

**Claritas Update
Demographics
Methodology**

© 2008 by Claritas Inc. All rights reserved.

Warning! The enclosed material is the intellectual property of Claritas Inc. (Claritas is a subsidiary of The Nielsen Company, a global leader in information services for the media and entertainment industries.) This document and its contents are protected under Federal Copyright Laws, Title 17 of the U.S. Code. Under this copyright law, unauthorized users may be subject to civil liability including an injunction, actual damages, infringer's profits, and statutory damages of up to \$100,000. Criminal penalties may include a fine of up to \$25,000 and/or up to one year's imprisonment. All reproduction of this document without the expressed written permission of Claritas Inc. is strictly prohibited. By reading further in this document, you are implicitly agreeing to these terms and conditions.

Claritas Update Demographics[™] is a trademark of Claritas Inc. All rights reserved.

TotalSource[™] is a trademark of Equifax.

Other company names and product names are trademarks or registered trademarks of their respective companies and are hereby acknowledged.

Contents

- Introduction 1**
 - Overview 1
 - Experienced Leadership 1
 - Evaluation and Support Materials 2
 - Annual Demographic Update 2
 - Methodological Changes Implemented in 2008 2
 - Enhanced Aggregation of Estimates to ZIP Codes 2
 - New Data Sources 3
 - American Community Survey (ACS) Data 3
 - Top-Down vs. Bottom-up Methodologies 3
 - Trending 3

- Variable Categories 4**

- Geography 5**

- Data Overview 6**
 - Base Counts 6
 - Population Characteristics 6
 - Population by Age/Sex 6
 - Population by Race/Ethnicity 6
 - Household Characteristics 7
 - Household Income 7
 - Household Size 7
 - Income by Age of Householder 7
 - Housing Unit Characteristics 8
 - Housing Value 8
 - Smoothed Data 8

- Base Counts 9**
 - Total U.S. 9
 - State 9
 - County 9
 - Place 10
 - Census Tract 10
 - Event Tracts 11
 - Military Base Closings 11
 - Rapid Change Review 11
 - Adjustment Process 11
 - Five-year projections 12
 - Block Group 12
 - ZIP Code Estimates and Projections 12
 - Census Data for ZIP Codes 13
 - ZIP Code Data from the Claritas Update Demographics 13

- Population Characteristics 15**
 - Population by Age/Sex 15
 - Accounting for Births 15

Exceptions to Cohort Survival.....	15
Five Year Projections	16
Population by Race and Ethnicity	16
Estimates and Projections of Race and Hispanic Ethnicity	16
Race Bridging	17
Five-Year Projections	17
All-Inclusive Race	17
Population by Age/Sex by Race/Ethnicity	18
Household Characteristics	19
Households by Income.....	19
Income Estimation Method.....	19
Five Year Projections	20
Family Household Income	21
Household Effective Buying Income	21
Income by Age of Householder	21
Income by Race and Ethnicity of Householder.....	22
Householders by Race and Ethnicity	22
Households by Size.....	23
Households by Year Moved Into Unit	23
Housing Unit Characteristics	25
Housing Value	25
Five Year Projections.....	25
Housing Units by Year Built.....	26
Smoothed Data	27
Additional Terminology	28
Block Group Parts.....	28
Consistency of Complete Count and Sample Census Totals.....	28
Adjustment Techniques.....	28
Ratio Adjustment	28
Iterative Proportional Fitting	29
Income Distributions.....	29
Ethnic Classifications	29
Extended Income and Pareto Interpolation.....	30
Inflation and Income	31

Introduction

Overview

This document will outline the methodology behind the Claritas Update Demographics™. With over 30 years of experience, Claritas has a proven track record as an industry leader producing quality data. Claritas has been able to maintain this position by continually looking for new ways to innovate and improve upon their update methodology, taking advantage of new external resources that can lead to increased accuracy, and continually keeping an eye to the future as well as changes that will impact the demographic estimation process (e.g. the replacement of the Census Long Form with the American Community Survey).

This document will outline the methodology process for the development of the 2008 update that includes household and population data covering over 4,000 demographic variables from the 1990 and 2000 census as well as current year estimates and five-year projections.

Experienced Leadership

The Claritas demographics team is headed by Dr Ken Hodges who brings a 25-year history of groundbreaking contributions to the field of applied demography. After earning his doctorate in Sociology and Demography from Cornell University, Dr. Hodges joined Donnelly Marketing Information Services (DMIS) in 1982 before moving to Claritas in 1993.

During his tenure, Dr. Hodges has developed and refined methods for using consumer database information as input to small area estimates, and adapted standard demographic methods for use in nationwide small area estimates.

In addition to his work at Claritas, Dr. Hodge's professional involvements include the following:

- Member, 2010 Census Advisory Committee
- Board Member, Association of Public Data Users
- Board Member, Council of Professional Associations on Federal Statistics
- Member, Population Association of America Committee on Population Statistics
- Member, ICSC North America Research Taskforce

In addition to the extensive experience of Dr Hodges, the Claritas Demographics Estimation Program draws upon the strengths of some of the industry's pioneering programs including: National Planning Data Corporation, Donnelly Marketing Information Services, National Decision Systems, and Market Statistics.

Evaluation and Support Materials

The Claritas Estimation Program is supported by extensive research and evaluation, with results often 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.
- Comparison of 1996 estimates and 2001 projections from alternative suppliers.
- Evaluation of geometric data retrieval methods.

Annual Demographic Update

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

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

The update begins 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.

Methodological Changes Implemented in 2008

With over three decades of experience to learn from, Claritas is always looking for ways to innovate and improve on estimation techniques in order to adapt to a changing demographic landscape and to take advantage of new resources. In 2008, the following changes were implemented:

Enhanced Aggregation of Estimates to ZIP Codes

In some very small, high growth ZIP Codes, rapid growth was successfully estimated for block groups, but was not being fully allocated to the correct ZIP Code. For example, growth that should have been allocated to ZIP Code A was allocated to adjacent ZIP Code B, thus leaving ZIP Code A inconsistent with residential address counts from the United States Postal Service (USPS). This enhanced aggregation provides more up-to-date factors by which to identify the percent of block group growth in ZIP Code A versus ZIP Code B.

Note: This enhancement did not reflect a change to the estimation method, but rather to the process used to aggregate block group estimates to ZIP Codes. For a more detailed discussion of how ZIP Code estimates are built, please see the “Zip Code Estimates and Projections” section.

New Data Sources

Effective with the 2008 Update, ADVO data on new housing units were introduced. This source provides a powerful indication of which small areas have experienced significant housing growth in recent years, and is making valuable contributions to the Rapid Change Review process.

Also effective with the 2008 update, Acxiom data on housing values (based on parcel-level records) are contributing to the Claritas estimates of home value in selected markets.

American Community Survey (ACS) Data

Although the Census Bureau's American Community Survey (ACS) will not provide block group data until 2010, it is already providing valuable estimates for states and large counties (with populations of 65,000 or more). And with new ACS estimates released every year, the Claritas Update continues to make use of ACS income and home value estimates to guide our estimates .

Top-Down vs. Bottom-up Methodologies

Over the years, some descriptions of small area estimation methods have emphasized a distinction between top-down and bottom-up methodologies. The distinction is overdone, as estimates produced with pure top-down or bottom-up methods are rare. A pure top-down method might produce estimates at the county level, and then distribute to tracts and block groups without reference to any post-census tract and block group data. This approach has obvious and severe limitations for small areas. On the other hand, a pure bottom-up method would use post-census block group data to produce estimates at that level, and then sum to all larger areas, without reference to independent estimates for larger areas such as cities, counties, and states. This approach has severe limitations for larger areas.

Claritas estimates combine the strengths of top-down and bottom-up methods. Post-census data for tracts and block groups drive the estimates at those levels, while Post-census data for cities, counties, and states drive the estimates at those levels. The small area estimates are then adjusted to conform with the independent estimates for larger areas. This approach is common because it is designed to achieve maximum accuracy at all levels. Thus it cannot be described as either top-down or bottom-up.

In a bottom-up approach Claritas identifies growth trends using small data sources such as USPS, local government agencies (where available), ADVO new construction data, and the Claritas Master Address File. The Master Address File contains over 135 million records and is informed by various household level sources such as Equifax, ADVO, InfoUSA, Targus, and TeleAtlas North America.

Once the trends have been identified and estimates have been generated at the block group level, a top-down approach is taken in applying controls that will prevent the generation of unrealistic estimates at larger geographic levels such as county or state. Claritas utilizes county level data provided annually by the Census Bureau and other federal agencies as control totals for the small area estimates.

Trending

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.

The target date for estimates and projections is January 1 of the relevant year.

Variable Categories

Claritas Update Demographics includes the categories and their respective data items listed below.

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 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 Households by income by race of householder Households by income by ethnicity of householder 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	Mean and median household income Mean and median family household income Mean and median Effective Buying 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 geographic areas including the following:

Standard Macro Geographies	Standard Micro Geographies	Industry Geographies
Country [1]	Place [25K+]	Cable [9K+]
Divison* [9]	Tract* [60K+]	Gas Utility [1K+]
State* [51]	Block Group* [200K+]	Electric Utility [2K+]
County* [3K+]	Block* [6M+]	Wire Centers [20K+]
Designated Market Area (DMA) [200+]	Zip Code [30K+]	Yellow Pages Directory (YPD) [7K+]
Core Statistical Area (CSA) [100+]	Minor Civil Division (MCD) [Includes Census Civil Divisions (CCD)*] [35K+]	Major & Basic Trading Area Wireless (MTA & BTA) [500 +]
Core Based Statistical Area (CBSA) [900+]		Metro/Rural Service Area for Wireless (RSA) [700+]
Three-Digit Zip (TDZ) [800+]		Federal Congressional Districts [400+]

* Census geographies

Data Overview

Base Counts

Base counts are the basic totals for population, households, family households, group quarters population, and housing units. At the national, state, county, and place geography levels, base count updates are based on estimates from the Census Bureau and, in some cases, state demographers. At the census tract and block group levels, base count information is based on sources including local estimates, trends in United States Postal Service (USPS) deliverable address counts, counts from the Claritas MAF, trends in consumer counts from the Equifax TotalSource™ database, and ADVOCOUNTS counts of new housing units.

Population Characteristics

Population is estimated for the following demographic characteristics:

- Age
- Sex
- Race
- Hispanic ethnicity
- Age by sex by race by Hispanic ethnicity

Population by Age/Sex

Age/sex distribution is estimated using 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 United States 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.

Household Characteristics

Households are estimated for the following characteristics:

- Household income
- Household size
- Age of householder
- Race and ethnicity
- Year householder moved into unit

Household Income

Income estimates and projections reflect the census money income definition, and are produced for current dollar values. Rates of change in median income are estimated first, 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 indicated by the Bureau of Economic Analysis (BEA) income estimates, income statistics from the Internal Revenue Service (IRS), and in large counties, income estimates from the Census Bureau's new American Community Survey (ACS).

Income change at the tract and block group levels is estimated based on a combination of:

- Change in consumer financial information from the Equifax Consumer Marketing database
- Change in income summarized from the TotalSource consumer household database
- Projections of inter-censal trends

Distributions of 2000 census income are 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.

Household 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 (IPF) is then used to ensure consistency with previously estimated household totals and average household size. For more information about [IPF](#), see the "Additional Terminology" section.

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 estimates 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 that not only sum to the income and age estimates, but also preserve the statistical relationship between income and age for each area as measured by the census.

Housing Unit Characteristics

Housing units are estimated for the following characteristics:

- Total count of owner-occupied units
- Value of owner-occupied units
- Year structure built

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 then used to advance the 2000 census distribution to current and projection year.

At the state and national levels, target rates of change in value are based on change in value estimated by the 2006 ACS, as well as change in the House Price Index from the Office of Federal Housing Enterprise Oversight (OFHEO).

At county level, the OFHEO data is combined with change in median sales price data from the National Association of Realtors to estimate change. An additional data source contributing to the estimates in counties with large populations, is the annual data from the ACS that is currently available for counties with populations in excess of 65,000. Tract rates of change are based on a combination of projected inter-censal trends and post-2000 change in average mortgage amounts from the Equifax Consumer Marketing database. And in some markets, home value data from the Axiom files contribute to the small area estimates of home value.

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 both controlled to the next higher level.

Smoothed Data

In addition to the annual demographic estimates and projections, the update provides a series of detailed census tables that have been ratio adjusted, or smoothed, to relevant current-year totals. These tables purport only to show the effect of applying decennial census distributions to estimated base count totals at the block group level.

Base Counts

For this document, base counts include basic totals such as population, households (occupied housing units), family households (households with two or more related persons), group quarters population (persons in dormitories, military quarters, prisons, nursing homes, and other non-household living arrangements), and housing units (a house, apartment, or group of rooms intended to serve as separate living quarters).

Total U.S.

Total U.S. population is estimated using Census Bureau estimates of total United States resident population (all persons residing in the United States, regardless of citizenship). The 2008 estimate was a short projection beyond the Census Bureau's most recent post-2000 estimate. Total group quarters population is estimated in a similar manner, based on the Census Bureau's most recent estimates for group quarters population.

Total estimated households are derived by subtracting the estimated group quarters population from the estimated total population to derive the total number of persons in households. This figure is then divided 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 household size data from the 2006 ACS.

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

State

State population counts are projections from the Census Bureau's most recent population estimates at the state level. (Census 2006 estimates were used for the 2008 update.) Household counts are estimated indirectly from the completed population estimates. Specifically, the group quarters population from the 2000 census is estimated forward to determine the estimated 2008 group quarters population. This number is then subtracted from the estimated population to determine estimated households. The result is divided by estimated average household size (based on inter-censal trends and ACS data) to determine estimated households. Total family households and housing units are estimated by applying Census 2000 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 are usually 18 months behind the Claritas estimation date, so a series of long- and short-term projections is produced for the target date (in this case, January 2008). The mean of these projections serves as the census-based county population estimate. Where state-produced estimates are available, and it has been shown that including these estimates increases estimation accuracy, these estimates also are projected to current year, and averaged with the census-based estimates. The resulting estimates are then adjusted to conform with the state population estimates described above.

As at the state level, household estimates were derived by subtracting estimated group quarters population (based on Census Bureau county group quarters estimates) from total population, and dividing by estimated persons per household. The county estimates of persons per household are

based on a combination of intercensal trends and, for large counties, a careful integration of household size data from the 2006 ACS.

Place

Population estimates for places and county subdivisions such as cities and towns are based on Census Bureau estimates for these geographic units. Population estimates for these units, as well as 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

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 as input for 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.

Among the data sources contributing to the 2008 tract level estimates are:

- Estimates produced by local governments or planning agencies
- Counts of deliverable residential addresses from the U.S. Postal Service
- Household counts from the Equifax TotalSource consumer database
- Counts from the Claritas Master Address File (MAF).
- ADVO counts of new housing units
- 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 is often the best information available on small area trends, Claritas contacts a large number of local agencies each year to obtain, review, and incorporate the work being done by local demographers and planners.

The local data does not come in a neat package. Methodologies, dates, and content all vary among the different organizations providing the data. For example, some local sources estimate population, while others estimate households, and still others count housing units. Once obtained, the data is then reviewed and prepared for input into 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 for areas where local estimates are not available. The most widely-used alternative source consists of deliverable residential address counts from the USPS. 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 Census 2000 household counts to establish preliminary tract-level household estimates. The conversion of postal address counts from carrier routes to tracts is subject to uncertainty in some areas, so separate conversions are made using 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.

Counts from the Claritas MAF provide another indication of change since the 2000 census. In contrast to address counts from the USPS Delivery Statistics files, the Claritas MAF combines, scrutinizes, and eliminates duplication among the addresses from a variety of sources, and geocodes these addresses to small area census geography. A byproduct of this process is a count of addresses that Claritas has verified for small areas.

The ADVO source is new to the Update effective with 2008, and provides counts of newly constructed housing units, thus identifying the extent of recent housing growth.

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

Event Tracts

The preparation of tract estimates is, by necessity, an exercise in demographic mass production, but Claritas demographers take the time to identify selected tracts that merit individual attention (e.g. Hurricane Katrina). It is during the review of local estimates and USPS address counts that Claritas demographers account for events such as earthquakes, fires, and hurricanes, which can have a dramatic impact on the population of specific areas. Local, USPS, Equifax, Claritas MAF, and now ADVO data all contribute to this effort, as does consultation with local demographers.

Such consultation was critical to the early efforts to estimate the impact of Hurricane Katrina example. Today, our data sources are reflecting the impact and subsequent recovery, but estimation in these areas still is not routine, so we continue to work closely with local data experts.

Military Base Closings

Claritas tracks military base closings and realignments, and estimates their impact on population at the census tract level. This effort is accomplished using military employment data from the Defense Manpower Data Center, which reports the timing and magnitude of downsizing by installation.

Rapid Change Review

In a process called the rapid change review, tracts where input data indicates dramatic growth or decline are reviewed individually. The review process involves a closer examination of the input data as well as cross checks among the alternative sources described above, with satellite images used for verification in some cases. The rapid change review identifies tracts where estimates are done by hand rather than relying on regular mass production efforts.

Adjustment Process

The preliminary household estimates described above are not adjusted directly to larger geographic levels, but are instead used as the basis for tract-level population estimates, which are then adjusted to the place and county subdivision estimates described above. Estimated group quarters population is then subtracted, and the remaining household population is 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 for small areas, but Claritas estimates do reflect changes in group quarters since 2000. Census Bureau estimates of group quarters population are the primary input source at the county level and above. A few local areas provide tract-level 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 that would affect these figures.

Family household estimates are obtained by applying the Census 2000 family households/households ratio to total estimated households. Housing units are similarly estimated by applying the Census 2000 housing units/households ratio to total estimated households.

Five-year projections

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 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. In most areas, USPS deliverable address counts are the primary input. As at the tract level, changes in USPS counts since 2000 are used to estimate Census 2000 household counts forward to the current year. 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 contained 15% of its tract's USPS count in 2000 might include 18% of the tract's households by current year. This percentage increase is then used as 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 then averaged with the USPS results to complete the block group household estimates. Counts from the Claritas MAF and ADV0 new housing counts provide additional input for the current distribution of households at the block group level. Population estimates are derived from the household estimates using methods similar to those outlined for tracts.

ZIP Code Estimates and Projections

Estimates and projections for ZIP Codes are aggregations of estimates already prepared for full and partial block groups. As such, there is not a distinct ZIP Code methodology. However, it is important to understand the process used to build ZIP Code estimates as well as the complications involved in analyzing ZIP Code data.

ZIP Code demographic data is widely used, but involves complications not encountered with other geographic areas. ZIP Codes are defined by the USPS for the delivery of mail, not for the presentation of data. They lack definitive boundaries, and change frequently at the determination of postal officials. In addition, ZIP Codes do not conform to the boundaries of other geographies such as counties, cities, census tracts, or census blocks.

Further complicating the specification of ZIP Code demographics is the imperfect relationship 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 specified location such as a post office in a nearby town. The boundaries of such general delivery and 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 in 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.

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 not widely used. With Census 2000, the Census Bureau included data for what it calls ZIP Code tabulation areas (ZCTAs). These ZCTAs approximate ZIP Code areas based on the allocation of whole census blocks. The Census Bureau points out that ZCTAs are not ZIP Codes, and users need to understand that ZCTA data does not constitute official ZIP Code estimates. Furthermore, because the Census Bureau does not update ZCTA definitions, these definitions are now eight years out of date.

ZIP Code Data from the Claritas Update Demographics

Claritas ZIP Code estimates and projections are aggregations of Claritas estimates for block groups and block group parts. The process used is similar to that for retrieving data for circles and polygons. Census data, including estimates and projections, already exists for block groups, and is aggregated to the current roster of ZIP Codes reflecting current definitions. Data for all years (including 1990 and 2000 census data) is 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 one or more times each year. This correspondence is based on the location of block centroids (latitude/longitude points) within current ZIP Code boundaries estimated by Tele Atlas. If a block's centroid falls within a ZIP Code boundary, it is allocated to that ZIP Code. These block-to-ZIP Code allocations determine the block groups (or partial block groups) that are included in a given ZIP Code. For block groups allocated to more than one ZIP Code, percent inclusion factors determine the percent of the block group allocated to each ZIP Code. In the past, these percent inclusion factors have been based upon 2000 census block population counts. However, in an effort to improve and innovate the update methodology, starting with the 2008 update, the Claritas Master Address File with over 135 million records has replaced 2000 Census block population counts as the base for determining percent inclusion factors. For all ZIP Codes with a GDT boundary, the resulting block group-to-ZIP Code relationship 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 or general delivery ZIP Codes. These ZIP Codes serve residents in rural areas where there is no mail delivery; residents pick up their mail at a central location such as a post office. 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 is 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 appears 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. Furthermore, 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 USPS. 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.

Population 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 2005 who survive another five years, will be age 40 in 2010. Accordingly, a population with a large proportion of 35 year olds in 2005 can expect to have large proportions of 40 year olds in 2010. 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 versus 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 (short of the January 1, 2008 estimate date), so a second survival is performed to 2010, and the results interpolated to January 2008. In the case of county estimates starting with July 2006 Census Bureau age/sex estimates, the single survival extends to July 2011, and the results are interpolated to January 2008.

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 facilities have high turnover rates, resulting in age/sex compositions that tend to be stable as a reflection of 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 2008 update, the 2008 estimates of population by age/sex were the starting point for five-year survivals to 2013. 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

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.

For more information about [Ethnic Classifications](#), see the “Additional Terminology” section.

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.

The application is not straightforward, since the Census Bureau’s race estimates reflect a modified definition, in which persons marking “**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% 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 inter-censal 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 it 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 suggests 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 suggests 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 suggests 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. This 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 methods similar to those used for the current-year estimates—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 the 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 with the previously completed estimates of population by age/sex and population by race/ethnicity. This application of IPF 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.

Household Characteristics

Households by Income

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 2008 estimates reflect 2008 income for 2008 households.

The method starts by estimating rates of change in median 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 was first produced for large areas, then for progressively smaller areas, with successive ratio adjustments ensuring consistency between levels. Aggregate, average, and median 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 following income ranges:

- \$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. 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. For more information about [Pareto](#) methods, see the “Additional Terminology” section.

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 ACS. 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 IRS wage and salary data, and BEA estimates of per capita income. Because national IRS and BEA income data tends 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 and BEA data, as well as a projection of the ratio based on 1990-2000 census trends.. For the 2008 update Claritas was able to incorporate Census ACS household income estimates

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 and BEA data, and the projection of 1990-2000 census trends. As at the state level, the ACS estimates are making a greater contribution as the ACS now provides income estimates for counties with populations of 65,000 or more.

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 ECMD data item “Average sum of credit limits for bank, national credit card, savings & loan, and credit union revolving accounts” 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 with 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

Five-year projections of income begin with the projection of current-year median household income to the projection year, and the advancing of the household income distribution to reflect the projected change. Income projections for sub-national areas are produced by projecting estimated change in median income 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 the national level, county projections to the state level, and so forth.

Family Household Income

A family household is one in which the householder is related to one or more persons living in the household. Family households also include any other non-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.

Household 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 provides 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 EBI.

Income by Age of Householder

The cross-tabulation of household income by age of householder is valuable because income and life cycle stage, when combined, 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 constitutes 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 is 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 with 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 the race and ethnicity of the 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 the current (or projected) year, and the resulting distributions are adjusted to conform with the householder by race/ethnicity estimates and projections described below.

Householders by Race and 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.

Households by Size

Working at the block group level, estimates of households by size (number of persons) are produced for the following categories:

- 1 person
- 2 persons
- 3 persons
- 4 persons
- 5 persons
- 6 persons
- 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.

Households by Year Moved Into Unit

Survival probabilities for “Year moved in” 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.

Housing Unit Characteristics

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 is 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 is 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 estimated based on change in the House Price Index data from the Office of Federal Housing Enterprise Oversight (OFHEO) as well as home value estimates from the ACS.

The ACS is now providing national and state home value estimates based on its full implementation sample, 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. Effective with the 2008 update, the home value estimates provided by the ACS for large counties provide further guidance to the Claritas county estimates.

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 variable "Average original balance on mortgage accounts" is strongly associated with home value. Claritas has compiled tract summaries of this variable for all census tracts for years dating 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 estimated home values. In selected markets, additional neighborhood level input is provided by Acxiom home value data aggregated from parcel level records.

Five Year Projections

Five-year projections of value are based on rates of change derived from change in median value from the 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 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 between the 1990 and 2000 censuses. For example, the number of units built between 1960 and 1969 surviving in 2000 is lower than the number of such units surviving 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.

Smoothed Data

In addition to the annual demographic estimates and projections, the Claritas Update provides a series of detailed census tables that are 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+ 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 is produced at the block group level on a bottom-up basis, percent distributions for aggregations (any area including more than one block group) can 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 is a valuable component of the annual Demographic Update. The list of smoothed data items is as follows:

- 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 Terminology

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 is 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 is provided as published, but has 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 that 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 sums to tract totals, which always sums 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 IPF. 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% ($561/552$) to achieve the target tract total. Similar adjustments are made to bring preliminary distributions (such as age and race) into conformity with population totals for each geographic unit.

Iterative Proportional Fitting

IPF methods are an elaborate form of ratio-adjustment, and are used when estimates must be adjusted to conform simultaneously with 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 by 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 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 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 sums, 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 is 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 2008 update provides estimates of 2008 households by income earned in 2008. When comparing such estimates against the census, note that total households represent a seven year change since 2000, while income represents an eight year change since 1999.

Ethnic Classifications

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.

Single Classification

The single classification option includes the following categories:

- 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 prior to the 2000 census, 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.

All-inclusive Classification

The total number of persons marking a given race category is revealed by the following all-inclusive categories:

- 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.

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% 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 was 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% 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 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.



TETRAD COMPUTER APPLICATIONS INC.

www.tetrad.com

PO Box 9754
Bellingham, WA 98227-9754

Tel 800-663-1334
Fax 360-734-4005

info@tetrad.com



TETRAD Computer Applications Inc. is a Licensed Distributor of Nielsen Claritas Data:

- Demographic Updates
- Business-Facts
- Consumer Buying Power
- Retail Market Power
- PRIZM Clusters