Evaluating Tract-level Intercensal Estimates of Neighborhood Demographics and Socioeconomics for U.S. Counties 2001-2009

Margaret M. Weden[1], Christine E. Peterson[1], Regina A. Shih[1]

[1]RAND Corporation

[PRELIMINARY MANUSCRIPT: DO NOT CITE WITHOUT PERMISSION OF THE AUTHOR]

Prepared for Session 70: Applying Demography to Population Health, of the Annual Meetings of the Population Association of America, May 1, 2014, Boston, MA. We gratefully acknowledge funding from the National Institute on Aging of the U.S. National Institutes of Health (NIA/NIH) through the grant R01-AG043960 (Regina Shih, Principal Investigator).

Evaluating Tract-level Intercensal Estimates of Neighborhood Demographics and Socioeconomics for U.S. Counties 2001-2009

ABSTRACT

The American Community Survey (ACS) multiyear estimation program has greatly advanced opportunities for U.S. research on small areas such as census tracts. Challenges remain, however, for researchers studying areas smaller than the thresholds for ACS annual estimates. We evaluate intercensal estimates of tract-level demographic and socioeconomic characteristics produced via linear interpolation between the 2000 and 2010 Census and 2005-2009 ACS. Discrepancies between interpolated estimates and comparison estimates from the Population Estimates Program (PEP), the Small Area Income and Poverty Estimates (SAIPE), and ACS are measured using the mean absolute percentage error (MAPE), mean algebraic percentage error (MALPE), and percentage difference thresholds. On average 80% of the interpolated estimates of population totals were within +/- 2% of the PEP estimate, and mean absolute error in gender and racial/ethnic distributions was less than 0.3 percentage-points. Error for socioeconomic measures was larger; however for nearly all measures mean absolute error was less than 1.75 percentage-points. Findings are discussed in relationship to the use of interpolated demographic and socioeconomic data in secondary research studies.

INTRODUCTION

Research on role of neighborhoods and communities in shaping the experiences of individuals has expanded rapidly over the last two decades buoyed both by a renewed interest in the role of place in public health and human development (Kearns 1993; Macintyre et al. 2002; National Reserach Council and Institute of Medicine 2000), as well as by innovations in data collection, computing, and statistics (Auchincloss et al. 2012; Entwisle 2007; Voss 2007). Among the recommended directions for future research on individuals and place is the need to rigorously embrace the consequences of time. This includes questions motivated by life course theories about historical context, critical (age) periods, timing, sequencing, and the accumulation of advantages and disadvantages of place as individuals age (Robert et al. 2010; Sampson et al. 2002). A critical obstacle to exploring these unanswered questions, however, has been the limited availability of contextual data for small areas (e.g. at the census-tract-level) that is regularly updated. Until 2000, the only publicly available source of spatially detailed, social and economic information with consistent measurement over long periods of time and for the entirety of the U.S. has been the decennial census long form (MacDonald 2006). In order to pursue research requiring small area contextual data at a periodicity greater than every 10 years – for example in studies considering neighborhood selection and neighborhood inequality (Crowder et al. 2012; Sampson and Sharkey 2008) and accumulated exposure to neighborhood disadvantage (Do 2009; Kling et al. 2007; Ludwig et al. 2012; Wodtke et al. 2011) – the standard approach has been to apply linear interpolation to produce data estimates for intercensal years.

A primary rationale to carry out the American Community Survey (ACS) was to provide annually updated estimates of population and housing characteristics (Torrieri 2007) that are not available in the decennial census long form. Although these data have greatly advanced opportunities for studying time and place, challenges remain for researchers interested in time trends for places (such as census tracts) that are smaller than the ACS population size thresholds for annual estimates and for researchers interested in incorporating trends prior to 2006 when the ACS was fully implemented. The primary challenge is that in order to be economically feasible, the increased periodicity of data has had to come at the cost of reduced precision (MacDonald 2006). This means that, although the ACS does provide 1-year multiyear estimates of nearly all (and some additional) demographic, social, economic, and housing characteristics previously covered in the census long-form, these estimates are only available for places with populations of at least 65,000 persons. Estimates for the full range of places covered by the decennial censuses, including census tracts are available annually, but only in 5-year multiyear estimates (e.g., 2005-2009, 2006-2010, and etc.).

It has been well documented that the ACS multiyear estimates are not equivalent to the decennial point-estimates and thus must be interpreted with caution (Beaghen and Weidman 2008; McElroy 2009; U.S. Census Bureau 2009a). In specific, the ACS estimates are period estimates that refer to a continuous window of time while decennial census estimates are point estimates of a snap-shot in time. In addition, although the ACS began to provide annual and multiyear estimates as early as 2001, full implementation of the ACS was not implemented until 2006 (U.S. Census Bureau 2009b). Prior to 2006, the ACS was not representative of the entire U.S. population; people living in group quarters (GQ) such as correctional facilities, nursing facilities and college residence halls were excluded with consequent sampling bias for populations over-represented in GQ (e.g., racial/ethnic minorities, older adults, young adults, and disabled populations). In addition, sampling was conducted at a lower rate, so that (with the exception of ACS test areas) 1-year estimates prior to 2006 are only available for populations of size 250,000 or larger. Recall that even for years 2006 forward, 1-year estimates are not available for small geographic regions like census tracts whose size is less than 65,000 persons. The first ACS estimate after 2000 for

census tracts and other small areas is the 2005-2009 5-year multiyear estimate. As a result, researchers requiring small area estimates for the period between 2000 and 2006 must employ interpolation between either the 2000 Census and a 5-year multiyear ACS estimate or the 2000 and 2010 Census. The latter option of interpolating between the 2000 and 2010 Census, is of course not possible for social, economic and housing questions now assessed exclusively in the ACS.

While linear interpolation between census point-estimates has been a well-established method for producing annual intercensal census tract estimates (Crowder, Pais and South 2012; Do 2009; Kling, Liebman and Katz 2007; Ludwig et al. 2012; Sampson and Sharkey 2008), it is unknown whether and how linear interpolation should be applied between the 2000 census (point-estimate) and ACS 2005-2009 (period-estimate) to obtain census tract estimates of social, economic and housing data for the first half of the 21st century. Although the Census Bureau advises against interpreting the multiyear estimates as a mid-year point-estimate (e.g., the 2005-2009 estimate as an estimate for 2007), there are no alternative census tract point-estimate data for socioeconomic variables after 2000, Thus, in all but one known research study in which small area estimates were employed for years between the 2000 Census and first 5-year multiyear ACS estimate, the mid-point is the assumption that has been employed (Do 2009; Do et al. 2012; Jarvis 2012; Ludwig et al. 2012). For the one exception, the period of study did not extend beyond 2005 and so the researchers avoided the issue of interpolating between a point-estimate and a period-estimate by conducting linear *projection* from the 2000 Census (Crowder, Pais and South 2012).

In this study we examine questions raised by the discontinuation of the decennial census long-form and the initiation of the annual, ACS multiyear estimation program in its place, with full implementation in 2006 and with census tract geographic specificity available only in 5-year multiyear estimates. In specific: 1) how well does linear interpolation perform for obtaining annual estimates of the demographic characteristics of U.S. census tracts for the years 2001 and 2009 between the 2000 and 2010 Census; and 2) how well does linear interpolation perform for obtaining annual estimates of socioeconomic characteristics of U.S. census tracts for the years 2001 through 2006 between the last decennial assessment in the 2000 Census and the mid-point year of 2007 for the first ACS 5-year multiyear tract-level estimate in 2005-2009? In addition, although it is beyond the scope of this paper to fully examine the midpoint assumption required to produce our annual, interpolated estimates of social and economic characteristics, we also examine whether annual trends in population characteristics across the 5-year period 2005-2009 support the decision to use 2007 at the endpoint for linear interpolation between the last Census long-form census tract estimates of social and economic characteristics in 2000 and the first ACS census tract estimates in 2005-2009.

METHODS

Interpolated Data

We evaluate a series of annual, intercensal estimates of the population demographic and socioeconomic characteristics of all U.S. census tracts for the years 2001 through 2009. These estimates were obtained using linear interpolation between the 2000 Census (Summary File 1 and Summary File 3, U.S. Census Bureau), 2010 Census (Summary File 1, U.S. Census Bureau) and 2005-2009 ACS (2005-2009 American Community Survey 5-year Estimates Summary File, U.S. Census Bureau). The interpolation methodology involved the following three steps. First, we harmonized the census tract data from the 2010 Census to the 2000 census tract boundaries using a transformation matrix we calculated from the Longitudinal Tract Data Base (LTDB)(Logan

et al. forthcoming)[1]. Second, we employed linear interpolation to estimate annual population counts by age, gender, and race/ethnicity between the 2000 and 2010 Census for every census tract in the U.S. Third, we employed linear interpolation to estimate annual population counts of four social and economic characteristics (including, persons with household income below the poverty level; persons by highest educational attainment; persons in the labor force by occupational; and annual median household income) between the 2000 Census (Summary File 3) and 2005-2009 ACS multiyear estimate. For this last set of annual interpolated estimates, we defined the endpoint as the mid-point of the 2005-2009 interval, i.e., 2007. Recall that this 2007 mid-point year assumption has also been the approach taken by all other studies conducting linear interpolation between the 2000 Census and 2005-2009 ACS (Do 2009; Do, Wang and Elliot 2012; Jarvis 2012; Ludwig et al. 2012). What these studies have not done, however, is to provide any information supporting the choice of 2007 as an endpoint. As described below, we evaluate the assumption by comparing 1-year ACS estimates to the 5-year estimate for each year within the 5-year period.

Comparison Data

The 2001-2009 database of annual, intercensal U.S. census tract data we have described above is evaluated with respect to annual intercensal estimates from the U.S. Census Bureau Population Estimates Program (PEP), ACS 1-year multiyear estimates, and U.S. Census Bureau Small Area Income and Poverty Estimates (SAIPE). The PEP is our primary validation data for demographic characteristics because it is the only known data source with intercensal estimates of population counts by gender, age, race, and Hispanic origin available for all counties in the U.S. for every year 2001 through 2009 period. The county is the smallest geographic unit for which all of these data are released[2]. ACS 1-year multiyear county estimates for the period 2006 through 2009 provide a secondary source of validation data for demographic characteristics. Recall that ACS 1-year estimates are available only for counties with population of at least 65,000 persons, a sample which comprises only 26 percentage of the counties in the U.S. (U.S. Census Bureau 2013). Recall also that ACS 1-year estimates of demographic characteristics (and selected social and economic characteristics) are available as early as 2001 for the ACS test site counties and for counties with populations of at least 250,000 persons; however these estimates exclude persons living in group quarters[3].

The SAIPE is our primary source of validation data for the annual, intercensal social and economic indicators for years 2001-2006 that we estimated by interpolating between the 2000 Census and the 2005-2009 ACS. The SAIPE is the only known data source that provides estimates of selected social and economic indicators (i.e., median household income and the percentage of the population with household income below the poverty level) for every county in the U.S. for each year, going back to 2001. The ACS 1-year multiyear county estimates for 2006 provide a secondary source of validation data for these two social and economic indicators, and

[1] The LTDB provides transformation coefficients and a tract correspondence matrix for harmonizing 2000 geographic boundaries to 2010 geographic boundaries. The methodology is similar to an earlier harmonization method developed for harmonizing 1990 boundaries to 2000 boundaries (Tatian 2003). We were able to use these data to produce transformation coefficients for (reverse) harmonizing from 2010 to 2000.

[2] Intercensal estimates for selected sub-county geographies exist, but no demographic detail is provided, only total population counts. (www.census.gov/popest/intercensal/cities/files/SUB_EST00INT.pdf)

[3] The decision to exclude these data from the comparison was supported by preliminary analyses in which it appeared that the exclusion of group quarters led to significant differences in population size and demographic characteristics (notably race). This made the pre-2006 ACS data a poor reference for comparisons to interpolations based on the data for all persons in housing units as well as group quarters.

they provide the only known data with which we can validate all of the social, economic, and housing characteristics that we interpolated between the 2000 Census and the 2005-2009 ACS. Among these additional characteristics (and in addition to poverty status and median household income), we employ 1-year ACS estimates of the population distribution by highest educational attainment and professional and managerial occupations for 2006.

Analytical Strategy

We selected the county as the unit of analysis for our evaluations of the 2000-2010 interpolated database because it was the smallest geographic unit for which we could evaluate a nationally comprehensive set of geographic units. Thus, for the purposes of validation, we aggregated the interpolated census tract data to the county-level. Our analytical strategy is to compare our annual, interpolated estimates (that were aggregated to the county-level) to the following series of annual, county-level reference data:

- PEP annual estimates of total population counts, percent female, percent non-Hispanic White, percent non-Hispanic Black and percent Hispanic for all counties in each year 2001-2009;
- SAIPE annual estimates of percent of population with household income below the poverty line and median household income for all counties in each year 2001-2009;
- ACS 1-year multiyear estimates of total population counts, percent female, percent non-Hispanic White, percent non-Hispanic Black and percent Hispanic for the subset of available counties (i.e., populations of at least 65,000 persons) in each year 2006-2009; and
- ACS 1-year multiyear estimates of the percent population distribution by education, professional and managerial occupations, and household income below the poverty line, and median income for the subset of available counties (i.e., populations of at least 65,000 persons) in 2006.

We evaluate error over all counties for all of the years possible for each of the above four comparisons. Thus, error is reported for county-years for the analyses of the demographic variables using the PEP and the ACS and for the analyses of social and economic variables using the SAIPE. And, error is reported for counties for analyses employing ACS data on social and economic variables (for which only one year of data is available). In addition, in order to produce a tractable number of estimates of error, we average the county-year estimates of error over the county population size in the 2000 Census (i.e., we produce estimates for counties with <5,000 persons; 5,000-9,999 persons; 10,000-24,999 persons; 25,000-59,000 persons; 60,000-149,999 persons; and 150,000 or more persons) and over all counties in U.S. combined.

Our measures of error depend on the outcome. For total population counts and median income, we compute two measures: the mean absolute percentage error (MAPE) and the mean algebraic percentage error (MALPE) with a county-year as the unit of comparison. The MAPE provides a measure of the absolute magnitude of error (whether positive or negative) and is one of the most commonly used measures for evaluating the accuracy of population count estimates. The formula for the MAPE is as follows:

$$\text{MAPE} = \left(\left(\sum_1^N \left(abs\left((interpolation - comparison)/comparison\right)\right)\right)/N\right) * 100,$$

where *'abs'* denotes absolute value and *'N'* denotes the number of county-years. A disadvantage of the MAPE is that it provides no information about the direction of the error (i.e., positive or negative) respective to the comparison data. The MALPE is an alternative measure that does

provide such information, albeit as an average of all the observed positive and negative error for county-years. It is calculated like the MAPE except that the actual algebraic difference between the estimate and comparison is employed rather than the absolute value of that algebraic difference, as follows:

MALPE = $\left(\left(\sum_1^N\left((interpolation\ -\ comparison)/comparison\right)\right)/N\right) * 100.$

In addition to the MAPE and MALPE, we compute the percentage of county-years where the absolute percent error (i.e., from above *'(abs(interpolation– comparison))/comparison)*100'*) is less than or equal to a set of selected threshold values (i.e., 1 percent, 2 percent, and 5 percent). These threshold-based measures of error offer the advantage of being a highly intuitive measure of the magnitude of error.

For the remaining social and demographic variables gender, race/ethnicity, education, poverty and occupation, our outcomes of interest are the percentage of the county with the given social or demographic characteristic. Because we are examining percentages rather than population counts, our measures of error entail a slight variation on those reported above. Instead of the MAPE and the MALPE, we computed the mean absolute error in percentage-points (MAE) and the mean algebraic error in percentage-points (MALE).[4] As above, we also compute threshold-based measures of error. In specific, we calculate the percentage of county-years where the absolute error in percentage-points was less than or equal to the following set of thresholds: 0.1 percentage-points, 0.5 percentage-points, and 1.0 percentage-points.

The above measures of error are used to evaluate the magnitude of error between our interpolated estimates and the respective comparison data sources. These analyses allow us to answer our first two research questions about the performance of linear interpolation between the 2000 and 2010 Census and between the 2000 Census and the 2005-2009 ACS.

Lastly, we employ the above measures of error to evaluate whether trends in the 1-year ACS estimates support the 2007 midpoint year assumption for the endpoint of the linear interpolation between the 2000 Census and the 2005-2009 ACS. The MAPE, MALPE, MAE and MALE allow us to assess the average difference between each of the nationally representative 1-year ACS estimates and the overall 5-year ACS estimate for the period 2005-2009 and determine whether – on average— the differences between the 1-year and 5-year estimate is minimized in 2007. Recall that 1-year ACS estimates did not become nationally representative until 2006, so the series of 1-year ACS estimates we compare are for 2006, 2007, 2008, and 2009.

FINDINGS

Interpolated Demographic Variables Compared to the PEP

Total Population

Comparisons between the annual interpolated census tract estimates of population counts aggregated to the county-level and the annual PEP county-level estimates for the period 2001 through 2009 are reported in Table 1. We observe 3,131 counties over 28,179 county-years and find that in 97 percent of the interpolated estimates are within 5 percent of the PEP estimate for a

[4] The mean absolute error is $\left(\left(\sum_1^N\left(abs(interpolation\ \% -\ comparison\ \%)\right)\right)/N\right)$ and the mean algebraic error is $\left(\left(\sum_1^N(interpolation\ \% -\ comparison\ \%)\right)/N\right)$.

given county and year. In addition, 90 percent of the estimates are within 3 percent of the PEP and 80 percent are within 2 percent. Only once we consider a threshold of plus or minus 1 percent do we find about half of the county-years to have this magnitude of error. Consistent with this, the mean absolute percent error of the county-years (MAPE) is 1.3 percent. On average the interpolated estimates are slightly more likely to overestimate than underestimate the PEP, with a MALPE of 0.20 percent.

[Table 1 About Here]

Analyses by the county population size in 2000 show that error is largest for the smallest counties and that these counties tend to overestimate the PEP while the larger counties are more evenly balanced between overestimation and underestimation, with a slight trend towards underestimation. Counties with fewer than 5,000 persons show nearly double the absolute error of the entire sample (MAPE of 2.50 percent), and on average overestimate the PEP (MALPE of 1.80 percent). Mean absolute error and mean algebraic error are both about a full percentage-point lower for counties with 5,000 to 9,999 persons, and mean absolute error is slightly more than a full percentage-point lower for counties with greater than 10,000 persons (MAPE of 1.10-1.20 percent). These larger counties are also more evenly balanced between overestimation and underestimation, with MALPE declining from a 0.10 percent overestimation to a -0.30 percent underestimation.

Gender and Race/Ethnicity

Comparisons between the interpolated estimates of the gender and race/ethnicity measures (i.e., percent female, percent non-Hispanic White, percent non-Hispanic Black, and percent Hispanic)[5] and the respective PEP county-level estimates for the period 2001 through 2009 are reported in Table 2. Of the 28,179 county-years, we find that 96-99 percent of the estimates for these four demographic characteristics are within 1.0 percentage-points of the respective PEP estimate for a given county and year. In addition, for all demographic characteristics other than percent non-Hispanic White, 92-95 percent of the county-years of interpolated estimates are within 0.5 percentage-points of the PEP. For non-Hispanic Whites, 86 percent of the interpolated estimates are within 0.5 percentage-points of the PEP. The mean absolute error in percentage-points ranges from 0.14 to 0.27 percentage-points, with the percent non-Hispanic Black and percent non-Hispanic White showing, respectively the smallest and largest error. Trends in mean algebraic error show that, underestimation and overestimation are fairly evenly balanced with interpolations on average being a small overestimate of the PEP for the percent female (by 0.04 percentage-points) and underestimate of the PEP for the percent non-Hispanic white (by -0.04 percentage-points), percent non-Hispanic black (by -0.05 percentage-points), and the percent Hispanic (by 0.06 percentage-points).

[Table 2 About Here]

As observed for the total population counts, error for the percent female, percent non-Hispanic white, and percent Hispanic is about twice as large for the smallest counties (with fewer than 5,000 persons), but is similar in magnitude for the larger counties. Trends in error for the percent non-Hispanic black are more similar by the county population size, and in the smallest counties actually show slightly less error than the largest counties. Irrespective of population ., however, the magnitude of error remains below the 1.0 percentage point threshold for nearly all counties,

[5]County-level percentages were calculated from tract-level data on population counts by gender and by race/ethnicity aggregated to the county-level.

with at least 86 percent of the counties (and typically at least 98 percent) showing no more than 1.0 percentage-points of error.

Interpolated Demographic Variables Compared to the ACS

Total population

Comparisons between the annual interpolated census tract estimates aggregated to county-level and the annual ACS county-level estimates for the period 2006 through 2009 are reported in Table 3. We observe 779 counties over 3,116 county-years. Because ACS annual estimates are only available for counties with 65,000 people or more, nearly all of the counties in this subsample fall into the two largest population groups (counties with 60,000-149,999 persons and counties with 150,000 or more persons in 2000).[6] Due to these differences in composition, in Table 3 we also present findings for the interpolated estimates compared to the PEP for the subset of counties with the ACS.

[Table 3 About Here]

For the 779 counties with ACS estimates for at least one year 2006-2009 (i.e., the 3,116 county-years of data for counties with at least 65,000 persons), we find greater error when comparing the interpolations to the ACS estimates than to the PEP estimates. For example, 82 percent of the interpolated estimates for a given county-year are within 2 percent of the PEP, while only 63 percent of those same county-years are within 2 percent of the ACS. Similarly the MAPE for the interpolated estimates compared to the PEP is 1.20 percent, while it is 2.07 percent compared to the ACS. Although the magnitude of error differs for comparisons to the PEP versus the ACS, the direction of the error is the same; in both cases the interpolated estimates underestimate the comparison data on average, with a MALPE of -0.51 percent respective to the PEP and -0.23 percent respective to the ACS. Note that these findings for the PEP on the magnitude and direction of the error are consistent with those reported in Table 1 for the subsample of larger counties.

Gender and Race/Ethnicity

Table 4 presents the comparison results respective to both the ACS and the PEP for the gender and race/ethnicity measures. Because there are some counties in which no population counts by race or ethnicity were available for one or two years between 2006 and 2009, the number of county-years for the race/ethnic measures is lower than that for percent female (i.e., 2,862 versus 3,116 county-years).

[Table 4 About Here]

As was observed for the total population counts, the interpolated estimates of have larger error for the percent female in comparison to the ACS than they do in comparison to the PEP for the same county-years. Mean absolute error is five times that seen with the PEP (0.4 percentage-points versus 0.08 percentage-points). In addition, the drop-off in the percentage of county-years that are within a given thresholds proceeds faster in the ACS than the PEP: almost all of the interpolations (99 percent) are within 0.5 percentage-points of the PEP, while only 71 percent are within that same threshold of the ACS; and 78 percent of the interpolations are within 0.1 percentage-points of the PEP, while only 17 percent are within that same threshold of the ACS.

[6] There were 15 counties with a population size of 25,000-59,999 in 2000 which had increased in population to meet the 65,000 person threshold by the time of the ACS assessment. Findings for these counties are not presented.

The interpolated estimates for the race/ethnic measures also show a greater error compared to the ACS than to the PEP; however, comparisons to both the ACS and PEP show more error for the percent Non-Hispanic white estimates than for the other two racial/ethnic measures. For example, with respect to the PEP, 98 to 99 percent of the interpolated county-year estimates for each of the racial/ethnic measures were within 1.0 percentage-points of the corresponding PEP estimates. By comparison, only respectively 63, 88 and 84 percent of the interpolated county-years estimates were within 1.0 percentage-points of the corresponding ACS estimates for the percent non-Hispanic white, non-Hispanic black, and Hispanic. Similarly, over 90 percent of the interpolated estimates of the percent non-Hispanic white, non-Hispanic black, and Hispanic were still within 0.5 percentage-points of the PEP (i.e., 91 percent, 94 percent, and 97 percent, respectively), while the rates had dropped considerably for comparisons to the ACS. Only 34 percent of the interpolations were within 0.5 percentage-points of the ACS for non-Hispanic White, and respectively 67 and 57 percent were within this threshold for non-Hispanic black and Hispanic estimates. In addition to these differences in the magnitude of error, the direction of the error differed in comparisons to the ACS versus comparisons to the PEP for non-Hispanic black and Hispanic estimates. Interpolated estimates were on average an underestimate of the PEP (mean algebraic error was -0.11 and -0.06 percentage-points, respectively for the non-Hispanic black and Hispanic estimates) while on average they were an overestimate of the ACS (mean algebraic error was 0.04 and 0.40 percentage-points for the non-Hispanic black and Hispanic estimates, respectively).

In summary, although the magnitude of error was larger for comparisons to the ACS versus comparisons to the PEP for the same set of county-years, the mean absolute error was still less than 1 percentage-point for all measures and for most counties and most years below about 0.5 percentage-points. The mean absolute error compared to the ACS was 0.40 percentage-points for the estimate of the percent female (0.08 compared to the PEP); 0.90 percentage-points for the estimate of the percent non-Hispanic white (0.22 compared to the PEP); 0.48 percentage-points for the estimate of the percent non-Hispanic black (0.15 compared to the PEP); and 0.56 percentage-points for the estimate of the percent Hispanic (0.13 compared to the PEP).

Interpolated Social and Economic Variables Compared to the SAIPE

Table 5 and Table 6 present the results of comparing the interpolated estimates of the percent of population with household income below the poverty line and median household income to comparison data from the SAIPE. We report on the 3,131 counties with SAIPE data for 2001 to 2006. The magnitude of the error for the percent of population with household income below the poverty line and for median household income is larger than that observed for gender and race ethnicity with ACS data. Overall, 48 percent of the interpolated estimates for the poverty measure are within 1.0 percentage points of the SAIPE estimate and mean absolute error is 1.6 percentage points. The magnitude of error decreases with county size with 69 percent of the largest counties having the interpolated estimate within 1.0 percentage points of the SAIPE.

[Table 5 About Here]

For median household income, 52 percent of the interpolated estimates are within 5 percent of the SAIPE estimate, with a mean algebraic error of 4.64 percent, we find that the interpolations more frequently overestimate than underestimate the SAIPE estimate.

[Table 6 About Here]

Interpolated Social and Economic Variables Compared to the ACS

Table 7 presents the results of comparing the interpolated estimates of the education, poverty, and professional and managerial occupational status to comparison data from the ACS and Table 8 presents the comparison for income. We report on the 779 counties with an ACS estimate for 2006 (i.e., with population at least 65,000 persons).[7]

[Table 7 About Here]

For the educational measures, 67 percent of the interpolated estimates of the percent less than high school are within 1.0 percentage-points of the ACS estimate, and the rate drops to 45 and 44 percent, respectively for the only high school and greater than high school measures. However, mean absolute error is still less than 1.5 percentage-points for each measure (i.e., 0.90 percentage-points for less than high school, 1.35 percentage-points for high school graduate, and 1.46 percentage-points for greater than high school). Error for the less than high school estimate is about evenly balanced between overestimation and underestimation, while the interpolations more frequently underestimate than overestimate the percentage of high school graduates and the opposite is true for the percentage with greater than high school education (i.e., mean algebraic errors are, respectively 0.01, -0.74, and 0.74 percentage-points).

The magnitude of error for below poverty measure is similar to that for the lowest educational category, with 55% of the interpolated estimates were within 1.0 percentage-points of the ACS estimate, the mean absolute error was 1.2 percentage-points and the mean algebraic error was about evenly balanced between overestimation and underestimation (with mean algebraic error of -0.02 percentage-points). Error for the percentage of the population with professional or managerial occupations was larger, with only about one third of the interpolated estimates within1.0 percentage-points of the ACS estimate and a mean absolute error of 1.73 percentage-points. The mean algebraic error of 0.27 percentage-points indicated that the interpolations were more likely to overestimate than underestimate the ACS comparison values.

Compared to the SAIPE, the interpolated percent of population with household income below the poverty line has better agreement, although it is a smaller set of counties, as the mean algebraic error in percentage points is -0.02, the mean absolute error is 1.2 percentage points, and 55 percent of the county-years are within 1.0 percentage points of the ACS measure. Since the ACS estimates are based on larger counties, looking at the SAIPE comparison by county size, we see that the interpolated estimates difference from the SAIPE estimates had a higher percentage within 1.0 percentage-points than with the overall ACS comparison; however, the mean absolute percent error is largely the same as the ACS while the mean algebraic error is higher than with the ACS.

[Table 8 About Here]

For median household income, only 50 percent of the interpolated estimates are within 5 percent of the ACS, with a mean algebraic error of 4.66 percent, we find, like with the SAIPE, that the

[7] In the 2009 ACS 5-year tract-level data, there are tracts that have no median income estimate yet appear to have households. No interpolations could be done for those tracts. When aggregating the tract-level interpolations to county-level, a weighted average median household income was not computed for the counties containing those tracts with no median household income value. This resulted in 676 counties in the median household income comparisons with the ACS versus 779 counties for the poverty, education and occupation comparisons.

interpolations more frequently overestimate than underestimate the ACS estimate. When looking at larger counties, the overall difference with the ACS is smaller than that with the SAIPE.

Comparison of 1-year ACS Estimates for 2006, 2007, 2008, and 2009 and 5-year ACS Estimate for 2005-2009

Our last set of analyses investigated decision to use the 2005-2009 5-year ACS data for our 2007 endpoint in interpolations. We compared county-level estimates for our demographic, social and economic measures from the annual ACS for 2006 thru 2009 to the 2005-2009 5-year estimates for the 779 counties appearing in all four years. Recall that the annual ACS only presents estimates for counties with a population greater than 65,000. As seen in Table 9, for all but two of the eleven measures, 2007 had the lowest mean absolute error, with 2006 as the year with the lowest difference for those other two. Appendix 1 presents the actual mean absolute and mean algebraic error values for 2006 to 2009 for each of the measures.

[Table 9 About Here]

DISCUSSION

Demographic Measures

We find that intercensal estimates of county population counts and percentage distributions by gender and race/ethnicity obtained via linear interpolation between tract-level data from 2000 and 2010 Census compare favorably with comparison data from the PEP and the ACS. Assessment of error for all census tracts in the U.S. was only possible in comparison to the PEP and after aggregating the interpolated estimates to the county-level. On average 80 percent of the interpolated estimates of population totals were within 2 percent of the PEP estimate, MAPE was 1.3 percent, and mean absolute error in gender and racial/ethnic distributions was less than 0.3 percentage-points.

Although we were able to use the PEP to evaluate the interpolated estimates of tract-level characteristics aggregated to the county-level for the entirety of the U.S. for all intercensal years, comparisons of the interpolated estimates to the ACS were only possible for a subset of larger counties for the years 2006-2009. Error was larger when the interpolated estimates were compared to ACS than it was when they were compared to the PEP for this same subset of counties and years. For example, the MAPE for the total population counts was 2.07 percent for comparison to the ACS but only 1.20 percent for the PEP for the same set of counties and years. And for all four of the demographic characteristics, the mean absolute error across county-years was below one-quarter of a percentage-point compared to the PEP, while respective error compared to the ACS ranges from about one-half to one percentage-point.

Our findings that the interpolated estimates align more closely to the PEP than the ACS are consistent with the reported documentation on the PEP and ACS estimation methodologies (U.S. Census Bureau 2012; U.S. Census Bureau 2009b). Specifically, although the PEP employs a cohort-component methodology that incorporates data on birth, migration and deaths not employed in our linear interpolation[8], the PEP methodology is similar to our interpolation

[8] In addition, we are also comparing interpolated estimates made at the tract-level and aggregated to the county-level against county-level PEP estimates conducted at the county-level.

methodology than the ACS in that the estimates are subsequently adjusted so that they fall within the bounds of the 2000 and 2010 Census. In contrast, the ACS estimates are based on continuous sampling of the U.S. population, albeit with adjustments in the weighting process which employ the PEP county-level estimates as population controls.

Social and Economic Measures

We find that the magnitude of error for the comparison of the interpolated estimates to the ACS and to the SAIPE on the selected socioeconomic measures is larger than that observed for the demographic measures compared to either the ACS or PEP. That said, for nearly all of the socioeconomic measures compared to the 2006 ACS (i.e., poverty, educational attainment, professional and managerial occupational status) and the 2001-2006 SAIPE (i.e., percentage poverty), the mean absolute error was less than 1.75 percentage-points. Error for the median household income estimates were larger for comparisons both to the 2006 ACS and 2001-2006 SAIPE.

Interpolation Methodology

Our analyses lend support to the choice of 2007 as an endpoint value for the interpolation between the 2000 Census and the 2005-2009 5-year ACS. The multi-year estimates for our comparison measures have, on average, the least difference with the 2007 ACS 1-year estimates in our county-level comparisons.

Conclusions

The impact of this magnitude of error we observe in this study will likely depend on the specific application of the interpolated estimates. For secondary data research applications employing census tract data on percent distribution of population characteristics (such as the longitudinal study of exposure to neighborhood socioeconomic disadvantage, migration and neighborhood inequality, or neighborhood influences on health and development), the error we observe of less than half a percentage-point is magnitudes smaller than the percentage-point range of these population characteristics.

We propose that simulation analyses that evaluate the consequences of error within the context of a specific model will provide the best assessment of the relative importance of this magnitude of error for the characteristics employed in a given research application. In addition, it is important to re-emphasize that because no tract-level reference data is available assessments of bias must be at the county-level. Thus, any conclusions about the consequences of interpolation for a given research question, must also take into consideration the fact that our estimates of error average-over potential tract-level fluctuations and that the (unidentifiable) magnitude of error at the tract-level may be larger (or smaller) than our county-level estimates of error.

REFERENCES

Auchincloss, A. H., S. Y. Gebreab, C. Mair, and A. V. Diez Roux. 2012. "A review of spatial methods in epidemiology, 2000-2010." *Annual review of public health* 33:107-22.

Beaghen, M, and L Weidman. 2008. "Statistical Issues of Interpretation of teh American Community Survey's One-, Three-, and Five-Year Period Estimates." Washington, D.C.: U.S. Census Bureau.

Crowder, K., J. Pais, and S. J. South. 2012. "Neighborhood diversity, metropolitan constraints, and household migration." *American sociological review* 77(3):325-53.

Do, D. P. 2009. "The dynamics of income and neighborhood context for population health: do long-term measures of socioeconomic status explain more of the black/white health disparity than single-point-in-time measures?" *Social science & medicine* 68(8):1368-75.

Do, D. Phuong, L. Wang, and M Elliot. 2012. "Investingating the relationship between neighborhood poverty and mortality risk: A marginal structural modeling approach." in *Population Studies Center Research Report 12-763*: University of Michigan, Institute for Social Research.

Entwisle, B. 2007. "Putting people into place." *Demography* 44(4):687-703.

Jarvis, B.F. 2012. "Neighborhood exerpiences and the influence of neighborhood racial composition in residential choice." Paper presented at the 2012 Meetings of the Population Association of America. San Francisco, CA.

Kearns, R. A. 1993. "Place and health - Towards a reformed medical geography." *Professional Geographer* 45(2):139-47.

Kling, J.R., J.B. Liebman, and L.F. Katz. 2007. "Experimental analysis of neighborhood effects." *Econometrica* 75(1):83-119.

Logan, John R., Zengwang Xu, and Brian Stults. forthcoming. "Interpolating US decennial census tract data from as early as 1970 to 2010: a longitudinal tract database " *Professional Geographer*.

Ludwig, J., G. J. Duncan, L. A. Gennetian, L. F. Katz, R. C. Kessler, J. R. Kling, and L. Sanbonmatsu. 2012. "Neighborhood effects on the long-term well-being of low-income adults." *Science* 337(6101):1505-10.

MacDonald, H. 2006. "The American Community Survey: warmer (more current), but fuzzier (less precise) than the decennial census." *Journal of the American Planning Association* 72(4):491-503.

Macintyre, S., A. Ellaway, and S. Cummins. 2002. "Place effects on health: how can we conceptualise, operationalise and measure them?" *Social Science and Medicine* 55(1):125-39.

McElroy, Tucker. 2009. "Incompatibility of trends in multi-year estimates from the American Community Survey." *The Annals of Applied Statistics* 3(4):1493-504.

National Research Council and Institute of Medicine. 2000. *From Neurons to Neighborhoods: The Science of Early Childhood Development*. Washington, D.C.: National Academy Press.

Robert, S.A., K.A. Cagney, and M.M. Weden. 2010. "A life-course approach to the study of neighborhoods and health." in *Handbook of Medical Sociology. Sixth Edition*, edited by Peter Conrad Chloe E. Bird, Allen M. Fremont, Stefan Timmermans. Nashville, TN: Vanderbilt University Press.

Sampson, R.J., and P. Sharkey. 2008. "Neighborhood selection and the social reproduction of concentrated racial inequality." *Demography* 45(1):1-29.

Sampson, Robert J., Jeffrey D. Morenoff, and Thomas Gannon-Rowley. 2002. "Assessing neighborhood effects: social processes and new directions in research." *Annual Review of Sociology* 28(1):443-78.

Tatian, P.A. 2003. *Neighborhood Change Database (NCDB) 1970-2000 Tract Data: Data Users Guide*. Washington, DC: Urban Institute.
Torrieri, Nancy K. 2007. "America is changing, and so is the census." *The American Statistician* 61(1):16-21.
U.S. Census Bureau. 2009a. *A Compass for Understanding and Using American Community Survey Data: What Researchers Need to Know*. Washington, D.C.: U.S. Government Printing Office.
—. 2009b. *Design and Methodology, American Community Survey*. Washington, D.C.: U.S. Government Printing Office.
—. 2012. "Methodology of the Intercensal Population and Housing Unit Estimates 2000-2010" (www.census.gov/popest/methodology).
—. 2013. "Geographic areas published, American Community Survey."
Voss, Paul R. 2007. "Demography as a spatial social science." *Population Research and Policy Review* 26(5-6):457-76.
Wodtke, G. T., D. J. Harding, and F. Elwert. 2011. "Neighborhood effects in temporal perspective." *American sociological review* 76(5):713-36.

Table 1. Annual Interpolated Estimates of the County Population Counts Compared to Estimates from the Annual Population Estimate Program (PEP) over the Years 2001-2009[1]

	Total number of counties	Total number of county-years (CY) over 2001-2009	Percentage of CY with absolute percent error less than threshold of: ≤1%	≤2%	≤5%	Mean absolute percent error of CY (MAPE)	Mean algebraic percent error of CY (MALPE)
All Counties in the U.S.	3,131	28,179	56	80	97	1.30	0.20
Counties by population size in 2000 (persons)							
<5,000	290	2610	28	52	89	2.50	1.80
5,000-9,999	402	3618	46	75	97	1.50	0.70
10,000-24,999	884	7956	58	84	98	1.20	0.10
25,000-59,999	755	6795	63	86	98	1.10	-0.10
60,000-149,999	445	4005	62	84	99	1.20	-0.20
150,000 and up	355	3195	60	84	98	1.20	-0.30

[1]In order to produce the annual interpolated estimates of county population counts, intercensal tract-level population counts are estimated via linear interpolation between the 2000 and 2010 Census and then aggregated to the county-level. Error is defined as (interpolated)-(comparison) where the comparison data come from the annual PEP estimates for counties.

Table 2. Annual Interpolated Estimates of the County Percent Population Distribution by Gender and Race/Ethnicity Compared to Estimates from the Annual Population Estimate Program (PEP) over the Years 2001-2009[1]

	Total number of counties	Total number of county-years (CY) over 2001-2009	Percentage of CY with absolute percentage point error less than threshold of: ≤0.1%	≤0.5%	≤1.0%	Mean absolute error of CY in percentage-points	Mean algebraic error of CY in percentage-points
Percent Female							
All Counties in the U.S. Counties by population size in 2000 (persons)	3,131	28,179	48	92	98	0.20	0.04
<5,000	290	2610	14	60	86	0.53	0.04
5,000-9,999	402	3618	26	86	98	0.27	0.07
10,000-24,999	884	7956	37	94	98	0.21	0.06
25,000-59,999	755	6795	56	98	99	0.13	0.05
60,000-149,999	445	4005	69	99	100	0.09	0.02
150,000 and up	355	3195	83	99	100	0.06	0.02
Percent Non-Hispanic White							
All Counties in the U.S. Counties by population size in 2000 (persons)	3,131	28,179	37	86	96	0.27	-0.04
<5,000	290	2610	24	75	90	0.43	0.04
5,000-9,999	402	3618	34	83	95	0.28	0.02
10,000-24,999	884	7956	38	87	96	0.26	-0.01
25,000-59,999	755	6795	45	90	98	0.22	-0.03
60,000-149,999	445	4005	38	88	98	0.24	-0.12
150,000 and up	355	3195	27	84	96	0.3	-0.16

(Continued Next Page)

Table 2. Continued

	Total number of counties	Total number of county-years (CY) over 2001-2009	Percentage of CY with absolute percentage point error less than threshold of: ≤0.1%	≤0.5%	≤1.0%	Mean absolute error of CY in percentage-points	Mean algebraic error of CY in percentage-points
Percent Non-Hispanic Black							
All Counties in the U.S. Counties by population size in 2000 (persons)	3,131	28,179	67	94	98	0.14	-0.05
<5,000	290	2610	81	97	99	0.09	-0.04
5,000-9,999	402	3618	74	94	98	0.13	-0.03
10,000-24,999	884	7956	65	93	98	0.15	-0.03
25,000-59,999	755	6795	67	96	99	0.13	-0.04
60,000-149,999	445	4005	64	96	99	0.13	-0.05
150,000 and up	355	3195	54	92	98	0.19	-0.14
Percent Hispanic							
All Counties in the U.S. Counties by population size in 2000 (persons)	3,131	28,179	58	95	99	0.15	-0.06
<5,000	290	2610	40	83	93	0.3	-0.12
5,000-9,999	402	3618	58	91	98	0.17	-0.08
10,000-24,999	884	7956	60	95	99	0.14	-0.07
25,000-59,999	755	6795	61	96	99	0.13	-0.05
60,000-149,999	445	4005	61	97	100	0.12	-0.02
150,000 and up	355	3195	59	95	99	0.14	-0.02

[1]In order to produce the annual interpolated estimates of the percentages, intercensal tract-level population counts by gender, race, and ethnicity are estimated via linear interpolation between the 2000 and 2010 Census and then aggregated to the county-level. Error is defined as (interpolated)-(comparison) where the comparison data come from the annual PEP estimates for counties.

Table 3. Annual Interpolated Estimates of the County Total Population Counts Compared to American Community Survey (ACS) 1-year Estimates and Annual Population Estimate Program (PEP) for 2006 to 2009[1]

	Total number of counties	Total number of county-years (CY) over 2006-2009	Percentage of CY with absolute percent error less than threshold of: ≤1%	≤2%	≤5%	Mean absolute percent error of CY (MAPE)	Mean algebraic percent error of CY (MALPE)
Comparison to ACS	779	3,116	35	63	92	2.07	-0.23
By County Population Size in 2000 (persons)[2]							
60,000-149,999	409	1636	34	59	92	2.07	-0.09
≥150,000	355	1420	37	68	93	1.99	-0.32
Comparison to PEP	779	3,116	61	82	98	1.20	-0.51
By County Population Size in 2000 (persons)[2]							
60,000-149,999	409	1636	61	83	99	1.12	-0.63
≥150,000	355	1420	62	84	98	1.23	-0.29

[1]In order to produce the annual interpolated estimates of county population counts, intercensal tract-level population counts are estimated via linear interpolation between the 2000 and 2010 Census and then aggregated to the county-level. Error is defined as (interpolated)-(comparison) where the comparison data come either from the 1-year ACS or annual PEP estimates for counties.

[2]Estimates by county population size exclude the 15 counties in the subset of counties with 1-year ACS estimates that had fewer than 60,000 persons in 2000 but at least 65,000 persons in one or more years between 2006 and 2009.

Table 4. Annual Interpolated Estimates of the County Population Distribution by Gender and Race/Ethnicity Compared to American Community Survey (ACS) 1-year Estimates and Annual Population Estimate Program (PEP) for 2006 to 2009[1]

	Total number of counties	Total number of county-years (CY) over 2006-2009	Percentage of CY with absolute percentage point error less than threshold of: ≤0.1%	≤0.5%	≤1.0%	Mean absolute error of CY in percentage-points	Mean algebraic error of CY in percentage-points
Percent Female							
Comparison to ACS	779	3,116	17	71	94	0.40	0.06
By County Population Size in 2000 (persons)[2]							
60,000-149,999	409	1636	15	66	92	0.45	-0.04
≥150,000	355	1420	19	79	97	0.33	0.15
Comparison to PEP	779	3,116	78	99	100	0.08	0.03
By County Population Size in 2000 (persons)[2]							
60,000-149,999	409	1636	74	99	100	0.08	0.03
≥150,000	355	1420	85	99	100	0.06	0.02
Percent Non-Hispanic White							
Comparison to ACS	779	2,862	7	34	63	0.90	-0.69
By County Population Size in 2000 (persons)[2]							
60,000-149,999	409	1395	7	36	67	0.88	-0.63
≥150,000	355	1416	6	32	60	0.91	-0.79

(Continued Next Page)

Table 4. Continued

	Total number of counties	Total number of county-years (CY) over 2006-2009	Percentage of CY with absolute percentage point error less than threshold of: ≤0.1%	≤0.5%	≤1.0%	Mean absolute error of CY in percentage-points	Mean algebraic error of CY in percentage-points
Percent Non-Hispanic White (Continued)							
Comparison to PEP	779	3,116	41	91	98	0.22	-0.04
By County Population Size in 2000 (persons)[2]							
60,000-149,999	409	1636	44	92	99	0.19	-0.05
≥150,000	355	1420	38	90	98	0.23	-0.05
Percent Non-Hispanic Black							
Comparison to ACS	779	2,862	17	67	88	0.48	0.04
By County Population Size in 2000 (persons)[2]							
60,000-149,999	409	1395	14	60	85	0.54	0.03
≥150,000	355	1416	21	74	91	0.40	0.06
Comparison to PEP	779	3,116	61	94	99	0.15	-0.11
By County Population Size in 2000 (persons)[2]							
60,000-149,999	409	1636	67	97	99	0.12	-0.07
≥150,000	355	1420	54	92	98	0.18	-0.15
Percent Hispanic							
Comparison to ACS	779	2,862	13	57	84	0.56	0.40
By County Population Size in 2000 (persons)[2]							
60,000-149,999	409	1395	14	58	84	0.54	0.43
≥150,000	355	1416	11	55	85	0.56	0.40

(Continued Next Page)

Table 4. Continued

	Total number of counties	Total number of county-years (CY) over 2006-2009	Percentage of CY with absolute percentage point error less than threshold of: ≤0.1%	≤0.5%	≤1.0%	Mean absolute error of CY in percentage-points	Mean algebraic error of CY in percentage-points
Percent Hispanic (Continued)							
Comparison to PEP	779	3,116	60	97	99	0.13	-0.06
By County Population Size in 2000 (persons)[2]							
60,000-149,999	409	1636	59	98	100	0.12	-0.05
≥150,000	355	1420	61	96	99	0.13	-0.06

[1]In order to produce the annual interpolated estimates of the percentages, intercensal tract-level population counts by gender, race, and ethnicity are estimated via linear interpolation between the 2000 and 2010 Census and then aggregated to the county-level. Error is defined as (interpolated)-(comparison) where the comparison data come either from the 1-year ACS or annual PEP estimates for counties.

[2]Estimates by county population size exclude the 15 counties in the subset of counties with 1-year ACS estimates that had fewer than 60,000 persons in 2000 but at least 65,000 persons in one or more years between 2006 and 2009.

Table 5. Annual Interpolated Estimates of the County Percent Population with Household Income Below the Poverty Level Compared to U.S. Census Bureau Small Area Income and Poverty Estimates (SAIPE) for 2001 to 2006[1]

	Total number of counties	Total number of county-years (CY) over 2001-2006	Percentage of CY with absolute percentage point error less than threshold of: ≤0.1%	≤0.5%	≤1.0%	Mean absolute error of CY in percentage-points	Mean algebraic error of CY in percentage-points
All Counties in the U.S.	3,131	18,783	6	26	48	1.60	0.58
Counties by population size in 2000 (persons)							
<5,000	290	1737	3	16	29	2.7	0.95
5,000-9,999	402	2412	4	20	38	2.0	0.42
10,000-24,999	884	5304	5	22	41	1.7	0.61
25,000-59,999	755	4530	5	27	50	1.4	0.67
60,000-149,999	445	2670	8	35	59	1.2	0.50
150,000 and up	355	2130	8	41	69	0.9	0.29

[1]In order to produce the annual interpolated estimates of the percentages, intercensal tract-level population counts of total number of persons with household income below the poverty level are estimated via linear interpolation between the 2000 Census and 2005-2009 ACS 5-year and then aggregated to the county-level. Error is defined as (interpolated)-(comparison) where the comparison estimates come from the SAIPE for counties.

Table 6. Annual Interpolated Estimates of the County Median Household Income Compared U.S. Census Bureau Small Area Income and Poverty Estimates (SAIPE) for 2001 to 2006[1]

	Total number of counties	Total number of county-years (CY) over 2001-2006	Percentage of CY with absolute percent error less than threshold of: ≤1%	≤2%	≤5%	Mean absolute percent error of CY (MAPE)	Mean algebraic percent error of CY (MALPE)
All Counties in the U.S.[2]	2996	17976	10%	20%	52	5.80	4.64
Counties by population size in 2000 (persons)							
<5,000	283	1668	8%	16%	43%	7.86	6.11
5,000-9,999	399	2382	10%	19%	51%	6.20	4.41
10,000-24,999	877	5172	13%	25%	59%	4.99	3.37
25,000-59,999	743	4386	12%	23%	58%	4.75	3.85
60,000-149,999	419	2406	9%	17%	48%	5.61	5.27
150,000 and up	275	1386	3%	7%	25%	8.75	8.66

[1]In order to produce the annual interpolated estimates of county population counts, intercensal tract-level population counts are estimated via linear interpolation between the 2000 and 2010 Census and then aggregated to the county-level. Error is defined as (interpolated)-(comparison) where the comparison data come from the annual PEP estimates for counties.

[2] In the 2009 ACS 5-yr tract-level data, there are tracts that have no median household income estimate but appear to have households. No interpolations could be done for those tracts. When aggregating the tract-level interpolations to county-level, a weighted average median household income was not computed for the counties containing those tracts with no median household income value.

Table 7. Annual Interpolated Estimate of County Population Distribution by Poverty, Education and Occupation in 2006 Compared to American Community Survey (ACS) 1-year Estimates of Educational Distribution in 2006[1]

	Total number of counties	Percentage of counties with absolute percentage point error less than threshold of: ≤0.1%	≤0.5%	≤1.0%	Mean absolute error of all counties in percentage-points	Mean algebraic error of all counties in percentage-points
Percent with Household Income Below Poverty Level						
Comparison to ACS	779	6	32	55	1.20	-0.02
By County Population Size in 2000 (persons)[2]						
60,000-149,999	409	4	21	40	1.55	-0.03
≥150,000	355	8	45	73	0.77	0.02
Percent with Educational Attainment Less than High School						
Comparison to ACS	779	8	37	67	0.90	0.01
By County Population Size in 2000 (persons)[2]						
60,000-149,999	409	6	31	55	1.11	0.03
≥150,000	355	10	45	81	0.63	-0.01
Percent with Educational Attainment High School Graduate						
Comparison to ACS	779	5	23	45	1.35	-0.74
By County Population Size in 2000 (persons)[2]						
60,000-149,999	409	6	19	39	1.58	-0.69
≥150,000	355	4	28	53	1.08	-0.80

(Continued Next Page)

Table 7. Continued

	Total number of counties	Percentage of counties with absolute percentage point error less than threshold of: ≤0.1%	≤0.5%	≤1.0%	Mean absolute error of all counties in percentage-points	Mean algebraic error of all counties in percentage-points
Percent with Educational Attainment Greater than High School						
Comparison to ACS	779	5	27	44	1.46	0.74
By County Population Size in 2000 (persons)[2]						
60,000-149,999	409	5	22	36	1.71	0.67
≥150,000	355	6	32	54	1.14	0.81
Percent with Professional/Managerial Occupations						
Comparison to ACS	779	3	19	35	1.73	0.27
By County Population Size in 2000 (persons)[2]						
60,000-149,999	409	2	15	28	2.11	0.12
≥150,000	355	3	23	41	1.35	0.36

[1]In order to produce the annual interpolated estimates of the percentages, intercensal tract-level population counts by poverty status, educational attainment status and occupational status are respectively estimated via linear interpolation between the 2000 Census and 2005-2009 ACS 5-year and then aggregated to the county-level. Error is defined as (interpolated)-(comparison) where the comparison estimates come from the 1-year ACS estimates for counties.

[2]Estimates by county population size exclude the 15 counties in the subset of counties with 1-year ACS estimates that had fewer than 60,000 persons in 2000 but at least 65,000 persons in 2006.

Table 8. Annual Interpolated Estimates of County Median Household Income for 2006 Compared to the American Community Survey (ACS) 1-year Estimate for 2006[1]

	Total number of counties	Percentage of counties with absolute percent error less than threshold of:			Mean absolute percent error of all counties (MAPE)	Mean algebraic percent error of all counties (MALPE)
		≤1%	≤2%	≤5%		
Comparison to ACS[2]	676	9	21	50	5.67	4.66
By County Population Size in 2000 (persons)[3]						
60,000-149,999	386	11	25	57	5.16	3.65
≥150,000	275	7	15	42	6.36	6.22

[1]In order to produce the annual interpolated estimates of county population counts, intercensal tract-level population counts are estimated via linear interpolation between the 2000 Census and 2005-2009 ACS and then aggregated to the county-level. Error is defined as (interpolated count)-(comparison count) where the comparison data come from the 1-year ACS estimates for counties. Income is in 1999 dollars for both the interpolation and the ACS 1-year estimate.

[2] In the 2009 ACS 5-yr tract-level data, there are tracts that have no median household income estimate yet appear to have households. No interpolations could be done for those tracts. When aggregating the tract-level interpolations to county-level, a weighted average median household income was not computed for the counties containing those tracts with no median household income value.

[3]Estimates by county population size exclude the 14 counties in the subset of counties with 1-year ACS estimates that had fewer than 60,000 persons in 2000 but at least 65,000 persons in 2006.

Table 9. Year of 1-year American Community Survey (ACS) County Estimates with Minimum Difference Compared to ACS 5-year, 2005-2009 Estimate for Demographic, Social and Economic Indicators

Demographic, Social and Economic Indicators	Measure of difference between the ACS 5-year estimate and the ACS 1-year estimates for 2006, 2007, 2008, and 2009			
	Mean Absolute Percent Error Over All Counties (MAPE)	Mean Algebraic Percent Error Over All Counties (MALPE)	Mean Absolute Error Over All Counties in Percentage- points	Mean Algebraic Error Over All Counties in Percentage-points
Demographic				
Total Population	2007	2007		
Percent Female			2007 and 2009	2007 and 2009
Percent Non-Hispanic White			2007	2007
Percent Non-Hispanic Black			2007	2006 and 2008
Percent Hispanic			2007	2007
Social and Economic				
Percent Below Poverty Level			2006	2006
Percent Less than High School			2007	2007
Percent High School Graduate			2006	2006
Percent Greater than High School			2007	2007
Percent Professional/Managerial			2007	2007
Median Household Income	2007	2007		

Appendix. 1. American Community Survey (ACS) 5-year Estimates of County Demographic, Social and Economic Measures in 2005-2009 Compared to Respective ACS 1-year Estimates of County Demographic, Social and Economic Measures in 2006, 2007, 2008, and 2009

Demographic Measure	Total number of counties	Mean Absolute Percent Error of CY (MAPE)	Mean Algebraic Percent Error of CY (MALPE)	Mean absolute error of all counties in percentage- points	Mean algebraic error of all counties in percentage-points
Total Population					
ACS 2005-2009 versus ACS 2006	779	1.70	1.04		
ACS 2005-2009 versus ACS 2007	779	0.63	-0.07		
ACS 2005-2009 versus ACS 2008	779	1.30	-1.00		
ACS 2005-2009 versus ACS 2009	779	2.21	-2.04		
Percent Female					
ACS 2005-2009 versus ACS 2006	779			0.29	-0.06
ACS 2005-2009 versus ACS 2007	779			0.28	-0.01
ACS 2005-2009 versus ACS 2008	779			0.28	-0.03
ACS 2005-2009 versus ACS 2009	779			0.27	-0.01
Percent Non-Hispanic White					
ACS 2005-2009 versus ACS 2006	779			0.57	-0.47
ACS 2005-2009 versus ACS 2007	779			0.28	-0.03
ACS 2005-2009 versus ACS 2008	779			0.45	0.36
ACS 2005-2009 versus ACS 2009	779			0.84	0.82
Percent Non-Hispanic Black					
ACS 2005-2009 versus ACS 2006	779			0.36	-0.04
ACS 2005-2009 versus ACS 2007	779			0.32	-0.10
ACS 2005-2009 versus ACS 2008	779			0.34	-0.05
ACS 2005-2009 versus ACS 2009	779			0.38	-0.07

(Continued Next Page)

Appendix. 1 (Continued)

Demographic Measure	Total number of counties	Mean Absolute Percent Error of CY (MAPE)	Mean Algebraic Percent Error of CY (MALPE)	Mean absolute error of all counties in percentage- points	Mean algebraic error of all counties in percentage-points
Percent Hispanic					
ACS 2005-2009 versus ACS 2006	779			0.41	0.38
ACS 2005-2009 versus ACS 2007	779			0.17	0.07
ACS 2005-2009 versus ACS 2008	779			0.25	-0.23
ACS 2005-2009 versus ACS 2009	779			0.53	-0.51
Social and Economic Measure					
Percent Below Poverty Line					
ACS 2005-2009 versus ACS 2006	779			1.23	0.20
ACS 2005-2009 versus ACS 2007	779			1.26	0.50
ACS 2005-2009 versus ACS 2008	779			1.26	0.30
ACS 2005-2009 versus ACS 2009	779			1.44	-0.83
Percent Less than High School					
ACS 2005-2009 versus ACS 2006	779			1.00	-0.53
ACS 2005-2009 versus ACS 2007	779			0.89	-0.09
ACS 2005-2009 versus ACS 2008	779			1.03	0.60
ACS 2005-2009 versus ACS 2009	779			1.07	0.73
Percent High School Graduate					
ACS 2005-2009 versus ACS 2006	779			1.32	-0.68
ACS 2005-2009 versus ACS 2007	779			1.38	-0.75
ACS 2005-2009 versus ACS 2008	779			1.37	0.78
ACS 2005-2009 versus ACS 2009	779			1.33	0.78

(Continued Next Page)

Appendix. 1 (Continued)

Social and Economic Measure	Total number of counties	Mean Absolute Percent Error of CY (MAPE)	Mean Algebraic Percent Error of CY (MALPE)	Mean absolute error of all counties in percentage- points	Mean algebraic error of all counties in percentage-points
Percent Greater than High School					
ACS 2005-2009 versus ACS 2006	779			1.70	1.21
ACS 2005-2009 versus ACS 2007	779			1.47	0.84
ACS 2005-2009 versus ACS 2008	779			1.76	-1.38
ACS 2005-2009 versus ACS 2009	779			1.82	-1.51
Percent Professional/Managerial					
ACS 2005-2009 versus ACS 2006	779			1.94	1.14
ACS 2005-2009 versus ACS 2007	779			1.58	0.47
ACS 2005-2009 versus ACS 2008	779			1.62	-0.51
ACS 2005-2009 versus ACS 2009	779			2.32	-1.76
Median Household Income					
ACS 2005-2009 versus ACS 2006	676	5.32	4.02		
ACS 2005-2009 versus ACS 2007	676	4.30	2.27		
ACS 2005-2009 versus ACS 2008	676	4.50	2.72		
ACS 2005-2009 versus ACS 2009	676	7.19	6.49		