THE METHODS AND MATERIALS USED TO GENERATE TWO KEY ELEMENTS OF THE HOUSING UNIT METHOD OF POPULATION ESTIMATION: VACANCY RATES (VR) AND PERSONS PER HOUSEHOLD (PPH)

David A. Swanson
Department of Sociology
University of California Riverside
Riverside, CA 92151 USA
David.swanson@ucr.edu

Abstract

The U.S. Census Bureau is considering the implementation of the Housing Unit Method (HUM) for use in its annual post-censal population estimates program. This paper provides an examination of methods that can be used to estimate the average number of persons per household (PPH) and vacancy rates (VR), variables used in the Housing Unit Method of population estimation. It compares the results of using the Census Bureau's current method for estimating PPH and VR with alternative methods. The report includes an evaluation of the methods along with description of each method, its underlying theory and logic and an example, where feasible, in the form of an illustrative calculation. Each method is assessed relative to a set of evaluation criteria. The review and evaluation of data and methods suggests that the Census Bureau continue with its development of a comprehensive system based on the HUM for the annual estimates of county populations nationwide and methods, data bases, and administrative arrangements are suggested that would facilitate this work.

Acknowledgments. This material is based upon work supported by the U.S. Census Bureau (Contract Order No. YA132307SE0374). Any opinions, findings, and conclusions are those of the author and do not necessarily reflect the views of the U.S. Census Bureau. The author is grateful for comments made by Jason Devine, Gregg Robinson, Tori Velkoff, and Jeff Tayman, among others.

I. Introduction

The U.S. Census Bureau is in the midst of evaluating the methods (and data) it uses in its program of annual post-censal population estimates. An important component of this evaluation is the Housing Unit Based Estimates Research Project“ (HUBERT) Program (U. S. Census Bureau, 2007a). This paper stems from research done in conjunction with the HUBERT Program. It deals with the Housing Unit Method (HUM) of population estimation (Bryan, 2004b: 550). Specifically, it provides an examination of methods that can be used to estimate the average number of persons per household (PPH) and vacancy rates (VR), particularly at the county level. The research underlying the report compares the results of using the Census Bureau’s current method for estimating these variables with alternative methods. The alternative methods proposed include using trended PPH and VR values as well as PPH and VR values informed by the American Community Survey. Methods selected for examination include the following model based methods: (1) simple and complex trend extrapolation from past decennial census counts (Swanson, Baker, and Van Patten, 1983; Lowe, Pittenger, and Walker, 1979; Smith, Tayman, and Swanson, 2001; Swanson, 1981); (2) regression and related model-based methods (National Research Council, 1980; Rives, 1982; Smith and Lewis, 1980; Smith, Nogle, and Cody, 2002; Swanson, 1980; Swanson and Beck, 1994; Voss, Palit, Kale, and Krebs, 1995); (3) survey-based methods (Griffin and Waite, 2006; Roe, Carlson, and Swanson, 1992; Swanson, 1989; Swanson, 2006); and (4) direct and indirect estimators from other sources (Swanson, Baker, and Van Patten, 1983; National Research Council, 1980; Lowe, Pittenger, and Walker, 1979; Rives, 1982; Smith and Lewis, 1980; Swanson, 1981; Swanson and Lowe, 1979).

This paper is organized into four sections and an Appendix. The following section (II) provides descriptions of the theory, logic, data, and assumptions found in the methods used to generate PPH and VR. It includes, as appropriate, critical commentary. Section III consists of illustrations of the calculations underlying the methods and the data to produce PPH and VR estimates, along with the theory and logic underlying them, and the steps they use. For some methods, such illustrations are not feasible (e.g., a windshield survey). Section IV provides an evaluation of the methods identified in the third section, along with recommendations and suggestion. Each method is assessed relative to a set of evaluation criteria (e.g., face validity, cost, ease of application and explanation, and accuracy) developed using ideas from, among others, Habermann (2006), Swanson, Baker, and Van Patten (1983), Hough and Swanson (2006), National Research Council (1980), Roe, Carlson, and Swanson (1992), Smith and Mandell (1984), Smith, Nogle, and Cody (2002), Smith, Tayman, and Swanson (2001: 279-299), Swanson (2006), Swanson (1989), Swanson (1981), Swanson and Lowe (1979), Swanson and Pol (2005), Swanson and Tedrow (1984), Swanson and Van Patten (1987), and Swanson, Tayman, and Barr (2000). The Appendix is a reproduction of the principles underlying the US Census Bureau's estimates and projections programs (Habermann, 2006; U.S. Census Bureau, n.d.).

While the paper is focused on the United States at least some of what it covers applies to certain other countries that, like the United States, have strong administrative record systems, but lack a population registry system and rely upon regular census counts for population information. Like the United States, these countries are largely English-speaking and include, among others, Australia, Canada, England, and New Zealand. It also is worth noting that the report goes beyond the HUM at times and discusses other methods for estimating population.

An important part of the context for this paper is the reason why estimates are done in the United States. The census is the most complete and reliable source of information on the number of people in the United States a census, however, is a time-consuming and costly endeavor. The 1990 U.S. census had a cost of about $2.5 billion (U.S. GAO, 2001). The cost of the 2010 U.S. census is estimated at $11.3 billion U.S. GAO, 2006). In the United States, a census of the population is done only once every ten years; in Australia, Canada, England and New Zealand, for example, it is once every five years. Because there is the potential for constant and sometimes quite rapid population change, especially at the sub-national level, census statistics for every tenth and even every fifth year are often inadequate for many purposes (Waldrop, 1995). To fill this gap, population estimates are used by government officials, market research analysts, public and private planners and others for determining national and sub-national fund allocations (Murdock and Ellis, 1991; Serow and Rives, 1995; Siegel, 2002), calculating denominators for vital rates and per capita time series, establishing survey controls, guiding administrative planning, marketing, and for descriptive and analytical studies (Long, 1993a; Pol and Thomas, 2001: 93-95; Swanson and Pol, 2005). In the United States, the Census Bureau is not the only provider of population estimates (Bryan, 2004b: 524-526), but it is the ultimate source of estimates and the data needed to develop them.

The development of methods of population estimation roughly corresponds to the development of censuses and vital statistics registries. For example, in the late 17th century, John Graunt estimated the population of London and then of the whole of England and Wales using what today is known as a censal-ratio method (Devlin 2008: 93-94). Not long afterward, in the 18th century, the French mathematician, Laplace, also used a censal-ratio method in combination with recorded births and a population sample to estimate the population of France (Stigler, 1986: 163-164). However, methodological development really only took off in the late 1930s and early 1940s, fueled in large part by the need for low-cost and timely information generated by the great depression of the 1930s and World War II (Bryan, 2004b; Eldridge, 1947; Hauser and Tepping, 1944; Shryock, 1938; Shryock and Lawrence, 1949). In the United States, the Census Bureau played a major role in this effort, but it was not alone. During the early 1940s, the Washington State Census Board, for example, developed a comprehensive program of annual population determinations based on estimation methods that are still used today (Swanson and Pol, 2005; Swanson and Pol, 2008). Around this same time, demographers also began developing estimation methods for what were then called "underdeveloped countries," (Brass, 1968, Chandrasekaran and Deming, 1949; Popoff and Judson, 2004; United Nations, 1969) and the use of sample surveys as a substitute for complete census counts took hold (Bryan, 2004b).

Today, population estimates are ubiquitous. They are done around the world by a host of governmental and non-governmental entities, as well as individual consultants (Bryan, 2004b; Siegel, 2002; Swanson and Pol, 2008). The widespread availability of data, methods, and technology has made is possible for many people not only to develop estimates, but to do so more quickly and less expensively than has ever been done before. This trend is not likely to abate, but it carries with it a cost in that estimates may both be made and used with little or no understanding of the issues involved, what constitutes good estimates, and how to identify them.

What is a Population Estimate?

A population estimate is the determination of the size or the characteristics of a population at a current or past date in the absence of census data for the same date. In the United Sates, they usually are made on a De jure basis, which means that people are estimated where they usually reside. This makes sense because the U.S. census is conducted on a De Jure basis. However, there also is a need for estimating the De facto population of a given place at a given time and researchers have developed these estimates (Swanson and Pol, 2005; Swanson and Pol, 2008). These estimates include vacationers (of interest, for example, to the casino industry in Las Vegas and the Hawaii Visitors Bureau), migratory workers (of interest, for example, to health care, school, and other social service providers), and the people who work in the central business district of a large city each day, but leave it largely vacant in the evenings (of interest to the San Francisco City Planning Office, for example). While estimates of de facto populations are of great interest, they are very difficult to make in the United States because of the lack of census benchmarks (Cook, 1996, Smith, 1994). This is an important estimation topic, but it is beyond the scope of my mandate to cover research needs for de facto populations in depth. I only suggest here that the U. S. Census Bureau is the logical agency to develop systematic and comprehensive estimates of De Facto populations in the United States – as are its sister agencies in other countries currently operating similar population estimation programs.

In this context, it is useful here to note that the American Community Survey uses what amounts to a De Facto population as its sample frame and therefore is essentially aimed at estimating the De Facto population and its characteristics earlier (National Research Council, 2006a). However, while the ACS is essentially based upon a De facto population, it is important to note that its raw numbers are then controlled to a De jure population (National Research Council, 2007; Swanson and Hough, 2007). I return to this topic later in the report.

An estimate can be prepared for a nation or a sub-national area such as a state, county, city, town, or census tract. An estimate also can be prepared for groups of sub-national areas, groups of nations, or even the world as a whole. The principal demographic characteristics for which an estimate is made include age and sex. However, in multiracial and multi-ethnic countries such as the United States and Canada, an estimate might be done not only by age and sex, but also by race and ethnicity. An estimate also can be made of social and economic sub-groups of the population, households, and families.

The term "population estimate" is frequently used in the public domain to refer to the determination of the size or the characteristics of a population at a future date. However, most demographers prefer to use the term projection when talking about the possible size and characteristics of a population in the future. In developing a portrait of a given population in the future, it is not uncommon for a series of projections to be made that incorporate a range of plausible assumptions (e.g., expected trends in fertility, mortality, and migration). However, when one of these projections is selected as representing the most likely future, it then becomes the forecast for the population in question. As opposed to a projection or a forecast, a population estimate is concerned with either the present or the past, but not the future (Smith, Tayman, and Swanson, 2001: 3-4). Thus, in this report, I make the following distinctions among the terms "estimate," "projection," and "forecast."

> ***Estimate*** – A calculation of a current or past population, typically based on symptomatic indicators of population change.
>
> ***Projection--*** The numerical outcome of a particular set of assumptions regarding future population trends.

Forecast – The projection deemed most accurate for the purpose of predicting future population.

Virtually all methods of population estimation can be categorized into one or the other of two traditions: (1) demographic (Bryan, 2004b); and (2) statistical (Kordos, 2000; Platek, Rao, Sarndal, and Singh, 1987; and Rao, 2003). Demographic methods are used to develop estimates of a total population as well as its ascribed characteristics, age, race, and sex. Statistical methods are largely used to estimate the achieved characteristics of a population, and include, for example, educational attainment, employment status, income, and martial status. As is the case in the national statistical agencies of other countries, the US Census Bureau produces estimates using both of these traditions, demographic and statistical.

Demographers and statisticians have developed estimation a wide range of methods designed to meet different information needs at varying levels of accuracy and cost. As noted earlier, for the most part they are based on the concept of a De jure population although there are exceptions (Swanson and Pol, 2005). The methods can be roughly placed into three categories: (1) analytical and statistical models that use data symptomatic of population and its changes; (2) mathematical models that use historical census data; and (3) sample surveys. Methods falling into the first category have generally been developed by and for applied demographers, many of whom work for national, state, and local governments. Methods falling into the second category have generally developed by and for academic demographers, most of whom work at universities and research institutes. The methods falling into the third category have generally been developed by and for statisticians and survey research scientists, but they also are widely used by demographers. Not surprisingly, there also are techniques that combine methods from two or even all three categories.

Population estimation methods also can be identified along a temporal dimension: (1) inter-censal estimates, which refer to a date between two census counts and usually take the results of both counts into consideration; (2) post-censal estimates, which refer to a date subsequent to the latest census count and usually into account one or more previous census counts; and (3) pre-censal estimates, which refer to a date prior to a census count, but usually take into account one or more subsequent census counts. This temporal classification is useful because different methods are typically employed in the development of inter-censal, post-censal, and pre-censal estimates (Bryan 2004b).[1] These definitions and distinctions fall into the demographic tradition.

Among survey statisticians, the demographer's definition of an estimate is generally termed an "indirect estimate" because unlike a sample survey, the data used to construct a demographic estimate do not directly represent the phenomenon of interest (Swanson and Stephan, 2004: 758 and 763). In this paper, I use the demographic tradition's definitions and distinctions unless specifically noted.[2]

There are other ways to classify estimation methods. John Long (1993a), for example, categorizes them generally into two types: (1) "flow" methods; and (2) "stock" methods. Flow methods are also known as component methods, because they require estimation of each component of population change (births, deaths, and migrants) since the last census. Stock methods relate changes in population size since the last census to changes in other measured variables: the number of housing units, automobile registrations, total number of deaths (and births), and tax returns. Long (1993a) also notes that stock and flow methods may be used in combination. Popoff and Judson (2004: 603), make the following useful distinctions between stocks and flows: "…stock data are the numbers of persons at a given date, classified by various characteristics…(and) are recorded from censuses….flow data are the collection of or summation of events. At the most basic level this includes births, deaths, and migration flows…." This distinction is useful for purposes of this paper because, as is discussed later in this section, there are population estimations methods that solely rely on "stock" data while others rely on a combination of "stocks" and "flows."

Finally, it is useful here to consider micro data and aggregated data in the context of population estimation methods. I take micro data to mean records for individual persons. These records are often linked by relationships to form family and household records and I use the term "micro data" to refer to these linked records as well. The "Public Use Microdata Sample" (PUMS) is such a file (Swanson and Stephan, 2004: 772). Aggregated data are summations of records of individuals (families and households) such as one would find in a table. The aggregations are often done to specific geographic areas, but they can also be done for types of people across different geographies. The life table constructed by Kintner and Swanson (1994) for retirees of General Motors is an example of such an aggregation.

The Housing Unit Method

The HUM is designed to generate estimates of the total population by focusing on the population residing in households. As such, it inherently fits within the demographic tradition. However, while the HUM is inherently demographic in nature, the two key HUM elements I cover in this report (PPH and VR) are generated using methods that fit within the statistical tradition. Thus, I cover both traditions in discussing the HUM. Given that The HUM is aimed at the population residing in households, it is easy to see that is used to generate estimates of the total “De Jure” population. This, of course, is the definition of population used by the US Census Bureau, which is based on place of “usual residence” (Cook, 1996; Wilmoth, 2004).

One of the first times that the HUM is mentioned in the literature is found in an article by Starsinic and Zitter (1968) who found that it made a “…surprisingly strong showing…” and that “…it may be worthwhile to devote considerably more effort to refining the input data for estimating the number of households in addition to dealing with the problem of deriving current estimates on the size of households” (Starsinic and Zitter, 1968: 484). The article mentions work by Carl Frisén (1958) on the HUM in the 1950s for the California Department” The work by Frisén involved testing methods of population estimation against special censuses done by the state of California during the 1950s. Earlier work along these lines by was reported by Frisén (1951) when he was at San Jose State University.

However, the HUM was used even before 1950. It was used as early as 1942 under the auspices of the Washington State Census Board, which utilized the sociology graduate program at the University of Washington to carry out a program of annual estimates for cities and towns under the overall direction of Professor Calvin Schmid (Lowe, 2009). In 1967, the operation was transferred to the Washington State government, where today it exists in the state’s Office of Financial Management (Lowe, 2009). Washington’s use of the HUM is done in conjunction with census counts that allows cities and towns to conduct a special ‘headcount’ census when disagreements over estimates arise (Washington Office of Financial Management, 1978). These census counts are conducted in accordance with residency and housing definitions used by the Census Bureau with training assistance and supervision (including auditing) from the Washington Office of Financial Management. In 1981, the Washington system of municipal population estimation was adapted by the state of Alaska (Alaska Department of Community and

Regional Affairs, 1981a, 1981b; Alaska Department of Labor, 1981, 1982; Swanson, Baker, and Van Patten, 1983). Today, the HUM is arguably the most commonly used method of population estimation in the United States (Bryan, 2004b).

The Housing Unit Method (HUM) is a "stock" method that describes a basic identity in the same way that the balancing equation does (Bryan, 2004b). In the case of the HUM, this identity is usually given as

$$P = H*(1-VR)*PPH + GQ \tag{1}$$

where P = Population,

H = Housing units,

VR= Vacancy Rate (Proportion Vacant),

PPH = Average number of persons per household, and

GQ = Population residing in "group quarters" and the homeless.

Like the balancing equation, the HUM equation can be expressed in less detail (i.e., P= HH*PPH + GQ, where HH=H*(1-VR), Smith and Cody, 2004: 2) or more detail - by structure type, for example (Swanson, Baker, and Van Patten, 1983). It also can be used in combination with sample data, which opens the door to developing measures of statistical uncertainty for the estimates so produced (Roe, Carlson, and Swanson, 1992).

The HUM is based on the assumption that virtually everyone lives in some type of housing structure. It is generally accepted that the HUM is the most commonly used method for making small area population estimates in the United States (Byerly, 1990; Smith, 1986; Smith, Nogle, and Cody, 2002). Because of how data are collected, the HUM had not been a method that could be used for all sub-national areas and the nation as a whole until recently. However, with the continuous “Master Address File” (MAF), it has now emerged as a method that can be used by the US Census Bureau for all sub-national areas and the nation as a whole (Wang, 1999). This is a new resource for the Census Bureau’s estimates program because in the previous “mail-out/mail-back censuses, the MAF was constructed from scratch before each census. As observed nearly 25 years ago by Pittenger (1982) and more recently by Wang (1999), this housing unit inventory is serves as a key resource in the Bureau’s ability to construct population estimates (Swanson and McKibben, 2009) .[3] Other resources in regard to the suitability of the MAF, include Perrone (2008), Reese (2006), Swanson and McKibben (2007), U.S. Census Bureau (2004a), U.S. Department of Commerce (2002), and U.S. GAO (2006). Related work on the development of housing unit information includes McDonald and McMillen (2000), Pittenger (2004), and the U.S. Census Bureau (2006, 2007b).

In testimony before Congress, Swanson (2006) advocated the use of the HUM by the Census Bureau. In so doing, he: (1) described what he believed was the major challenge faced by the Census Bureau in providing timely, accurate, and cost-effective estimates; (2) provided a suggestion for dealing with this challenge; and (3) identified the major issues presented by his suggestion that need to be resolved.

In regard to the major challenge faced by the Census Bureau in providing timely, accurate, and cost-effective estimates, Swanson observed that this stemmed from the proliferation of federal programs distributing benefits using decennial census data and the knowledge that federal courts were now willing to consider apportionment cases. He noted in this regard that several lawsuits were filed against the Census Bureau following the 1970 census, a practice that has proliferated over the past thirty years and threatens to move into other areas of the Census Bureau's work such as the annual estimates program. Swanson argued that the reason for much of this conflict is clear: Billions of federal dollars are allocated each decade to states and local governments using census counts and inter-censal estimates and these funds are allocated in a "zero-sum" fashion. This situation will lead to even more litigation and other forms of conflict as the states, cities, and counties struggle to get their "populations" counted in the decennial censuses and estimated during the inter-censal periods.

It is this "atmosphere of conflict" that Swanson(2006) believes is the major challenge facing the Census Bureau's decennial census and inter-censal estimates programs.   Within the Census Bureau it not only serves to foster a "defensive" working environment, but also takes important resources away from production and research activities. As the defensive climate within the Bureau hardens, states and local governments feel even more frustration in their attempts to work cooperatively with the Bureau and turn to more confrontational forms of communication. This is particularly attractive for the local governments in states lacking strong demographic centers.

Swanson's suggestion for dealing with the challenges facing the Census Bureau revolves around the MAF. He noted in his testimony that, breaking with the past, the Census Bureau decided to retain and update its Master Address File – the MAF - for the 2000 Census. The MAF is a critical resource for the American Community Survey and its retention facilitates the planning and conduct of an accurate and cost-effective 2010 census. The continuously updated MAF and the related TIGER improvements are a fundamental element of success for an accurate 2010 census. Importantly, Swanson argues, the continuously updated MAF also represents an untapped resource for inter-censal estimates. It leads directly to the potential to have timely, accurate, and cost-effective estimates done using a method that is not only simple to apply and explain, but one that offers the potential for a meaningful role for states and local governments to play in the development of these estimates. He identified this method as the HUM, but to be successful this approach he argued that it needs a nationwide system of state demographic centers that participates in a meaningful partnership with the Census Bureau. He also notes that the state demographic centers, in turn, would need an active and meaningful partnership with the local governments within their respective states. Swanson (2006) argued that MAF-based population estimates would contribute toward having more timely, comprehensive, and internally consistent demographic and housing data for the U. S. as a whole and its sub-areas. In regard to geography, he notes that MAF- based data are extremely flexible in that they can be geo-coded to a specific location (as opposed to being assigned to an area defined by administrative or statistical boundaries). This also means that the MAF-based system can be overlaid with other features using GIS capabilities. The TIGER street address file comes to mind first in this regard.

Swanson (2006) testified that this approach to inter-censal population estimation would lead to an entirely new way of looking at the concept of a "small area," in that boundaries could be drawn that are much finer than those allowed by the census defined block. However, he noted that this would allow much higher precision in defining areas for purposes of marketing, site location and micro-simulation analysis, and modeling. Once up and running, this would also allow for greater ease in producing a consistent time series for areas in which administrative boundaries changed over time. The estimates would also provide population controls for the American Community Survey.

Swanson (2006) identified three major issues that needed to be resolved if his suggestion were to be successfully implemented. The first was confidentiality (e.g., Title 13 requirements), not an insignificant problem. However, he argued that this problem is not insurmountable in regard to his proposal for a MAF-based population estimation system. The National Research Council (2006b) has issued recommendations to reconcile access and confidentiality and the Census Bureau itself has appointed a Chief Privacy Officer and worked to put effective procedures in place regarding this reconciliation. Thus, he believes that the Census Bureau is capable of creating a national MAF-based population estimation system that meets confidentiality concerns.

Another important obstacle identified by Swanson (2006) is the financial cost of developing a national system of state demographic centers such that each state center functions according to accepted standards. States need to shoulder a share of these costs. After all, it is to their benefit to have high quality state demographic centers. As such, he proposed that a funding mechanism involving federal-state matching funds be considered.

Swanson (2006) also rhetorically asked if the proposed MAF-based population estimation system could provide accurate data and noted that the GAO identified MAF/TIGER problems that needed to be solved in order to have a good census in 2010. These problems include: (1) resolving address related issues such as duplication, omission, deletion, and incorrect locations in the MAF; and (2) implementing GPS-based geo-coding of housing units. Swanson noted that these same two problems represent sources of error in the proposed MAF-based system. Consequently, if the Census Bureau solves these problems in regard to the 2010 census, Swanson testified that it would do much in regard to the accuracy of the proposed MAF-based population estimation system.

Swanson (2006) also identified some lesser problems, ones largely already known to Census Bureau staff and others in regard to using the HUM effectively, to include: (1) tracking new housing units, converted housing units, and deleted housing units; (2) dealing effectively with seasonal populations and seasonal housing. He also notes that with the implementation of the ACS, the seasonality problem is compounded because of differences between the ACS and the decennial census in regard to what constitutes the de jure population. As such, he observed that an accurate MAF-based population estimation system will need to deal with the seasonal housing issue and the differences in the definition of the de jure population found in the ACS and the decennial census. However, Swanson(2006) testified that given the experience being gained by Census Bureau in regard to the MAF/TIGER system, the widespread knowledge use of the Housing Unit Method, and the capabilities of the best of the State Demographic Centers – Alaska, California, Florida, Texas, and Washington, for example, the timeliness and accuracy of MAF-based population estimates based on a comprehensive system of state demographic centers functioning at the level of the best state demographic centers would be sufficient for purposes of resource allocation, research, decision-making, and planning for the national, state and local levels. He also testified that he believed that it would also prove to be cost-effective and equitable and noted that the conflict-free system used in Finland to produce annual population data has the type of state-national participation and cooperation that he proposed, even though Finland uses population data to distribute funds and other resources in a zero-sum fashion to regional and local governments.

Swanson (2006) concluded his testimony with the observation that with the exception of the issue of confidentiality, all of the challenges facing the development of a national MAF-based population estimation system are in the form of costs, technical problems, or a combination of both. The major technical tasks in building and maintaining a MAF-based population estimation system come down to two areas - address data collection and MAF/TIGER update. The feasible way to effect a solution to these problems is to enhance the federal-state-local cooperative programs already part of Census Bureau activities such that local entities are compensated for helping to maintain the system. There are data collection activities in the United States that already follow this model, such as the vital registration system. He also noted in conclusion that by exploiting a functional MAF with the HUM that the Census Bureau and its state and local government partners would have universal means of population estimation for all areas of geography, administrative and statistical, and that state demographic centers should be developed to a uniform level of capability. Thus, he suggested that this proposal be supported by state-federal matching funds. He summarized his testimony bay stating that he believed that this system would lead not only to timely, accurate and cost-effective inter-censal population estimates, but also to greater equity in that there would be a uniformly higher level of demographic human capital in the country.

Data Foundations for Population Estimates.

As can be surmised from the preceding discussion of the MAF and other sources of housing unit data, all estimates, including the HUM, rely on one or more censuses and use administrative record systems on which different estimation methods for census-defined populations rely – vital events, tax returns, housing permits, assessor parcel files, utility hookups, licensed drivers, covered employment, K-12 enrollment, Medicare, and child support payments, among others (Bryan, 2004a; Bryan, 2004b; Bryan and Heuser, 2004). It is important to note that there is some variation in availability and quality of administrative records systems by state and by local jurisdictions in the US as well as variation among countries. For example in many areas of the United States, Kindergarten through 8th grade enrollments are used in the calculations of population estimates to avoid mistaking students who drop out of high school as out migrants from the area (McKibben, 2006).

It also is important to note that the U. S. Census Bureau maintains as much consistency in data sources and methods as it can because among other desirable features, it wants to have a consistent set of estimates for a given "vintage" year (See, Habermann, 2006; Appendix A of this report, U.S. Census Bureau, n. d.).

General Concepts and Methods

Although it is not used directly in the HUM, the fundamental demographic identity known as the balancing equation forms the conceptual framework for most other methods of population estimation (and forecasting). This identity is defined as $P_t = P_0 + I – O$, where P_t is the given population at time 0 + t, P0 is the given population at time 0, I is the number of persons entering the population through birth and in-migration during the period 0 – t, and O is the number of persons exiting the population through death and out-migration during the period 0 – t (Swanson and Stephan, 2004: 753). This identity can be phrased in more detail to separate recognize births, deaths, in-migration, and out-migration and is used as a point of departure to discuss in detail the concept of "stocks and flows" and the measurement thereof encompassed in the following five general types of methods, not including the HUM.

Simple Interpolation and Extrapolation Methods. Although no longer widely used in their own right, interpolation methods (see, .g., Judson and Popoff, 2004) and extrapolation methods (see, e.g., Smith, Tayman, and Swanson, 2001) represent ways to construct, respectively, intercensal estimates and post-censal estimates. These methods range from being relatively simple (e.g., linear trending) to very complex (ARIMA models). Both interpolation and extrapolation are based on mathematical formulas that are applied to “stock” data to produce “flows” that, in turn, generate estimates. As such, the principles underlying these methods, particularly extrapolation, are often found in other estimation methods (e.g., regression methods).

Regression Methods. This approach to population estimation represents a “stock” method in which measures of change in the ratios of indicators to population are used as “flow” estimates that are extrapolated to generated population estimates (Bryan, 2004b). The flow estimates serve as independent variables in these forms, while the dependent variable is a measure of population change. Measures of change can be in the form of ratios, lagged ratios, and differences (Bryan 2004b). These regression methods require a nested set of geographies (e. g., the counties within a given state) and they are inherently embedded in statistical inference (Swanson, 2004). As observed by Prevost and Swanson (1985), the “ratio-correlation” form can be viewed as a regression-based version of the so-called “synthetic” method of estimation.[4]

Component Methods. These are directly based on the fundamental demographic identify known as the balancing equation. As such, they are stock and flow methods. Included in this set are “Component Method II,” “Cohort-Component Method,” and the “Tax Return Method,” each of which is described by Bryan (2004b). The stock data are comprised of census counts in each of these methods, which use administrative records (e. g, vital events) to develop flow estimates.

Administrative Records. So-called direct estimates can be acquired from selected types of administrative records systems, namely the national population registration systems found in the Nordic countries (Bryan, 2004a: 31-33; Statistics Finland, 2004). Although the United States lacks a national population registration system, it has several national administrative record systems that serve as partial population registers, including those relating to social insurance and welfare and the payment of income taxes (Bryan, 2004a; Judson, 2000).[5] It is worthwhile at this point to consider the MAF, which represents a national housing registration system that can be used to generate estimates using the Housing Unit Method (Swanson, 2006).

Other Methods (Not including the HUM). Here, I include the economic-demographic models and urban systems models described by Smith, Tayman, and Swanson (2001: 185-237) as well as the iterative proportional fitting, log-linear, and multiregional methods described by Judson and Popoff (2004). To this list can be added the methods developed for statistically underdeveloped countries and those for estimating wildlife populations (briefly discussed in Endnote # 2) as well as the imputation methods used by the US Census Bureau to compensate for missing data (see, e.g., Swanson and Stephan, 2004: 762).

In concluding this brief overview of methods of population estimation other than the HUM, I note that it is often the case that various data adjustments must be made to effectively operate the preceding methods and that these adjustments serve as "other methods" in themselves (Wang, 1999). For example, the presence of non-household populations, such as found in prisons, school dormitories, and long-term care facilities, can affect the accuracy of virtually all of the methods just described, as can the presence of seasonal populations, undocumented aliens, and the occurrence of disasters, natural and otherwise.[6]

II. Methods used to Generate VR and PPH Data

In the preceding section, I mentioned the HUM study reported by Starsinic and Zitter (1968) who suggested that it was likely to be worthwhile to devote some effort to improving estimates of the number of households and their average size. There are two variables required to estimate the number of households: (1) housing units; and (2) a vacancy rate. By multiplying the number of housing units by the vacancy rate and subtracting this product from the number of housing units one arrives at the number of occupied housing units, which is synonymous with the number of households. So, I start with vacancy rate estimation.

Vacancy Rates

Following up on the suggestion by Starsinic and Zitter (1968) to work on ways to generate household numbers, Lowe, Pittenger, and Walker (1977) describe a method for generating vacancy rates using "windshield surveys." This method is labor-intensive and time-consuming (Alaska Department of Labor, 1981; Lowe, Pittenger, and Walker, 1977); Swanson, Baker, and Van Patten, 1982; Washington Office of Financial Management, 1978), but given the sample survey operations already being conducted by regional offices, it is potentially feasible for the Census Bureau. I return to this in more detail in the next section and in the final section

There are three other potential methods for obtaining vacancy rates that also are viable for the Census Bureau: (1) holding vacancy rates from the most recent decennial census constant until the subsequent census; (2) the use of US Postal Service (USPS) delivery data; and (3) modeling, which could also use the vacancy rates from the most recent decennial census and the USPS delivery data as inputs. In terms of holding vacancy rates constant since the last census, there is not much to describe, other than being aware of boundary changes and the use of structure type classifications.

In regard to USPS delivery data, Theresa Lowe (1988) examined the accuracy of postal survey data in reporting residential housing unit occupancy estimates against vacancy rates found in the 1970 and 1980 U.S. decennial census counts for 26 Washington State cities. The postal surveys were conducted by the Federal Department of Housing and Urban Development in the 1970s within 2 months of collection of census data. She found that postal surveys almost always show lower vacancy rates than census data because they do not include unfinished or new units, or concealed unoccupied conversions in single family homes. Suburban single family housing generally had the highest occupancy rates. However, she also found that postal data were much more accurate than census data in areas where occupancy rates were subject to high variation, as is found, for example, in cities near military bases, and in multi-unit structures. Because of this variation, Lowe (1988) argued that it was difficult to model vacancy rates in such areas.

Lowe (2000b) subsequently examined real estate vacancy surveys, which are aimed at the market for apartment rentals (multi-unit structures) and found that because they do not use random sample procedures, they did not match up well with the vacancy rates found in a decennial census, the later typically showing higher rates of vacancy. She suggested that the tendency of real estate vacancy surveys to be lower was primarily due to two factors: (1) many 'rented' units are not 'occupied' in the same manner that the census defines occupancy; and (2) the surveys only cover apartment units that are currently on the rental market (excluding, among other things for examples, newly constructed units that according to census definitions are unoccupied housing units). In her examination, she found that real estate vacancy rates tended to be around five percentage points lower than equivalent census vacancy rates and she provides adjustment factors for real estate vacancy surveys of multi-unit structures so that they more closely match the equivalent census vacancy rates.

Lowe, Mohrman, and Brunink (2003) examined postal delivery data within a context of factors affecting vacancy rates in general using the 2000 census of Washington as a benchmark.

Acknowledging that United States Postal Service (USPS) delivery data recognize postal deliveries rather than housing units, they found that for the state of Washington as a whole, residential postal delivery data exceeded the 2000 census count of 2,451,075 housing units by 7.6 percent. However, when post office box deliveries were excluded, they found that residential deliveries fell about 7.1 percent units short of the 2000 census count of housing units for the state as a whole. When looking at Washington's 39 counties, they found, however, that metropolitan counties had lower differences than did non-metropolitan counties. Considering the 2,001,325 housing units counted in the 2000 census for metropolitan counties, the postal delivery data were 7.3 percent higher when all deliveries were included and -3.3 percent lower when post office box deliveries were excluded. Considering the 449,750 housing units counted in the 2000 census for non-metropolitan counties, they found that the postal delivery data were 9.2 percent higher when all deliveries were included and 24.1 percent lower when the post office box deliveries were excluded.

Moving on to vacancy rates themselves, Lowe, Mohrman and Brunink (2003: 5) note that postal delivery data recognize deliveries as "possible" and "active." "Active" deliveries are reported within "possible" deliveries so by subtracting "active" from "possible," to obtain a residual set, "possible, but not active" that corresponds roughly to vacant units (Lowe, Mohrman, and Brunink, 2003: 5). Carrying out these operations, they compare the "possible, but not active" set to vacant housing units by county in Washington using 2000 data. They find that for the state as a whole, the 2000 census found a housing unit vacancy rate of 7.33 percent while the comparable rate from the USPS data was 1.78 percent when all deliveries are used and 1.33 percent when post office box deliveries are excluded. Following the state as a whole, Lowe, Mohrman, and Brunink (2003: 6-8) the 2000 USPS data produce, on average, estimates of vacancy rates that are 11 to 12 percentage points lower than the vacancy rates from the 2000 census for Washington's 39 counties. They find that the USPS data are only about 2 percentage points lower than the census vacancy rates in metropolitan counties, however. The largest arithmetic differences are found in counties that have substantial seasonal housing, which are non-metropolitan and that USPS data are, on average, 12-13 percentage points lower than the census vacancy rates across all non-metropolitan counties. Lowe, Mohrman, and Brunink (2003) also examined changes in the USPS data subsequent to the 2000 test and found that they were in accordance with expected vacancy rate changes due to population and housing changes. They conclude that USPS data may be a useful tool for 'adjusting" (modeling) decennial census vacancy rates at the county level. However, they advise that counties be examined individually in accordance with metropolitan/non-metropolitan classifications and the presence of substantial seasonal housing stock, among other variables.

Lowe and Mohrman (2003) extend the research reported by Lowe, Mohrman, and Brunink (2003) by examining the consistency of 2002 HUM-based county population estimates using USPS adjustments with 2002 population estimates made by the US Census Bureau. They used all possible residential deliveries, including post office boxes and vacancy rates from the 2000 census were adjusted on a county-by-county basis. They fond that the mean algebraic percent difference (MALPE, which includes the sign of the percent difference) across all 39 counties in 2002 was only 0.14 percent and that in 17 counties the HUM-based estimates exceeded the Census Bureau estimate and that in 22 counties they were lower. The highest positive difference (2.80%) was for the non-metropolitan county of Garfield, which is not adjacent to a metropolitan county; the highest negative difference (-5.39%) was for Island County, which is adjacent to the metropolitan county of Snohomish. Lowe and Mohrman (2003) also ‘backed-into” the USPS adjustments that would be required to match the “most likely” populations of these counties in 2002 (which were a combination of state and Census bureau estimates, accounted for the population in group quarters, and maintained 2000 PPH values). They found that at the state level, a 25.2 percent change was required of the USPS data and that most counties required between a 20 and 40 percent change. They concluded that the process they used needed to be extended to more years subsequent to 2000 to assess the stability of the relationship between the (assumed) underlying actual vacancy rates and the rates derived from the USPS data.

Moving away from USPS delivery data, Swanson, Baker, and Van Patten (1983) discuss vacancy rates within the context of an overall assessment of the HUM by state demographic centers. They point out that the HUM is optimal when it is done in conjunction with an active “headcount” census program, which can be used to update the elements of the HUM, including VR and PPH, if there is a dispute between the state demographic center and local agency preparing a population estimate using the HUM. As was described earlier, they also point out that reasonably estimates of vacancy rates can be obtained from “windshield surveys, for which detailed procedures are given in the Housing Unit Method Manual produced by the Alaska Department of Labor (1981), which was adapted from a similar manual developed by the Washington Office of Financial Management (1978).

Smith and House (2007) suggest the use of the ACS, but not for directly estimating vacancy rates. Instead, they argue for using it to develop estimates of ‘temporary’ migrants, a topic that affects areas with seasonal populations and, hence, vacancy rates of seasonal housing units. An important issue in considering the ACS as a source of VR data is the fact that it does not use the same residency definition as the decennial census (Swanson and Hough, 2007;

Swanson and McKibben, 2009). The decennial census (along with the CPS, and SIPP, among other products), use the De jure rule of residency while the ACS uses what amounts to a De facto residency rule. However, it is useful to note here that when the micro-level ACS data are aggregated to geographic areas, they are controlled to number produced for these areas by the Census Bureau's annual population estimates program, which are produced on a De Jure basis. As Swanson and McKibben (2009) observe, this may not be a huge issue at the national level, but at sub-state levels, the effects of these different residency rules could be substantial.

Most of the "modeling" techniques for developing estimates of occupied and vacant residential units (and the corresponding rates) come from the field of housing economics. Edelstein and Tang (2007) develop and test a theoretical model for residential housing market cyclical dynamics. The model employs an interactive supply and demand framework to engender housing price dynamics. Under our set of assumptions, the two equation system is econometrically identified: the first equation, housing demand, relates rent, property values, and capitalization rates with demand fundamentals. The second equation, housing supply, relates housing investment and property values with supply fundamentals. Edelstein and Tang (2007) use their model to analyze empirically the cyclical dynamics for residential properties in Los Angeles, San Francisco, San Diego and Sacramento for the 1988–2003 time period. The authors argue that their empirical analyses suggest that fundamentals, such as employment growth and interest rates, are key determinants of residential real estate cycles, but that local fundamentals tend to have greater cyclical impacts than those of national or regional fundamentals.

Gabriel and Nothaft (2001) make use of inter-metropolitan and time-series data from the BLS to model the incidence and duration of rental vacancies and to assess their importance to the price adjustment mechanism for rental housing. They find that duration varies with measures of

MSA housing costs and housing stock heterogeneity, while incidence varies with measures of population mobility, public housing availability, and population growth. Results support a more general specification of rental price adjustment in which the rate of real rent change reflects deviations in observed vacancy incidence and duration from their equilibrium levels.

Hendershott, MacGregor and Tse (2002) note that rental adjustment equations have been estimated for a quarter century and that in the U.S., models have used the deviation of the actual vacancy rate from the natural rate as the main explanatory variable, while in the UK, drivers of the demand for space have dominated the estimation. They derive a more general model incorporating both supply and demand factors and find that it is greatly superior to the vacancy rate model. They also construct a variant of the general model with a separate vacancy rate equation and find that it also performs better than those produced earlier. .

Hsueh, Tseng, and Hsieh (2007) develop a model using 1990 and 2000 data for Taiwan that simultaneously looks at housing price, vacancy rate, and moving rate. The results for 1990 show that in a booming market situation, both expected housing price and current housing price had a strong, positive impact on the vacancy rate; however, the housing vacancy rate did not display a negative impact on housing price as expected. The results for 2000 show that housing price did not significantly affect the vacancy rate; however, the vacancy rate had a negative impact on housing price that was highly statistically significant. This result reflected the fact that housing market operation had swung to another extreme after the real estate bubble that started in the late 1980s and burst in the mid-1990s.

Khor, Ming, and Yuan (2000) look at the concept of a “natural” vacancy rate, which they define as an equilibrium level of inventory of space, in the sense that both the matching process between landlord and tenant is facilitated, and that building owners hold an optimal buffer stock of inventory to meet future leasing contingencies. They find that when vacancy rates are above the natural vacancy rate, rents will fall and vacancies will drift upward toward equilibrium. The determination of the natural vacancy rate is therefore significant in that it can facilitate the monitoring of the market conditions since a vacancy rate below the natural vacancy rate signifies a tight market. The authors find that the converse is true if the vacancy rate is above the equilibrium level.

Extrapolation models are discussed in the following section on PPH. However, they are not discussed in regard to VR because they are not useful. As is discussed in regard to PPH, extrapolation models match up well with the demographic determinants underlying changes in PPH and because of the inertia in these determinants, the models have a good track record in regard to post-censal PPH estimates. Extrapolations do not match up well with the determinants underlying changes in VR, however, so I neither describe nor evaluate them for purposes of generating VR values.

Persons Per Household

The development of PPH data has been examined by demographers more than VR data development has. Similar to the examination of VR modeling by economists, more demographic theory comes to bear on PPH values than on VR values. I describe some of the theory here and return to this topic in Section IV.

Akkerman (1980) finds that a household composition matrix by age of head and age of other household members operates as a linear transformation from the vector of household distribution by age of head to the vector of population age distribution. He continues this observation by showing that the first row of the matrix may be interpreted as representing a vector of average household fertility rates. Akkerman (1980) observes that if a linear relationship between household and population distributions is fully implemented, then a relationship between household fertility and the size of the youngest age group can be derived. He concludes by noting that an extension of this result enables the simultaneous projection of population and households. Following up on these observations, Akkerman (2000) shows that the formal relationship between age of head to age of other household members is equivalent to the input-output relationship in the Leontief model of the open economy. Thus, he argues that the notions of household composition and household accommodation, which have emerged independently over the past two decades, are formally linked within this relationship.

Akkerman (2004) applies his earlier work to a case study of the Czech Republic during its post-communist transition to market economy. He uses the household composition matrix as a demographic gauge to the behavioral response of households to Czech housing markets and policy. He observes that the age-specific household size shown for the various regions of the Czech Republic follow a Gamma function, with anomalies detected in the trajectory of age-specific household size for Prague confirming unique housing market conditions and a commensurate demographic response. Acs and Nelson (2004) also examine the effect of policies on PPH values, but do so in the US and without the household composition matrix. They suggest that welfare policies may have contributed to the decline in single parenting and the rise in cohabitation between 1997 and 1999, a consequence of which is an increase in PPH. Brown (1999) also found household compositional matrices to be useful in examining PPH variation across small areas.

Burch (1967) examined census data from a range of countries and found that PPH mainly reflected past fertility rates rather than changes in family structure. Specifically, he finds that smaller PPH values had little to with the then-common view that urbanization was linked “breakdowns" of the extended residential family that had been more common in less urbanized societies. Burch (1970) followed with a broader examination that looked at the influence of demographic variables (viz., mortality, fertility, and age at marriage) on average household size under different family systems--nuclear, extended and stem. Using the model he developed, Burch found that under all three family systems, average household size is positively correlated with fertility, life expectancy, and average age at marriage. Other studies identifying the demographic determinants of PPH include Bumpass (1990), Burch et al. (1987) and Coale (1965). Ellen and O’Flaherty (2007) extend the work on the demographic determinants of PPH by examining the effect of government policies. Using data from a survey of households in New York City, they find that these policies appear to have an impact.

Using census data from 1790 to 1970, Korbin (1976) examined the long-term fall in household size in the United States and found that household changes were due to demographic changes. Sweet (1984) decomposed the growth of households during the 1970s and found that about one-third of the increase in the number of households was due to increased age by marital status propensity to form households while the remaining two-thirds was due to shifts in the age by marital status distribution and population growth. Specifically, he found that the increased propensity to form households had its major impact at ages under 35, and primarily among never-married persons while the composition component had its primary impact at ages 25-44 as a result of the baby boom, and also because of the increased fractions never married and separated and divorced during this period.

Myers and Doyle (1990) develop a demographic framework for determining PPH and propose models for estimating these values with the age structure of the household population. Mason and Racelis (1992) consider four models that explicitly incorporate the impact of changes in the number of men and women on the number and joint age distribution of husband-wife households. The models are applied to the Philippines using data from the 1988 National Demographic Survey to project households to 2010. The models are also evaluated by 'backcasting' and comparing the results with special tabulations from the 1970 and 1980 censuses and the 1975 National Demographic Survey. Van Imhoff et al. (1995) provide a wide range of demographic models that can be used to forecast and estimate households. Zeng, Land, Wang, and Gu (2006) propose a household model that uses demographic rates as input and projects more detailed household types, sizes, and living arrangements for all members of the population.

Although, it contains less demographic theory than the studies just reviewed, work by Smith, Nogle, and Cody (2002) found that regression models using symptomatic indicators of PPH change performed very well in estimating county-level PPH values in four states (Florida, Illinois, Texas, and Washington). Kimpel and Lowe (2007) and Washington Office of Financial Management (2007) found less encouraging results in Washington for PPH estimates developed from regression models using symptomatic indicators. Swanson and Lowe (1979) examined the use of IRS data on average exemptions per return as a symptomatic indicator of PPH and found that while it had deficiencies, it held promise.

Even less demographic theory (i.e., virtually none) is found in the trend extrapolation models used in Washington and Alaska, among other places to develop post-censal PPH estimates (Alaska Department of Labor, 1981; Lowe, Pittenger, and Walker, 1977; Swanson, Baker and Van Patten, 1982; Washington Office of Financial Management, 1978). Work by Findley and Reinhardt (1980) found that PPH change during the 1970s was non-linear. This work serves to confirm that simple geometric trend extrapolation has the potential to develop accurate PPH estimates.

Sample surveys offer a means of obtaining current PPH data (Alaska Department of Labor, 1981; Hogan, 2007; Washington Office of Financial Management, 1978; Swanson and Hough, 2007). The obvious survey of choice for the Census Bureau is the ACS (Hogan, 2007).

III. Illustrations of Methods used to Generate VR and PPH Data

Vacancy Rates

There is not much in the way of illustrations in this section. The models described earlier, the ones developed largely by economists, are also not suitable for illustration. Holding vacancy rates constant since the last census does not require an illustration. For USPS delivery data, one needs to obtain the data (see, e.g. http://www.usps.com/ncsc/addressinfo/deliverystatistics.htm), learn about what they represent, and then apply the data to given areas with the appropriate adjustments. Adjustments are needed because of what the delivery data represent relative to vacancy rates (as explained earlier). An illustration of such an adjustment is given by Lowe and Mohrman (2003), who apply the rate of change found in the USPS delivery data from the date of the previous census to the current (estimate) date to the vacancy rate found in the previous census. This effectively brings forward the census VR to the current estimate date. This can be done for all housing units combined, with subsequent adjustments by structure type, as needed.

The Windshield surveys mentioned in the preceding section can be done on a full count basis or a sample basis (Alaska Department of Labor, 1981; Washington Office of Financial Management, 1978). This type of field work generally works best in towns of less than 100,000 that are not characterized by massive numbers of multi-unit housing units. It can be done in urban areas with concentrations of multi-unit housing, but typically this requires a fair amount of work on foot, which adds to costs. The detailed procedures are not presented here; rather, an overview is given.

A windshield represents field work, which is expensive. However, a windshield survey is far less expensive than door-to-door canvassing. The first step is to map out the area for which VR is desired in terms of blocks (Alaska Department of Labor, 1982). This map then serves as the basis for a full count and as the frame for a sample. If a full count is to be done, these blocks can then be organized into enumeration areas. If a sample is to be done, the blocks serve as PSUs, which can be randomly selected (or more preferably, randomly selected with variations, such as the more cost-effective random selection of clusters of blocks and then full counting of the blocks within these clusters, where the clusters are equivalent to "enumeration areas." Once the areas are mapped, two-person teams systematically canvass the blocks in automobiles using the same in a set manner (e.g., driving such that only turns are made to the right, starting in the northeast corner of a block and ending up where the canvassing began) with the driver focused on driving and the second person focused on counting, classifying, and recording.

Basic training is required, but generally this can be done in a single day. The training includes census definitions of housing units and vacancies, as well as practical field procedures Evidence from Washington, suggests that in areas with less than 20,000 people, the full count was preferable to a sample survey. When areas had between 20,000 and 100,000 then a sample was preferable because of the cost savings. Many more of the nuances and details (e.g., identifying large clusters of certain types of housing structure during the mapping phase such as mobile home parks and where possible, obtaining vacancy information from resident managers, once they understood the nature of a census-defined vacancy) are found in the HUM Manuals produced by the Alaska Department of Labor (1981) and the Washington Office of Financial Management (1978), with supplemental information found in the State of Alaska Census Administrators Manual (Alaska Department of Labor, 1982) and the State of Alaska Census Enumerator's Manual (Alaska Department of Labor, 1981).

Persons Per Household

Current PPH values have often been obtained by using the value from the most recent census or extrapolating trends found from the two most recent decennial censuses, with the geometric model being typically used (Alaska Department of Labor, 1981; Washington Office of Financial management, 1978; Bryan, 2004b; Lowe, Pittenger, and Walker, 1977; Swanson, Baker, and Van Patten, 1983). In this approach, the rate of change is benchmarked to two most recent successive census counts and then applied to the PPH value found in the most recent census count, which is then extrapolated beyond the most recent census by applying the rate of change to it.

The process takes place in two steps. The first is the calculation of the rate of change in PPH:

$$r = (PPH_l/PPH_b)^{(1/y)} - 1 \qquad [2]$$

where

r = rate of change
PPH = Persons Per Household
l = Launch Year (most recent census)
b = Base Year (Census preceding launch Year
y = Number of years between l and b (10 years)

The second step is applying the rate to the launch year to find PPH values:

$$PPH_t = (PPH_l) [(1 + r)^{(y)}]$$

where

r = rate of change (from step 1)
PPH = Persons Per Household
t = Target Year
l = Launch Year (most recent census)
y = number of years between t and l

Table 1 shows the results of this process for counties in Washington. The base year is 1980 and the launch year is 1990. As an example of the two steps described above, consider Kitsap County for which the first step yields a rate of -0.0013, which is equal to $[(2.6469/2.6820)^{(1/10)} – 1]$, and for which the second step yields a 2000 PPH estimate of 2.6123, which is equal to $\{(2.6469) [(1 – 0.0013)^{(10)}]\}$

Table 1 also provides a comparison of the PPH values estimated for 2000 with the reported PPH values from the 2000 Census. As the summary statistics at the bottom of Table 1 show, the estimated PPH values are accurate, appoint to which I return in the following section.

Other forms of extrapolation can be used, with the caution that a linear extrapolation model is not as desirable as non-linear ones such as the geometric model (Findley and Reinhart, 1980). For example, a double-ratio geometric model could be used, although it is a bit more complicated than the simple geometric model. It requires three time points of data. The first step is to calculate separate rates of change for the two time periods (e.g., 1980 to 1990 and 1990 to 2000) as is done in Equation [2] for the simple geometric model. If the rate of change for second (more current) period is less than the rate of change for the first period then one calculates the ratio of the absolute rate of change from the first period to the absolute rate of change for the second time period and, which is then used to modify the launch year PPH; If the rate of change for second period is greater than the rate of change for the first period then one uses PPH as given in the launch year (most recent census).

(TABLE 1 ABOUT HERE)

Other forms of extrapolation can be used, with the caution that a linear extrapolation model is not as desirable as non-linear ones such as the geometric model (Findley and Reinhart, 1980). For example, a double-ratio geometric model could be used, although it is a bit more complicated than the simple geometric model. It requires three time points of data. The first step is to calculate separate rates of change for the two time periods (e.g., 1980 to 1990 and 1990 to 2000) as is done in Equation [2] for the simple geometric model. If the rate of change for second (more current) period is less than the rate of change for the first period then one calculates the ratio of the absolute rate of change from the first period to the absolute rate of change for the second time period and, which is then used to modify the modify the launch year PPH; If rate of change for second period is greater than the rate of change for the first period then one uses PPH as given in the launch year (most recent census)from the most recent census

IRS data on the number of exemptions per return have been examined with an eye toward using them to model current PPH values (Swanson and Lowe, 1979; Voss and Krebs, 1979). One problem noted by Swanson and Lowe (1979) is that the use of the IRS data effectively collapses all structure types into one category. This is a problem because of the different PPH levels associated with different structure types (Swanson and Lowe, 1979), which lead to their suggestion to use IRS data in conjunction with PPH data from the Current Population Survey. Their report preceded the ACS by 20 years, so an appropriate revision of their suggestion would be to use the IRS data in conjunction with PPH values taken from the ACS.

Stan Smith, June Nogle, and Scott Cody (2002) considered four models when they examined the accuracy of a regression approach for estimating PPH using county data from four states, Florida, Illinois, Texas, and Washington (N = 462).

Model 1 (Basic model) is defined as:

$$PPH_{it} = a + b_1{*}Births_{it} + b_2{*}School_{it} + b_3{*}Medicare_{it}$$

where
a = the intercept term and b_1, b_2, and b_3 = regression coefficients
i = county (i = 1 to 462)
t = time (a given year, 1970, 1980, and 1980)
PPH = Persons Per Household
Births = Births Per Household
School = K-12 School Enrollment Per Household
Medicare = Medicare Enrollment (age 65+) Per Household

Smith and his colleagues found that the basic model had the following characteristics for 1970, 1980, and 1990, respectively:

$$\hat{PPH}_{i70} = 2.477 + 10.199{*}Births_{i70} + 0.359{*}School_{i70} - 0.487{*}Medicare_{i70}$$

Adj. $R^2 = 0.756$ and all coefficients are statistically significant ($p \leq .01$)

$$\hat{PPH}_{i80} = 2.072 + 11.192{*}Births_{i80} + 0.512{*}School_{i80} - 0.164{*}Medicare_{i80}$$

Adj. $R^2 = 0.744$ and all coefficients are statistically significant ($p \leq .01$)

$$\hat{PPH}_{i90} = 2.077 + 12.213{*}Births_{i90} + 0.473{*}School_{i90} - 0.288{*}Medicare_{i90}$$

Adj. $R^2 = 0.763$ and all coefficients are statistically significant ($p \leq .01$)

Model 2 (Ratio model) is defined as:

$$PPH_{ikt} = a + b_1*(Births_{ik}/Births_k)_t + b_2*(School_{ik}/School_k)_t + b_3*(Medicare_{ik}/Medicare_k)_t$$

where
a = the intercept term and b_1, b_2, and b_3 = regression coefficients
ik = county i in state k (i = 1 to 67, k=1; i=1 to 102, k=2; i=1 to 254, k=3; i=1 to 39, k=4)
k = state (1 = Florida, 2 = Illinois, 3 = Texas, and 4 = Washington)
t = time (a given year, 1970, 1980, and 1980)
PPH = Persons Per Household
Births = Births Per Household
School = K-12 School Enrollment Per Household
Medicare = Medicare Enrollment (age 65+) Per Household

Smith and his colleagues found that the ratio model had the following characteristics for 1970, 1980, and 1990, respectively:

$$\hat{PPH}_{ik70} = 0.727 + 0.124*(Births_{ik}/Births_k)_{70} + 0.180*(School_{ik}/School_k)_{70} - 0.035* (Medicare_{ik}/Medicare_k)_{70}$$

Adj. $R^2 = 0.843$ and all coefficients are statistically significant ($p \leq .01$)

$$\hat{PPH}_{ik80} = 0.751 + 0.086*(Births_{ik}/Births_k)_{80} + 0.180*(School_{ik}/School_k)_{80} - 0.021* (Medicare_{ik}/Medicare_k)_{80}$$

Adj. $R^2 = 0.821$ and all coefficients are statistically significant ($p \leq .01$)

$$\hat{PPH}_{ik90} = 0.738 + 0.125*(Births_{ik}/Births_k)_{90} + 0.154*(School_{ik}/School_k)_{90} -0.021* (Medicare_{ik}/Medicare_k)_{90}$$

Adj. $R^2 = 0.836$ and all coefficients are statistically significant ($p \leq .01$)

Model 3 (Change model) is defined as:

$$(\mathrm{PPH}_{it} - \mathrm{PPH}_{ic}) = a + b_1{*}(\mathrm{Births}_{it} - \mathrm{Births}_{ic}) + b_2{*}(\mathrm{School}_{it} - \mathrm{School}_{ic}) + b_3{*}(\mathrm{Medicare}_{it} - \mathrm{Medicare}_{ic}) + b_4{*}(\mathrm{PPH}_{ic})$$

where
a = the intercept term and b_1, b_2, and b_3 = regression coefficients
i = county (i = 1 to 462)
t = time (a given year, 1970, 1980, and 1990)
c = preceding census (1970, 1980)
PPH = Persons Per Household
Births = Births Per Household
School = K-12 School Enrollment Per Household
Medicare = Medicare Enrollment (age 65+) Per Household

Smith and his colleagues found that this model had the following characteristics for 1970-80, and 1980-90, respectively:

$$\widehat{(\mathrm{PPH}_{i80} - \mathrm{PPH}_{i70})} = 0.241 + 1.851{*}(\mathrm{Births}_{i80} - \mathrm{Births}_{i70}) + 0.225{*}(\mathrm{School}_{i80} - \mathrm{School}_{i70}) - 0.621{*}(\mathrm{Medicare}_{i80} - \mathrm{Medicare}_{i70}) - 0.162{*}(\mathrm{PPH}_{70})$$

Adj. $R^2 = 0.559$ and all coefficients are statistically significant ($p \leq .01$)

$$\widehat{(\mathrm{PPH}_{i90} - \mathrm{PPH}_{i80})} = 0.014 + 1.544{*}(\mathrm{Births}_{i90} - \mathrm{Births}_{i80}) + 0.934{*}(\mathrm{School}_{i90} - \mathrm{School}_{i80}) - 0.219{*}(\mathrm{Medicare}_{i90} - \mathrm{Medicare}_{i80}) - 0.033{*}(\mathrm{PPH}_{80})$$

Adj. $R^2 = 0.490$ and all coefficients are statistically significant ($p \leq .01$) except the intercept

Model 4 (Ratio Change model) is defined as:

$$[(PPH_{ik} / PPH_{k})_{t} - (PPH_{ik} / PPH_{k})_{c}\,] = \; a + b_{1}*[(Births_{ik}/Births_{k})_{t} - (Births_{ik}/Births_{k})_{c}\,]$$

$$+\, b_{2}*[(School_{ik}/School_{k})_{t} - (Births_{ik}/Births_{k})_{c}\,]$$

$$+\; b_{3}*[(Medicare_{ik}/Medicare_{k})_{t} - (Medicare_{ik}/Medicare_{k})_{c}\,]$$

$$+\, b_{4}*(PPH_{i})_{c}$$

where

a = the intercept term and b_1, b_2, and b_3 = regression coefficients
ik = county i in state k ($i = 1$ to 67, $k=1$; $i=1$ to 102, $k=2$; $i=1$ to 254, $k=3$; $i=1$ to 39, $k=4$)
k = state (1 = Florida, 2 = Illinois, 3 = Texas, and 4 = Washington)
t = time (a given year, 1980, and 1990)
c = preceding census (1970, 1980)
PPH = Persons Per Household
Births = Births Per Household
School = K-12 School Enrollment Per Household
Medicare = Medicare Enrollment (age 65+) Per Household

Smith and his colleagues found that this model had the following characteristics for 1970-80 and 1980-90, respectively:

$$\widehat{[(PPH_{ik} / PPH_{k})_{80} - (PPH_{ik} / PPH_{k})_{70}\,]} = \; 0.138 + 0.029*[(Births_{ik}/Births_{k})_{80} - (Births_{ik}/Births_{k})_{70}\,]$$

$$+\, 0.070*[(School_{ik}/School_{k})_{80} - (Births_{ik}/Births_{k})_{70}\,]$$

$$-\; 0.047*[(Medicare_{ik}/Medicare_{k})_{80} - (Medicare_{ik}/Medicare_{k})_{70}\,]$$

$$-0.041*(PPH_{i})_{70}$$

Adj. $R^2 = 0.490$ and all coefficients are statistically significant ($p \leq .01$)

$$[(PPH_{ik}/PPH_k)_{90} \overset{\wedge}{-} (PPH_{ik}/PPH_k)_{80}] = 0.011 + 0.040*[(Births_{ik}/Births_k)_{90} - (Births_{ik}/Births_k)_{80}]$$

$$+ 0.132*[(School_{ik}/School_k)_{90} - (Births_{ik}/Births_k)_{80}]$$

$$- 0.022*[(Medicare_{ik}/Medicare_k)_{90} - (Medicare_{ik}/Medicare_k)_{80}]$$

$$- 0.006*(PPH_i)_{80}$$

Adj. $R^2 = 0.470$ and all coefficients are statistically significant ($p \leq .01$) except the intercept and b4 (-0.006, for lagged persons per household, $(PPH_i)_{80}$)

While the preceding models are a bit dated and cover only four states (Florida, Illinois, Texas, and Washington), they offer guidance on the construction of regression models for estimating PPH using more current data for all counties in the U.S. It appears that the state-specific models offer more promise than those that cross state boundaries

IV. Evaluation of Methods used to Generate VR and PPH Data

Evaluation Criteria

Without question, an estimate should be accurate, but accuracy is not the only criterion by which an estimate should be judged. Following the argument presented by Swanson and Tayman (1995), I suggest that attention be focused on the broader concept of utility. As described in Section I, there are many methods that in principle can be used to estimate a population, and improvements are a regular feature of these methods. Further, there is a wide range of decision-making situations in which population estimates are used. It follows, therefore, that no method should be universally judged to be superior to others and, by the same token, neither should any method be judged universally inferior to all others. I suggest instead, that relative to a given use, utility is gained by selecting a method that provides a sufficient amount of information for the purpose(s) at hand, while keeping cost and time to a minimum. In the case of an estimate, the sufficiency of the information provided is judged on the ability of using it to make good decisions. So, if an estimate is produced at minimal cost but provides timely information sufficient to make good decisions, then it has high utility. If an estimate does not meet these conditions then it has low utility. An important underlying component of sufficiency is "transparency." That is, the ability of a decision-maker to understand how an estimate was done so that he or she can determine if the assumptions, methods, and data are reasonable. Another important component of utility that affects agencies responsible for (hierarchically-structured) sets of estimates across geographic areas is consistency, especially when different methods are used at different levels of geography. This is sometimes referred to as a "one-number roll-up" and it is obvious that this is an important component for the Census Bureau.

Like most other methods for generating population estimates (See, e.g., Bryan, 2004b), it has been possible to evaluate methods for estimating total population estimates by using decennial census counts (recognizing that unlike the total population counts and the ascribed characteristics such as age, race, and sex, virtually all of the ascribed population characteristics in the decennial census counts were derived from the sample-based "long form"). Fortunately, this will not change in terms of VR and PPH in 2010, when the Census Bureau abandons the long form in favor of the ACS. Thus, unlike many population and housing characteristics, the decennial census rather than the ACS will remain the future "gold standard" against which methods for estimating VR and PPH can be evaluated. This is fortuitous, because traditional frameworks and criteria for evaluating the accuracy of VR and PPH estimates will not require modification However, it is worthwhile to note in this regard that even though decennial census data provide the most convenient and accurate standard against which to evaluate population estimation methods, there always have been several important considerations that were taken into account before estimates are compared with census values. First, there was often a tendency to assume that earlier and later censuses are completely consistent, but such consistency cannot be taken for granted. Second, subnational areas often differ in geography and populations covered, and census definitions may have changed as well. Third, where a method was based on a past census that differs from a more recent census against which estimates resulting from the method in question were compared, in any significant way, an accurate evaluation was compromised.

All of this leads up to the point that comparisons of estimates resulting from different methods against the census have to be considered "measures of difference" rather than "measures of error" because it is virtually impossible to precisely determine the degree to which error in the census and error in the estimate contribute to the overall difference. This affects the evaluation of methods used to estimate VR and PPH.

Similarly, it is important to note that a direct "method-to-method" comparison is rarely possible when attempting to make a population estimate. Often, what might be the most accurate method may not be practicable due to excessive time, cost and resources. Other hindrances may include unavailability or inconsistency of necessary data. Furthermore, it will be seen, certain methods are better suited to particularly large or small areas of geography. While a certain method may generate "good results" at a national level, they may be wholly inadequate for other levels of geography. Thus the amount of resources available, the level of geography as well as historical accuracy of each method must always be considered. General criteria that apply at all levels for evaluating methods for estimating the total population (and its ascribed characteristics , include continuity; timeliness of information; refinement; production; cost, and replication. Generally speaking, these criteria can be applied to data, methods, and administrative structures, topics to which I now turn in regard to VR and PPH.

Evaluation of VR and PPH Data and the Methods for Obtaining Them

This section starts with modeling, moves to administrative data and ends with an evaluation of the uses of surveys as sources of VR and PPH. The ACS is treated separately from other sources and I evaluate the ACS in terms of both VR and PPH data toward the end of this section, just before the summary. One caution before starting to look at the individual elements making up an HUM estimate is the interaction effect observed by Lowe, Weisser, and Myers (1988). The authors observe that of some of the terms in the equation for the Housing Unit Method are themselves correlated and point out that an accuracy improvement in the data used in a given term may end up having an adverse impact on other terms such that the resulting estimate is less accurate. This is an important point because it underscores the need to examine the HUM and its data elements as a system rather than in terms of its individual terms and their data elements. Lowe, Weisser, and Myers (1988) provide suggestions for dealing with this issue.

Vacancy Rates

Holding rates constant over the last census is feasible, but this approach to determining current VR levels should be done in a system such that doing so appears warranted, rather than simply using this approach to the exclusion of others. If other data suggest little change since the last census, then holding PPH values constant may be appropriate. One way to judge where constant rates would work is to examine the IRS migration data and USPS data. If little change is indicated then constant rates may be appropriate. The process of determining if past census VR levels should be changed can be largely automated, with thresholds and decisions points empirically established by examining IRS and USPS data changes relative to past census counts.

Similar to the suggestion for setting up an automated system to determine if it appears feasible to hold VR constant since the last census and assuming that USPS delivery data would be part of it, then the same system could be used to apply USPS trends to previous census VR values as suggested by the work of Lowe (1988), Lowe and Mohrman (2003), and Lowe, Mohrman, and Brunink (2003). Using the USPS data for this purpose would appear to fit the Census Bureau's needs in regard to national coverage, consistency of definitions, and having single source of data. There would be much work to do in regard to the findings of Lowe and her colleagues concerning metropolitan and metropolitan counties, the presence of special populates (e.g., military bases) and the like. However, the Census Bureau is well equipped to undertake the needed analyses, given it has basically all of the data on hand already. Refinements to this general approach – a research agenda – are found in work by Lowe, Weisser, and Myers (1984) as well as Bousfield (1977), Clogg, Schockey, and Elisason (1990), and Weidman et al. (2008). An important potential refinement can be found in Fonseca and Tayman (1989) who develop and evaluate a method for deriving postcensal estimates of household income distributions for counties. The modified lognormal probability curve they use as a model of income distribution cold be adopted for use with VR, with the USPS delivery data and census

VR data serving the role that IRS income data and decennial income data do for Fonseca and Tayman (1989), respectively.

The windshield surveys mentioned in the preceding section could become a tool in the "automated review" system I have just proposed. If indications are such that neither holding rates constant nor adjusting previous census data using trends shown from USPS data looks viable for a given area, then a windshield survey might be considered, given the size and type of area (as described in the previous section. If any are used, it would be useful to have areas 'clustered" such that findings from one windshield survey (e.g., changes since the last census) could be applied to other areas within the same cluster. The variables used to identify such clusters could include size, metropolitan status, presence of a downtown core, percent of multiple unit structure, and so forth. Given the size training, and experience of the Census Bureau's field operations and staff in regard to collecting data, conducting windshield surveys as indicated by the automated review system may be a tractable solution.

Persons Per Household

As was suggested in regard to VR values, holding PPH values constant over the last census is feasible, but it should be done in a system such that doing so appears warranted, rather than simply using this approach to the exclusion of others. Thus, it appears preferable to use administrative data (e.g., IRS migration, USPS deliveries, vital statistics, K-12 school enrollment, Medicare, and covered employment) to determine if PPH change is likely. If little change is indicated then holding PPH values constant may be appropriate. Like the process of determining if VR values could be held constant, the process of determining if past census PPH levels should be held constant can be largely automated, with thresholds and decisions points empirically established by examining administrative data changes relative to past census counts.

Trend extrapolation is inexpensive and easy to implement. It also has been found to provide accurate projections when used judiciously (Smith , Tayman, and Swanson, 2001: 167-183). That is, by not extending the past too far into the future and by understanding the demographic dynamics being modeled by an extrapolation technique such as the geometric model in conjunction with PPH data from successive and recent census counts. It is worthwhile here to point out that while trend extrapolation may not be well suited to capture certain aspects of demographic change (e. g, cohort effects on the total population), the general approach used to generate the estimated 2000 PPH values found in Table 1 is well suited to capture PPH changes. It is so, because two successive decennial census points provide a good time frame for the effects of demographic determinants on PPH to be observed and a geometric model used to extrapolate this change effectively exploits the inertia underlying it. Recall that Burch (1967) and others (Bumpass, 1990; Burch et al., 1987; and Coale, 1965 have identified these determinants and found that their effects on PPH are not played out over a short period of time. These aspects generally mean that extrapolation methods such as found in the geometric model can work well, as is shown in Table 1.

IRS data on the number of exemptions per return have been examined with an eye toward using them to model current PPH values (Swanson and Lowe, 1979; Voss and Krebs, 1979). One problem noted by Swanson and Lowe (1979) is that the use of the IRS data effectively collapses all structure types into one category. This is a problem because of the different PPH levels associated with different structure types (Swanson and Lowe, 1979) , which leads to their suggestion to use IRS data in conjunction with PPH data from the Current Population Survey.

Their report preceded the ACS by 20 years, so an appropriate revision of their suggestion would be to use the IRS data in conjunction with PPH values taken from the ACS, which could be massaged using, for example, techniques described by Bousfield (1977) , Clogg, Schockey, and Eliason (1990), and Lowe (2000a).

The regression models for estimating PPH that were constructed by Stan Smith, June Nogle, and Scott Cody (2002) show that this approach has considerable promise for the Census Bureau, with model 2 (Ratio model) standing out. Recall that this model is state-specific and defined as follows:

$$PPH_{ikt} = a + b_1*(Births_{ik}/Births_k)_t + b_2*(School_{ik}/School_k)_t + b_3*(Medicare_{ik}/Medicare_k)_t$$

where
a = the intercept term and b_1, b_2, and b_3 = regression coefficients
ik = county i in state k (i = 1 to 67, k=1; i=1 to 102, k=2; i=1 to 254, k=3; i=1 to 39, k=4)
k = state (1 = Florida, 2 = Illinois, 3 = Texas, and 4 = Washington)
t = time (a given year, 1970, 1980, and 1980)
PPH = Persons Per Household
Births = Births Per Household
School = K-12 School Enrollment Per Household
Medicare = Medicare Enrollment (age 65+) Per Household

The results of the models constructed by Smith and his colleagues (2002) for 1970, 1980, and 1990 are very good in terms of their statistical properties, as is shown below.

1970 Model

$$\hat{PPH}_{ik70} = 0.727 + 0.124*(Births_{ik}/Births_k)_{70} + 0.180*(School_{ik}/School_k)_{70} - 0.035* (Medicare_{ik}/Medicare_k)_{70}$$

Adj. $R^2 = 0.843$ and all coefficients are statistically significant ($p \leq .01$)

1980 Model

$$\hat{PPH}_{ik80} = 0.751 + 0.086*(Births_{ik}/Births_k)_{80} + 0.180*(School_{ik}/School_k)_{80}$$

$$- 0.021^{*} (\mathrm{Medicare}_{ik}/\mathrm{Medicare}_{k})_{80}$$

Adj. $R^2 = 0.821$ and all coefficients are statistically significant ($p \leq .01$)

1990 Model

$$\widehat{PPH}_{ik90} = 0.738 + 0.125^{*}(\mathrm{Births}_{ik}/\mathrm{Births}_{k})_{90} + 0.154^{*}(\mathrm{School}_{ik}/\mathrm{School}_{k})_{90}$$

$$-0.021^{*} (\mathrm{Medicare}_{ik}/\mathrm{Medicare}_{k})_{90}$$

Adj. $R^2 = 0.836$ and all coefficients are statistically significant ($p \leq .01$)

However, even though the characteristics do not seem as favorable as this is the case with Model 2 (Ratio Model), I also would not discount Model 4 (Ratio Change Model). I make this suggestion based on findings about ratios combined with change in regression models used to estimate population by Swanson (2004) and the fact that Model 4 could be modified using ideas from Swanson (1980), Swanson and Beck (1994), Swanson, Tayman, and Beck (1995), and Swanson and Tedrow (1984). While I cannot say for certain if Model 4 is sufficient, I believe it would be worth examining it further, with the idea that it may be sufficient as is, or that revisions may make it so.

In regard to the work of Smith and his colleagues (2002) in developing and examining their models, it should be noted that the “traditional” methods against they were evaluated did not include geometric and other non-linear models. Instead, the models were assessed relative to the following three ‘traditional methods: (1) holding PPH constant for each county from the last census; taking the percent change in PPH for each during the previous decade, applying it to the most recent census PPH value and bringing it forward accordingly; and (3) using the percent change in PPH at the state level since the most recent census and applying this uniformly to the most recent census PPH value in each county and bringing them each forward accordingly. Thus, I suggest that if the Census Bureau examines regression methods for estimating PPH relative to other methods that the geometric model and its non-linear relatives be included.

One important finding that emerges from the work of Smith, Nogle and Cody (2002) is that it is consistent with other work that stresses that 'no-size fits all' when it comes to the HUM and other methods of population methods. The use of specific local area (in this case, county) data is important as is using state-specific models. Thus, I suggest that the Census Bureau take these two related issues in to account as it considers how to implement the HUM nationwide. In addition, in regard to the HUM itself, it is wise to consider the findings of Lowe, Weisser, and Myers (1988) in regard to the interactions of terms in the HUM. It is important to realize that knowledge that improvement in the accuracy of a given data element such as housing units may adversely impact the accuracy of the HUM overall, which suggests that improvements to data used in the HUM need to be assessed in a comprehensive manner rather than simply assuming that accuracy improvements in one element will automatically lead to improvement in the accuracy of the overall population estimates.

An important factor in regard to regression models, is that they can be used with sample data (e. g, Ericksen , 1973, 1974), which brings in the ACS as a data source, a point to which I now turn.

The ACS

Background

The American Community Survey (ACS) is a U. S. Census Bureau product designed to provide accurate and timely demographic and economic indicators on an annual basis for both large and small geographic areas within the United States (Citro and Kalton, 2007; U. S. Census Bureau, 2004b). Operational plans call for ACS to serve not only as a substitute for the decennial census long-form, but as a means of providing annual data at the national, state, county, and sub-county levels (Cork, Cohen, and King, 2004; Smith, 1998; U. S. Census Bureau, 2001a, 2001b, 2003, 2004b) . In addition to being highly ambitious, this approach represents a major change in how data are collected and interpreted (Citro and Kalton, 2007; Hough and Swanson, 1998, 2006). Two of the major questions facing the ACS are its functionality and usability (Citro and Kalton, 2007).

As has been noted earlier, the Decennial Census, (along with the CPS, and SIPP, among other products) uses the De jure rule of residency while the ACS uses what amounts to a De facto residency rule. However, when the micro-level ACS data are aggregated to geographic areas, they are controlled to number produced for these areas by the Census Bureau's annual population estimates program. This may not be a huge issue at the national level, but at sub-state levels, the effects of these different residency rules could be substantial. Where ACS micro-level data are used in conjunction with micro-level data from the CPS, SIPP, or the (2000) census, it is highly likely that mismatches occur. Add to this difference, there is evidence that ACS values at the sub-state level are subject to high levels of variation in general (Fay, 2007; Van Auken, Hammer, Voss, and Veroff, 2006). Moreover, there is now evidence that PPH values taken from the ACS are affected by both of these issues (Swanson and Hough, 2007) .

Putting these two concerns together leads to a high level of uncertainty in regard to the applicability of ACS-derived PPH values top given areas. As an illustration of the potential magnitude of this uncertainty, consider Table 2 (in four parts, a, b, c, and d), which compares "official population estimates" produced by the Census Bureau for 115 cities in California with "ACS population estimates" for these cities (San Francisco is shown in the table, but excluded from the analysis because the city boundaries are the same as the county boundaries). Table 2 also shows the margins of error around the ACS population estimates and shows which cities have official estimates that are not within the respective margins or error. Even with relatively wide margins of error the official estimates are not contained within them in 20 of the 115 cities. That is, 17.4% of the cities have official estimates that are beyond the margins of error of the corresponding ACS estimates. It likely to be discouraging for all users that an agency would be producing two sets of estimates for cities nationwide. Particularly disturbing is the fact that given there are two sets of city numbers, over 17 percent of the 2007 official estimates for California cities are beyond the ACS margins of errors. Hopefully, on-going research within the Census Bureau (e.g., Robinson and Dixon, 2009; Weidman et al., 2008) along with administrative cognizance of the adverse effects of having two sets of estimates will lead to resolutions of these two related issues.

(TABLES 2a through 2c ABOUT HERE)

This following subsection serves as an introduction to an evaluation of the use of ACS data in regard to VR and PPH. In the following section, I focus specifically on VR and then on PPH.

ACS Vacancy Rates

In regard to ACS vacancy rates, it is highly likely that they have too much variance to use. However, this serves as a research question rather than a conclusion because VR itself is subject to higher levels of variation than PPH values over a given interval of time. In addition to the issue of statistical variance, the interaction of seasonality and residency differences between the ACS and the decennial census also must be considered. It is more difficult to assess the effects of statistical uncertainty and the interaction of seasonality and residency differences on VR than it is on PPH because changes in the determinants of VR lack the inertia associated with changes in the demographic determinants of PPH. With these point in mind, suggestions for analysis of the suitability of ACS VR values for use in an HUM estimation system include: (1) conducting a broad scale comparison, taking note of county-level 'market conditions' that are likely to have impacts on VR levels and their changes; and (2) making adjustments to ACS VR values (deriving model-based PPH values from the ACS) that may provide more statistical stability.

ACS Persons Per Household[7]

The data used in this exploration of the usability and functionality of ACS PPH data are taken from 18 counties that were in the 1999 ACS test sites (See Exhibit 1). The examination proceeds by comparing ACS PPH values for these 18 counties to PPH values generated using a geometric model based on PPH change from Census 1990 to Census 2000. The ACS PPH values represent what could be called the "statistical perspective" because variations in the values of specific variables over time and space are viewed largely by statisticians with an eye toward sample (and non-sample) error (Citro and Kalton, 2007; Fay, 2005; Kish, 1998; Purcell and Kish, 1979, 1980; U. S. Census Bureau, 2001a, 2001b, 2003, 2004b). The model-based PPH values represent a "demographic perspective" because PPH values are largely viewed by demographers as varying systematically, an orientation stemming from theory and empirical evidence that PPH values respond to demographic and related determinants (Burch, 1967, 1970; Burch et al., 1987; Coale, 1965; Goldsmith, Jackson, and Shambaugh, 1982; Kimpel and Lowe, 2007; Korbin, 1976; Myers and Doyle, 1990; Smith, Nogle, and Cody, 2002).

(EXHIBIT 1 ABOUT HERE)

As noted earlier, the HUM is based on the assumption that virtually everyone lives in some type of housing structure. It is generally accepted that the HUM is the most commonly used method for making small area population estimates in the United States (Byerly, 1990; Smith, Nogle, and Cody, 2002). One of the reasons for this is that current data for two of its elements are generally available, the number of households and the group quarters population (Smith, Nogle, and Cody, 2002). The other remaining element needed to get the household population is PPH. Until the full implementation of the ACS, current PPH values were obtained by using the value from the most recent census or extrapolating trends found from the two most recent decennial censuses (Bryan, 2004b; Smith, Nogle, and Cody, 2004; Swanson, Baker, and Van Patten, 1983). With the expansion of the ACS to its full design in 2005 (Griffin and Waite, 2006), it is not surprising that among the large number of HUM users, more than a few are interested in seeing if the ACS can provide usable PPH values. Thus, this evaluation.

Evaluation Data

The U. S. Census Bureau established the operational structure for the ACS in 1994 when it put in place the "Continuous Measurement Office," which implemented the first operational test of the ACS in four test sites in 1995 (Griffin and Waite, 2006). These test sites were subsequently expanded, and by 1999, operational tests took place in 36 counties spread across 26 states (Griffin and Waite, 2006). Three year ACS averages centered on 2000 were set up for these counties to support comparisons with Census 2000. Relevant among the many findings of these tests was that the arithmetic mean (2.63) of the PPH values found in the ACS for these 36 counties was the same as that found in Census 2000 and that there were no statistically significant differences for PPH (U. S. Census Bureau, 2004b: 17). It was also noted that this result was not unexpected because the total household population and the total number of housing units found in Census 2000 are used as control variables in ACS weighting (U. S. Census Bureau 2004b: 17).

Among the 36 ACS test counties, annual PPH values estimated from single-year ACS collections are available online for 21 of them for the period 2001 to 2006; annual PPH values estimated from three-year ACS collections are available online for 18 of these same 21 counties for the period 1999-2001 to 2003-2005. (See Exhibit 1). It is for these 18 counties that both single-year and three-year ACS PPH values are used in our comparison with model-based PPH values.

The analytical method for generating the model-based PPH values is one method commonly used by applied demographers for this purpose, namely, the geometric rate of change (Lowe, Pittenger, and Walker, 1977; Smith, Nogle, and Cody, 2002; Smith, Tayman, and Swanson, 2001; Swanson, Baker, and Van Patten, 1983). In this approach, the rate of change is benchmarked to two most recent successive census counts and then applied to the PPH value found in the most recent census count, which is then extrapolated beyond the most recent census by applying the rate of change to it.

The process takes place in two steps. The first is the calculation of the rate of change in PPH, which is described in the preceding section as equation, along with the second step (applying the rate to the launch year to obtain current PPH values), so I will not repeat them here.

Results

The data for the 18 counties are shown in exhibits 2 through 19. Each of these exhibits is divided into two parts. The first part shows the single-year ACS PPH values for each year from 2001 to 2006 while the second part shows the three-year ACS PPH values for each year from 2001 to 2005, the latter corresponding to the ACS collections from 1999-2001 to 2003-2005. Both parts of each of the exhibits also show the annual ACS values generated using the geometric model. The ACS PPH values are labeled as “ACSPPH” in each of the two parts and the model-generated PPH values are labeled as “ADPPH” (where “AD” stands for Analytically Derived).

(EXHIBITS 2 through 19 ABOUT HERE)

Before discussing the results, it is important to note that the PPH values generated by geometric trend extrapolation are used as benchmarks not because they are inherently more accurate than those derived from other models or from samples such as the ACS, but, rather, because they represent the type of systematic change demographers expect to see in PPH values. However, in order to provide evidence that county level PPH values generated by the geometric trend extrapolation method are reasonably accurate, refer to Table 1 which is found in Section II.

In this test, Census 1980 and 1990 PPH values are used as input to the geometric model, which is applied to the Census 1990 PPH values to generate PPH values for 2000. These estimated PPH values are then compared to Census 2000 PPH values. The results support the argument that the geometric method is capable of generating PPH values sufficiently accurate for use in post-censal HUM estimates: (1) The mean error is 0.068; (2) the mean absolute percent error is 2.97; (3) the mean algebraic percent error is -2.60; and (4) the number of absolute percent errors that are 5.0 or greater is six.

In comparing the single-year ACS PPH values to the model-based PPH values, the ACS PPH values are above the model-based PPH values in seven counties for the entire period, 2001-2006, that they are below the model-based values in two counties for the entire period and cross over the model-based values in nine counties (three of which (Bronx, Multnomah, and Schuylkill) have two crossovers each and one of which (Jefferson) has three crossovers). In terms of directional changes, the single-year ACS PPH values change direction three or more times in three counties, twice in nine counties, and once in six counties.

The three-year ACS PPH values remain above the model-based values for the entire period , 1999-2001 to 20003-2005 in nine counties, while in only one county (Yakima) they remain below the model-based values, and cross over the model based values nine times. The three-year ACS PPH values change direction twice in two counties and once in seven counties. In the remaining nine counties no directional changes are observed, although there are some in which trends become flattened for some of the time. The model-based PPH values show a secular decline in 11 counties and an increase in seven.

In some of the counties with declining model-based PPH values, the trends are very slight (e.g., Pima County, Arizona) and in others, more distinct (Schuylkill County, Pennsylvania). Similarly, some of the counties with increasing model-based PPH values have a very slight upward trend (e.g., San Francisco County, California), in others they are much more pronounced (e.g., Tulare County, California).

Table 2 provides mean PPH values across the 18 counties (and their standard deviations) by year. Not surprisingly, the single-year ACS PPH values exhibit the least systematic change over time and the most variation each year. In two of the six years, these values are less than the model-based PPH values while in the remaining four years they exceed the model-based values The means of the three-year ACS PPH values show a systematic decline over time with annual variations comparable to the model-based PPH values.

Discussion of ACS PPH Data

As noted earlier, the U. S. Census Bureau found encouraging results for the three-year ACS PPH values among the set of 1999 test counties when it compared the 1999-2001 numbers to the PPH values of the 2000 Census (U. S. Census Bureau 2004b). As also was noted earlier, this finding was no surprise because the total household population and the total number of housing units found in Census 2000 are used as control variables in ACS weighting. Given this, the results found here are a bit discouraging, given that these same variables are also used as control variables in ACS weighting – with one major change – once beyond the 2000 census, the total household populations and housing units are not enumerated directly, but, instead, estimated.

(TABLE 2 ABOUT HERE)

Not surprisingly, it is the single-year ACS PPH estimates that are the most discouraging. They jump around a great deal from year to year in many of the counties, a characteristic that is not desirable for both demographers who use the HUM and the stakeholders for whom HUM estimates are done. This is because there is an expectation on the part of both these demographers and the stakeholders that PPH values should exhibit systematic changes unless there is compelling substantive evidence (e.g., the PPH values jumped because of a surge of in-migrants with high fertility and large family sizes) to the contrary. If such PPH values are used in the absence of compelling substantive evidence justifying their temporal instability then it appears to me that the risk of challenges and related administrative and legal actions increases (See, e.g., Walashek and Swanson, 2006), especially when these estimates are used to allocate resources, which is often the case (National Research Council, 1980, 2003; Scire, 2007).

In considering the three-year ACS PPH values, the results are not as discouraging, as those for the single-year values, but neither are they strongly encouraging. These values change more systematically than do the single-year ACS PPH values, but they still exhibit temporal instability. However, how one uses the three-year data is not very clear. Can one use them for the first year of the three-year interval as well as the second and third years? The Census Bureau stresses that they are interval rather than point estimates, but the fact that they are linked to point estimates (e.g., controlled to annual population estimates) and are needed in return for point estimates (e.g., annual population estimates) leaves many questions about their use.

In addition to the temporal instability issue, which is itself partly a function of statistical variance over time, one must ask what causes some of the substantial differences observed between the mean ACS PPH values and the mean model-based PPH values. For example, in 2001, the mean ACS PPH is 2.503 while the model-based mean is 2.627. This is a substantial difference, one likely beyond the scope of simple sampling error. Is this difference partly due to the ACS residency rule? After all, it is not the same as the Decennial Census residency rule, the one that is inherent in the model-based ACS PPH values. With a two-month rule, the ACS clearly will tend to have higher PPH values in areas in which seasonal migrants are currently residing than would be the case with the "majority of your time" rule used by the Decennial Census. This might explain in part the higher ACS PPH values found in Pima County, Arizona, However, if this were the case, one would expect that the ACS PPH values would consistently be higher than the model-based PPH values in Tulare County, California, but they are not.

The ACS as a Source of PPH Data

As described at the start of this section, the ACS provides annual PPH estimates that are subject to sample (and non-sample) error. This means that they can fluctuate from year to year in a given population, which reflects a "Statistical Perspective." Demographers, however, tend to view PPH as a population attribute that has demographic determinants. This implies that demographers view PPH as an attribute that changes systematically over time - the "Demographic Perspective." The comparisons suggest that the ACS PPH values exhibit too little systematic change over time for a given area to be usable by demographers and others preparing post-censal population estimates.

The finding that the ACS PPH values are not particularly usable for purposes of making HUM-based population estimates is preliminary in nature. More work needs to be done not only to confirm this finding, but also to figure out if the ACS PPH values can be modified so that they could be used if the finding is confirmed. With this in mind, our suggestions for further analysis include: (1) conducting a broader scale comparison, taking into account the full range of counties; (2) examining ACS PPH values that are not controlled; (3) consideration of a way to utilize sample error (i.e., confidence intervals) in determining ACS PPH changes over time; (4) an examination of 5-year ACS PPH values when at least five years of data become available; and (5) making adjustments to ACS PPH values (deriving model-based PPH values from the ACS) that may provide more temporal stability.

The ACS is a resource of high potential value to all stakeholders and ACS PPH values represent the same type of resource to demographers making population estimates and their stakeholders (See, e.g., Smith, 1998). The goal of our suggestions for further research is to see if the ACS PPH values can become usable in terms of the demographic perspective, especially as implemented in HUM-based estimates.

In conclusion, I note that differences between statisticians and demographers are stressed n this section. However, a demographic perspective is not incompatible with a statistical perspective. At one level, the demographic perspective can be viewed as a model-based approach, a perspective that is shared with statistics (Hill, 1990; Jiang and Lahiri, 2006). Further, as noted throughout this report, demographers view PPH as a variable that responds to demographic and related determinants. Thus, at another level, the demographic perspective described here represents 'causality.' This also is a perspective that is shared with statistics (Cox and Wermuth, 2004). Finally, at a third level, the demographic perspective is empirical, which also is a perspective that is shared with statistics – Stigler (1986: 1) observes, for example, that ""..Modern statistics provides a quantitative technology for empirical science;…"

In short, the view that PPH is a variable that responds to demographic and related determinants is not only worthy of consideration, but one that is compatible with statistics, broadly speaking. I have identified three shared commonalities - a model-based perspective, a causal perspective, and an empirical perspective - that support this argument.

Summary

This review and evaluation of data and methods suggests that the Census Bureau continue with its development of a comprehensive system based on the HUM for the annual estimates of county populations nationwide. Ideally, this system would include current and historical housing unit data from previous census counts and the MAF as well as: (1) VR and PPH data from at least the two most recent decennial census counts; (2) postal deliveries for the same census years and the years since the most recent census; (3) for the same census years and years and all other available years, IRS data, Medicare enrollment, covered employment, school enrollment, births and deaths and other data available for all counties that have consistent definitions nationwide; and (4) ACS data, such as PPH and VR. I suggest that the system be set up to handle HUM estimates by structure type, knowing that for some data elements the structure type data are not available (e.g., the IRS data). This system would also include the standard geographic identifiers used by the Census Bureau along with derived classifications such as metropolitan and non-metropolitan status. In addition, it could include markers for the presence of seasonal populations and special populations by type (military, college dormitory populations, prisons, etc.).

With these and other data (as determined in a more rigorous examination), the system could be set up so that it has default actions (e.g., carry the VR and PPH values forward from the previous census) that are subject to automated checks that would ‘flag’ areas if the indicators used I the checks suggested that the default procedures may not be optimal (e.g., since the last census the IRS data show substantial in-migration as do the school enrolment data and the MAF shows added single unit housing structures) . If the flags are triggered, then a series of subsequent flags could indicate which other data (and their underlying methods) might be better suited (e.g., use the change in USPS delivery data to modify VR). Having this capability, of course, implies that the various methods need to be actively implemented in the system. For example, if PPH for a given county is flagged such that holding it constant since the last census is not optimal, then the geometric method or a ratio–regression model should be ready to provide alternatives, one of which could be selected by an analyst using established procedures.

In addition to developing the internal data, methods, and administrative procedures for such a system, I suggest that the Census Bureau continue to pursue external links. The HUBERT program serves as an excellent starting point. It could be the case, for example, that some states in the FSCPE have developed excellent systems for capturing changes stock that the Bureau could use. I also suggest that the regional offices be considered as active partners. Here, I cam thinking of the possibility of doing windshield and other surveys to update data for counties in which all of the 'flags' and subsequent analyses suggest that changes are of a magnitude that field work is required to obtain good VR and PPH data. If this was implemented, field staff training would be needed relative to optimal data capture methods and the channels for sending these data back into Suitland.

As noted by Swanson(2006) in his congressional testimony, the Census Bureau and its state and local government partners (as well as the private sector) would all benefit from a universal HUM system employing a functional MAF. He also stated that be believed that the problems he identified in achieving this goal were solvable. His suggested approach requires new thinking, new arrangements, and new relationships, not only by the Census Bureau and its traditional partners in the public sector, but also by the private sector. In developing such a system, I suggest that in addition to the Bureau's own experience with developing and maintaining such systems (e.g., Devine and Coleman, 2003; Judson, 2000; Long, 1993a, 1993b; Marquis, Wetrogan, and Palacios, 1996; Prevost, 1996; Prevost and Leggieri, 1999; Wetrogan, 2007) , it consider the ideas presented by Tayman (1986) in regard to the integrated system used by the San Diego Association of Governments for preparing census tract level estimates, among others (e.g., Alaska Department of Labor, 1981; Kimpel and Lowe, 2007; Lowe and Mohrman, 2003; Lowe, Pittenger, and Walker, 1977; Swanson, Baker, and Van Patten, 1983)

Finally, there is the ACS. In its current stage, the PPH data it generates are not suitable as inputs into an HUM system "as is." Substantial massaging needs to be done to iron out the temporal instabilities and the large variances found for PPH values in many areas. I have not examined them, but I suspect that there high levels of variance associated with VR estimates from the ACS and, as noted earlier, it is much more difficult to sort out variance from change for VR than PPH because of the difference in the determinants underlying them. However, ACS data should be an element in the comprehensive system I am proposing. Using the VR, PPH, and other data it produces will require among other things, a good understanding of the effects of the differences in residency definitions between the decennial census and the ACS. For many counties, these differences may substantially interfere with the use of "raw" and even "controlled" ACS data. As the ACS matures along with the analysts who use data from it, the residency and other issues will likely become resolved. As this occurs, some of the methods I have proposed will become less likely to be used (e.g., holding VR constant since the last census) while others become normal (e.g., using regression based methods in conjunction with census and ACS data to update PPH).

Endnotes

1. The US Census Bureau document distributed at this conference uses the term "inter-censal estimate" in its title, while the document itself clearly makes reference to post-censal estimation work (U.S. Census Bureau, n.d.). I believe, however, that the distinction between inter-censal and post-censal is worth maintaining.

2. For the record, one can also construct estimates for a point in time that predates a census. I have not run across the term "pre-censal," however and so do not use it here. Here it also is useful to note that there is a large body of literature on how to make estimates of populations and their characteristics for countries that lack censuses and good registration systems (Popoff and Judson, 2004). There are also methods developed for the estimation of wildlife populations that can be used with special populations such as the homeless – "capture-recapture" and "transit surveys," for example (Williams, Nichols and Conroy, 2002). However, as is the case with the "statistical" tradition, I do not cover the estimation methods associated with "statistically underdeveloped areas" and wildlife populations

3. The MAF is already being used for "direct estimation" because it forms the sample frame for the Census Bureau's "American Community Survey."

4. The synthetic method of estimation is defined by Swanson and Stephan (2004: 776) as "a member of the family of ratio estimation methods used to estimate characteristics of a population in a sub-area (e. g., a county) by re-weighting ratios (e.g., prevalence rates or incidence rates) obtained from a survey or other data available at a higher level of geography (e.g., a state) that includes the sub-area in question." As alluded to in the preceding definition, the synthetic method is usually viewed as belonging to the statistical tradition because of its frequent use with survey data. For a description of the synthetic method see Judson and Popoff (2004: 681-683). I also note that the "composite" method (Bryan, 2004b: 550-551) is a type of synthetic estimation.

5. While the United States lacks a national population registration system there are, as noted in the body of the report, administrative records in the private sector that contain information on people that is used for commercial purposes (e.g., credit reporting systems such as those operated by Equifax, Experian, and TransUnion). Experian also conducts consumer marketing activities (See endnote # 9). These systems can be used to generate population estimates. However, using them requires money and the accuracy of such estimates is hard to judge because of the proprietary nature of the data.

6. Although their discussion of such adjustments is in the context of making projections rather than estimates, Smith, Tayman, and Swanson (2001: 239-277) provide a comprehensive description that covers many of the same issues found in developing estimates.

7. This section is adopted from Swanson and Hough (2007).

References

Alaska Department of Community and Regional Affairs. 1981a. *The Housing Unit Method of Population Estimation: A Manual for Municipal Personnel Responsible for Annual Population Estimates.* Juneau, AK: Alaska Department of Community and Regional Affairs, Division of Local Government Assistance and Alaska Department of Labor, Research and Analysis Section.

Alaska Department of Community and Regional Affairs. 1981b. *Standards for Conducting a Population Census in Small Alaskan Cities.* Juneau, AK: Alaska Department of Community and Regional Affairs, Division of Local Government Assistance.

Alaska Department of Labor. 1981. *State of Alaska Census Enumerator's Manual*. Juneau, AK: Alaska Department of Labor, Research and Analysis Section.

Alaska Department of Labor. 1982. *State of Alaska Census Administrator's Manual*. Juneau, AK: Alaska Department of Labor, Research and Analysis Section.

Brass, W., A. 1968. *The Demography of Tropical Africa*. Princeton, NJ: Princeton University Press.

Bryan, T. 2004a. "Basic Sources of Statistics." pp. 9-41 in J. Siegel and D. Swanson (Eds.) *The Methods and Materials of Demography, 2nd Edition*. New York, NY: Elsevier Academic Press.

Bryan, T. 2004b. "Population Estimates." pp. 523-560 in J. Siegel and D. Swanson (Eds.) *The Methods and Materials of Demography, 2nd Edition*. New York, NY: Elsevier Academic Press.

Bryan, T. and R. Heuser. 2004. "Collection and Processing of Demographic Data." pp. 43-63 in J. Siegel and D. Swanson (Eds.) *The Methods and Materials of Demography, 2nd Edition*. New York, NY: Elsevier Academic Press.

Burch, T. 1970. "Some Demographic Determinants of Average Household Size: An Analytic Approach." *Demography* 7 (1): 61-69.

Burch, T. 1967. "The Size and Structure of Families: A Comparative Analysis of Census Data." *American Sociological Review* 32 (3): 347-363.

Burch, T., S. Halli, A. Madan, K. Thomas, and L. Wai. 1987. "Measures of household composition and headship based on aggregate routine census data." pp. 19-33 in J. Bongaarts, T. Burch and K. Wachter (eds.) *Family Demography: Methods and Their Application*. Oxford: Clarendon Press.

Byerly, E. 1990. “State and Local Agencies Preparing Population and Housing Estimates.” *Current Population Reports, Series P-25-1063*. Washington, DC: U. S. Government Printing Office.

Chandrasekaran (Chandra Sekar), C. and W. E. Deming. 1949. “On a Method of Estimating Birth and Death rates and the Extent of Registration.” *Journal of the American Statistical Association* 44: 101-115.

Citro, C. and G. Kalton (Eds.). 2007. *Using the American Community Survey: Benefits and Challenges.* Panel on the Functionality and Usability of Data from the American Community Survey. National Research Council. Washington, DC: The National Academies Press.

Coale, A. 1965. “Estimates of Average Size of Household.” pp. 64-69 in A. Coale, L. Fallers, M. Levy, D. Schneider, and S. Tompkins (Eds.) *Aspects of the Analysis of Family Structure.* Princeton, NJ: Princeton University Press.

Cook, T. 1996. “When ERPs Aren’t Enough: A Discussion of Issues Associated with Service Population Estimation.” *Working Paper 96/4*. Demography Section, Australian Bureau of Statistics, Belconnen, ACT, Australia.

Cork, D., M. Cohen, and B. King (Eds.). 2004. *Reengineering the 2010 Census: Risks and Challenges*. Panel on Future Census Methods, National Research Council. Washington, DC: National Academies Press.

Cox, D. and N. Wermuth. 2006. “Causality: A Statistical View.” *International Statistical Review* 72 (3): 285-305.

Devine, J. and C. Coleman. 2003. “People Might Move but Housing Units Don’t: An Evaluation of the State and County Housing Unit Estimates.” *Population Division Working Paper Series No. 71*. Washington, DC: U. S. Census Bureau (available online at http://www.census.gov/population/www/documentation/twps0071/twps0071.html., last accessed October, 2008).

Devlin, K. 2008. *The Unfinished Game: Pascal, Fermat, and the Seventeenth-Century Letter that Made the World Modern.* New York, NY: Perseus Publishers.

Edelstein, R. and D. Tang. 2007. “Dynamic Residential Housing Cycles Analysis.” *The Journal of Real Estate Finance and Economics* 35 (3): 295-313.

Eldridge, H. 1947. “Problems and Methods of Estimating Post-censal Population.” *Social Forces* 24: 41-46.

Erickson, E. 1974. “A Regression Method for Estimating Population Changes of Local Areas.” *Journal of the American Statistical Association* 69: 867-875.

Ericksen, E. 1973. “A Method for Combining Sample Survey Data and Symptomatic Indicators to obtain Population Estimates for Local Areas.” *Demography* 10: 137-160.

Fay, R. 2005. “Model-assisted estimation for the American Community Survey.” pp. 3016-2023 in *Proceedings of the Joint Statistical Meetings, ASA Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.

Fay, R. 2007. “Variance Reduction for Sub-County Estimates in the American Community Survey.” Paper presented at the Spring Meeting of the Census Advisory Committee of Professional Associations. Suitland, MD: April 19th.

Frisėn, C. 1958. "Effectiveness of Our Tools for Estimating Population Change in Small Areas," pp. 229-232 in *Proceedings of the American Statistical Association, Social Statistics Section*. Alexandria, VA: American Statistical Association.

Frisén, C. 1951. “Symptomatic Data in Population Estimates: Problems and Limitations.” Paper presented at the annual meeting of the Pacific Sociological Association, Berkeley, CA (published in *Research Studies of the State College of Washington*, vol. 19, pp. 107-110).

Gabriel, S. A., and F. E. Nothaft. 2001. “Rental Housing Markets, The Incidence and Duration of Vacancy, and the Natural Vacancy Rate.” *Journal of Urban Economics* 49 (1): 121-49.

Goldsmith, H. D. Jackson, and J. Shambaugh. 1982. “A Social Analysis Approach.” pp. 169-190 in E. Lee and H. Goldsmith (eds.) *Population Estimates: Methods for Small Area Estimates*. Beverly Hills, CA: Sage Publications.

Griffin, D. and P. J. Waite. 2006. “American Community Survey Overview and the Role of External Evaluations.” *Population Research and Policy Review* 25: 201-223.

Habermann, H. 2006. “Research to Improve Population Estimates” Part of a presentation by H. Habermann at the Spring (May 18th -19th) 2006 meeting of the Census Advisory Committee for Professional Associations.

Hauser, P., and B. Tepping. 1944. "Evaluation of Census Wartime Population Estimates and of Predictions of Postwar Population Prospects for Metropolitan Areas.” *American Sociological Review* 9: 473-480.

Hendershott, P. H., B. D. MacGregor and R. Y. C. Tse. 2002. "Estimation of The Rental Adjustment Process." *Real Estate Economics* 30 (2):165-183.

Hill, J. 1990. “A General Framework for Model-Based Statistics.” *Biometrika* 77 (1): 115-126.

Hogan, H. 2007. “Measuring Population Change using the American Community Survey.” pp. 13-30 in S. Murdock and D. Swanson (eds.) *Applied Demography in the 21st Century.* Dordrecht, The Netherlands: Springer.

Hough, G., and D. Swanson. 1998. “Towards an assessment of continuous measurement: A comparison of returns with 1990 census returns for the Portland test site,” *Journal of Economic and Social Measurement* 24: 295-308.

Hough, G. and D. Swanson. 2006 “An Evaluation of the American Community Survey: Results from the Oregon Test Site.” *Population Research and Policy Review* 25(3): 257-273.

Hsueh, L., H. Tseng, and C. Hsieh. 2007.. “Relationship Between the Housing Vacancy Rate, Housing Price, and the Moving Rate at the Township Level in Taiwan, in 1990 and 2000.” *International Real Estate Review* 10 (1): 119-150.

Jiang, J. and P. Lahiri. 2006. “Mixed Model Prediction and Small Area Estimation.” *Test* 15 (1): 1-96.

Judson, D. 2000. “The Statistical Administrative Records System and Administrative Records Experiment 2000.” Paper presented at the National Institutes of Statistical Sciences Data Quality Workshop, Morristown, NJ, November 30th – December 1st .

Judson, D. and C. Popoff. 2004. “Selected General Methods.” pp. 677-732 in J. Siegel and D. Swanson (Eds.) *The Methods and Materials of Demography, 2nd Edition*. New York, NY: Elsevier Academic Press.

Khor, A., Y Ming, and L. Yuan. 2000. “The Natural Vacancy Rate of the Singapore Office Market.” *Journal of Property Research* 17 (4): 329-338.

Kimpel, T. and T. Lowe. 2007. “Estimating Household Size for Use in Population Estimates*.” Population Estimates and Projections, Research Brief no. 47.* Olympia, WA: Washington State Office of Financial Management.

Kintner, H. and D. Swanson. 1994. “Estimating Vital Rates from Corporate Databases: How Long will GM’s Salaried Retirees Live?” pp. 265 – 295 in H. Kintner, T. Merrick, P. Morrison, and P. Voss (Eds.) *Demographics: A Casebook for Business and Government*. Boulder, CO: Westview Press.

Kish, L. 1998. “Space/Time Variations and Rolling Samples.” *Journal of Official Statistics* 14 (1): 31-46.

Kordos, J. (Ed.). 2000. “Special issue on Small Area Estimation.” *Statistics in Transition: Journal of the Polish Statistical Association* 4 (4).

Korbin, F. 1976. “The Fall in Household Size and the Rise of the Primary Individual in the United States.” *Demography* 13 (1): 127-138.

Long, J. 1993a. “An Overview of Population Estimation Methods.” pp 20-1 to 20-5 in D. Bogue, E. Arriaga and D. Anderton (Eds.) *Readings in Population Research Methodology* (Vol. 5), Chicago: United Nations Population Funds and Social Development Center.

Long, J. 1993b. *Postcensal Population Estimates: States, Counties and Places*. Technical Working Paper No. 3. Washington DC: U.S. Bureau of the Census.

Lowe, T. 2009. Personal Communication.

Lowe, T. 2000a “Developing Trends in Household Size for use in Population Estimates.” *Population Estimates and Projections Research Brief No. 10*. Olympia, WA: Washington State Office of Financial Management.

Lowe, T. 2000b “Occupied versus Rented: Census and Real Estate Occupancy Rates and Their Use in Population Estimates.” *Population Estimates and Projections Research Brief No.* 9 (May). Olympia, WA: Washington Office of Financial Management.

Lowe, T., and M. Mohrman. 2003. “Use of Postal Delivery Data in the Population Estimate Process.” *Population Estimates and Projections Research Brief No.* 18 (January). Olympia, WA: Washington Office of Financial Management.

Lowe, T., M. Mohrman and D. Brunink. 2003. “Developing Postal Delivery Data for Use in Population Estimates.” *Population Estimates and Projections Research Brief No.* 17 (January). Olympia, WA: Washington Office of Financial Management.

Lowe, T., L. Weisser, and B. Myers. 1984. “A Special Consideration in Improving Housing Unit Estimates: The Interaction Effect. Paper presented at the annual meeting of the Population Association of America, Minneapolis, MN.

Lowe, T., D. Pittenger, and J. Walker. 1977. “Making the Housing Unit Method Work: A Progress Report. Paper presented at the Annual Meeting of the Population Association of America, St. Louis, MO.

Marquis, K., S. Wetrogan, and H. Palacios. 1996. “Towards a U.S. Population Database from Administrative Records.” Unpublished Paper, U.S. Census Bureau (Available online, http://www.census.gov/srd/papers/pdf/km9601.pdf, last accessed May, 2008).

McDonald, J. and D. McMillen. 2000. “Residential Building Permits in Urban Counties: 1990–1997.” *Journal of Housing Economics* 19 (3): 175-186.

McKibben, J. 2006. “School District Planning and the 2010 Census: Data Uses and Needs”. *Journal of Economic and Social Measurement*, 31, (3) 221-232.

Murdock, S. and D. Ellis. 1991. *Applied Demography: An Introduction to Basic Concepts, Methods, and Data*. Boulder, CO: Westview Press.

Myers, D. and A. Doyle. 1990. “Age-specific Population-per-Household Ratios: Linking Population Age Structure with Housing Characteristics.” pp. 109-130 in D. Myers (Ed.) *Housing Demography: Linking Demographic Structure and Housing Markets*. Madison, WI: University of Wisconsin Press.

National Research Council. 2007. *Using the American Community Survey: Benefits and Challenges*. Washington, DC: National Academies Press.

National Research Council. 2006a. *Once, Only Once, and in the Right Place: Residence Rules in the Decennial Census*. Washington, DC: National Academies Press.

National Research Council. 2006b. *Expanding Access to Research Data: Reconciling Risks and Opportunities.* Washington, D.C: National Academies Press.

National Research Council. 2003. *Statistical issues in Allocating Funds by Formula*. Washington, DC: National Academies Press.

National Research Council. 1980. *Estimating Population and Income of Small Areas*. Washington, D.C.: National Academies Press.

Perrone, S. 2008. “Address Coverage and Improvement and Evaluation Program – 2005 National Estimate of Coverage of the Master Address File.” pp. 37-85 in S. Murdock and D. Swanson (Eds.) *Applied Demography in the 21st Century*. Dordrecht, The Netherlands, Springer.

Pittenger, D. 2004. “Use of Census Data and City Housing Reports in Creating Housing-Unit Demolition Rates.” *Population Estimates and Projections Research Brief No. 28*. Olympia, WA: Washington State Office of Financial Management.

Pittenger, D. 1982. “Critique of Administrative Record Procedures.” pp. 39-42 in E. S. Lee and H. F. Goldsmith (Eds.), *Population Estimates: Methods for Small Area Analysis*. Beverly Hill, CA: Sage Press.

Platek, R., J. Rao, C. Sarndal, and M. Singh (Eds.) (1987). *Small Area Statistics: An International Symposium*. New York, NY: John Wiley.

Pol, L. and R. Thomas. (2001). *The Demography of Health and Health Care, 2nd Edition*. New York, NY: Kluwer Academic/Plenum Press.

Popoff, C. and D. Judson. (2004). “Some Methods of Estimation for Statistically Underdeveloped Areas.” pp. 603-641 in J. Siegel and D. Swanson (Eds.) *The Methods and Materials of Demography, 2nd Edition*. New York, NY: Elsevier Academic Press.

Prevost, R. (1996). “Administrative Records and the New Statistical Era.” Paper presented at the 1996 Annual meeting of the Population Association of America. New Orleans, LA, May 9th - 11th.

Prevost, R., and C. Leggieri. 1999. “Expansion of Administrative Records Uses at the Census Bureau: A Long-Range Research Plan.” Paper presented at the conference of the Federal Committee on Statistical Methodology (available online, http://www.fcsm.gov/99papers/prevost.pdf, last accessed, May 2008).

Prevost, R. and D. Swanson. (1985). “A New Technique for Assessing Error in Ratio-Correlation Estimates of Population: A Preliminary Note.” *Applied Demography* 1 (November): 1-4.

Purcell, N. J. and L. Kish. 1980. “Postcensal Estimates for Local Areas (or Domains).” *International Statistical Review* 48: 3-18.

Purcell, N. J. and L. Kish. 1979. “Estimation for Small Domains.” *Biometrics* 35: 365-384.

Rao, J. (2002). *Small Area Estimation*. San Francisco, CA: Jossey-Bass.

Reese, A. J. 2006. “A Comparison of Housing Unit Estimates to the American Community Survey’s Aggregated Master Address File.” Paper Presented at the Annual meeting of the Southern Demographic Association, Durham, NC.

Rives, N.W. 1982. “Assessment of a Survey Approach.” pp. 79-96 in E. Lee and H. Goldsmith (Eds.) *Population Estimates: Methods for Small Area Analysis*. Beverly Hills, CA: Sage Publications.

Robinson, J.G. and D. Dixon. 2009. “An Evaluation of the Consistency of Subcounty Estimates of Population Based on the American Community Survey and the Population Estimates Program.” Paper (to be) presented at the annual meeting of the Population Association of America, Detroit, MI: May 2nd.

Roe, L, J. Carlson, and D. Swanson. 1992. “A Variation of the Housing Unit Method for Estimating the Population of Small, Rural Areas: A Case Study of the Local Expert Method.” *Survey Methodology* 18(1): 155-163.

Serow, W. and N. Rives. 1995. “Small Area Analysis: Assessing the State of the Art.” pp. 1-9 in N. Rives, W. Serow, A. Lee, H. Goldsmith, and P. Voss (eds.). *Basic Methods for Preparing Small Area Population Estimates.* Madison, WI: Applied Population Laboratory, Department of Rural Sociology, University of Wisconsin.

Shryock, H. 1938. “Methods of Estimating Post-censal Population.” *American Journal of Public Health* 28: 1042-1047.

Shryock, H. and N. Lawrence. 1949. “The Current Status of State and Local Population Estimates in the Census Bureau.” *Journal of the American Statistical Association* 44 (246): 157-173.

Scire, M. 2007. “2010 Census: Population Measures are Important for Federal Funding Allocations.” *Testimony Before the Subcommittee on Information Policy, Census, and National Archives, Committee on Oversight and Government Reform, House of Representatives. Publication GAO-08-230T.* Washington, DC: General Accountability Office

Siegel, J. 2002. *Applied Demography: Applications to Business, Government, Law, and Public Policy.* San Diego, CA: Academic Press.

Smith, A. S. 1998. “The American Community Survey and Intercensal Population Estimates: Where are the Crossroads?” *Population Division Technical Working Paper No. 31.* U. S. Census Bureau.

Smith, S. K. 1994. “Estimating Temporary Populations: The Contributions of Robert C. Schmitt.” *Applied Demography* 9 (1): 4-7.

Smith, S. 1986. “A Review and Evaluation of the Housing Unit Method of Population Estimation.” *Journal of the American Statistical Association* 81: 287-296.

Smith, S., and S. Cody. 2004. “An Evaluation of Population Estimates in Florida: April 1, 2000.” *Population Research and Policy Review* 23 (1): 1-24.

Smith, S. K., and M. House. 2007. “Temporary Migration: a Case Study of Florida.” *Population Research and Policy Review* 26 (4): 437-454.

Smith, S., and B. Lewis, 1980. “ Some New Techniques for Applying the Housing Unit Method of Local Population Estimation.” *Demography* 17: 323-339.

Smith, S., and M. Mandell. 1984. “A Comparison of Population Estimation Methods: Housing Unit Versus Component II, Ratio Correlation, and Administrative Records.” *Journal of the American Statistical Association* 79: 282-289.

Smith, S. K., J. Nogle, and S. Cody. 2002. "A Regression Approach to Estimating the Average Number of Persons per Household." *Demography* 39 (4): 697-712.

Smith, S., J. Tayman, and D. Swanson. (2001). *State and Local Population Projections: Methodology and Analysis*. New York, NY: Kluwer Academic/Plenum Publishers.

Statistics Finland. 2004. *Use of Register and Administrative Data Sources for Statistical Purposes: Best Practices of Statistics Finland*. Handbook Series, No. 45. Helsinki, Finland: Statistics Finland.

Stigler, S. 1986. *The History of Statistics: The Measurement of Uncertainty before 1990.* Cambridge, MA: Belknap Press of Harvard University Press.

Swanson, D. 2006. Statement before the Subcommittee on Federalism and the Census Oversight Hearing Committee on Government Reform, U.S. House of Representatives. Hearing Topic: "Two Plus Two Should Never Equal Three: Getting Intercensal Population Estimates Right the First Time." Wednesday, 6 September 2006 Room 2247 Rayburn House Office Building.

Swanson, D. 2004. "Advancing Methodological Knowledge within State and Local Demography: A Case Study." *Population Research and Policy Review* 23 (4): 379-398.

Swanson, D., and D. Beck. 1994. "A New Short-term County Population Projection Method." *Journal of Economic and Social Measurement* 20: 25-50.

Swanson, D. 1989. "Confidence Intervals for Post-censal Population Estimates: A Case Study for Local Areas." *Survey Methodology* 15: 271-280.

Swanson, D. 1981. "Allocation Accuracy in Population Estimates: An Overlooked Criterion With Fiscal Implications." pp. 13-21 in *Small Area Population Estimates and Their Accuracy, Series GE-41, No. 7.* U. S. Bureau of the Census. Washington, DC: US Government Printing Office.

Swanson, D. 1980. "Improving Accuracy in Multiple Regression Estimates of Population Using Principles from Causal Modeling." *Demography* 17: 413-428.

Swanson, D. 1977."An Evaluation of the 1977 Round of Municipal Vacancy Estimates for 1 & 2 Unit Structures." *Staff Document No. 42*. Office of Financial Management, Population, Enrollment and Economic Studies Division. Olympia, WA: Washington State Office of Financial Management.

Swanson, D. and G. Hough. 2007. "An Evaluation of Persons Per Household (PPH) Data Generated by the American Community Survey: A Demographic Perspective." Paper presented at the Annual Meeting of the Southern Demographic Association. Birmingham, AL. (Available online at http://www.census.gov/acs/www/AdvMeth/Multi_Year_Estimates/presentations.html, last accessed, October, 2008).

Swanson, D. and T. Lowe. 1979. “Comparative Analysis of Change in Average household Size with reference to IRS Data on Average Exemptions Per Return: Census Results from Selected Municipalities in Washington, 1970, 1977, and 1978.” *Special Report No. 22.* Office of Financial Management: Olympia, WA. Washington State Office of Financial Management.

Swanson, D. and J. McKibben. 2009. “Developing Annual Population Data in the United States: New Possibilities for the 21st Century.” (To be) Presented at the 2009 Conference of the International Union for the Scientific Study of Population, September 26-October 3, Marrakech, Morocco (Session 2405).

Swanson, D. and L. Pol. 2005. “Contemporary Developments in Applied Demography within the United States. *Journal of Applied Sociology* 21 (2): 26-56.

Swanson, D. and L. Pol. 2008. “Applied Demography: Its Business and Public Sector Components.” in Yi Zeng (ed.) The Encyclopedia of Life Support Systems, Demography Volume. UNESCO-EOLSS Publishers. Oxford, England. (with L. Pol).(Online at http://www.eolss.net/).

Swanson, D. and G. E. Stephan. 2004.”Glossary.” pp. 751-778 in J. Siegel and D. Swanson (Eds.) *The Methods and Materials of Demography, 2nd Edition*. New York, NY: Elsevier Academic Press.

Swanson, D., and J. Tayman. 1995 “Between a Rock and a Hard Place: The Evaluation of Demographic Forecasts.” *Population Research & Policy Review* 14: 233-249.

Swanson, D., and L. Tedrow. 1984. “Improving the Measurement of Temporal Change in Regression Models used for County Population Estimates.” *Demography* 21: 361-372.

Swanson, D., B. Baker, and J. Van Patten. 1983. “Municipal Population Estimation: Practical and Conceptual Features of the Housing Unit Method.” Paper Presented at the 1983 Annual Meeting of the Population Association of America, Pittsburgh, PA, April 14th - 16th. (Reprinted as pp. 99-105 in *Alaska Population Overview 1982*. Alaska Department of Labor, Juneau, AK.

Swanson, D., J. Tayman, and D. Beck. 1995. “On the Utility of Lagged Ratio-Correlation as a Short-term County Population Estimation Method: A Case Study of Washington State.” *Journal of Economic and Social Measurement* 21:1-16.

United Nations.1969. *Methodology and Evaluation of Population Registers and Similar Systems. Studies in Methods, No. 15*. Sales No. E.69.XVII.15. New York City, NY: United Nations.

U. S. Census Bureau. (No Date). "The U.S. Census Bureau's Intercensal Population Estimates and Projections Program: Basic Underlying Principles." Unpublished document provided to participants of this conference (reproduced as Appendix A in this paper).

U.S. Census Bureau. 2007a. "Housing Unit Based Estimates Research Project Draft Research Agenda." Washington, DC: U.S. Census Bureau, Population Division.

U. S. Census Bureau. 2007b. "County Housing Unit Estimates, April 1, 2000 to July 1, 2006." Washington, DC: Population Division, U. S. Census Bureau. (available online, http://www.census.gov/popest/housing/files/HU-EST2006_US.CSV, last accessed January, 2009).

U. S. Census Bureau. 2006. Methodology for County Housing Unit Estimates for Vintage 2006." Washington, DC: Population Division, U.S. Census Bureau(available online, http://www.census.gov/popest/topics/methodology/2006_hu_meth.html, last accessed January 2009).

U. S. Census Bureau. 2004a. "The Census Bureau's Master Address File (MAF): Census 2000 Address List Basics." Washington, DC. Geography Division, U. S. Census Bureau (available online http://www.census.gov/geo/mod/maf_basics.pdf , last accessed September, 2008).

U.S. Census Bureau. 2004b. Meeting 21st Century Demographic Needs: Implementing the American Community Survey." *Report 8: Comparisons of the American Community Survey Three-Year Averages with the Census Sample for a Sample of Counties and Tracts*. Washington, DC: U.S. Census Bureau.

U. S. Census Bureau. 2003. "American Community Survey Operations Plan. Release 1: March 2003." Washington, DC: U. S. Census Bureau.

U. S. Census Bureau. 2001a. "Meeting 21st Century Demographic Data Needs – Implementing the American Community Survey: July 2001." *Report 1: Demonstrating Operational Feasibility.* Washington, DC: U. S. Census Bureau.

U. S. Census Bureau. 2001b. The American Community Survey, Updated Information for America's Communities, November, 2001 (ACS/01-BLKT).

U.S. GAO. 2006. *2010 Census: Census Bureau Needs to take Prompt Actions to Resolve Long-Standing and Emerging Address and Mapping Challenges.* GAO-06-272. Washington, DC: U.s. Government Accountability Office (Note: Effective July 7, 2004, the GAO's legal name changed from the Government Accounting Office to the Government Accountability Office).

U.S. GAO. 2001.2000 *Census: Significant Increase in Cost Per Housing Unit Compared to 1990 Census.* GAO-02-31. Washington, DC: U.s. Government Accountability Office (Note: Effective July 7, 2004, the GAO's legal name changed from the Government Accounting Office to the Government Accountability Office).

U. S. Department of Commerce. 2002. "A Better Strategy is needed for Managing the Nation's Master Address File." *Inspection Report No. OSE-12065*. Washington, D. C.: U. S. Department of Commerce.

Van Auken, P., R. Hammer, P. Voss, and D. Veroff. 2006. "The American Community Survey in Counties with 'Seasonal' Populations." *Population Research and Policy Review* 25 (3): 275-292.

Voss, P., and H. Krebs. 1979. "The use of Federal Revenue Data for Improving Current Estimates of Average Household Size for Minor Civil Divisions: An Evaluation." *Technical Series 70-5.* Applied Population Laboratory. Madison, WI: University of Wisconsin.

Voss, P., C. Palit, B. Kale, and H. Krebs. 1995. "Censal Ratio Methods." pp. 70-88 in N. W. Rives, W. Serow, A. Lee, H. Goldsmith, and P. Voss (Eds). *Basic Methods for Preparing Small-Area Population Estimate*s. Madison, WI: Applied Population Laboratory, University of Wisconsin.

Walashek, P. and D. Swanson. 2006. The Roots of Conflicts over US Census Counts in the late 20th Century and Prospects for the 21st Century." *Journal of Economic and Social Measurement* 31 (3, 4): 185-206.

Waldrop, J. 1995. "Preface." pp. v-vi in N. Rives, W. Serow, A. Lee, H. Goldsmith, and P. Voss (eds.) *Basic Methods for Preparing Small-Area Population Estimates*. Madison, WI: Applied Population Laboratory, Department of Rural Sociology, University of Wisconsin.

Wang, C. 1999. "Development of National Accounting of Address and Housing Inventory: The Baseline Information for Post-censal Population estimates." Paper prepared for the Estimates Methods Conference, U.S. Bureau of the Census, Federal Office Building #3, Suitland, Maryland, June 8th. (available online at http://www.census.gov/population/www/coop/popconf/paper.html, last accessed 28 June 2006).

Washington Office of Financial Management. 1978. *The Housing Unit Method: A Manual for Municipal Personnel Responsible for Making Annual Population Estimates.* Olympia, WA: Washington Office of Financial Management.

Washington Office of Financial M anagement. 2007. "Developing Trends in Household Size for use in Population Estimates." *Population Estimates and Projections, Research Brief no. 10.* Olympia, WA: Washington State Office of Financial Management.

Weidman, L., R. Creecy, D. Malec, and J. Tsay. 2008. "Exploration of the Use of Empirical Bayes Procedures for Estimating Changes in Occupancy Rate and Persons Per Household." *Research Report Series (Statistics #2008-07*). Suitland, MD: U.S. Census Bureau.

Wetrogan, S. 2007. “Evaluating County Population Estimates Using a Housing Unit Based Approach.” Paper presented at the Annual Meeting of the Southern Demographic Association. Birmingham, AL.

Williams, B., J. Nichols, and M. Conroy. 2002. *Analysis and Management of Wildlife Populations.* San Diego, CA. Academic Press.

Wilmoth, J. (2004). “Population Size.” pp. 65-80 in J. Siegel and D. Swanson (Eds.) *The Methods and Materials of Demography, 2nd Edition*. New York, NY: Elsevier Academic Press.

Zeng, Y., K. Land, Z. Wang, and D. Gu. 2006. “U. S. family household momentum and dynamics: an extension and application of the ProFamy Method.” *Population Research and Policy Review* 25: 1–41

Appendix. Principles underlying the US Census Bureau's estimates and projections programs.

I. Background

The U.S. Census Bureau's Population Estimates and Projections program is designed to fulfill the mandates of Title 13, Section 181, of the U.S. Code.

> During the intervals between each census of population required under section 141 of this title, the Secretary, to the extent feasible, shall annually produce and publish for each State, county, and local unit of general purpose government which has a population of fifty thousand or more, current data on total population and population characteristics and, to the extent feasible, shall biennially produce and publish for other local units of general purpose government current data on total population. Such data shall be produced and published for each State, county, and other local unit of general purpose government for which data is compiled in the most recent census of population taken under section 141 of this title. Such data may be produced by means of sampling or other methods, which the Secretary determines will produce current, comprehensive, and reliable data.

A. To satisfy this mandate, the program of population estimates has grown over the years to produce the following products annually:

1. Monthly estimates of the national population of the United States by age, sex, race, and Hispanic origin
2. Annual estimates of the population of states by age, sex, race, and Hispanic origin
3. Annual estimates of the population of counties by age, sex, race, and Hispanic origin
4. Annual estimates of the total population of functioning governmental units
5. Annual estimates of the number of housing units for states and counties.

B. In addition to meeting the mandates of Title 13, these estimate products are used for a variety of purposes, including the following:

1. Controls for federally sponsored surveys, including the Current Population Survey (CPS) and the American Community Survey (ACS)
2. Allocation of federal dollars totaling over $200 billion annually
3. Denominators for various indicators, including vital statistics, per capita income, and cancer incidence rates
4. Calculation of the number of clerks the Senate hires
5. Requirements of the Federal Election Commission

6. Denominators for poverty rate estimation at selected levels of geography
7. Program planning by federal, state, local, and private entities

II. Implicit Assumptions

Implementation of the annual program of intercensal estimates is guided by several implicit assumptions.

A. Timely release of the annual products is critical

   1. The maximum lag time between estimate date and dissemination of last data product is 12 months.
   2. Annual national and state population totals must be released within 6 months of estimate date to meet requirements of IRS Bonding Authority.
   3. State estimates of the population aged 18 and older must be available within 6 months of estimate date to satisfy requirements of the Federal Election Commission.
   4. National and state population controls to be used for the new calendar year CPS must be available by late January of the new calendar year.
   5. Estimates of state and county characteristics must be available within 9 months to meet requirements for use as population controls for the American Community Survey.
   6. Estimates of functioning governmental units should be available within 12 months of estimate date for use by HUD in funds allocation.

B. Each annual production consists of a time series of estimates from the last decennial census date to the estimate date and is produced using the latest available data and the current approved methodology.

   1. Current-year data products contain revisions to the prior year's estimates that are caused by incorporating:
      a. Improved methodology.
      b. New data inputs.
      c. Revisions to prior year data inputs.

   2. The term "vintage" is used to refer to the reference date of an estimates cycle. Estimates released with a reference date of July 2005 are referred to as the "vintage 2005" set of population estimates and will include a consistent time series back to April 2000.

C. Within any vintage, all products use the same vintage of input data and must sum to the earlier released products of the same vintage for the same measurement.

   1. Since the national and state population totals are the first to be released, all subsequent estimate products must sum to the national and state totals

that already appear for that vintage. This insures consistency within any vintage and means that the sum of the "parts" will always equal the previously released U.S., state, or county total.

2. Since the national population estimates tabulated by characteristics are the first characteristics to be released, the sum of the state and county characteristics must equal the national characteristics of the same vintage.

D. Only one consistent set of products and related materials is developed within a vintage. That set of products is intended to serve all customers' needs and uses.

1. The methodology and data inputs used to develop the population estimates used as denominators for vital statistics rates are consistent with those used to develop the population controls for the CPS and ACS.

2. Custom data products are consistent with the publicly released data products. For example, the annual race estimates for counties use a bridged race algorithm developed by NCHS. However, while the race data conform to the bridging algorithms developed by NCHS, the estimates of total populations and populations by age and sex generally agree with the publicly released data products.

E. The population estimates begin with the most recent decennial-census enumerated count updated to July 1 of each year, and as such, are based on the usual-residence concept used in the most recent decennial census.

1. The population estimates base for each estimate date is updated to include Count Question Resolution (CQR) changes to the decennial census base as well as geographic updates due to annexation and other geographic program changes.

2. The components of population change used to update the most recent census will be consistent with the best set of components available. Ongoing evaluation indicates that the coverage and the consistency of vital statistics and other administrative records data differ from those of decennial census data. Therefore, in the annual estimates, the size of the population based mainly on administrative records data differ from the size based mainly on census data.

F. States, counties, and units of local government have the right to challenge the population estimates prepared by the Census Bureau under the provisions of Title 15, The Code of Federal Regulations, Part 90. The results of accepted challenges will be incorporated into the following year's population estimates as long as the challenge is received by October 1 of the year in which the estimate was released.

III. Current Broad Methodological Assumptions

A. Prior to incorporating a new methodology or data set, it is desirable to thoroughly evaluate a set of estimates that use this new methodology or data set and compare it with the most recent decennial census results. When this is not possible, the methods are judged by the following criteria.

1. Soundness: The method should be based on solid reasoning – i.e., the formulas that embody the method should be mathematically valid and respect the attributes of the input data as they relate to the estimation task.

2. Integrity: A strategy that consistently applies the declared method is preferred to one that uses ad-hoc fixes to address particular challenges of the estimation task.

3. Parsimony: A simpler strategy is preferred to a more complex one.

4. Robustness: The method that produces the most reasonable estimates (defined below) across the full range of potential input-data values and in the presence of the random variation normally associated with those values while maintaining the orthodoxy and consistency of the estimates (also defined below) is preferred.

5. Adaptability: A technique that can be applied more broadly (e.g., across geographic summary levels), thus promoting the integration of the Census Bureau's estimates system, is preferred to a more product-specific remedy.

6. Transparency: A strategy that is more readily understandable and replicable by external parties is preferred. Moreover, a strategy that provides some explanatory information (i.e., how did the size or distribution of the population come to be this way) is preferred over one that is merely predictive.

7. Usability: The method must be executable along with all other current projects under current staffing levels in a way that allows the Census Bureau to meet current deadlines.

8. Flexibility: The preferred method will allow the production of estimates when a specific instance of the input data normally required by the method is unavailable or deemed unsuitable.

B. As a final test, the method should produce output data that have the following qualities.

   1. Orthodoxy: The values of the population estimates should be appropriate (e.g., no negative population numbers, all population estimates in whole numbers).

   2. Consistency: The values of the population estimates for all universes (e.g., resident, civilian, civilian non-institutionalized), geographies (e.g., national, state, county), and characteristics (e.g., age, sex, race, Hispanic origin) should not contradict one another.

   3. Reasonableness: The values of the population estimates should approximate the real values as determined by the following assessments.
      a. Post-Censal Change: The reasonableness of the total change in the population since the last decennial census.
      b. Time-Series Change: The reasonableness of the annual change in the estimates since the last census.
      c. Demographic Appropriateness: The values of the estimates and the demographic rates they imply fall within acceptable limits when evaluated by general demographic principles (e.g., the appropriateness of the sex ratios, age progression, implied family size, life expectancies, total fertility rates, etc.).
      d. Comparability: The estimates appear realistic when compared with other indicators of the size and distribution of the population (e.g., Medicare enrollment, school enrollment, housing unit estimates, etc.).

C. A consistent method is used for entities at the same level of geographic aggregation.

   1. The method adopted for state totals must be used for all states.

   2. The method adopted for counties within a state must be used for all counties within that state.

D. The Census Bureau develops the basic estimates for the nation, states, and counties by disaggregated race groups in order to meet the various custom race aggregations needed by users.

E. The cohort-component method is the preferred method for development of the national, state, and county-level total population estimates and population estimates by characteristics.

F. The distributive housing-unit method is the preferred method for the development of the functioning subcounty governmental-unit-level estimates.

G. State total population estimates are not developed independently. National population estimates are first developed; then county total population estimates are developed and controlled to the national total population estimates. The state total population estimates are the sum of the "nationally controlled" county total population estimates for the state.

H. Data on vital statistics and group quarters provided by members of the Federal State Cooperative Program for Population Estimates (FSCPE) are included in the process of developing state and county population estimates.
I. Although state members of the FSCPE are provided the opportunity to review the state and county population totals prior to final production, they must follow strict criteria and provide objective evidence when requesting modifications.

IV. Current Specified Methodologies

A. National level estimates will use the cohort-component technique applied to data from the latest decennial census as the base, data on births and deaths provided by the National Center for Health Statistics, and estimates of net international migration derived from data from the American Community Survey (ACS) See the url

<http://www.census.gov/popest/topics/methodology/2003_nat_char_meth.html http://www.census.gov/popest/topics/methodology/v2005_nat_char_meth.html>

For a detailed discussion of the methodology used to develop the most recent set of national population estimates by demographic characteristics.

B. State and county population estimates are developed using a demographic procedure called an "administrative records component of population change" method. A major assumption underlying this approach is that the components of population change are closely tracked by administrative data in a demographic change model. In order to apply the model, Census Bureau demographers estimate each component of population change separately. For the population residing in households, the components of population change are births, deaths, and net migration, including net international migration. For the non-household population, change is represented by the net change in the population living in group-quarters facilities.

Each component in our model represents data that are symptomatic of an aspect of population change. For example, birth certificates indicate additions to the population resulting from births, so these data are used to estimate the birth component for a county. Other components are derived from death certificates, Internal Revenue Service data (IRS), Medicare enrollment records, Armed Forces data, group-quarters population data, and data from the American Community Survey.

For a more detailed discussion of the development of county population totals see
<http://www.census.gov/popest/topics/methodology/2003_st_co_meth.html
http://www.census.gov/popest/topics/methodology/2005_st_co_meth.html>

C. State population characteristics are currently developed in a two-stage process. Estimates by age and sex are developed first using a cohort-component procedure whereby estimates of net migration are developed using school enrollment data. These estimates are controlled both to the national-level estimates by age and sex as well as the previously developed state population totals.

The second step in the process distributes the state age and sex estimates into race by Hispanic origin categories. This is done by preparing an initial set of state estimates by age, sex, race, and Hispanic origin that are controlled to the state age and sex estimates prepared in the first step and to the previously developed national estimates by age, sex, race, and Hispanic origin.

For a more detailed discussion of the development of the state population characteristics by age, sex, race, and Hispanic origin see

<http://www.census.gov/popest/topics/methodology/2003_st_char_meth.html
http://www.census.gov/popest/topics/methodology/2004_st_char_meth.html>

D. County population characteristics are developed using a proportional distribution method beginning with previously developed resident county population estimates by age (0-64 and 65+) and resident state population estimates by age, sex, race, and Hispanic origin. Then county-level estimates of age, sex, race, and Hispanic origin distributions are developed using information about post-censal change in the corresponding populations. Third, these distributions are applied to the original county estimates by age and state characteristics.

A detailed discussion of this method is provided at

<http://www.census.gov/popest/topics/methodology/2004_co_char_meth.htm>

V. Enhancement Priorities

A. Improve estimates of net international migration

1. Provide up-to-date, useful statistics and methodologies on the size, characteristics, and demographic impact of international migration to and from the United States for use in policy-making decisions and demographic and economic research.

2. Goals of immigration research
   a. Produce annual estimates of international migration
   b. Improve current migration-related survey questions on the ACS.
   c. Conduct extensive evaluations to determine the best method to incorporate ACS data into the population estimates.

3. Activities

   a. Evaluate reasonableness of estimates of annual change in the foreign-born data from ACS at the national level.

   b. Produce revised estimates of net international migration at the national level.

   c. Produce new demographic and geographic distributions for migrants.

   d. Construct algorithms to estimate the migrant status of the foreign-born populations.
   e. Produce estimates of international migrants by migrant status (legal migrants, temporary migrants, quasi-legal migrants, unauthorized migrants, and emigrants).

E. Improve Estimates of Internal Migration

   1. Improve the accuracy of the annual migration estimates by age, sex, race, and Hispanic origin for counties by maximizing the efficient use of available administrative data files, Census 2000 data, and the American Community Survey (ACS) data.

   2. The ultimate goal is to implement a person-based migration model incorporating administrative data from files such as the IRS 1040 and 1099 records, Medicare records, a derived person-characteristic file developed from the Social Security Administrative NUMIDENT file, and other administrative data that can be merged into the database. The database will enable analysts to match administrative data with Census 2000 (100% and sample data), CPS, and ACS data in order to develop models that correct possible demographic and geographic biases inherent in the use of an administrative records database when estimating migration rates for counties.

F. Develop a new methodology for estimating subnational population characteristics

   1. Replace the methodology that develops state estimates by age and sex based on school enrollment data with a method that is consistent with the best set of administrative data available and exploits the power of current computing capacity.

   2. Develop a method that addresses current deficiencies in the age distributions of the population in selected states and counties, especially the age distribution of the population aged 18 to 24.

   3. Develop a new method to estimate county population by age, sex, race, and Hispanic origin.

G. Develop procedures to systematically incorporate participation by State FSCPE Agencies in the production of state and county population estimates

   1. Address issues of consistency

2. Establish criteria for incorporating state participation

VI. Other Enhancements

A. Improve the distributive housing unit approach at the subcounty level.
   1. Develop procedures to update Census 2000 measures of vacancy and numbers of people per household (PPH or the Person Per Household measure) used in the estimates process.

   2. Improve estimates of housing units.

   3. Address inconsistencies between estimates developed using the distributive housing unit approach and those developed using the component approach.
      a. Develop improved procedures to estimate housing unit loss.
      b. Integrate enhancements from the Master Address File.

B. Address inconsistencies between data from the decennial census base and data on components of change from administrative records databases.

   1. Address inconsistencies between Census 2000 data and NCHS data on race and Hispanic-origin characteristics.

   2. Address unreasonable results from pairing NCHS mortality data with decennial census data and estimate results.

VII. Administrative Constraints

A. The methods developed must be capable of being implemented with current resources and within the current time frame for estimate production.

B. Production of the complete set of estimates must continue during any development stages.

C. Methods must be Transparent and Reproducible

Table 1. Accuracy Test of the Geometric Method of Estimating PPH Values for Counties: Washington State 2000

Washington State PPH Values By County, 1980, 1990, and 2000

	1980 Persons Per Household	1990 Persons Per Household	2000 Persons per Household	1980-1990 Geometric Rate of Change	Estimated 2000 Persons Per Household	Absolute Error	Percent Error	MAPE
STATE	2.6086	2.5348	2.5349	-0.0029	2.4631	-0.0718	-2.83%	2.83%
Adams	2.9113	2.9405	3.0949	0.0010	2.9700	-0.1249	-4.03%	4.03%
Asotin	2.5662	2.4727	2.4162	-0.0037	2.3826	-0.0336	-1.39%	1.39%
Benton	2.7971	2.6516	2.6795	-0.0053	2.5137	-0.1658	-6.19%	6.19%
Chelan	2.4827	2.4863	2.6192	0.0001	2.4899	-0.1293	-4.93%	4.93%
Clallam	2.5374	2.4007	2.3066	-0.0055	2.2714	-0.0353	-1.53%	1.53%
Clark	2.7625	2.6625	2.6900	-0.0037	2.5661	-0.1239	-4.61%	4.61%
Columbia	2.5254	2.4368	2.3628	-0.0036	2.3513	-0.0115	-0.49%	0.49%
Cowlitz	2.6619	2.5588	2.5531	-0.0039	2.4597	-0.0934	-3.66%	3.66%
Douglas	2.7591	2.6769	2.7554	-0.0030	2.5971	-0.1583	-5.74%	5.74%
Ferry	2.8567	2.6978	2.4938	-0.0057	2.5477	0.0539	2.16%	2.16%
Franklin	2.8817	3.034	3.2637	0.0052	3.1943	-0.0693	-2.12%	2.12%
Garfield	2.5955	2.3948	2.3911	-0.0080	2.2096	-0.1815	-7.59%	7.59%
Grant	2.7986	2.7407	2.9204	-0.0021	2.6840	-0.2364	-8.09%	8.09%
Grays Harbor	2.5966	2.4813	2.4826	-0.0045	2.3711	-0.1115	-4.49%	4.49%
Island	2.6706	2.6149	2.5223	-0.0021	2.5604	0.0381	1.51%	1.51%
Jefferson	2.4537	2.3089	2.2122	-0.0061	2.1726	-0.0395	-1.79%	1.79%
King	2.4868	2.3982	2.3905	-0.0036	2.3128	-0.0777	-3.25%	3.25%
Kitsap	2.682	2.6469	2.6007	-0.0013	2.6123	0.0115	0.44%	0.44%
Kittitas	2.3976	2.3251	2.3314	-0.0031	2.2548	-0.0766	-3.29%	3.29%
Klickitat	2.7211	2.6409	2.5361	-0.0030	2.5631	0.0270	1.06%	1.06%
Lewis	2.6732	2.5997	2.5690	-0.0028	2.5282	-0.0408	-1.59%	1.59%
Lincoln	2.5726	2.4308	2.4233	-0.0057	2.2968	-0.1265	-5.22%	5.22%
Mason	2.5458	2.5162	2.4891	-0.0012	2.4869	-0.0022	-0.09%	0.09%
Okanogan	2.6674	2.5877	2.5762	-0.0030	2.5104	-0.0658	-2.56%	2.56%
Pacific	2.4465	2.3499	2.2711	-0.0040	2.2571	-0.0140	-0.62%	0.62%
Pend Oreille	2.8088	2.6029	2.5074	-0.0076	2.4121	-0.0953	-3.80%	3.80%
Pierce	2.6586	2.6231	2.6047	-0.0013	2.5881	-0.0166	-0.64%	0.64%
San Juan	2.2946	2.2489	2.1587	-0.0020	2.2041	0.0454	2.10%	2.10%
Skagit	2.5656	2.5495	2.6032	-0.0006	2.5335	-0.0697	-2.68%	2.68%
Skamania	2.7896	2.6921	2.6120	-0.0036	2.5980	-0.0140	-0.54%	0.54%
Snohomish	2.7606	2.67935	2.6547	-0.0030	2.6005	-0.0542	-2.04%	2.04%
Spokane	2.5789	2.4747	2.4646	-0.0041	2.3747	-0.0899	-3.65%	3.65%
Stevens	2.907	2.7318	2.6439	-0.0062	2.5672	-0.0768	-2.90%	2.90%
Thurston	2.6441	2.553	2.4987	-0.0035	2.4650	-0.0337	-1.35%	1.35%
Wahkiakum	2.7724	2.4762	2.4243	-0.0112	2.2116	-0.2127	-8.77%	8.77%
Walla Walla	2.5411	2.4955	2.5388	-0.0018	2.4507	-0.0880	-3.47%	3.47%
Whatcom	2.5902	2.5324	2.5113	-0.0023	2.4759	-0.0354	-1.41%	1.41%
Whitman	2.4668	2.3868	2.3115	-0.0033	2.3094	-0.0021	-0.09%	0.09%
Yakima	2.7711	2.8039	2.9576	0.0012	2.8371	-0.1205	-4.08%	4.08%

COUNTY LEVEL SUMMARY STATISTICS	
Mean Error	**-0.0680**
MAPE	**2.97%**
MALPE	**-2.60%**
N ABS % ERROR >5	**6**

Table 2. Mean ACS Values by Year and Their Standard Deviations

Year	Mean 1-Year ACS PPH Values*	Mean Model-Based PPH Values*	Mean 3-Year ACS PPH Values*	Year
2001	2.503 (0.295)	2.627 (0.281)	2.648 (0.290)	1999-2001
2002	2.509 (0.287)	2.625 (0.286)	2.647 (0.286)	2000-2002
2003	2.642 (0.294)	2.622 (0.289)	2.642 (0.289)	2001-2003
2004	2.647 (0.319)	2.620 (0.300)	2.644 (0.300)	2002-2004
2005	2.623 (0.323)	2.618 (0.303)	2.635 (0.312)	2003-2005
2006	2.717 (0.312)	2.625 (0.309)	N/A	N/A
*The value shown in parentheses is the standard deviation (N=18)				

Table 2.a. Comparison of 2007 Official Population Estimates for California Cities with ACS estimates, Alemeda through Fullerton.

City	ACS TOTAL POP 2007	REGULAR ESTIMATE TOTAL POP 2007	Numeric Difference	ACS MARGIN OF ERROR, TOTAL POP	DIFFERENCE WITHIN ACS MARGIN OF ERROR?
Alameda city, California	75,642	70,272	5,370	+/-6,407	YES
Alhambra city, California	88,393	86,352	2,041	+/-7,561	YES
Anaheim city, California	342,856	333,249	9,607	+/-16,199	YES
Antioch city, California	104,426	99,619	4,807	+/-7,963	YES
Apple Valley town, California	69,835	70,322	-487	+/-7,377	YES
Bakersfield city, California	324,540	315,837	8,703	+/-11,127	YES
Baldwin Park city, California	76,945	77,800	-855	+/-8,390	YES
Bellflower city, California	69,477	73,434	-3,957	+/-7,658	YES
Berkeley city, California	111,680	101,377	10,303	+/-5,974	NO
Buena Park city, California	85,992	79,281	6,711	+/-8,490	YES
Burbank city, California	96,972	103,286	-6,314	+/-7,251	YES
Carlsbad city, California	95,796	95,439	357	+/-7,106	YES
Chico city, California	83,460	83,128	332	+/-4,963	YES
Chino city, California	83,914	82,830	1,084	+/-8,228	YES
Chula Vista city, California	227,336	217,478	9,858	+/-11,597	YES
Citrus Heights city, California	88,576	84,469	4,107	+/-7,880	YES
Clovis city, California	92,987	90,808	2,179	+/-7,419	YES
Compton city, California	100,037	94,425	5,612	+/-9,928	YES
Concord city, California	124,300	120,844	3,456	+/-8,089	YES
Corona city, California	156,394	150,308	6,086	+/-11,341	YES
Costa Mesa city, California	114,057	108,978	5,079	+/-8,421	YES
Daly City city, California	104,752	100,882	3,870	+/-7,752	YES
Downey city, California	109,920	108,109	1,811	+/-11,536	YES
El Cajon city, California	97,964	92,533	5,431	+/-7,993	YES
Elk Grove city, California	138,072	131,212	6,860	+/-9,718	YES
El Monte city, California	113,308	122,272	-8,964	+/-9,809	YES
Escondido city, California	128,819	136,246	-7,427	+/-8,744	YES
Fairfield city, California	111,007	103,992	7,015	+/-7,979	YES
Folsom city, California	74,795	67,401	7,394	+/-5,299	NO
Fontana city, California	193,716	183,502	10,214	+/-11,369	YES
Fremont city, California	214,957	201,334	13,623	+/-10,482	NO
Fresno city, California	476,460	470,508	5,952	+/-11,446	YES
Fullerton city, California	126,955	132,066	-5,111	+/-8,303	YES

Table 2.b. Comparison of 2007 Official Population Estimates for California Cities with ACS estimates, Garden Grove through Norwalk.

City	ACS TOTAL POP 2007	REGULAR ESTIMATE TOTAL POP 2007	Numeric Difference	ACS MARGIN OF ERROR, TOTAL POP	DIFFERENCE WITHIN ACS MARGIN OF ERROR?
Garden Grove city, California	145,923	165,610	-19,687	+/-13,426	NO
Glendale city, California	200,859	196,979	3,880	+/-10,101	YES
Hawthorne city, California	92,321	84,422	7,899	+/-9,239	YES
Hayward city, California	129,885	140,943	-11,058	+/-8,203	NO
Hemet city, California	77,001	70,288	6,713	+/-7,235	YES
Hesperia city, California	90,312	85,515	4,797	+/-8,245	YES
Huntington Beach city, California	188,056	192,885	-4,829	+/-10,437	YES
Indio city, California	70,791	83,937	-13,146	+/-7,026	NO
Inglewood city, California	106,581	113,376	-6,795	+/-9,709	YES
Irvine city, California	205,813	201,160	4,653	+/-8,408	YES
Lake Forest city, California	78,130	75,688	2,442	+/-8,483	YES
Lakewood city, California	89,289	78,956	10,333	+/-7,840	NO
Lancaster city, California	155,902	143,616	12,286	+/-11,940	NO
Livermore city, California	79,213	79,532	-319	+/-6,366	YES
Long Beach city, California	458,302	466,520	-8,218	+/-18,630	YES
Los Angeles city, California	3,806,003	3,834,340	-28,337	+/-43,027	YES
Lynwood city, California	69,537	70,336	-799	+/-6,745	YES
Merced city, California	73,224	76,879	-3,655	+/-6,075	YES
Milpitas city, California	66,494	66,770	-276	+/-5,593	YES
Mission Viejo city, California	92,673	94,586	-1,913	+/-5,823	YES
Modesto city, California	198,456	203,955	-5,499	+/-10,352	YES
Moreno Valley city, California	190,990	188,936	2,054	+/-11,306	YES
Mountain View city, California	70,000	70,436	-436	+/-5,467	YES
Murrieta city, California	89,885	90,555	-670	+/-7,722	YES
Napa city, California	71,664	74,247	-2,583	+/-4,183	YES
Newport Beach city, California	89,125	79,554	9,571	+/-5,726	NO
Norwalk city, California	112,001	103,720	8,281	+/-10,988	YES

Table 2.c. Comparison of 2007 Official Population Estimates for California Cities with ACS estimates, Oakland through San Leando.

City	ACS TOTAL POP 2007	REGULAR ESTIMATE TOTAL POP 2007	Numeric Difference	ACS MARGIN OF ERROR, TOTAL POP	DIFFERENCE WITHIN ACS MARGIN OF ERROR?
Oakland city, California	358,829	401,489	-42,660	+/-13,801	NO
Oceanside city, California	168,814	168,602	212	+/-9,661	YES
Ontario city, California	156,027	170,936	-14,909	+/-11,593	NO
Orange city, California	142,097	134,299	7,798	+/-11,764	YES
Oxnard city, California	167,412	184,725	-17,313	+/-8,354	NO
Palmdale city, California	132,266	140,882	-8,616	+/-10,047	YES
Pasadena city, California	136,936	143,400	-6,464	+/-9,751	YES
Pleasanton city, California	69,348	66,544	2,804	+/-5,983	YES
Pomona city, California	142,111	152,631	-10,520	+/-11,043	YES
Rancho Cucamonga city, California	157,777	170,266	-12,489	+/-12,011	NO
Redding City	87,130	89,780	-2,650	+/-5,302	YES
Redlands city, California	73,539	69,941	3,598	+/-8,059	YES
Redondo Beach city, California	70,948	67,019	3,929	+/-6,838	YES
Redwood City city, California	69,559	73,603	-4,044	+/-5,891	YES
Rialto city, California	108,969	98,713	10,256	+/-9,628	NO
Richmond city, California	97,279	101,454	-4,175	+/-9,020	YES
Riverside city, California	316,154	294,437	21,717	+/-14,637	NO
Roseville city, California	114,958	108,759	6,199	+/-6,578	YES
Sacramento city, California	451,404	460,242	-8,838	+/-15,995	YES
Salinas city, California	140,499	143,517	-3,018	+/-8,046	YES
San Bernardino city, California	203,691	199,285	4,406	+/-11,585	YES
San Buenaventura (Ventura) city, California	105,673	103,219	2,454	+/-6,800	YES
San Diego city, California	1,276,740	1,266,731	10,009	+/-22,810	YES
San Francisco city, California	764,976	764,976	0	*****	N/A
San Jose city, California	922,389	939,899	-17,510	+/-16,294	NO
San Leandro city, California	96,186	77,725	18,461	+/-9,192	NO

Table 2.d. Comparison of 2007 official Population Estimates for California Cities with ACS estimates, San Marcos through Yorba Linda.

City	ACS TOTAL POP 2007	REGULAR ESTIMATE TOTAL POP 2007	Numeric Difference	ACS MARGIN OF ERROR, TOTAL POP	DIFFERENCE WITHIN ACS MARGIN OF ERROR?
San Marcos city, California	75,217	78,286	-3,069	+/-7,344	YES
San Mateo city, California	91,461	91,768	-307	+/-5,967	YES
Santa Ana city, California	327,780	339,555	-11,775	+/-14,085	YES
Santa Barbara city, California	89,959	86,204	3,755	+/-6,453	YES
Santa Clara city, California	105,591	109,756	-4,165	+/-6,451	YES
Santa Clarita city, California	177,740	169,951	7,789	+/-13,076	YES
Santa Maria city, California	86,160	85,685	475	+/-6,581	YES
Santa Monica city, California	86,857	87,212	-355	+/-6,317	YES
Santa Rosa city, California	147,516	154,241	-6,725	+/-8,276	YES
Simi Valley city, California	127,053	120,464	6,589	+/-8,047	YES
South Gate city, California	104,031	97,110	6,921	+/-8,719	YES
Stockton city, California	295,070	287,245	7,825	+/-10,999	YES
Sunnyvale city, California	135,548	131,140	4,408	+/-8,627	YES
Temecula city, California	93,743	94,767	-1,024	+/-8,318	YES
Thousand Oaks city, California	128,519	123,349	5,170	+/-7,821	YES
Torrance city, California	143,628	141,420	2,208	+/-8,316	YES
Tracy city, California	82,383	79,705	2,678	+/-6,540	YES
Turlock city, California	69,330	68,133	1,197	+/-6,246	YES
Tustin city, California	63,524	70,869	-7,345	+/-6,624	YES
Union City city, California	73,212	70,075	3,137	+/-6,373	YES
Upland city, California	78,260	72,464	5,796	+/-9,002	YES
Vacaville city, California	93,795	92,084	1,711	+/-6,076	YES
Vallejo city, California	106,608	115,552	-8,944	+/-6,297	NO
Victorville city, California	97,534	107,221	-9,687	+/-9,214	NO
Visalia city, California	115,899	118,603	-2,704	+/-8,732	YES
Vista city, California	97,977	90,839	7,138	+/-10,065	YES
West Covina city, California	103,154	106,388	-3,234	+/-8,704	YES
Westminster city, California	91,994	88,678	3,316	+/-6,606	YES
Whittier city, California	82,755	82,850	-95	+/-8,684	YES
Yorba Linda city, California	57,550	65,434	-7,884	+/-4,415	NO

EXHIBIT 1. The 18 COUNTIES USED IN THE ACS ANALYSIS

Pima County, AZ
Jefferson County, AR
San Francisco County, CA
Tulare County, CA
Broward County, FL
Lake County, IL
Black Hawk County, IA
Calvert County, MD
Hampden County, MA
Madison County, MS
Douglas County, NE
Bronx County, NY
Rockland County, NY
Franklin County, OH
Multnomah County, OR
Schuylkill County, PA
Sevier County, TN
Yakima County, WA

EXHIBIT 2.1*

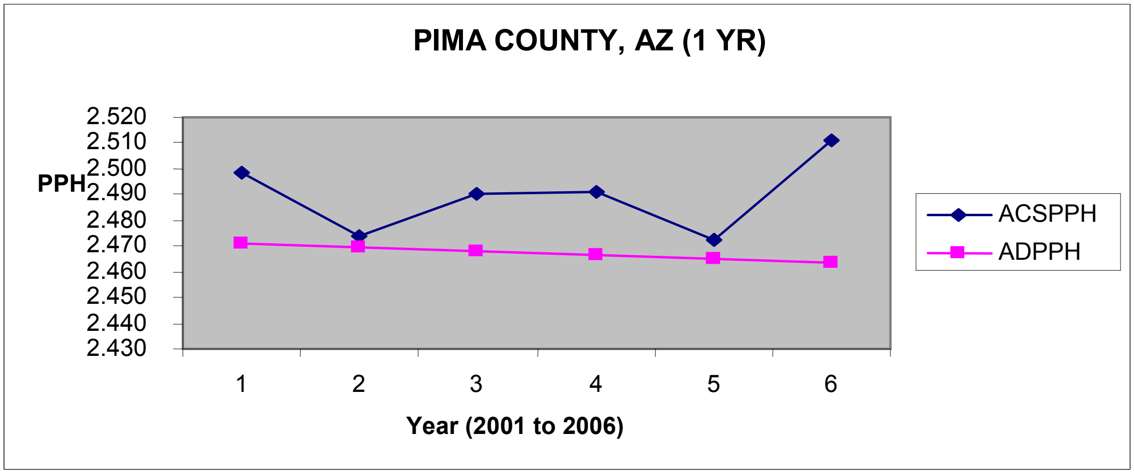
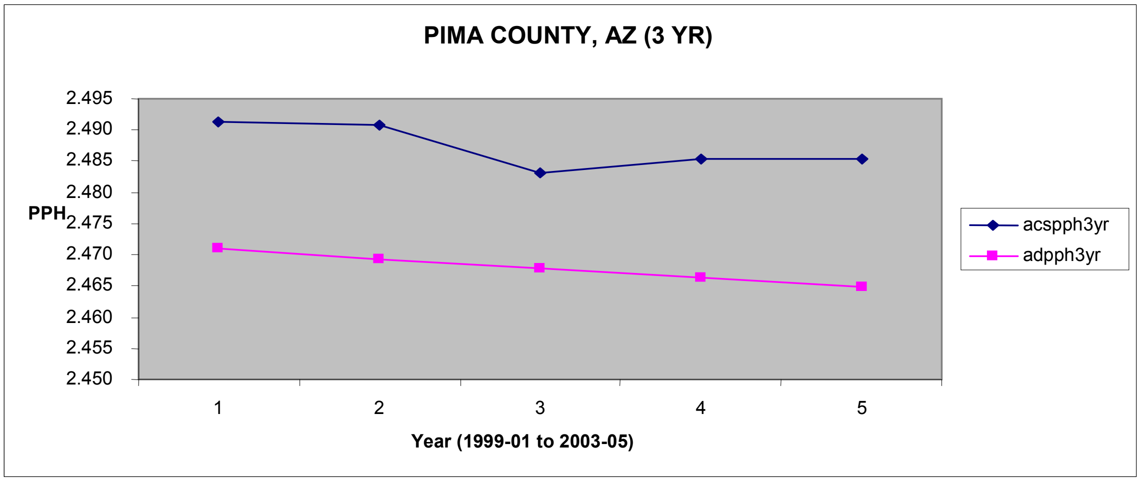


EXHIBIT 2.2*



*The ACS PPH values are labeled as "ACSPPH" and the model-generated PPH values are labeled as "ADPPH" (where "AD" stands for Analytically Derived). "(1 YR)" stands for single year ACS data and "(3 YR)" stands for 3 year ACS data.

EXHIBIT 3.1*

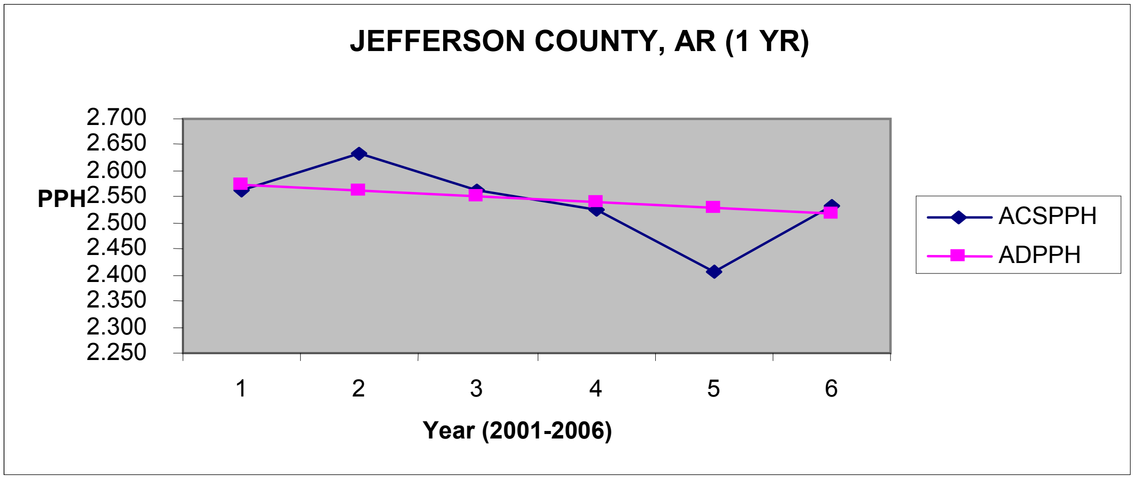
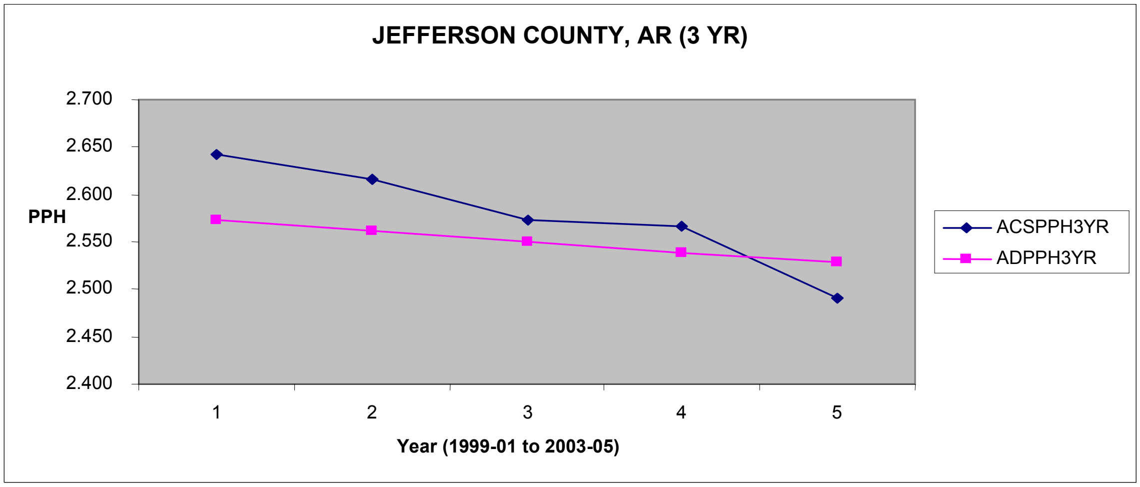


EXHIBIT 3.2*



*The ACS PPH values are labeled as “ACSPPH” and the model-generated PPH values are labeled as “ADPPH” (where “AD” stands for Analytically Derived). “(1 YR)” stands for single year ACS data and “(3 YR)” stands for 3 year ACS data.

EXHIBIT 4.1*

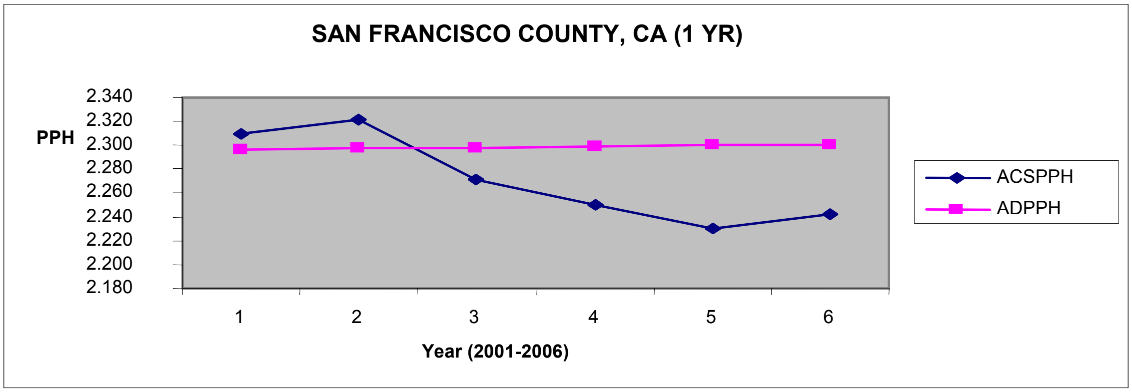
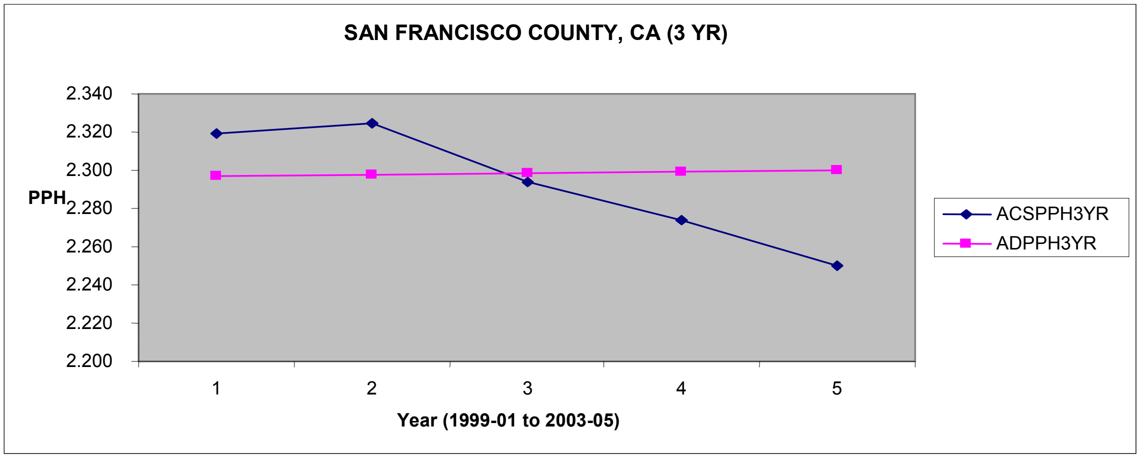


EXHIBIT 4.2*



*The ACS PPH values are labeled as “ACSPPH” and the model-generated PPH values are labeled as “ADPPH” (where “AD” stands for Analytically Derived). “(1 YR)” stands for single year ACS data and “(3 YR)” stands for 3 year ACS data.

EXHIBIT 5.1*

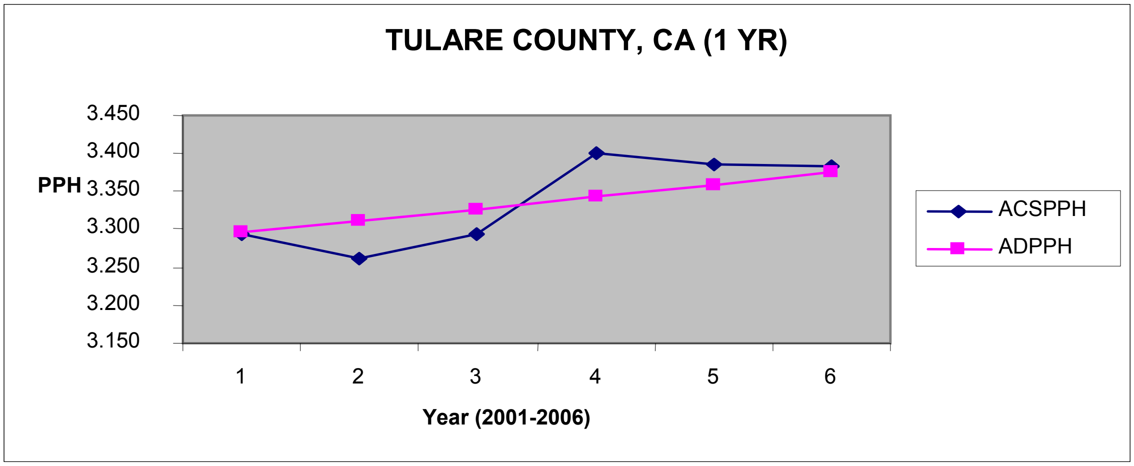
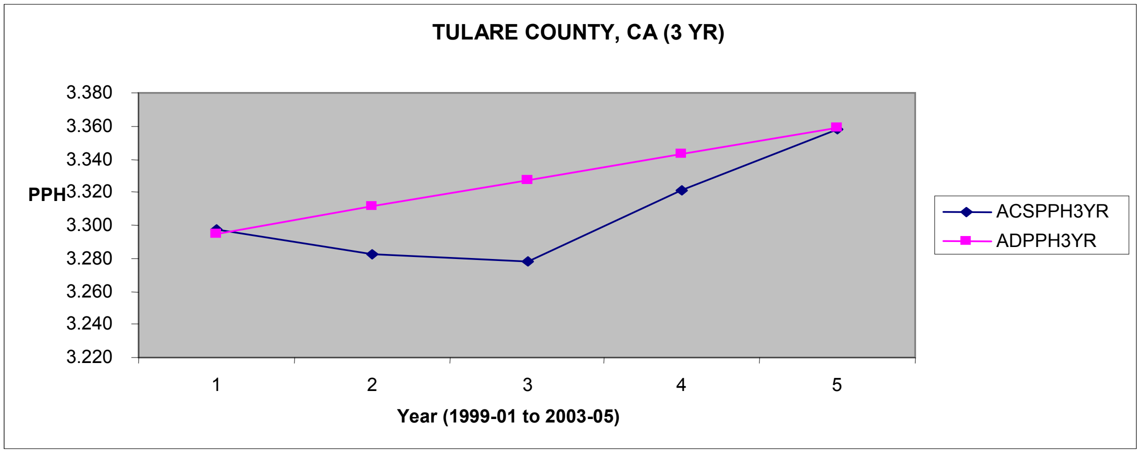


EXHIBIT 5.2*



*The ACS PPH values are labeled as “ACSPPH” and the model-generated PPH values are labeled as “ADPPH” (where “AD” stands for Analytically Derived). “(1 YR)” stands for single year ACS data and “(3 YR)” stands for 3 year ACS data.

EXHIBIT 6.1*

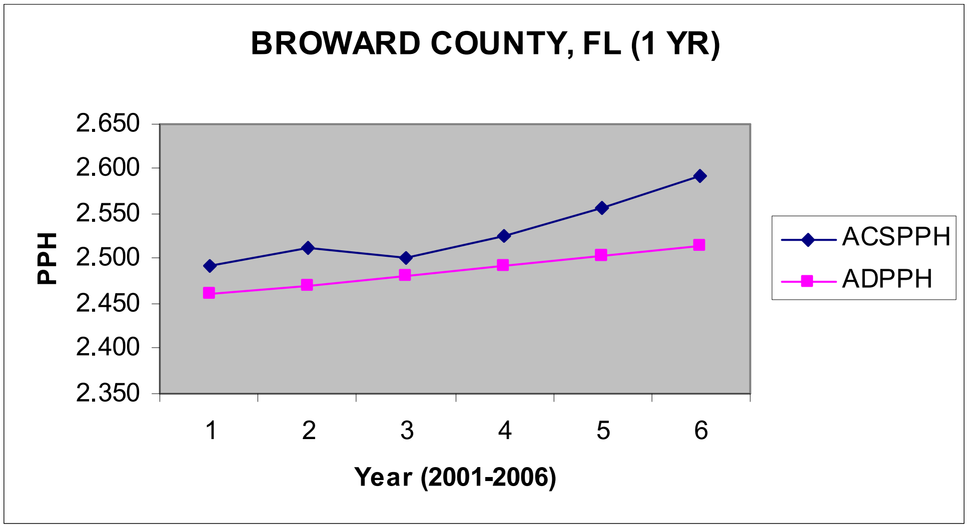
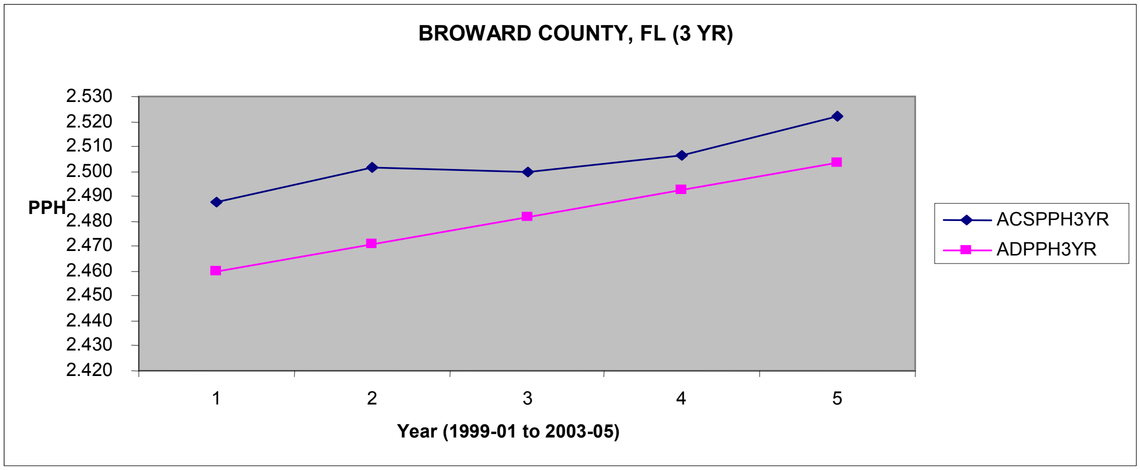


EXHIBIT 6.2*



*The ACS PPH values are labeled as "ACSPPH" and the model-generated PPH values are labeled as "ADPPH" (where "AD" stands for Analytically Derived). "(1 YR)" stands for single year ACS data and "(3 YR)" stands for 3 year ACS data.

EXHIBIT 7.1*

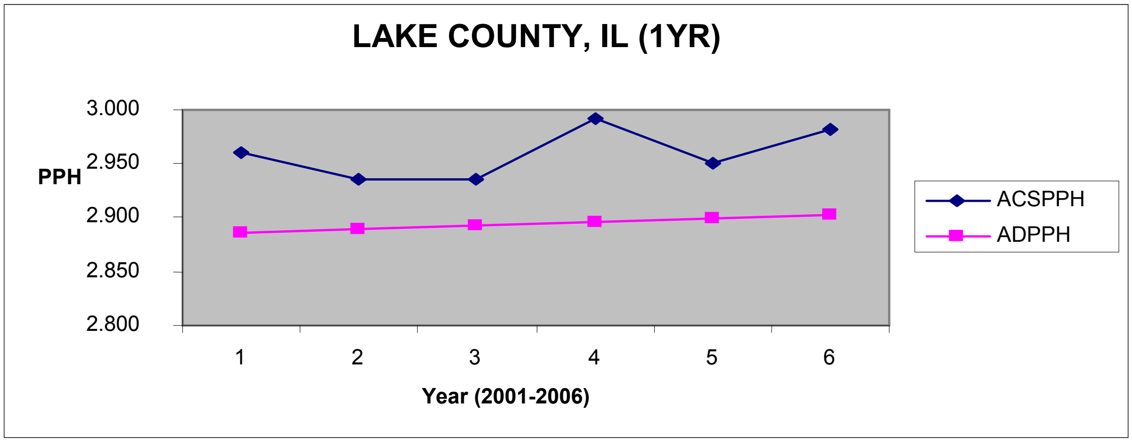
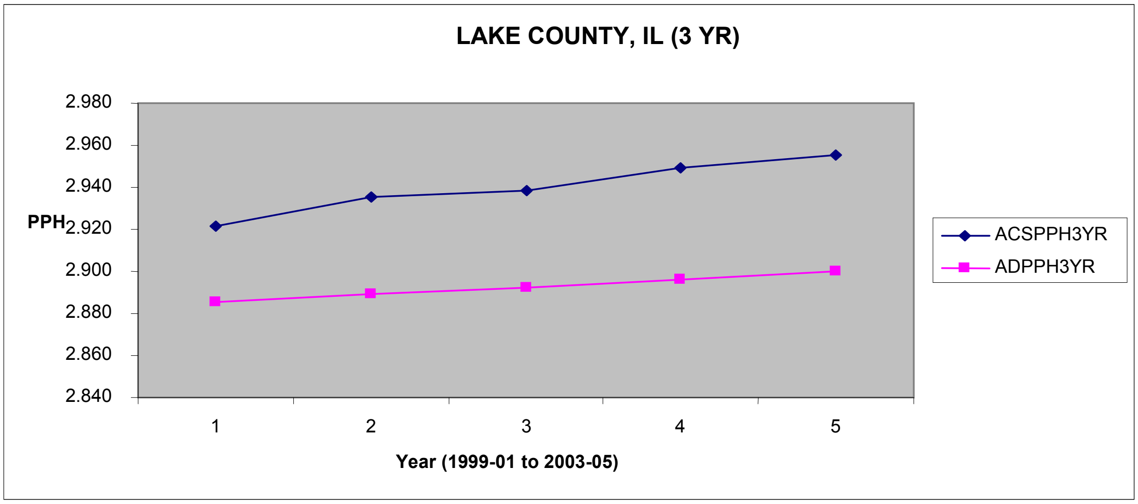


EXHIBIT 7.2*



*The ACS PPH values are labeled as "ACSPPH" and the model-generated PPH values are labeled as "ADPPH" (where "AD" stands for Analytically Derived). "(1 YR)" stands for single year ACS data and "(3 YR)" stands for 3 year ACS data.

EXHIBIT 8.1*

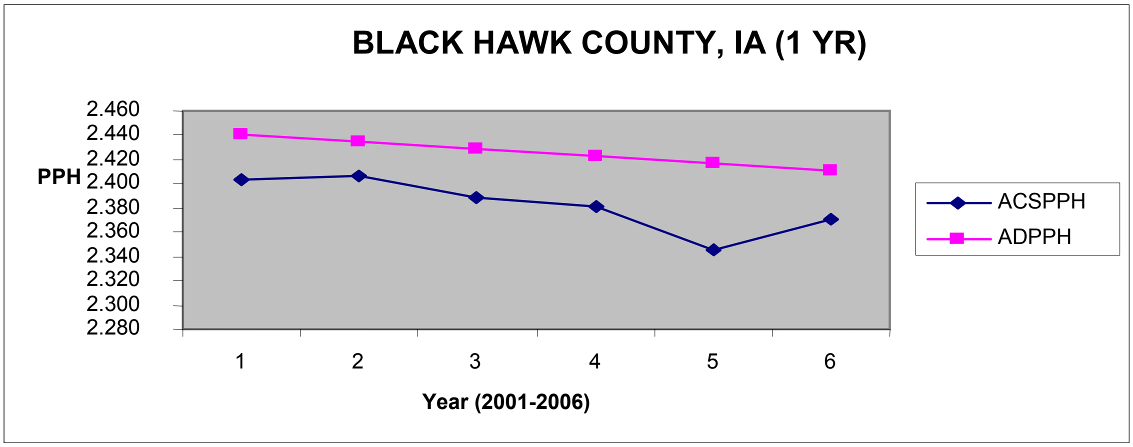
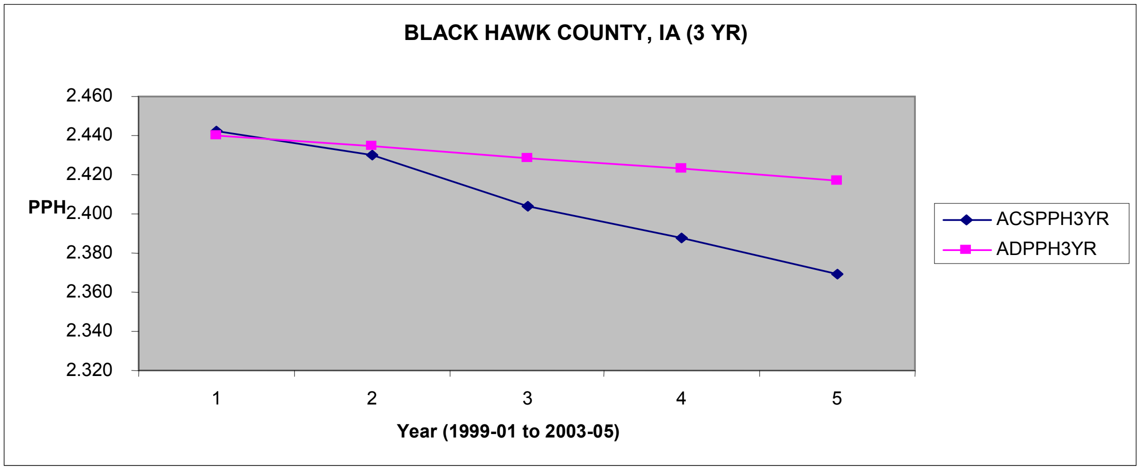


EXHIBIT 8.2*



*The ACS PPH values are labeled as "ACSPPH" and the model-generated PPH values are labeled as "ADPPH" (where "AD" stands for Analytically Derived). "(1 YR)" stands for single year ACS data and "(3 YR)" stands for 3 year ACS data.

EXHIBIT 9.1*

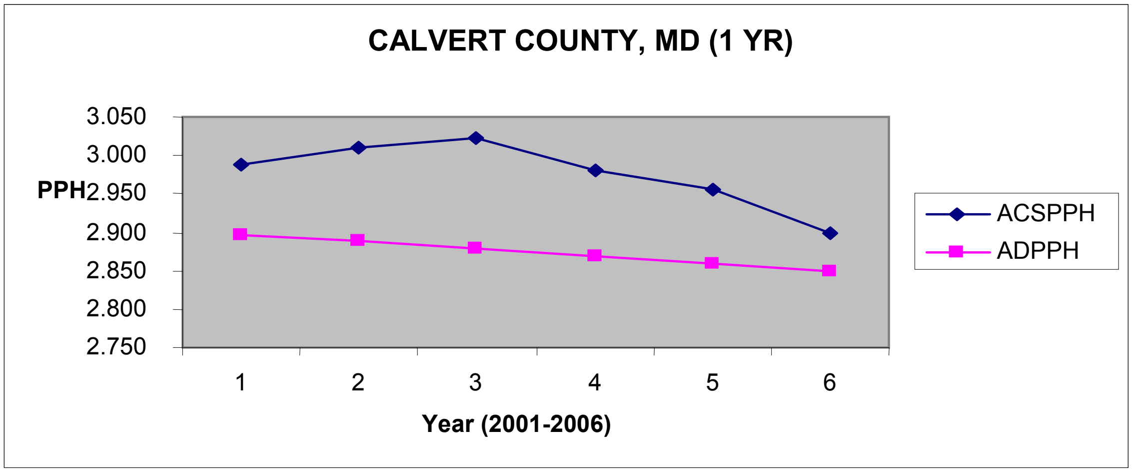
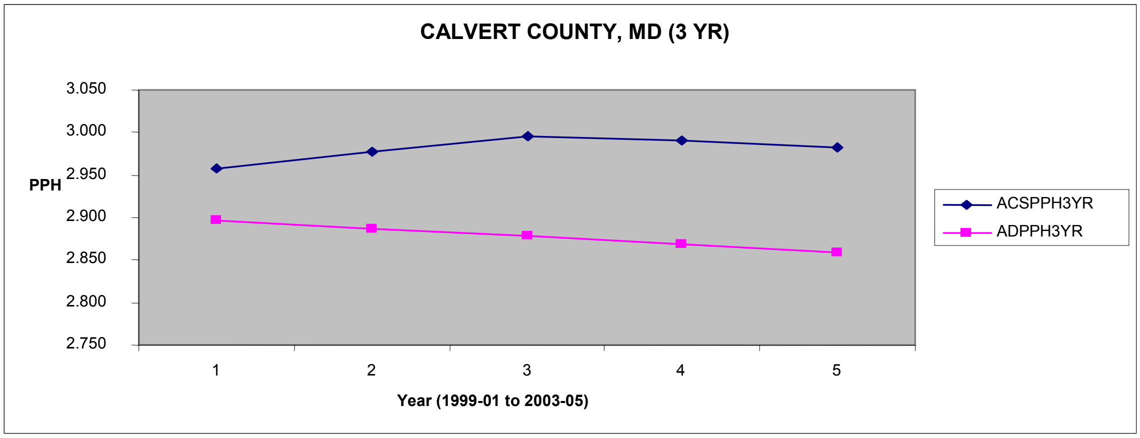


EXHIBIT 9.2*



*The ACS PPH values are labeled as "ACSPPH" and the model-generated PPH values are labeled as "ADPPH" (where "AD" stands for Analytically Derived). "(1 YR)" stands for single year ACS data and "(3 YR)" stands for 3 year ACS data.

EXHIBIT 10.1*

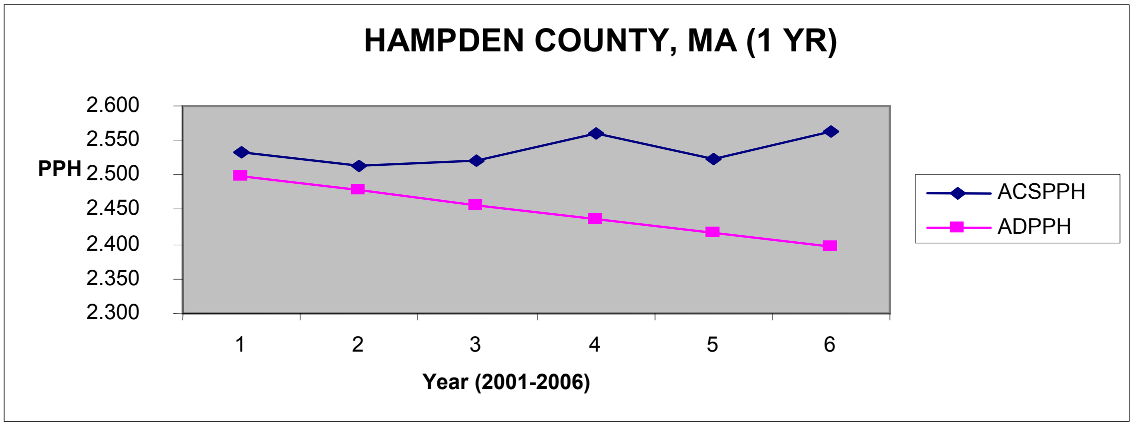
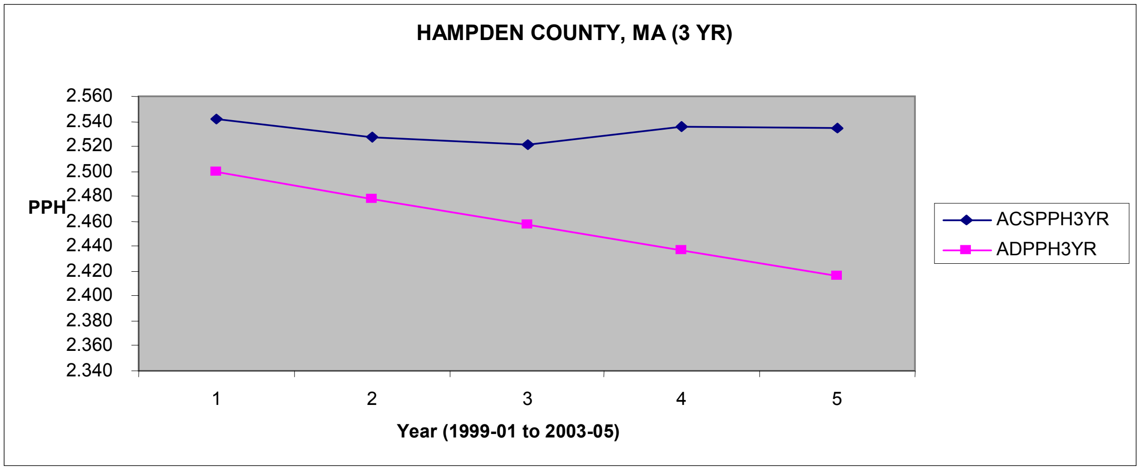


EXHIBIT 10.2*



*The ACS PPH values are labeled as "ACSPPH" and the model-generated PPH values are labeled as "ADPPH" (where "AD" stands for Analytically Derived). "(1 YR)" stands for single year ACS data and "(3 YR)" stands for 3 year ACS data.

EXHIBIT 11.1*

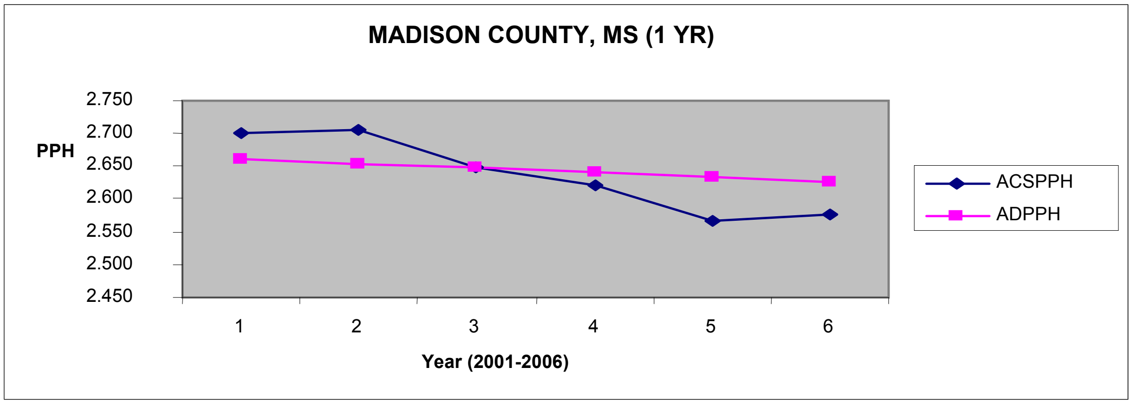
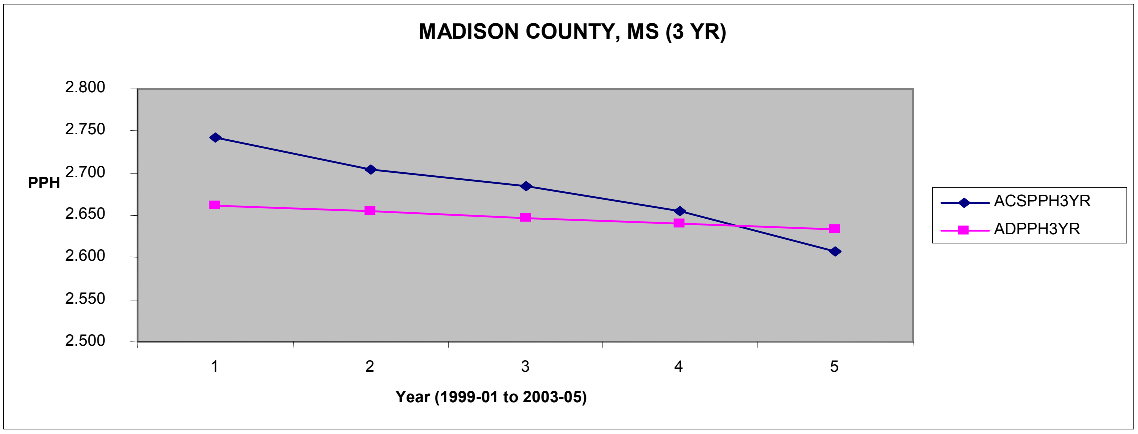


EXHIBIT 11.2*



*The ACS PPH values are labeled as "ACSPPH" and the model-generated PPH values are labeled as "ADPPH" (where "AD" stands for Analytically Derived). "(1 YR)" stands for single year ACS data and "(3 YR)" stands for 3 year ACS data.

EXHIBIT 12.1*

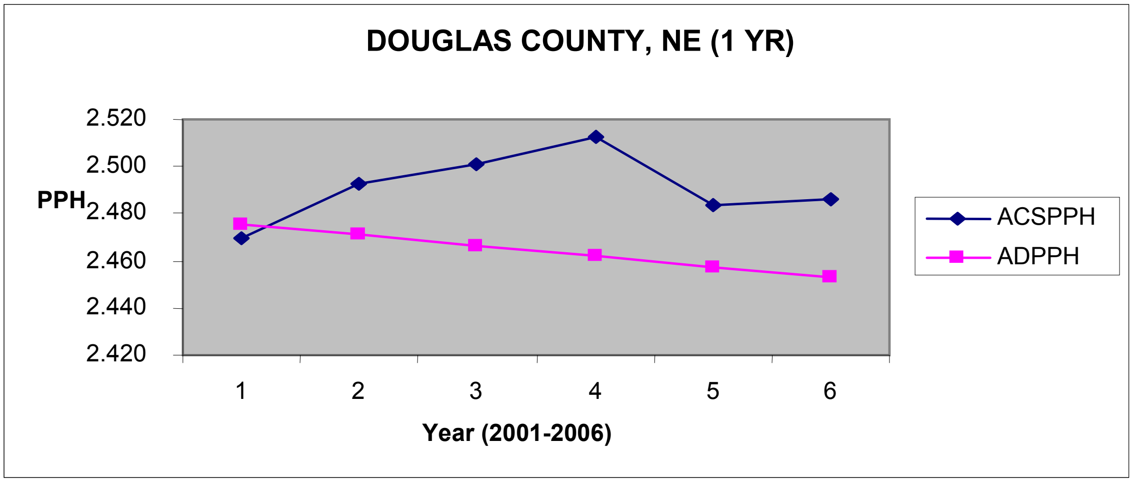
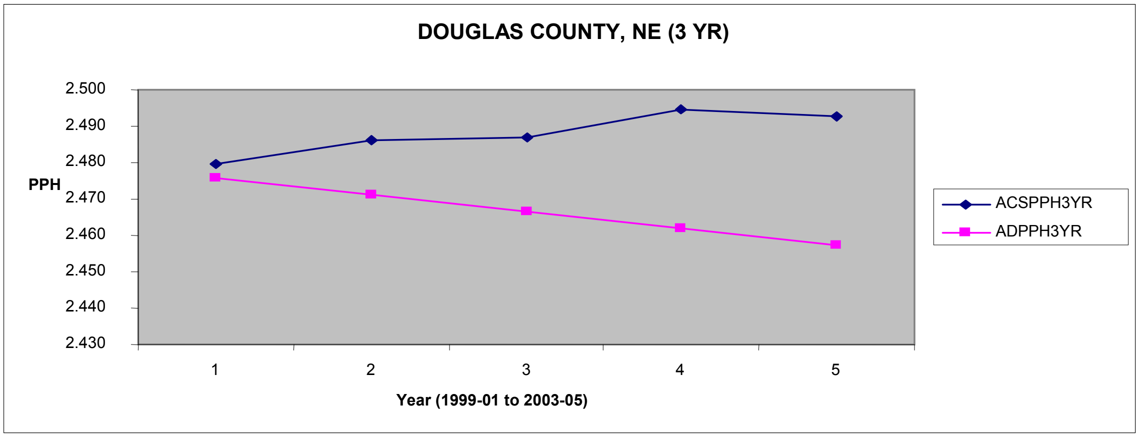


EXHIBIT 12.2*



*The ACS PPH values are labeled as "ACSPPH" and the model-generated PPH values are labeled as "ADPPH" (where "AD" stands for Analytically Derived). "(1 YR)" stands for single year ACS data and "(3 YR)" stands for 3 year ACS data.

EXHIBIT 13.1*

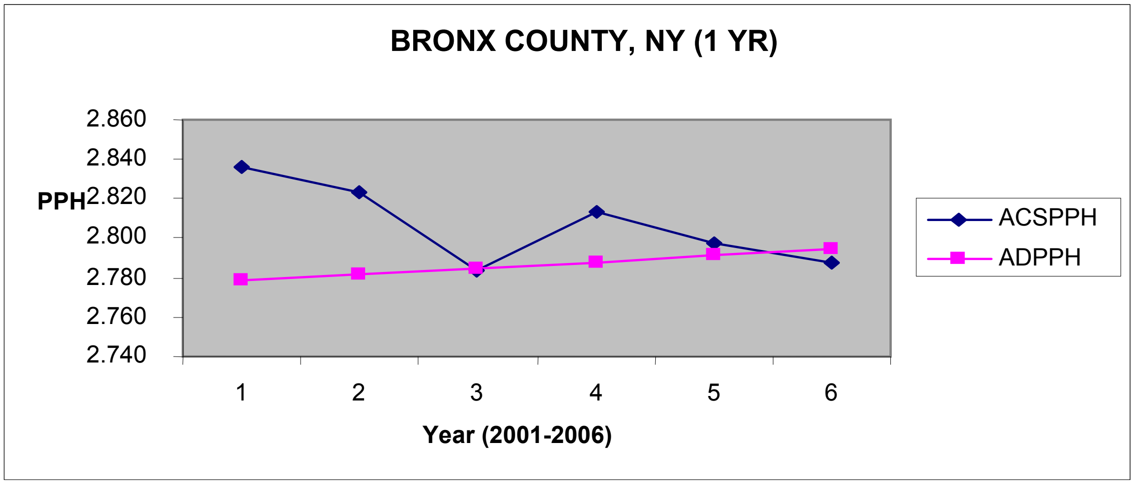
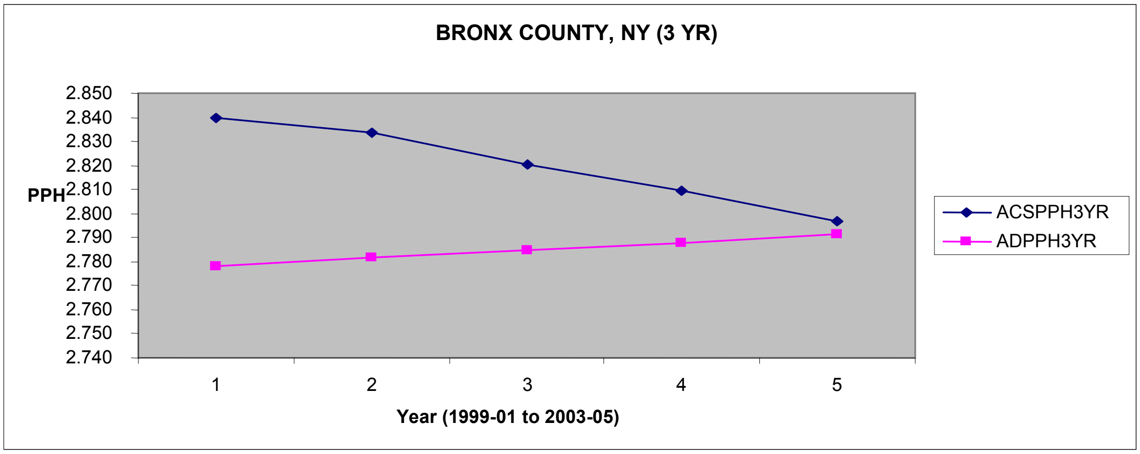


EXHIBIT 13.2*



*The ACS PPH values are labeled as "ACSPPH" and the model-generated PPH values are labeled as "ADPPH" (where "AD" stands for Analytically Derived). "(1 YR)" stands for single year ACS data and "(3 YR)" stands for 3 year ACS data.

EXHIBIT 14.1*

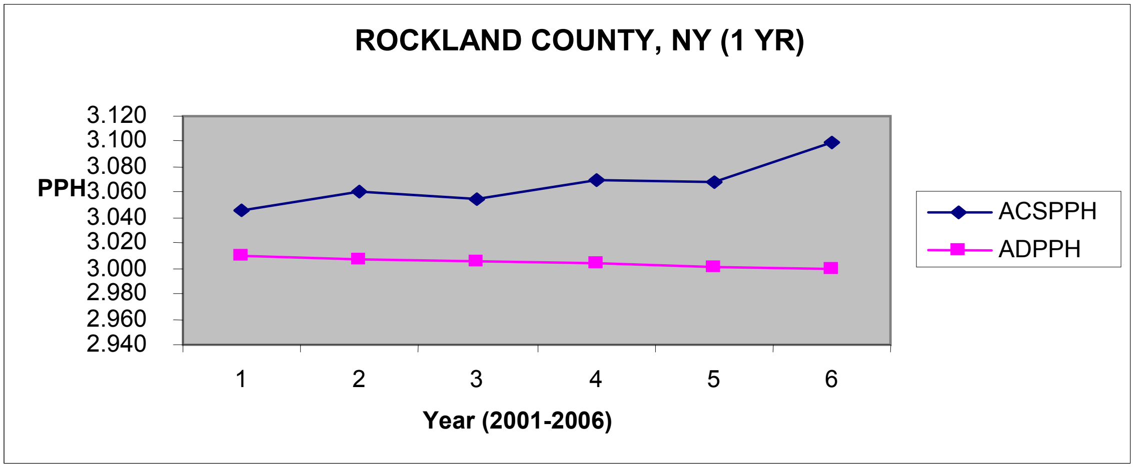
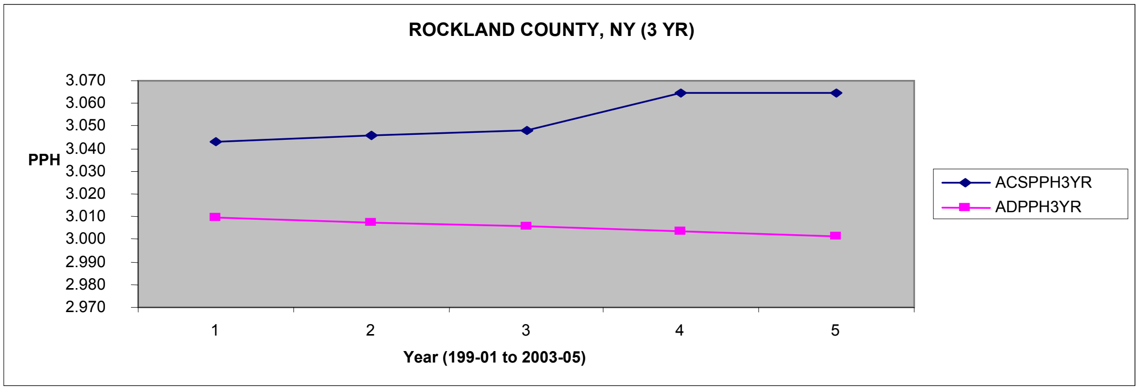


EXHIBIT 14.2*



*The ACS PPH values are labeled as "ACSPPH" and the model-generated PPH values are labeled as "ADPPH" (where "AD" stands for Analytically Derived). "(1 YR)" stands for single year ACS data and "(3 YR)" stands for 3 year ACS data.

EXHIBIT 15.1*

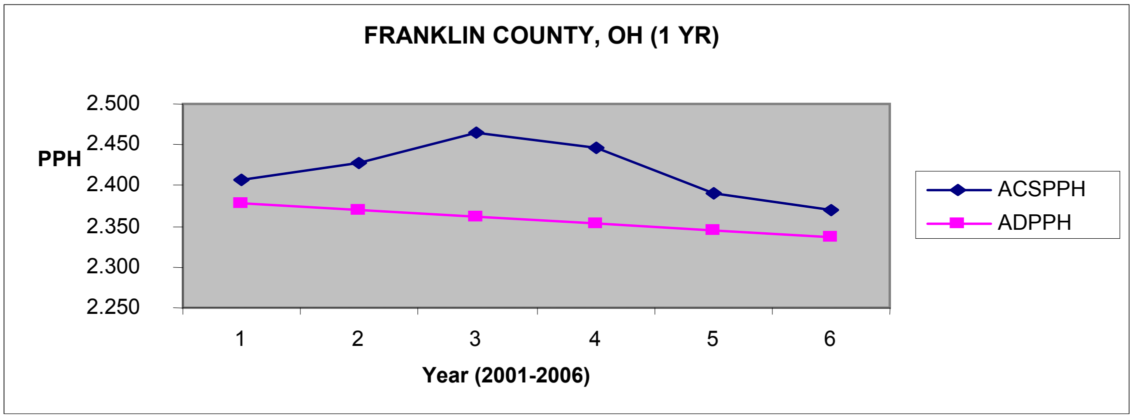
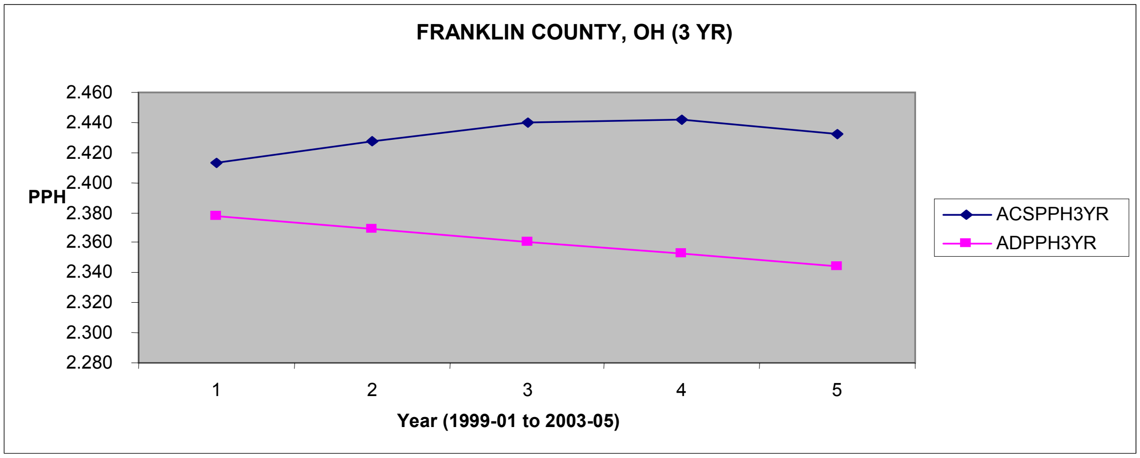


EXHIBIT 15.2*



*The ACS PPH values are labeled as "ACSPPH" and the model-generated PPH values are labeled as "ADPPH" (where "AD" stands for Analytically Derived). "(1 YR)" stands for single year ACS data and "(3 YR)" stands for 3 year ACS data.

EXHIBIT 16.1*

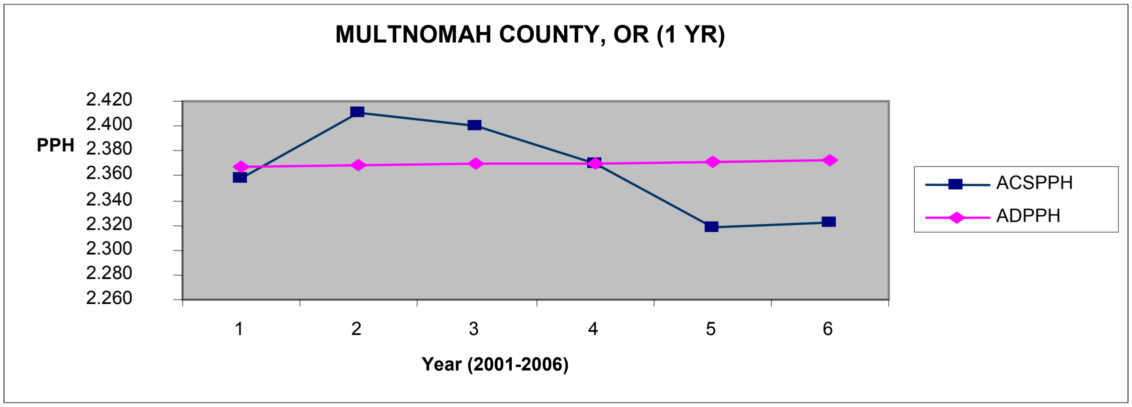
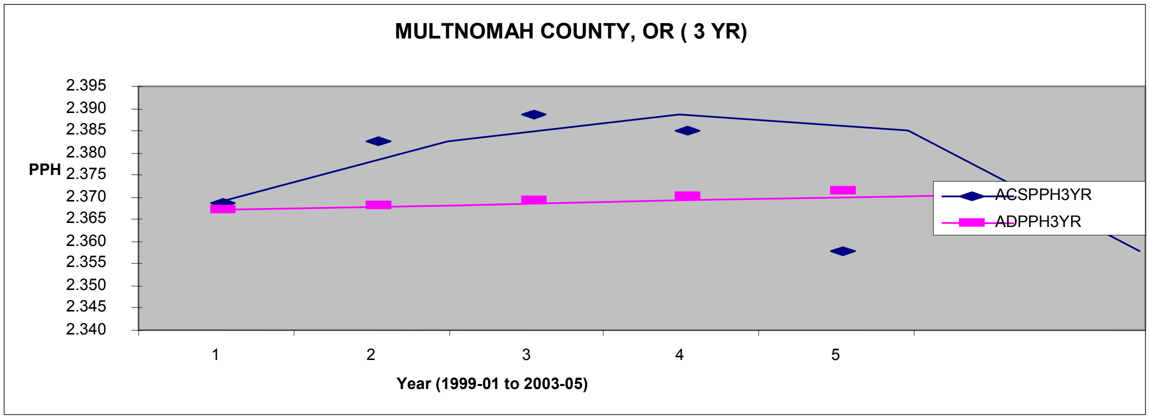


EXHIBIT 16.2*



*The ACS PPH values are labeled as “ACSPPH” and the model-generated PPH values are labeled as “ADPPH” (where “AD” stands for Analytically Derived). “(1 YR)” stands for single year ACS data and “(3 YR)” stands for 3 year ACS data.

EXHIBIT 17.1*

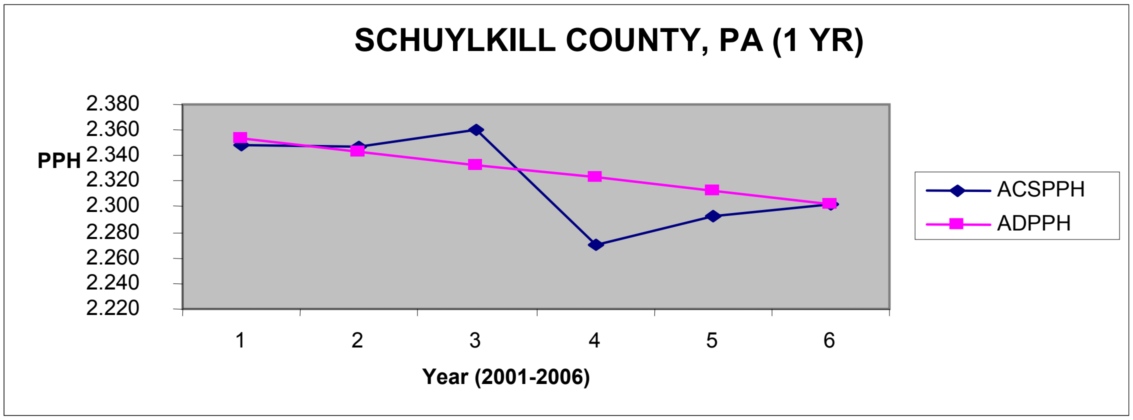
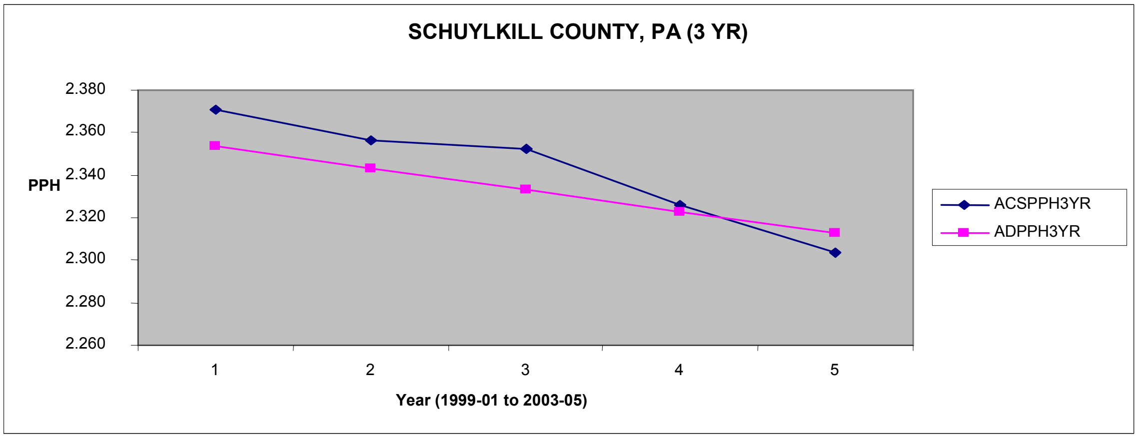


EXHIBIT 17.2*



*The ACS PPH values are labeled as "ACSPPH" and the model-generated PPH values are labeled as "ADPPH" (where "AD" stands for Analytically Derived). "(1 YR)" stands for single year ACS data and "(3 YR)" stands for 3 year ACS data.

EXHIBIT 18.1*

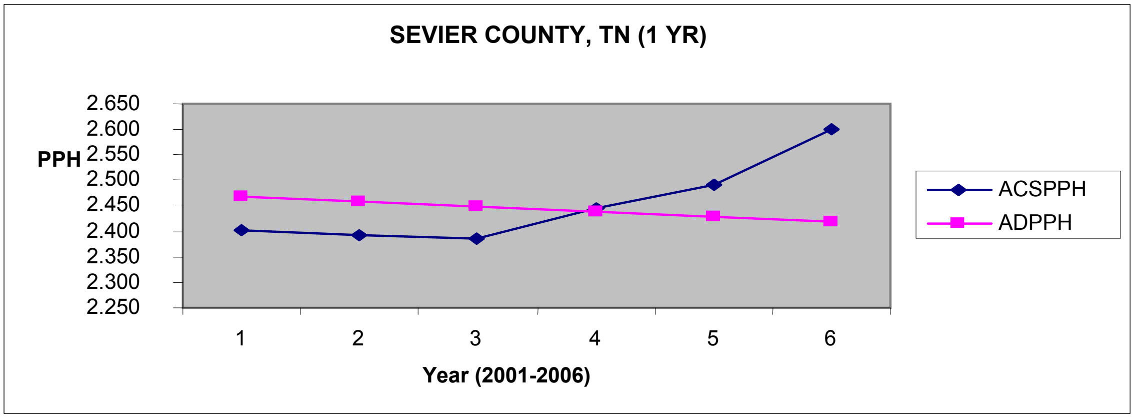
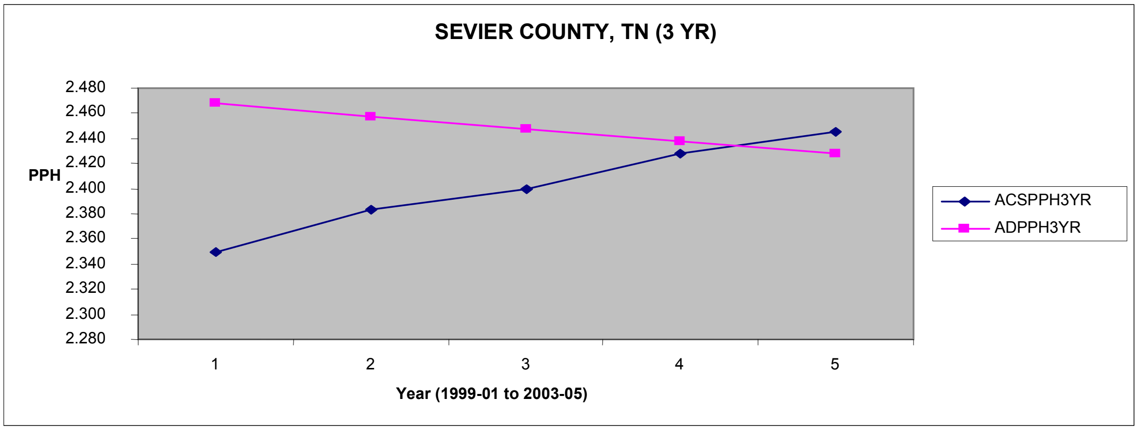


EXHIBIT 18.2*



*The ACS PPH values are labeled as "ACSPPH" and the model-generated PPH values are labeled as "ADPPH" (where "AD" stands for Analytically Derived). "(1 YR)" stands for single year ACS data and "(3 YR)" stands for 3 year ACS data.

EXHIBIT 19.1*

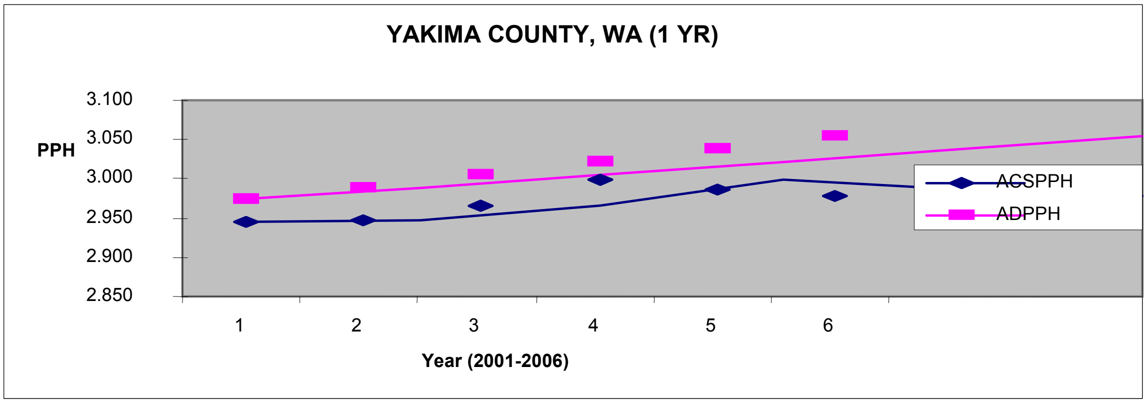
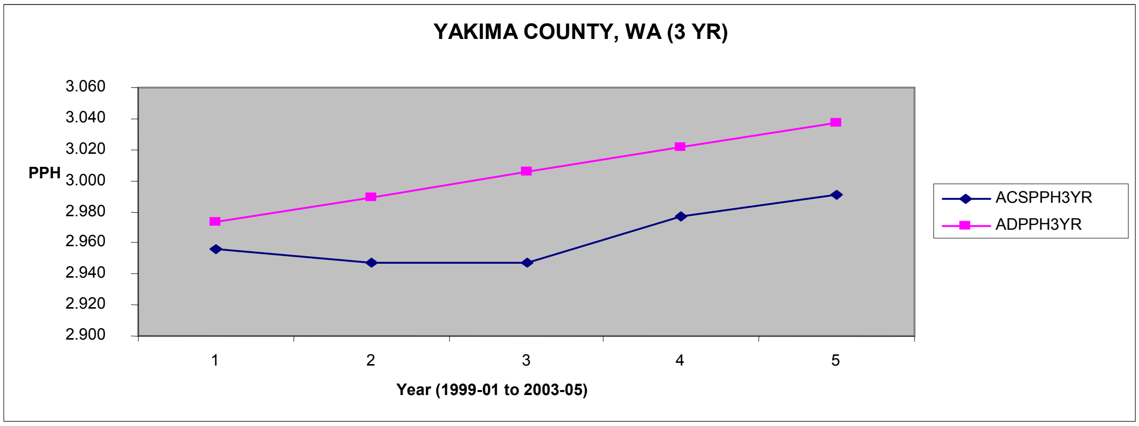


EXHIBIT 19.2*



*The ACS PPH values are labeled as "ACSPPH" and the model-generated PPH values are labeled as "ADPPH" (where "AD" stands for Analytically Derived). "(1 YR)" stands for single year ACS data and "(3 YR)" stands for 3 year ACS data.