatial definition estimates for “parent ZIP Codes” are often higher than delivery counts since they also include the populations served by P.O. box ZIP Codes.

To assist users in identifying areas where spatial and list definition data would show significant differences, Claritas ZIP Code products also provide counts of deliverable addresses reported by the U.S. Postal Service. When combined with the spatial definition estimates, these counts indicate where different ZIP Code definitions would result in the greatest differences in ZIP Code household and population totals.

---

## Demographic Characteristics

### Population by Age/Sex

Population by age/sex composition is estimated and projected using cohort survival methods. Cohort survival is a major factor in changing age structures, and is driven by the reality that, for example, persons age 35 in 2000 who survive another five years, will be age 40 in 2005. Accordingly, a population with a large proportion of 35 year olds in 2000 can expect to have large proportions of 40 year olds in 2005. It is this process that has swelled the U.S. age structure at progressively older age categories as the baby boom generation (or birth cohort) has aged.

The Claritas cohort survival method is executed first at county level, then for tracts, and finally block groups, with each set of estimates controlled to the results at the next higher geographic level. To enhance consistency with Census Bureau age/sex estimates, the county estimates begin with the Census Bureau’s most recent county age/sex estimates. Tract and block group estimates begin with Census 2000 age/sex estimates. At all levels, the method starts with five-year age/sex categories—separating persons in households from those in group quarters. Because Census 2000 data (and the Census Bureau age/sex estimates) do not provide full age/sex detail for household vs. group quarters populations, Claritas estimates the detail required to execute the cohort survival method.

The cohort survival process is set into motion with the application of age/sex-specific five-year survival rates to the Census age/sex data described above. Each round of cohort survival ages the population of each block group ahead five years. For example, the process projects the number of 30-34 year olds in a block group who will survive to become 35-39 years old (and so on for all five-year age categories) by 2005. The initial survivals yield projections of age/sex composition for April 2005 (or July 2007 in the case of county estimates starting with 2002 Census estimates), so interpolations (based on percent age/sex) are made to the January 2005 estimation date.

#### ***Accounting for Births***

As part of each round of cohort survival, the population less than age five is “survived” to age “5-9,” so an estimate of births is required to fill the vacated “0-4” category. Births are estimated using the child/woman ratio—defined as the population age 0-4 divided by females age 15-44 (childbearing age).

The child/woman ratio is an indirect measure of fertility specific to each small area, but more important, it is sensitive to projected changes in the number of women of child bearing age—itsself a byproduct of the cohort survival process. An increase in the number of child bearing women will result in an increased number of births even if fertility rates (or child/woman ratios) remain constant. The child/woman ratios applied in the Claritas age/sex estimates and projections are derived from the 2000 census, but reflect slight increases evident in the Census Bureau’s post-2000 estimates.

***Exceptions to Cohort Survival***

The cohort survival process is at work in all areas, but in some areas its effects are overridden by migration. In the absence of authoritative age-specific migration data for small areas, the Claritas method defaults to the assumption that the age/sex composition gained or lost through migration is similar to the area's "survived" population.

However, because of migration, the cohort survival process is often not applicable to populations living in group quarters facilities such as dormitories, military quarters, prisons and nursing homes. These populations have high turnover, and therefore age/sex compositions, which tend to be stable, reflecting the nature of the facility. For this reason, cohort survivals are applied only to the population living in households. Group quarters populations are estimated separately and their age/sex compositions held constant with those measured in the census.

Claritas also identifies segments of the household population (such as concentrations of college students in off-campus housing) for which cohort survival is not applicable. Concentrations of these "hidden group quarters" populations are identified through their distinctive imprint on small area age compositions, and are similarly exempted from the cohort survival process.

***Five-Year Projections***

Five-year projections of age/sex composition are produced with the same method used for the current year estimates. In the 2005 Update, the 2005 estimates of population by age/sex were the starting point for five-year survivals to 2010. As with the current year estimates, age/sex projections are produced first for counties, followed by tracts and block groups, with adjustments ensuring consistency between geographic levels.

**Population by Race and Ethnicity**

***Definitions***

There are no universally accepted definitions of race and Hispanic ethnicity. The census currently defines "Hispanic or Latino" as an ethnicity, not a race. Race and Hispanic ethnicity are separate census questions, so in census tabulations, persons of Hispanic ethnicity can be of any race. Hispanics are included in each race category, and the race categories alone sum to total population.

The race definitions used by the 2000 census (and Claritas estimates) include the following basic categories.

- White
- Black or African American
- American Indian and Alaska Native
- Asian
- Native Hawaiian and Other Pacific Islander
- Some other race

However, because the current race standards permit respondents to mark one or more race categories, there are actually 63 categories—the six basic races plus 57 possible combinations of two or more races. When cross tabulated by Hispanic/non-Hispanic, there are 126 race by Hispanic categories.

Short of presenting data for all 63 race categories, there are two basic tabulation options—"single classification" and "all-inclusive." The single classification option is illustrated below.

- White *alone*
- Black or African American *alone*
- American Indian and Alaska Native *alone*
- Asian *alone*
- Native Hawaiian and Other Pacific Islander *alone*
- Some other race *alone*
- Two or more races

This option identifies the number of persons marking each race category by itself, and then provides a seventh category identifying the number marking two or more races. The tabulation is similar to those used in the past, and sums to total population. However, it provides no information about the race of persons in the “two or more” category, so it is not possible to determine the total number of persons identifying with a given race. The total number of persons marking a given race category is revealed by the all-inclusive categories illustrated below.

- White alone or in combination
- Black or African American alone or in combination
- American Indian and Alaska Native alone or in combination
- Asian alone or in combination
- Native Hawaiian and Other Pacific Islander alone or in combination
- Some other race alone or in combination

This option identifies the total number of persons marking each race category—either by itself or as part of a combination of two or more races. However, because persons marking two or more races are counted two or more times, the table sums to totals larger than total population.

The Claritas Update provides estimates and projections for both the single-classification and all-inclusive tabulations. Estimates for the seven single-classification categories (by Hispanic/not-Hispanic ethnicity) are produced first, and all-inclusive estimates are then derived from the single-classification numbers.

## **Estimates and Projections of Race and Hispanic Ethnicity**

At the county level and above, estimates of race and Hispanic ethnicity are based on the Census Bureau’s estimates of population by race and ethnicity at the county level. When the 2005 Update was produced, the Census Bureau had released county race estimates for July 2003.

The application is not straightforward, since the Census Bureau’s race estimates reflect a “modified” definition, in which persons of “Some other race” were re-assigned (with imputation techniques) to a “specified” race category. This reassignment increases the numbers in the specified categories, making them inconsistent with the “census definition” race counts reported in standard Census 2000 products.

For this reason, the Claritas method applies the Census Bureau’s estimated rates of change from the most applicable “modified” race category to the relevant Census 2000 race counts. For example, the Census estimates might suggest a 4.2 percent increase in the percent of a county’s population that is (modified) “Asian not-Hispanic.” The Claritas estimate is established by applying this rate of change to “percent Asian not-Hispanic” from the 2000 census. Estimates are produced for the seven not-Hispanic race categories. Percent Hispanic or Latino population is estimated separately based on the rate of change in percent Hispanic population suggested by the Census Bureau estimates. The Hispanic or Latino estimates are then distributed to race based on county specific percentages from the 2000 census. The estimates for the 14 race/ethnicity categories are then finalized by applying estimated percent race/ethnicity to the previously completed estimates of total population for each county.

Race/ethnicity estimates below the county level are based on 1990-2000 census trends in the percent of population in each race/ethnicity category. Again, the method focuses on the percent of population in each category. Estimates are produced first for tract level (with adjustments to county level), then for block groups (with adjustments to tract level).

The projection of intercensal trends is not a preferred method, but the approach was an achievement made possible by the conversion of 1990 data to 2000 geography, and the bridging of 1990 race to 2000 race definitions.

### ***Race Bridging***

The current race definitions make it impossible to identify definitive race trends between the 1990 and 2000 censuses. However, to estimate 1990-2000 trends, Claritas “bridged” 1990 census race data to the 2000 definitions. Specifically, Claritas estimated what the 1990 census race data might have looked like had they been collected using 2000 categories, and the option of marking two or more races.

All race bridging was accomplished separately for the Hispanic or Latino and not-Hispanic populations (preserving race by Hispanic cross tabulation options) for all block groups nationwide. The first step was the bridging of 2000 race to 1990 definitions. After combining the Asian and Native Hawaiian and Other Pacific Islander categories (whether alone or part of combinations) to the 1990 Asian or Pacific Islander Category, counts from the remaining multiple-race categories were distributed to single 1990 race categories. This distribution was accomplished with equal fractions assignments in most cases (combinations of two races distributed half to one category and half to the other, combinations of three races distributed by thirds, and so forth), but National Health Interview Survey proportions were used for selected combinations. These include:

- White and Black or African American
- White and American Indian or Alaska Native
- White and Asian
- Black or African American and American Indian or Alaska Native.

The bridged 2000 race data suggest how many persons would have been added to each “race alone” category had multiple-race response not been an option in 2000. For example, bridging 2000 data to 1990 definitions added some persons from multi-race categories to “Black or African American alone” to estimate the 1990 “Black” category. From the reverse perspective, the data suggest the proportion of the bridged “Black” population that would be lost to race combinations when transitioning back to the 2000 “Black or African American alone” definition. The 2000 bridged data suggest such percentages for all 1990 race categories, and these percentages were applied to the 1990 census race data (converted to 2000 block groups) to estimate the number that would have been lost from each category to multiple race responses in 1990. Census 2000 patterns then were used to distribute the estimated 1990 “two or more races” population to the 57 categories reflecting combinations of two or more 2000 census race categories.

The bridging project produced a set of 1990 census population data distributed to the 126 Census 2000 race by Hispanic categories, and converted to 2000 census block groups. These data—collapsed to single-assignment race—provided a basis for estimating race/Hispanic population trends for census tracts and block groups.

#### ***Five-Year Projections***

Five-year projections of race/ethnicity are produced with similar methods—projecting the current year estimates (of percent race/ethnicity) to the five-year projection date. Again, projections are made for percent race/ethnicity distributions, and applied to previously completed projections of population. Counties are projected first, followed by tracts and block groups, with adjustments ensuring consistency between geographic levels.

#### ***All-Inclusive Race***

Estimates and projections for all-inclusive race/ethnicity are produced as derivatives of the single-classification estimates and projections. For each race/ethnicity category, the 2000 census ratio of all-inclusive race/single-classification race is identified, and applied to the estimate or projection of single-classification race—with adjustments made in some areas to ensure consistency with the number of persons estimated (or projected) to be of two or more races. Because the all-inclusive estimates and projections are derived from data already adjusted to county controls, the all-inclusive estimates and projections are produced only at block group level, and summed to higher levels.

#### ***Population by Age/Sex by Race/Ethnicity***

Estimates and projections also are provided for the cross tabulation of population by age/sex/race/ethnicity. These estimates start with the completed estimates of population by age/sex and population by race/ethnicity at the block group level. Census 2000 does not provide age/sex/race/ethnicity detail at the block group level. For this reason, age/sex/race/ethnicity distributions for census tracts are used as “seed values” for component block groups, and iteratively adjusted to conform to the previously completed estimates of population by age/sex and population by race/ethnicity. This application of iterative proportional fitting (see appendix) produces block group estimates consistent with estimated age/sex and race/ethnicity, as well as the statistical relationship between these characteristics observed for the census tract in the 2000 census.

## **Householders by Race and Hispanic Ethnicity**

Estimates and projections of householders by (single assignment) race and Hispanic ethnicity are based on the estimates and projections of population by race/ethnicity.

For each block group, the 2000 census ratio of householders by race/Hispanic to population by race/ethnicity is identified, and applied to the current year estimate of population by race/ethnicity. This ratio indicates the percent of persons in each race/ethnicity category who were householders in the 2000 census. The final ratio is modified somewhat through the adjustment of householders by race to total households for each area, and it is the final current year ratio that is applied to the five-year projections.

## **Income Estimates**

All Claritas income estimates are expressed in current year dollars using the “money income” definition reported in the 2000 census. The estimates and projections reflect household income, which includes the income earned by all persons living in a housing unit (i.e., all household members). In contrast to the 2000 census, which reported income for the previous calendar year (1999), Claritas income estimates are for the calendar year relevant to each set of estimates and projections. For example, the 2005 estimates reflect 2005 income for 2005 households.

The method starts by estimating rates of change in mean, or average, household income for each area. Based on this rate of change, household income distributions from the 2000 census are advanced to current (or projection) year. As with the population estimates and projections, data were produced first for large areas, then for progressively smaller areas, with successive ratio adjustments ensuring consistency between levels. Aggregate and per capita income numbers were derived from the resulting income distributions.

Claritas estimates household income for all 16 income categories reported by the 2000 census in Summary File 3 (SF3). In addition, Claritas has extended the standard “\$200,000 or more” category to the income ranges indicated below.

- \$200,000 to \$249,999
- \$250,000 to \$499,999
- \$500,000 or more

Although few households had incomes this high at the time of the 2000 census, the expanded detail is important for analyses in affluent markets. And with income growth, incomes in excess of \$200,000 are not as exceptional as they were in 1999, and the five-year projections extend to years when incomes at these levels will be even more commonplace.

The extended income categories were estimated first for 1999 (2000 census), and are part of the 2000 census data from which the current year estimates and five-year projections are produced. Pareto methods, which involve an assumption of exponential decay, were applied to the 2000 census income distribution in each block group to estimate the number of households in each of the extended income categories.

## **Income Estimation Method**

Income change at the national level is estimated based on national estimates of income change from the Current Population Survey and the American Community Survey. The final estimates reflect an average of estimates based on the two sources. The national income distributions serve as a target for the state estimates, rather than a control total.

State income estimates are based on Internal Revenue Service (IRS) wage and salary data, and Bureau of Economic Analysis (BEA) estimates of per capita income. Because national IRS and BEA income data tend to reflect more rapid income growth than the national estimate, these sources are used to monitor each state’s income growth relative to the national level—change in the ratio of state income to national income. The final rates of change reflect the average of such ratios based on IRS, BEA, as well as a projection of the ratio based on 1990-2000 census trends.

County income rates of change are estimated with similar procedures—this time applying county/state ratios of IRS and BEA income data to 2000 census county/state income ratios. Again, the final estimated rates of change reflect the average of ratios based on IRS, BEA, and the projection of 1990-2000 census trends.

Income change at the tract level also is estimated with alternative sources, with the final estimated rate reflecting the average of these rates. The first estimate is based on historical performance. Specifically, tracts were estimated to outpace or lag behind county income growth in proportion to their performance relative to county during the 1990 to 2000 census period. The second is based on post-2000 trends in income estimates aggregated from the Equifax TotalSource consumer household database. The TotalSource income estimates are modeled for all individual household records on the database. The third is based on trends in the Equifax ACE-Geosummary database, which provides tract level summaries of consumer financial information from the Equifax Consumer Marketing Database (ECMD). Although not a direct measure of income, the “Average sum of credit limits for bank, national credit card, savings & loan, and credit union revolving accounts” ECMD data item is strongly associated with income at the tract level, so change in this variable is used to track change in income at the tract level.

The approach with all three sources is to track change in the tract/county ratios—or the performance of tract income relative to county level. Income change at the block group level is estimated with sources and methods similar to those described for census tracts above.

For all geographic levels, the estimated rate of income change is used to advance, or “shift,” the 2000 census distribution of households by income forward to current year. This procedure involves the estimation of the number of households advancing from one income category to another—based on the area's estimated rate of income growth.

The resulting current year distribution is adjusted to conform to that estimated for the next higher geographic level. For example, the estimated household income distribution for states is adjusted to the national distribution, the county estimates are adjusted to the final state distributions, and so forth.

Five-year projections of income begin with the projection of current year mean household income to the projection year, and the advancing of the household income distribution to reflect the projected change. Mean incomes for sub-national areas are produced by projecting estimated income change to the projection year, and advancing the current year estimated income distribution to reflect that rate of change. As with the current year estimates, the initial income distributions are adjusted to the final distributions for the next higher geographic level. State projections are adjusted to national, county to state, and so forth.

## **Family Household Income**

A family household is one in which the householder is related to one or more other persons living in the household. Family households also include any other not-related persons living in the same housing unit. Family household income includes the income of all persons living in a family household.

Family household income is estimated by applying the final estimated household income growth rates (1999 to current year) to the 2000 census distribution of family households by income—advancing the family household income distribution to reflect the relevant rate of income growth. Five-year projections were produced by trending the estimated rate of family income change out five years, and advancing the current year distribution to reflect the projected change. Because the estimates and projections of family household income are derivatives of the completed household income estimates—which already reflect control totals for large areas—they are estimated and projected at block group level only, and summed directly to higher levels.

## **Effective Buying Income**

Effective Buying Income (EBI) reflects income after taxes. Because EBI is not provided by the census, the estimates are computed as derivatives of household income, based on the correspondence between before-tax and after-tax income identified for each state from the Current Population Survey (CPS). For each state, three-year combinations of CPS data identify the mean before-tax income of households within the income in ranges estimated for the Claritas Update. The CPS also identifies mean income for these same households after deductions for

federal income taxes, state taxes, FICA, annual property taxes, and federal retirement payroll deductions. Where relevant, earned income credits were added to refine the measure of after-tax income.

The CPS data provide before-tax to after-tax income rates of change specific to households with before-tax income in selected income ranges—and specific to each state. These rates of change are applied to the current year estimated and five-year projected distributions of households by before-tax income (the standard Claritas income estimates and projections) to estimate the movement of households to lower income ranges after deductions for taxes. The resulting distribution of households by after-tax income provides the basis for computing mean, median and aggregate Effective Buying Income.

## **Income by Age of Householder**

The cross tabulation of household income by age of householder is valuable because income and life cycle stage, together, are so strongly associated with consumer needs and behavior. The Claritas “income by age” updates are produced after the estimates of population by age and households by income have been completed. The data constitute a 198-cell table defined by 18 categories of household income and 11 categories of householder age. The row and column totals from these tables (the “income” and “age” totals) are commonly referred to as the “marginal totals.”

The estimates of households by income serve as the income “marginals,” but population by age estimates must be converted to householder by age for use as the age “marginals.” For each area estimated, 2000 census data are used to determine age-specific “headship rates,” or the percent of persons in specific age categories who are householders. These headship rates are then applied to estimated population by age to produce estimated householders by age. A final adjustment to total households ensures consistency with that critical base count.

With the income and age (row and column) marginal totals estimated, the final step is to estimate the full cross tabulation of income by age of householder. In other words, values must be determined for each of the 198 income by age categories, or cells. Block group-level income by age cell values from the 2000 census provide the initial input (or seed values). Within each age category, the 2000 census income distributions are advanced to reflect the block group's (previously) estimated rate of income growth. This adjustment expresses the 2000 census income by age distribution in current dollar values. The resulting table is then adjusted to conform to both the income and age of householder totals estimated for current year. These adjustments are accomplished through “iterative proportional fitting,” which adjusts the 2000 table to conform simultaneously with the household income and householder by age estimates, while preserving the block group specific statistical relationship between income and age reflected in the 2000 census income by age data.

The income by age estimates are produced at the county, tract, and block group levels, with adjustments ensuring consistency between levels.

Five-year projections are produced using similar methods. Projected households by income serve as the income marginal totals, and Census 2000 headship rates again are used to convert projected population by age to projected householders by age. The income by age table is then advanced to projection year dollar values, and iteratively adjusted to the projected income and age marginal totals.

## **Income by Race and Ethnicity of Householder**

Estimates and projections of households by race and ethnicity of householder are produced by applying the estimated/projected rates of change in income for each area to the income distribution for each race/ethnicity group in the area. The rates of change are used to project each distribution forward to current (or projected) year, and the resulting distributions are adjusted to conform to the householder by race/ethnicity estimates and projections described above.

## **Households by Size**

Working at the block group level, estimates of households by size (number of persons) are produced for the categories 1, 2, 3, 4, 5, 6 and 7 or more persons. The distribution of households by size from the 2000 census serves as the base from which the current year estimates are derived. The 2000 distribution is advanced to current year

based on estimated change in persons per household (average household size). Iterative proportional fitting is then used to ensure consistency with estimated household totals and average household size.

Projections of households by size are based on the 2000 census and current year estimated distribution of households by size. The current year distribution is shifted to reflect the growth or decline in average household size during the projection interval. Iterative proportional fitting is then used to ensure consistency with projected household totals and average household size.

## **Housing Value**

Housing value (often referred to as “home value”) is estimated and projected for all owner-occupied housing units, and is based on the 2000 census measure, which reflects census respondents’ estimates of how much their dwellings would sell for, or the asking price of units currently for sale. Median value is estimated and projected as well as the distribution of units among the 24 categories of value reported by the 2000 census.

The total number of owner-occupied housing units is estimated by applying the relevant 2000 census percentage to the completed estimate of total occupied housing units. The basic rate of change in value is estimated first, and used to advance the 2000 census distribution of units by value to reflect this rate of change.

At the national and state levels, the rate of change in home value is monitored through the Census Bureau’s American Community Survey (ACS), and “House Price Index” data from the Office of Federal Housing Enterprise Oversight (OFHEO). Even in its test phase, the ACS was collecting data on home value from a nationwide sample of 700,000 households. And the OFHEO House Price Index is a measure of post-2000 changes in housing value derived from Fannie Mae and Freddie Mac mortgage transaction data.

County rates of change in home value are derived from two sources at the metropolitan area level. The first is data indicating the change in median sales price from the National Association of Realtors (NAR). Changes in sales price reflect only units sold during the time in question, but are strongly associated with overall change in home value. The second source is change in the OFHEO House Price Index described above.

At the census tract level, change in home value is tracked with ACE-Geosummary data from the Equifax Consumer Marketing Database. The Equifax files do not measure home value directly, but the “Average original balance on mortgage accounts” variable is strongly associated with home value. Claritas has compiled tract summaries of this variable for all census tracts for years back to 2000, and uses trends to track small area changes in home value. The greater the increase in mortgage amounts, the greater the increase in home values.

Five-year projections of value are based on rates of change derived from change in median value from 2000 census to the current year estimate.

For each geographic level, the estimated rates of change are used to advance the 2000 census home value distribution to current year. Estimates and projections are produced first at state and national levels, but these estimates serve as targets for the county estimates and projections, rather than control totals. Starting at the county level, the estimates and projections serve as control totals for smaller areas.

## **Housing Units by Year Structure Built**

Estimates and projections of housing units by year built start with the distributions from the 2000 census. These distributions are advanced to current year (and five year) targets based on housing loss patterns exhibited in the 1990 and 2000 censuses. For example, the number of units built from 1960-1969 in 2000 is lower than the number of such units in 1990, and suggests a 10-year “survival probability” for units of that age. The method establishes a set of such probabilities at the national level, and applies them to the 2000 census housing data to generate the current year estimates and five-year projections. In most areas, “surviving” units are fewer than total units (estimated separately), and the excess units form the estimate of units built after 2000. Thus, areas with rapid housing growth will show the greatest concentrations of new housing.

## Households by Year Moved Into Unit

The estimates of households by Year Moved into Unit follow a similar method. Year moved in “survival probabilities” are computed from 1990 and 2000 census data (in this case reflecting the loss of residents of specific lengths of residence). These national level probabilities are applied to the 2000 census distribution of households by Year Moved In to establish estimated and projected distributions. Households in excess of those “surviving” (staying in place) to longer lengths of residence are those estimated to have moved in following the 2000 census. Thus, areas with rapid household growth will show the greatest concentrations of new movers.

---

## Smoothed Data

In addition to the annual demographic estimates and projections, the Claritas Update provides a series of detailed census tables ratio-adjusted, or “smoothed,” to relevant current year totals. For example, the 2000 census table on marital status was adjusted for conformity with estimated population age 15 and above by sex. These “smoothed” tables are not estimates, and do not purport to show anything beyond the effect of applying decennial census distributions to estimated base count totals at the block group level.

Nevertheless, such data can be quite valuable. While percent distributions of characteristics are not estimated beyond the 2000 census, the “smoothed” totals within categories are often more accurate than those from the census—especially in areas where population growth or decline has been rapid. And because “smoothed” data are produced at the block group level on a “bottom-up” basis, percent distributions for aggregations (any area including more than one block group) will differ from those observed in 2000. This bottom-up effect can be advantageous. For example, if the most rapidly growing block groups in a county tend to have relatively high concentrations of married couple households, the “smoothed” result will indicate an increased proportion of married couple households in that county for current year. Taken for what they are, and used with an understanding of their limitations, the Claritas “smoothed” data are a valuable component of the annual Demographic Update. The list of “smoothed” data items is indicated below:

- Persons by detailed single classification race
- Hispanic or Latino population by specific origin
- Persons by ancestry
- Households by household size, household type, and presence of own children
- Households by presence of persons under 18 years and household type
- Households by household type and household size
- Group quarters population by group quarters type
- Persons 5+ by language spoken at home
- Persons 15+ by sex and marital status
- Working Population 16+ by means of transportation to work
- Working population 16+ by travel time to work
- Population 25+ by educational attainment
- Hispanic or Latino population 25+ by educational attainment
- Persons 16+ by sex and employment status
- Employed civilian population 16+ by industry
- Employed civilian population 16+ by occupation
- Employed civilian population 16+ by class of worker
- Aggregate household income by type of income
- Families by poverty status, family type, and presence of related children under age 18
- Housing units by units in structure
- Occupied housing units by tenure and vehicles available

---

## Additional Information

### Note About Block Group Parts

Many Claritas methods are executed at what is technically the “block group and block group part” level of geography. Block group parts are defined where block groups are split by place and/or MCD boundaries, and census data reported for block groups are reported for these block group parts. Thus, block group parts function as a geographic level between block group and block. Because it is more familiar, the term “block group level” is used throughout this document. However, it is worth keeping in mind that Claritas “block group level” applications usually refer to data and methodologies executed for block groups and block group parts.

### Consistency of Complete Count and Sample Census Totals

Because much census information was collected on a sample basis using the census “long form,” the Census Bureau used weighting techniques to present such data in “complete count” form. The weighted sample totals presented in SF 3 often differ from the SF 1 complete count totals by small amounts. For example, a census tract with 1,200 (SF 1) households might have an income table (from SF 3) summing to 1,206 or 1,197 households. The differences are statistically inconsequential.

Claritas products provide 2000 census data as published by the Census Bureau. The 1990 census data also are provided as published, but have been converted to 2000 census geography. Thus, for both 1990 and 2000 census, the usually minor discrepancies between sample and complete count totals are preserved.

### Adjustment Techniques

The adjustment process is essential to the production of estimates, which use input data at various geographic levels, and are consistent across all levels of geography. The Claritas Updates are geographically consistent, meaning that for each data item, block group data always sum to tract totals, which always sum in turn to county, state and national totals. Adjustment techniques also ensure that characteristic distributions sum to base count totals (e.g., households by income always sums to total households). The simultaneous adjustment of characteristics to higher level control totals and to total persons or households within each smaller area is achieved with iterative proportional fitting. The basic techniques are described below.

#### ***Ratio Adjustment:***

Ratio adjustment is used to bring small area data into conformity with large area totals. For example, if preliminary block group population estimates sum to a tract total of 552, but the independent tract estimate is 561, the preliminary block group estimates are adjusted upward by 1.63 percent ( $561/552$ ) to achieve the target tract total. Similar adjustments are made to bring preliminary distributions (such as age and race) into conformity with population totals within each geographic unit.

#### ***Iterative Proportional Fitting***

Iterative Proportional Fitting (IPF) methods are an elaborate form of ratio-adjustment, and are used when estimates must be adjusted to conform simultaneously to two sets of “marginal” control totals--often referred to as the dimensions of a two-dimensional table. Income by age of householder is a good example. The estimates must sum to both households by income and householders by age.

IPF methods begin with a table with target row and column totals, referred to as the row and column marginal totals. For example, one might have 12 categories of households by income as the row totals and 11 categories of householders by age as the column totals established for a 132 cell (12 x 11) table. The objective is to produce estimates for the table’s 132 cells that sum to both the row and column marginals.

The execution of IPF methods requires initial or “seed” cell values. In the case of income by age of householder, seed values are obtained from the 1990 census. This matrix of cell values reflects an intricate set of probabilities defining the relationship between income and age--as measured for the specific geography in the census. However, these 1990 census figures sum to neither estimated households by income nor estimated householders by age.

Iterative proportional fitting achieves this conformity through a series of ratio adjustments to the row and column marginal totals. Each round (or iteration) of row and column adjustments brings the seed values closer to conformity with the marginal totals. The number of iterations required varies by area, but the values eventually “converge” on a result that sum, within rounding error, to the marginal totals. The resulting estimates not only sum to the desired marginal totals, but preserve the statistical relationship between the two variables (income and age) measured for the area by the census.

## **Income Distributions**

A source of occasional confusion is the fact that the 2000 census reported income earned during calendar year 1999. This is the case whether the data are described as “1999 income” or “2000 census income.” The one-year census lag is logical, since no one had yet received their 2000 income in April 2000 when the census was taken. The Claritas Update is not constrained by this reporting limitation, and therefore presents income for the calendar year corresponding to the household estimate or projections. For example, the 2005 Update provides estimates of 2005 households by income earned in 2005. When comparing such estimates against the census, note that total households represent a five-year change since 2000, while income represents a six-year change since 1999.

## **Extended Income and Pareto Interpolation**

Income tabulations from the 2000 census top out at the “\$200,000 or more” category. This reporting limit made sense for census products since in 1999, only 2.4 percent of all households had incomes in excess of \$200,000. However, higher income breaks are important in affluent areas, and are becoming more important as incomes in excess of \$200,000 become more common. Claritas has therefore “extended” the 2000 census income distributions to include categories of: \$200,000 to \$249,999, \$250,000 to \$449,999, and \$500,000 and over.

Vilfredo Pareto (1848-1923), creator of the unrelated “80/20 rule,” also is credited for creating a method used to approximate the upper end of an income distribution. Pareto's distribution is an exponential decay curve. The Pareto distribution is typically used to extend income ranges for very large areas, such as whole countries, where income distributions are smooth. The application of Pareto methods for small areas, where distributions can be irregular, requires some care.

Claritas applied the Pareto extension to the 2000 census income data only. Estimated and projected income for the extended categories were produced with standard methods applied to the extended 2000 census base.

## **Inflation and Income**

A common question is how the effect of inflation is accounted for in the Claritas income estimates. Inflation, as commonly measured by the Consumer Price Index, reflects changing prices, and a corresponding change in the value of a dollar. For example, items that would have cost \$100 in 1983 would have cost about \$147 by 1993--a 47 percent inflation in prices. Thus \$100 was not the same in 1993 as it was in 1983.

Inflation is not a measure of income change, but the two are related. Some income sources (such as Social Security and some union contracts) are “indexed” by inflation, and workers typically require and demand more pay to cover the increased costs of living. Although income tends to follow inflation, it does not move at the same rate. There are periods when income growth outpaces inflation, and periods when it lags behind. These income changes relative to inflation are referred to as “real” income growth.

## ***The Claritas Demographic Methodology***

The Claritas income estimates and projections are expressed in current dollar values, which reflect how many dollars are being received at the relevant year. As such, they reflect both “real” income growth (or decline) and the change due to the effect of inflation. Rather than estimating the effects separately, Claritas measures the combined or net effect through input sources (such as the Bureau of Economic Analysis income estimates) which themselves estimate income change in current dollars. The inflation effect built into these estimates is implicitly incorporated into the Claritas estimates. Note that “accounting for inflation” in this manner is different from “controlling for inflation”, which requires removing the effect of inflation, to produce estimates in constant dollar values.

**This page left intentionally blank.**