2013 POPULATION FILE FOR USE WITH AHRQ QUALITY INDICATORS™
Version 4.5

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
http://www.qualityindicators.ahrq.gov

Contract No. HHSA290201200001C

Prepared by:
Battelle
505 King Avenue
Columbus, OH 43201

May 2013

Table of Contents

List of Tables

List of Figures

1.0 Overview

The Agency for Healthcare Research and Quality (AHRQ) Quality Indicators (QI)™ include 36 area-level indicators (Table 1). These indicators are intended to measure healthcare quality across the population in a geographic area rather than for a single facility or provider. With a few exceptions, as noted in Table 1, the denominators for area-level indicators are the population of the area being examined, subset by age or (for some indicators) gender. The denominators for these indicators must be constructed from an outside source rather than drawn from a subset of discharges in the user's input file.

The objective of this document is to describe how the population data estimates are derived from public-use census data for use with the SAS QI Software Version 4.5 (SAS QI v4.5) and Windows QI Software Version 4.5 (WinQI v4.5). Population figures through 2013 for use with SAS QI v4.5 are provided in the file POP95T13.txt, available as a separate download on the AHRQ QI website. Population data are built into the installation package for WinQI v4.5.

Table 1. AHRQ QI Area-Level Indicators

IQI #26 Coronary Artery Bypass Graft (CABG) Rate	PQI #10 Dehydration Admission Rate
IQI #27 Percutaneous Coronary Intervention (PCI) Rate	PQI #11 Bacterial Pneumonia Admission Rate
IQI #28 Hysterectomy Rate	PQI #12 Urinary Tract Infection Admission Rate
IQI #29 Laminectomy or Spinal Fusion Rate	PQI #13 Angina Without Procedure Admission Rate
PSI #21 Retained Surgical Item or Unretrieved Device Fragment Rate	PQI #14 Uncontrolled Diabetes Admission Rate
PSI #22 Iatrogenic Pneumothorax Rate	PQI #15 Asthma in Younger Adults Admission Rate
PSI #23 Central Venous Catheter-Related Blood Stream Infection Rate	PQI #16 Lower-Extremity Amputation among Patients with Diabetes Rate
PSI #24 Postoperative Wound Dehiscence Rate	PQI #90 Prevention Quality Overall Composite
PSI #25 Accidental Puncture or Laceration Rate	PQI #91 Prevention Quality Acute Composite
PSI #26 Transfusion Reaction Rate	PQI #92 Prevention Quality Chronic Composite
PSI #27 Perioperative Hemorrhage or Hematoma Rate	PDI #14 Asthma Admission Rate
PQI #1 Diabetes Short-Term Complications Admission Rate	PDI #15 Diabetes Short-Term Complications Admission Rate
PQI #2 Perforated Appendix Admission Rate[1]	PDI #16 Gastroenteritis Admission Rate
PQI #3 Diabetes Long-Term Complications Admission Rate	PDI #17 Perforated Appendix Admission Rate1
PQI #5 Chronic Obstructive Pulmonary Disease (COPD) or Asthma in Older Adults Admission Rate	PDI #18 Urinary Tract Infection Admission Rate
PQI #7 Hypertension Admission Rate	PDI #90 Pediatric Quality Overall Composite
PQI #8 Heart Failure Admission Rate	PDI #91 Pediatric Quality Acute Composite
PQI #9 Low Birth Weight Rate[1]	PDI #92 Pediatric Quality Chronic Composite

[1]These indicators use discharge data from the input data file to estimate the denominator rather than demographic data from the population file.

2.0 Data and Methodology

Every year, the Census Bureau releases postcensal population estimates[†] (as of July 1 of each year) that are generated with the assistance of the Federal State Cooperative Program for Population Estimates (FSCPE) using residence, total births, total deaths, and net migration. With each new issue of July 1 estimates from the Census Bureau, the Census Bureau makes revisions to all years back to the last decennial census. Each decade, after a decennial census, the Census Bureau produces a set of intercensal estimates that provide annual population estimates that are adjusted to smooth the transition from one decennial census to the next. These estimates are used to derive the AHRQ QI Population File to be used with the AHRQ QI software.

2.1 Census Data Files

Public-use files of intercensal and postcensal estimates of county-level population by five-year age group, sex, race, and Hispanic origin were acquired from the Census Bureau (http://www.census.gov/popest/) covering the years 1995 through 2011. Table 2 presents detailed information and sources for the specific files acquired and used to generate the POP95T13.txt file for use within the AHRQ QI software.

[†] *"Estimates are for the past, while projections are based on assumptions about future demographic trends. Estimates generally use existing data collected from various sources, while projections must assume what demographic trends will be in the future"* (http://www.census.gov/population/www/projections/aboutproj.html)

Table 2. Census Dataset Descriptions and Sources.

DATA NAME	YEARS	BASE DECENNIAL YEAR	TYPE	SOURCE
Intercensal Estimates of the Resident Population by Five-Year Age Groups, Sex, Race, and Hispanic Origin for Counties	2000-2010	2010	Intercensal	http://www.census.gov/popest/data/intercensal/county/CO-EST00INT-alldata.html
Annual Estimates of the Resident Population by Age, Sex, Race, and Hispanic Origin for Counties	2010-2011	2010	Postcensal	http://www.census.gov/popest/data/counties/asrh/2011/CC-EST2011-alldata.html
Intercensal Estimates of the Resident Population by Single Year of Age and Sex for States and the United States	2000-2010	2010	Intercensal	http://www.census.gov/popest/data/intercensal/state/state2010.html
State Single Year of Age and Sex Population Estimates	2010-2011	2010	Postcensal	http://www.census.gov/popest/data/state/asrh/2011/index.html
State and County Intercensal Estimates by Demographic Characteristics	1990-1999	2000	Intercensal	http://www.census.gov/popest/data/intercensal/st-co/characteristics.html

2.1.1 Notable Differences of Population Estimates from 2000 Census to 2010 Census

There are four counties that existed for the 2000 Census, but not for the 2010 Census (http://www.census.gov/geo/www/tiger/tgrshp2010/usernotes.html):

- 02201 - Prince of Wales-Outer Ketchikan Census Area, AK
- 02232 - Skagway-Hoonah-Angoon Census Area, AK
- 02280 - Wrangell-Petersburg Census Area, AK
- 51560 - Clifton Forge city, VA

In the 2010 Census, the populations from these four counties are distributed to other surrounding counties. This means that while the POP95T13.txt file contains estimates for these four defunct counties for the years 1995-1999, the POP95T13.txt file estimates for the years 2000-2013 are listed as "0" since they are based on 2010 Census county boundaries.

2.1.2 Modifications to Census Estimates for use in the POP95T13.txt File

Modifications to the census estimates were required to fit the specifications of the AHRQ QI software. The first is the categorization of race and Hispanic origin. Table 3 depicts how the race categories used by the AHRQ QI software were defined from the census race and Hispanic origin groupings. This set of race categorizations captures the entire US population.

Table 3. Race Category Aggregations Based on Census Reporting Categories.

RACE CATEGORY	DESCRIPTION
1	Non-Hispanic, White Alone
2	Non-Hispanic, Black Alone
3	Hispanic
4	Non-Hispanic, Asian Alone OR Non-Hispanic, Native Hawaiian and Other Pacific Islander Alone
5	Non-Hispanic, American Indian and Alaska Native Alone
6	Non-Hispanic, Two or More Races

In addition, the population of interest for the area-level indicators in the Pediatric Quality Indicator (PDI) module is the population ages 17 and under, while the population of interest for the other indicator modules is the population ages 18 and older. The default five-year age groups reported by the Census Bureau are 15-19 years of age and 20-24 years of age. To capture the separation between the pediatric and adult populations, the POP95T13.txt file contains an age range that spans the ages of 18-24 that is constructed using the two default census age groups. To generate the 18-24 year old age group, state-level estimates of population by sex and single year of age (see Table 2) were used to calculate the percent of the population between 15 and 19 years old (the age grouping for the county-level data) that are between 18 and 19 years old. Then, the county-level population of 18-19 year olds was subtracted from the census-defined age group of 15-19 (to form the 15-17 age group) and added to the 20-24 age group (to form the 18-24 age group).

2.1.3 Census Data File Mapping to AHRQ QI Population File

The POP95T13.txt file population estimates for 1995 through 1999 are based on intercensal estimates by demographic characteristics (Table 2). Since these data are adjusted to the 2000 Census, they are no longer updated by the Census Bureau with more recent postcensal estimates and the estimates are unchanged from version release to version release of the AHRQ QI software.

The POP95T13.txt file population estimates for 2000 through 2010 are based on intercensal estimates by demographic characteristics that are adjusted to the 2010 Census. The POP95T13.txt file population estimates for 2011 are based on postcensal estimates by demographic characteristics that use the 2010 Census as the base.

Public-use files of postcensal population estimates from the Census Bureau are currently available only through 2011. The POP95T13.txt file contains population estimates for 2012 and 2013 based on linear projections of the population counts for each county, sex, age group, and race combination. The projections were made according to the following model:

$$y_{ijt} = \alpha_{ij} + \beta_{ij}t,$$

where i is the county (1, 2,…, 3147), j is an indicator of demographics representing a combination of sex, age group. and race (1, 2,…, 216), and t is the year (2000, 2001,…, 2011). That is, we fit a county-specific linear growth model for each demographic group. The population estimates for each county and demographic combination, $\hat{y}$, for 2012 and 2013 were calculated using the following equations:

$$\hat{y}_{ij2012} = \hat{\alpha}_{ij} + \hat{\beta}_{ij}2012$$

$$\hat{y}_{ij2013} = \hat{\alpha}_{ij} + \hat{\beta}_{ij}2013.$$

where $\hat{\alpha}_{ij}$ and $\hat{\beta}_{ij}$ are the coefficients estimated from the linear regression models.

2.2 Version History

The population file released with each version of the software is generated with the most recent data available at the time of software development. As such, this file will change from version to version (including the filename) as data are updated and released by the Census Bureau. The differences between population files for AHRQ QI software release versions can be caused by changes in population estimates themselves and/or changes in methodology. Table 4 summarizes the population files for AHRQ QI software release versions. Note that data for population files included with previous releases of the AHRQ QI software are not updated with each new release.

Table 4. Population Files Used With Various Versions of AHRQ QI Software

SOFTWARE RELEASE (FILENAME)	YEARS	BASE DECENNIAL YEAR	DATA SUMMARY	METHODOLOGY SUMMARY
v4.5 (POP95T13.TXT)	Estimates: 1995-1999	2000	(1) Sex/Age/Race by County (2) Age 18-24 by State	Permutated file of sex/age/race by county Used state estimate of population from 18-24 to break 15-19 and 20-24 age groups into 15-17 and 18-24
	Estimates: 2000-2011 Projections: 2012-2013	2010	(1) Age/Sex/Race by County (2) Age (single year) by State	Permutated file of sex/age/race by county Used state estimate of single year of age to break 15-19 and 20-24 age groups into 15-17 and 18-24
v4.4 (POP95T12.TXT)	Estimates: 1995-1999	2000	(1) Sex/Age/Race by County (2) Age 18-24 by State	Permutated file of sex/age/race by county Used state estimate of population from 18-24 to break 15-19 and 20-24 age groups into 15-17 and 18-24
	Estimates: 2000-2010 Projections: 2011-2012	2010	(1) Sex/Age by County (2) Sex/Race by County (3) Age (single year) by State	Combined sex/age and sex/race files by county to get estimates of sex/age/race Used state estimate of single year of age to break 15-19 and 20-24 age groups into 15-17 and 18-24
v4.3 (POP95T11.TXT)	Estimates: 1995-2009 Projections: 2010-2011	2000	(1) Sex/Age/Race by County (2) Age 18-24 by State	Permutated file of sex/age/race by county Used state estimate of population from 18-24 to break 15-19 and 20-24 age groups into 15-17 and 18-24

2.2.1 Comparison of v4.4 and v4.5

At the time of the AHRQ QI v4.4 development, the Census Bureau had not yet released the intercensal estimates of population by age, sex, race, and Hispanic origin at the county level for the years 2000 through 2010 that were updated to be consistent with the 2010 Census. In order to use the most recent data available, two separate, county-level files (one containing sex and age and the other containing sex and race) were merged to generate the estimates by sex, age, and race. In this fashion, the distribution of age group categories was applied evenly across all race categories (e.g. the percentage of non-Hispanic white males estimated to be between 0-4 years old was equal to the percentage of Hispanic males estimated to be between 0-4 years old for a given county). A comparison of the v4.4 and v4.5 files (with the v4.5 files based on updated 2010 Census data) revealed that this assumption is not necessarily true for all counties and races.

For counties where the age group by race distribution is approximately equal to the total age group distribution (i.e., not race dependent), there are not large differences between the population file used in v4.4 and that used in v4.5. However, for counties that have significantly different age group distributions for different races, large differences may be observed. For example, Figure 1 shows a comparison of the population estimates for two small counties (total populations less than 30,000). Each individual symbol (n=216) on the plot represents a gender, age group, race observation for the county. The blue, dashed lines indicate a ±10% deviation from the one-to-one line indicating perfect agreement between the v4.4 and v4.5 estimates.

Madison County, Iowa has a predominantly non-Hispanic white population (>97% in 2009), resulting in estimates that agree very well between the two population files, while the estimates for Morehouse Parish, Louisiana, which has almost equal non-Hispanic white (51%) and African American (47%) populations, demonstrate some large differences. These differences between v4.4 and v4.5 occur because the two race categories in this county have different age group distributions, while the v4.4 methodology applied a single distribution across all races.

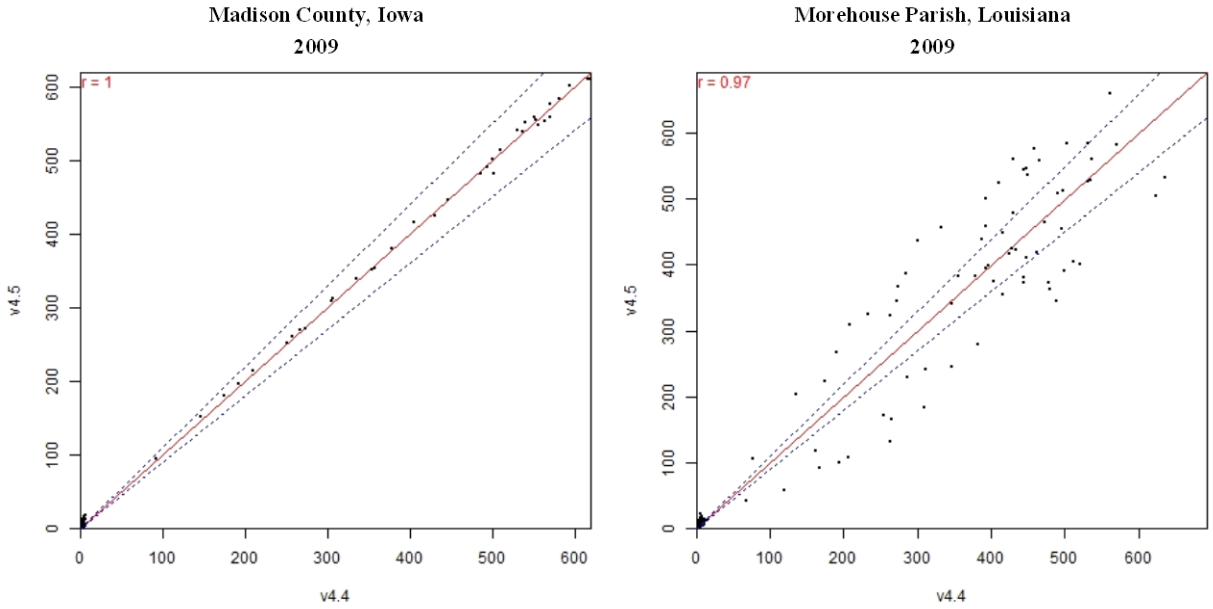

Figure 1. Comparison of population estimates for v4.4 and v4.5 by gender, age group and race for Madison County, Iowa (left) and Morehouse Parish, Louisiana (right) for the year 2009. The blue dashed lines represent ±10% deviation from the one-to-one line (red line).

While the previous example was for two small counties, similar results are observed for large counties. Figure 2 shows the comparison of population estimates by gender, race, and age group for Miami-Dade County, Florida and Maricopa County, Arizona (Phoenix). The estimates for Miami-Dade County are more similar between population file versions than for Maricopa County. The age group distributions for the three races that contribute most to the total population of Miami-Dade County, Hispanic (65%), African American (17%) and non-Hispanic white (16%) all follow a very similar pattern, resulting in smaller deviations in the v4.5 estimate from the v4.4 estimate. However, in Maricopa County, the distributions for the two largest-contributing race groups, non-Hispanic whites (59%) and Hispanics (29%), follow very different patterns, meaning that the age group distribution applied in v4.4 for all races was not representative of the age group distributions for both of these races.

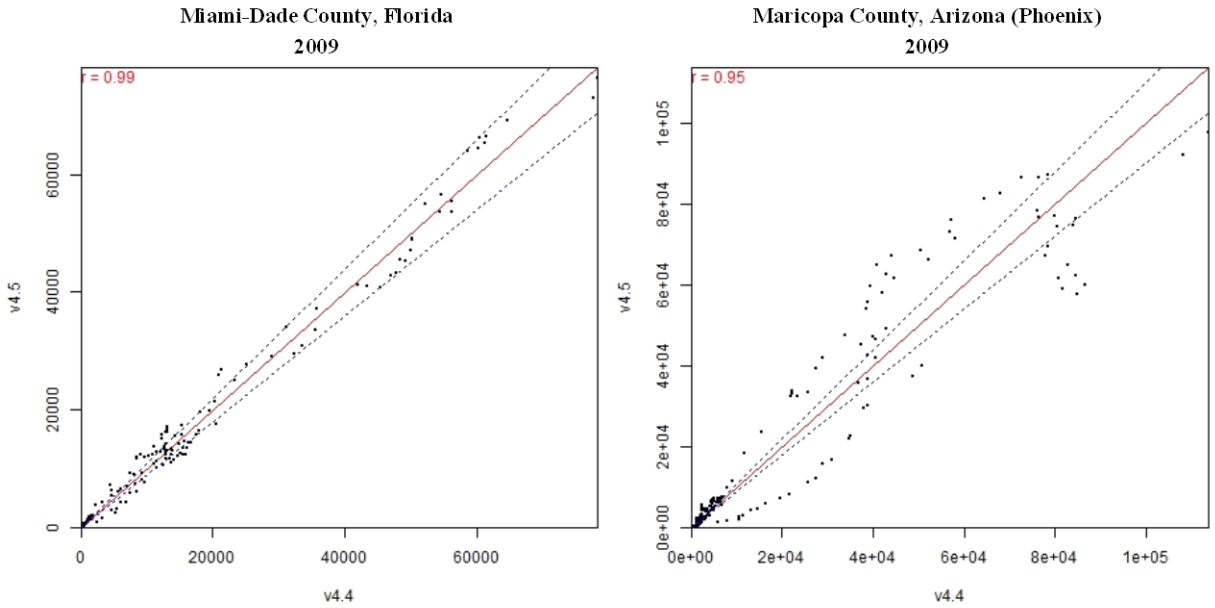

Figure 2. Comparison of population estimates for v4.4 and v4.5 by gender, age group and race for Miami-Dade County, Florida (left) and Maricopa County, Arizona (right) for the year 2009. The blue dashed lines represent ±10% deviation from the one-to-one line (red line).

To summarize, the largest effects of this change are realized in large counties that have two or more race groups that contribute large proportions to the total population, but have different age group distributions, such as Maricopa County, Arizona and Los Angeles, California (not shown here). Small counties with similar demographics (e.g. Morehouse Parish, Louisiana) will also see a difference, though the absolute differences (i.e. numbers of people) are not as large. Counties that have one dominant race category and those with more than one that have very similar age group distributions will still have changes to the population estimates, but they are likely to be relatively small.

3.0 POP95T13.txt File Specification

The POP95T13.txt file is an ASCII-based text file containing 679,752 records with a fixed logical record length of 150 bytes. It is in fixed column format. Table 5 presents the file's specific fields and the code schema used for each field.

The file is structured for use with AHRQ QI programs PQSASA2.SAS, PQSASA3.SAS, PSSASA2.SAS, IQSASA2.SAS, IQSASA3.SAS, PDSASA2.SAS, and PDSASA3.SAS, as well as the Windows QI (WinQI) software. As such, any modification to this file will affect the operation of these programs.

A given county is identified by the Federal Information Processing Standards code (FIPS code) for the state in which it is located and by the county's FIPS code. For each county within the U.S., the file contains 216 records: a record for each unique combination of gender, eighteen age groups, and six race groups. Each physical record represents a gender, age group, and race

group combination for that county and contains population estimates (rounded to integer values) for that combination for each year from 1995 through 2013.

The file has data for 3,147 counties or "equivalent areas", defined to constitute primary divisions of their states. "Equivalent areas" include the independent cities of Baltimore, Maryland; St. Louis, Missouri; Carson City, Nevada; and 39 independent cities in Virginia. Because they are independent of any contiguous county, they are treated as separate counties with their own population records. Population figures for surrounding counties exclude them. Differences in the record count from previous population files are due to changes in county definitions or such independent cities. Definitions for state and county FIPS codes can be found at http://quickfacts.census.gov/qfd/meta/long_fips.htm.

Table 5. Data Fields in POP95T13.txt

FIELD	VARIABLE	COLUMN POSITION	FORMAT	CODES
1	State	1-2	Zero Filled Numeric	FIPS Code
2	County	3-5	Zero Filled Numeric	FIPS Code
3	Sex	7	Numeric	1=Male, 2=Female
4	Age Group	9-10	Numeric	1=0-4 years 2=5-9 years 3=10-14 years 4=15-17 years 5=18-24 years 6=25-29 years 7=30-34 years 8=35-39 years 9=40-44 years 10=45-49 years 11=50-54 years 12=55-59 years 13=60-64 years 14=65-69 years 15=70-74 years 16=75-79 years 17=80-84 years 18=85+ years
5	Race	12	Numeric	1=White, 2=Black, 3=Hispanic, 4=Asian & PI, 5=Amer. Indian, 6=Other
6	1995 Population	13-19	Numeric	Integer Totals
7	1996 Population	20-26	Numeric	
8	1997 Population	27-33	Numeric	
9	1998 Population	34-40	Numeric	
10	1999 Population	41-47	Numeric	
11	2000 Population	48-54	Numeric	
12	2001 Population	55-61	Numeric	
13	2002 Population	62-68	Numeric	
14	2003 Population	69-75	Numeric	
15	2004 Population	76-82	Numeric	
16	2005 Population	83-89	Numeric	
17	2006 Population	90-96	Numeric	
18	2007 Population	97-103	Numeric	
19	2008 Population	104-110	Numeric	
20	2009 Population	111-117	Numeric	
21	2010 Population	118-124	Numeric	
22	2011 Population	125-131	Numeric	
23	2012 Population	132-138	Numeric	
24	2013 Population	139-145	Numeric	