

**Developing Annual Population Data in the United States:
New Possibilities for the 21st Century ***

David A. Swanson
Department of Sociology and Center for Sustainable Suburban Development
University of California Riverside
Riverside CA USA 92521
David.swanson@ucr.edu

Jerome N. McKibben
McKibben Demographic Research
P.O. Box 2921
Rock Hill, SC USA 29732
j.mckibben@mckibbendemographics.com

*An earlier version of this paper was presented at the 2007 Conference on Small Area Estimation, September 3-5, at the University of Pisa, Faculty of Economics. The authors are grateful for the comments of conference participants and others, particularly those by anonymous reviewers, as well as those by Steve Murdock, Stan Smith, and Paula Walashek.

**Developing Annual Population Data in the United States:
New Possibilities for the 21st Century ***

Abstract

The advent of a continuously updated Master Area File (MAF) following the 2000 census represents an information resource that can be tapped for purposes of developing timely, cost-effective, and precise population estimates for even the smallest of geographical units (e.g., census blocks). We argue that the MAF can be enhanced (EMAF) for these purposes. In support of our argument we describe a set of activities needed to develop EMAF, each of which is well within the current capabilities of the U.S. Census Bureau and discuss various costs and benefits of each. We also describe how EMAF would provide population estimates containing a wide range of demographic (e.g., age, race, and sex) and socio-economic characteristics (e.g., educational attainment, income, and employment). As such, it would largely negate the need for many of the traditional demographic methods of population estimation and possibly reduce the number of sample surveys. We identify important challenges that must be surmounted in order to realize EMAF and make suggestions for doing so. We conclude by noting that the idea of the EMAF could be of interest to other countries with MAF files and strong administrative records systems that, like the United States, are facing the challenge of producing good population information in the face of increasing census costs.

Key Words: Population Estimates, Housing Units, Master Address File, EMAF

I. Introduction

In the 1990 and earlier censuses, the U.S. Census Bureau prepared a Master Address File, a geographically referenced nationwide address list, as part of its preparations for each census. After each of these censuses, the existing Master Address File (MAF) was discarded and a new MAF was constructed as the next census approached. With the passage of Public Law 104-30, “The Census Address List Improvement Act of 1994,” the legal and administrative groundwork was laid for an on-going MAF. Following the enactment of this law, the Census Bureau started the development of a MAF that would not only be used for the 2000 Census, but continuously updated thereafter. This continuously up-dated MAF is now a fact of life at the Census Bureau.

We believe that the advent of this continuously updated MAF represents an information resource that can be tapped for purposes of developing timely, cost-effective, and precise population estimates for even the smallest of geographical units (e.g., census blocks). To accomplish this, we propose that the MAF be extended to what we term the Enhanced Master Address File (EMAF). In support of our argument, we describe a set of activities needed to develop EMAF, each of which is well within the current technical and administrative capabilities of the U.S. Census Bureau. We further describe how EMAF could provide demographic (e.g., age, race, and sex) and socio-economic characteristics (e.g., educational attainment, income, and employment). We also identify challenges facing the construction of EMAF and discuss how these may be overcome.

As a means of providing a context for this effort it is important to recall 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 – as well as in Australia, Canada, England, and New Zealand. In addition to actually conducting census counts, there are three other characteristics that link the United States with these other countries: (1) well-developed administrative records systems (e.g., vital events registration); (2) regular census counts; and (3) no population registration system, such as those found in the Nordic countries (see, e.g, Statistics Finland, 2004). A census is a time-consuming and costly endeavor. 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, developing marketing, and for descriptive and analytical studies (Long, 1993; 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.

In order to meet the need for current population figures, many estimation methods have been developed, virtually all of which can be categorized into one or the other of two traditions: (1) demographic (Bryan, 2004b); and (2) statistical. The former is characterized by a range of methods and data sources (Bryan, 2004b; Lee and Goldsmith, 1982; National Research Council,

1980; Rives, et al., 1995; and Swanson and Pol, 2005) while the latter tends to be confined to sample surveys and the methods developed to “extend” sample surveys (Fay, 2005; Ghosh and Rao, 1994; Kordos, 2000; National Research Council, 1980; Platek, Rao, Sarndal, and Singh, 1987; Rao, 2003; and Subcommittee on Small Area Estimation, 1993). Demographic methods are used to develop estimates of a total population as well as its demographic characteristics – age, race, and sex, for example (Bryan, 2004b; Lee and Goldsmith, 1982; National Research Council, 1980; Rives, et al., 1995; Siegel, 2002: 489-508; and Swanson and Pol, 2005).

Although there are exceptions (Bousfield, 2002), statistical methods are largely used to estimate the socio-economic characteristics of a population – educational attainment, income, and employment, for example (Bryan, 2004b; National Research Council, 1980; National Research Council, 2007; Siegel, 2002: 489-508). As is the case in the national statistical agencies of other countries, the U.S. Census Bureau produces estimates using both of these traditions (Bryan, 2004a, 2004b; Siegel, 2002: 489-508). We focus the discussion on methods that fit within the demographic tradition and only touch on those that fit within the statistical tradition. However, we identify links among selected methods in both traditions. This discussion provides a point of departure for our recommendations in regard to the production of population estimates.

Our discussion primarily is aimed at the development of “de jure” population, which is the definition used by the U.S. Census Bureau and is based on place of usual residence (Cook, 1996; Cork and Voss, 2006; Wilmoth, 2004). We note that “de facto” populations are also of importance (Cook, 1996; Happel and Hogan, 2002; Schmitt, 1975; Smith, 1994; and Smith and House, 2007). They include vacationers (of interest, for example, to the casino industry in Las Vegas and the Hawai’i Visitors Bureau), migratory workers (of interest, for example, to health care, school, and other social service providers), temporary migrants such as “snowbirds” (of

interest to city of Palm Beach for purposes of providing services) 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 interest, they are very difficult to make in the United States because of the lack of census type benchmarks (Cook, 1996, Smith, 1994). As such, discussing the development of de facto population information is beyond the scope of our paper. We 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.

The remainder of this paper consists of six sections, endnotes and references. The following section provides an overview of basic concepts, data sources, and methods used to estimate populations in the U. S. The third section discusses the needs of users, with a focus on researchers. The fourth section describes EMAF, our suggestion for meeting the needs of users while the fifth section describes some of its benefits. The sixth section discusses the obstacles associated with this EMAF and how they might be overcome. The seventh and final section asks if EMAF is feasible.

II. Basic Concepts, Data Sources, and Methods

In this section, our intention is not to cover concepts, data sources, and methods related to population estimates in depth. Rather, it is to generally describe them while providing citations to more detailed descriptions and discussions.

Basic Concepts. 1. Following Smith, Tayman, and Swanson (2001: 16), we 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.

In regard to an estimate, demographers traditionally distinguish between “post-censal” and “inter-censal,” where the former refers to an estimate for a date between two censuses that takes the results of these censuses into account and the latter refers to an estimate for a date subsequent to the most recently available census (Bryan, 2004b: 523).¹ 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, 763).²

Another useful set of concepts is the notion of “stocks and flows”. As defined by Popoff and Judson (2004: 603), “...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 define micro data and aggregated data. We take micro data to mean records for individual persons. These records are often linked by relationships to form family and household records and we 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

Basic Data Sources. All estimates, including post-censal ones, 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). 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 U.S. 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).

With the development of the continuously updated MAF for Census 2000, the Census Bureau has introduced an important new source of data. As observed nearly 25 years ago by Pittenger (1982) and more recently by Wang (1999), this “living” housing unit inventory could serve as a key resource in the Bureau’s ability to construct population estimates.³ Not surprisingly, the Census Bureau explicitly recognizes the potential of the MAF and has embarked on a series of evaluations into using it for a range of activities related to estimation, both direct and indirect (Hakanson, 2007; Liu, 2007, 2008; Reese, 2006; Swanson, 2009; U.S. Census Bureau, 2007).

Methods. Although it is not used directly in any of the standard population estimation methods used at the sub-national level, the fundamental demographic identity known as the balancing equation forms the conceptual framework for most of these same methods. This identity is defined as $P_t = P_0 + I - O$, where P_t is the given population at time $0 + t$, P_0 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 methods.

1. 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, inter-censal 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).

2. Housing Unit Method

The Housing Unit Method (HUM) is a “stock” method that describes a basic identity in the same way that the balancing equation does. In the case of the HUM, this identity is usually given as $P = H * O * PPH + GQ$, where P = Population, H = housing units, O = Proportion occupied, PPH

= average number of persons per household, and GQ = the population residing in “group quarters” and the homeless (Bryan, 2004b). Like the balancing equation, the HUM equation can be expressed in less detail (i.e., $P = HH * PPH + GQ$, where $HH = H * O$, Smith and Cody, 2004: 2) or more detail - by structure type, for example (Devine and Coleman, 2003; 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). 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 continuously updated MAF, the HUM has now emerged as a method that can be used by the U.S. Census Bureau for all sub-national areas and the nation as a whole (Swanson, 2009; Wang, 1999).

3. Regression Methods

Regression approaches to population estimation are basically “stock” methods in which measures of change in the ratios of indicators to population are used as “flow” estimates that are extrapolated to generate population estimates (Bryan, 2004b). The flow estimates serve as independent variables in these forms, which result in a dependent variable that represents 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

4. Component Methods

Component methods are directly based on the fundamental demographic identity 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.

5. 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 again bring up the MAF, which represents a national housing registration system that can be used to generate estimates using the Housing Unit Method.

6. Other Methods

Here, we 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 found in the “statistical tradition” (see, e.g., Platek, Rao, Sarndal, and Singh, 1987). Others include those developed for statistically underdeveloped countries (see, e.g., Popoff and Judson, 2004) and those for estimating wildlife populations (see, e.g., Williams, Nichols and Conroy, 2002) as well as the imputation and other methods used to compensate for missing data (see, e.g., Judson and Popoff, 2004; and Longford, 2005). Finally, there are “agent

based models,” which generally come under the rubric of “microsimulation methods” (see, e.g., Statistics Canada, 2009). “Microsimulation” is relatively new to most demographers, but it represents an approach that we believe shows great potential and we return to it later in the paper.

In concluding this brief overview of the methods of population estimation, we 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 (See, e.g., Cork and Voss, 2006; Smith, Tayman, and Swanson, 2001).

III. The Needs of Users

Virtually all users desire accurate, timely and accessible data, with cost-effectiveness often ,but not always, being an issue (Swanson, Burch, and Tedrow, 1996). Many tend to use aggregated data (Clark, 1986; Coale and Demeny, 1966; Dharmalingam, 2004; Li and Tuljapurkar, 2005; Pollard, 1973; Rogers, 1995; Rogers, Hummer, and Nam, 2000; Stockwell, Goza, and Balistreri, 2005; Suchindran, 2004; Treyz, Rickman, Hunt, and Greenwood, 1993), However, some users, particularly academic researchers, would prefer to use micro data. This is because many of these basic researchers are interested in hypotheses concerning individuals (Brandon and Hogan, 2004; Livingston, 2006; Mutchler and Baker, 2004; Ryan, Manlove, and Hofferth, 2006) and in using aggregated data to addresses their hypotheses about individuals,

they have to deal with problems such as aggregation bias and the ecological fallacy (Freedman, 2004; and King, Rosen, and Tanner, 2005). Because micro level data can be aggregated and aggregated data are not generally amenable to being dis-aggregated, what we believe is needed by all users is a data system that provides current and historical sets of sub-county estimates of populations and their characteristics that can be rolled up to all higher administrative and statistical geographies for a given vintage to produce a “one number” hierarchy. It should be consistent not only with data both from decennial census counts and sample surveys done by the Census Bureau, but also with the principles underlying the Bureau’s estimates program (U.S. Census Bureau, n. d.). Further, the ideal foundation of these estimates would, we believe, be comprised of individual data on persons that are linked to households and other living arrangements in specific locations. What we have just described, of course, is something that does not exist for the United States – a national population register, a system that contains micro level data that can be rolled up and linked both across time and with other data, such as the case found in Finland (Statistics Finland, 2004).

We do not believe that there are many who would argue against the utility of a national population file. We believe that this observation applies not only to researchers, but also to users in general. The issue here, of course, is that “utility” is not the over-riding factor. American traditions and values are not in favor of such a system, given concerns about government intrusion into privacy (El-Badry and Swanson, 2007; Seltzer and Anderson, 2000; Siefert and Reylea, 2004). So, why have we bothered to discuss this ideal but unachievable data source? The reason is that the MAF is a file that could, with some enhancements, yield such information when coupled with the Bureau’s record matching, extant data collection, and other capabilities. It is to this subject we now turn.

IV. EMAF: A Suggestion for the Production of Population Estimates

We believe that EMAF would contribute toward having not only population estimates that are timely, comprehensive, and internally consistent, but also estimates of housing, as well as demographic and socio-economic characteristics for the U. S. as a whole and its sub-areas. However, before we offer our suggestion regarding the enhancement of the MAF and its potential for meeting the needs of researchers and other users, it is important to acknowledge that others have thought along similar lines. Here, we are thinking primarily of research into the development of an “administrative records census,” which has been going on (and off) for at least 20 years (Alvey and Scheuren, 1982; Kliss and Alvey, 1984, Scheuren, 1999). Initially, much of this work was done within the U.S. Internal Revenue Service, but this broadened to include other agencies, including the Census Bureau (Prevost, 1996, 1999; Prevost and Leggieri, 1999; Judson, 2000, 2003; Judson and Bauder, 2002). Research and other activities in the U. S. related to administrative records censuses have also been commented on by researchers outside of the country (Redfern, 1986). However, it is still the case that the U.S. Census Bureau had not attempted to conduct a full-blown administrative records census (Bryan 2004a, Bryan 2004b, Bryan and Heuser 2004).

We also again acknowledge that our suggestion is largely based on the call by Wang (1999) for greater recognition of the utility of the MAF in regard to population estimates. Wang provided specific suggestions on how to overcome the problems associated with maintaining and updating the MAF such that the data were of high quality. Wang’s (1999) suggestions, along with the ideas underlying an administrative records census provided by Judson (2003), lead

directly to the idea of viewing the MAF as the basis for developing the EMAF, which is a housing unit register with population information. Exhibit 1 provides an overview of how EMAF might be developed and maintained.

(EXHIBIT 1 ABOUT HERE)

As can be seen at the lower far left of Exhibit 1, the MAF/TIGER file is an input into EMAF that goes through a geocoding process. Other inputs into the Geocoding process include processed (“Address Processing” in Exhibit 1), as well as edited, and unduplicated addresses (“Editing and Unduplication” in Exhibit 1) that originate from the following sources: IRS individual Master 1040 File (“IRS IMF” in Exhibit 1); IRS Information Returns Master File (“IRS IRMF” in Exhibit 1); Medicare enrollment database (“Medicare” in Exhibit 1); Selective Service File (“Selective Service” in Exhibit 1); Tenant Rental Assistance file from the Department of Housing and Urban Development (“HUD TRACS” in Exhibit 1); Indian Health Service patient file (“Indian Health Service” in Exhibit 1); and HUD’s Tenant Rental Assistance Certification System (“HUD MTCS” in Exhibit 1). These same files also feed “Person Processing,” where after being processed (“Person Processing” in Exhibit 1) they are fed into “SSN Validation” as shown in Exhibit 1 and matched with the Census Bureau’s extract (“Census NUMIDENT” in Exhibit 1) from the Social Security Administration’s “Numerical Identification System” file (“Social Security NUMIDENT” in Exhibit 1), which contains the name of the applicant, place and date of birth, and other information since the first social security cards were issued in 1936. The valid “Matched Person-Numident” records are then unduplicated (Unduplication) and, as indicated at the lower center of Exhibit 1, merged with the address records and enter EMAF. The records that fail the validation processing of the “Person-Numident” merger, enter into a file that requires further processing (“Invalid SSNs” in Exhibit 1)

with the idea that additional work would yield additional valid data to be merged with the address records so that they could enter EMAF.

The Census Bureau's NUMIDENT file also feeds into a Persons Characteristics File ("PCF" in Exhibit 1) that itself is informed by Census Bureau data sources, including the decennial census, the ACS, and modeling, which taken altogether represent the "Demographic Characteristics Model" and the "Socio-economic Characteristics Model" data files, as shown in Exhibit 1. While the merged "Person-Address-Numident" file would be powerful, it needs information from the PCF so that the potential of EMAF is fully realized. There are significant technical challenges facing not only the development of a functional PCF, but also its merger with the Person-Address-Numident file.

Initial data from the "Demographic Characteristics Model" could be provided directly by census 2000 short form data while the "Socio-economic Characteristics Model" data could be provided by a combination of census 2000 long form data and imputation/modeling/ methods so that they are characteristics assigned to the short form records. In turn, they would be informed by the Census Numident Records, which would result in the PCF. From the PCF they would, in turn, inform the "Person-Address-Numident" so that individual and household/group quarters characteristics be assigned to individual addresses in the MAF. Once this initial EMAF is constructed, it can be brought forward in time on a regular basis (e. g, once each year) using the processes identified in Exhibit 1. Here, it is useful to think about the possibility of using microsimulation methods (see, e.g., Statistics Canada, 2009) as the means to accomplish bringing the EMAF forward in time. The microsimulation system would yield aggregated data that could be calibrated against aggregated ACS and other empirical data that are regularly collected by the Census Bureau. This means that the parameters being used in the

microsimulation would be adjusted until data from the EMAF matched (with given tolerance levels) the empirical data. The re-calibration could include direct substitution in EMAF addresses appearing in the ACS sample for a given vintage (i.e., a given year), and imputation, simulation, and related estimation methods for those EMAF addresses in the same vintage and area that are not in the ACS. Data for addresses in the “old” EMAF version could be so identified and remain attached to each record so that measures of change could be computed for individual address and person records. Thus, EMAF would be an address register containing a combination of collected and estimated data centered on demographic characteristics (i.e., age, sex, race, household relationships) distinguished, as appropriate, by year. When a year ending in zero is reached, EMAF would be updated (and calibrated) using data from the decennial census.

In concluding this section, we emphasize that we are providing a conceptual roadmap rather than a work plan in terms of constructing EMAF. The files and processes identified in Exhibit 1, for example, are likely to look different than those identified by the Census Bureau if it embarks on the construction of EMAF and develops a full scale work plan for this task.

V. Potential Benefits

What are some of the specific benefits of EMAF? Here are some examples. To begin, we believe it would assist the Census Bureau in solving four of the problems facing its estimates program identified by Habermann (2006). First, “short form” data from EMAF would serve well as the population controls for the ACS. This could be particularly important for small pieces of geography. Second, the combination of short and long form data in EMAF could serve to improve estimates of internal migration as well as emigration and immigration. Third, EMAF

could serve as a platform onto which bringing additional data sources could be brought into the sub-national population estimates beyond the ACS. These data sources could include, for example, administrative data sources on employment and taxes in a manner similar to what is done by Statistics Finland (2004). And, fourth, EMAF would allow for research needed to improve methods to achieve integrated and consistent population estimates at different levels of geography. In this regard, Habermann (2006) observes that the current approach begins at the county level, with the estimates controlled only at the national level.

Although the Census Bureau recently benefited from increased funding from the Economic Stimulus Package, its history is one of under-funding (Lowenthal, 2009). For example, The U.S. Census Bureau was confronted with a shortfall of more than \$50 million in the budget proposed by the Executive Branch for its FY 2007 operations (Lowenthal, 2006). This is not a new phenomenon and much of the impetus for reduced and otherwise tight budgets comes from the high costs of collecting data. In this regard, we believe that EMAF would also be of benefit. For example, Statistics Finland (2004: 26) reports that it was pressured by the Ministry of Finance to move to a register-based system because of the recurring high costs associated with taking a census. After it made the change following its 1980 census, Statistics Finland (2004: 26) reports that in terms of 2003 euros, terms the cost of its 2000 register-based census was less than one million euros while the traditional 1980 census costs were approximately 35 million euros. This evidence strongly suggests that EMAF would assist the U.S. Census Bureau in containing costs.

We believe that EMAF would not only reduce costs in the long run, but also contribute toward having more timely, comprehensive, and internally consistent demographic, housing, and socio-economic data for the U. S. as a whole and its sub-areas. In regard to geography, we note that register-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 EMAF can be overlaid with other features using GIS capabilities. The TIGER street address file comes immediately to mind in this regard. This 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 and more precise than those allowed by the zip code tabulation area. This would allow much higher precision in defining areas for purposes of marketing, site location. 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 (e.g., school attendance zones).

It is also worthwhile to note that if geo-coded group quarters, commercial establishments, and public buildings (e.g., fire stations) were included in the EMAF, the result would be a tremendous data source for applied researchers and users. Imagine being able to map not only existing, but also historical and potential “future” service areas and their populations using such a system. Here, it is useful to note that is precisely the situation that exists currently in Finland (Statistics Finland, 2004: 41-44). We also note that this proposal also is in line with recommendations made by the National Research Council’s Committee on the Human Dimensions of Global Change (National Research Council, 2005a).

We also note that another benefit of EMAF is that it would negate the need for many of the traditional demographic methods for developing population estimates. The demographic methods largely use aggregate data and include the Housing Unit Method, regression methods, and component methods. Depending on how it is configured, EMAF might also reduce the need for at least some of the sample surveys being done (e.g., the CPS, SIPP). As can be implied from the discussion of how EMAF might be developed, there would likely be a need for accurate,

efficient, and cost-effective record matching methods, as well as imputation and microsimulation methods. Of course, in addition to the benefit of reducing the number of methods needed to produce population estimates, there is the cost of migrating to new methods. These costs include acquiring new equipment, building new data files, creating new administrative, regulatory, and legal arrangements, and developing and extending new forms of technical expertise.

To summarize, we picture EMAF as an integrated file that contains not only existing MAF variables (e.g., geocode, address, and structure type), but also information on the occupancy status of housing units and the people within these units and non-household living arrangements (group quarters). Occupancy status and the demographic and socio-economic characteristics would be generated using a combination of decennial census and ACS and administrative records data largely in conjunction with a combination of record matching, imputation and microsimulation methods.

VI. Obstacles and How They Might Be Overcome

Turning now to the obstacles associated with our EMAF proposal, we begin with the issue of confidentiality. The National Research Council's Panel on Data Access for Research Purposes (2005b) has identified the lack of resources and structural incentives for making data more readily available as major contributors to the difficulty of reconciling access to data with the need to preserve confidentiality.⁶ The issue of confidentiality is not an insignificant problem. As the U.S. Census Bureau recently learned, even the perception of a breach of confidentiality can become a major outcry (Clemetson 2004a, 2004b, 2004c; Lipton, 2004). One can see that the development by the U.S. Census Bureau of any type of file containing information on individuals

can run into public and political resistance due to confidentiality concerns. This was noted over twenty years ago by Pittenger (1982). However, we believe that this problem is not insurmountable in regard to our proposal. The National Research Council (2005b) has issued recommendations to reconcile access and confidentiality and the U.S. Census Bureau itself has appointed a Chief Privacy Officer and worked to put effective procedures in place regarding this reconciliation. There are recommendations for going even further (El-Badry and Swanson, 2007) as well as the ideas provided by the highly effective laws, rules, and procedures, developed by Statistics Finland (2004) to effect the reconciliation of access to data and the preservation of confidentiality.⁷ Taken altogether, we believe that the U.S. Census Bureau is capable of creating an EMAF that would be useful to researchers (and ultimately other users) while also being subject to strong confidentiality safeguards.

What about the issue of privacy? What may be ideal from a researcher's point of view may not be ideal from the perspective of others. For example, those concerned about the intrusion of the Federal Government into private lives would not be pleased at the prospect of what amounts to a national individual data base even no major outcry has been raised in regard to the three "lightly" regulated, non-mandated, de facto private sector registration systems maintained by Equifax, Experian, and TransUnion for purposes of determining credit worthiness. We believe that this may be a more difficult obstacle for the U.S. Census Bureau to overcome than that represented by concerns over confidentiality. Much of this has to do with privacy being intertwined with the mix of constitutional mandate, case law, executive orders, and general tradition that calls for an actual count of the population rather than the development of a database such as EMAF (Anderson, 1988; U.S. GAO, 2003; Walashek and Swanson, 2006; Weinjert,

2003). Thus, the U.S. Census Bureau and its allies would have to mount a dedicated effort to build public and institutional trust in order to have EMAF.

Another obstacle is the financial cost of development. An idea of these costs is given by Redfern (1986) in his discussion of the cost of converting from a traditional census to an administrative records census. However, once developed (or converted, as the case may be), it appears that the costs for a national housing register could be less than the system currently being used in the U.S. for developing post-censal estimates and decennial census counts. We use here the information from Statistics Finland (2004: 26) discussed earlier in regard to the comparative costs of registries and censuses. It also is worth noting here that local officials in Finland update the country's population and housing registries (Statistics Finland, 2004: 21). Thus, we see no major cost obstacle in following Wang's (1999) suggestion that state and local governments be funded to assist in maintaining EMAF under the general supervision of the Census Bureau. Before such a major step is taken, however, it would be wise to research the various forms this could take. El-Badry and Swanson (2007) call for research on such a recommendation in terms of public involvement in administrative oversight of the Census Bureau.

What about accuracy? Could EMAF provide accurate data? In a recent report, the Government Accounting Office (U.S. GAO, 2006) 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. These same two problems represent sources of error in the proposed housing register. Consequently, if the U.S. Census Bureau solves these problems in regard to the 2010 census, it will essentially do so in regard to EMAF.

There are problems already known in regard to using the housing unit method of population estimation that would affect the MAF and therefore the accuracy of the proposed EMAF. They include tracking new housing units, converted housing units, and deleted housing units. Many of these are known to the U.S. Census Bureau staff already dealing with MAF updates (Perrone, 2008; Reese, 2006; U.S. Census Bureau, 2004a, 2004b, 2007). One problem worth mentioning here involves seasonal populations and seasonal housing. In areas with substantial seasonal changes in population, great care must be taken to get an estimate of the de jure population. Since the implementation of the ACS, this problem will be compounded. This is because of differences between the ACS and the decennial census in regard to what constitutes the de jure population (CACPA/PAA, 2005; Cork and Voss, 2006: 254-266). As such, an accurate EMAF 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 (Cork and Voss, 2006: 254-266).

A second issue has to do with the quality of the U.S. Postal Service's delivery data for purposes of updating the MAF, particularly for rural area. The Census Bureau has been studying this issue with an eye toward improving the quality of the MAF (Liu, 2008; Perrone, 2008; Reese, 2006; U.S. Census Bureau, 2004a, 2004b, 2007). As it gains more understanding of these issues and resolves the problems in regard to the MAF, the EMAF, of course, benefits.

A third issue regarding accuracy is accounting for the populations that do not have a standard address, such as the institutionalized and homeless or transient populations (Cork and Voss, 2006: 146-151). It is true that these types of groups would be missed in any estimate using the MAF and separate methods and practices need to be developed to accurately estimate these populations. However, it is this same population that the decennial census itself has problems with (Cork and Voss, 2006: 146-151). Fortunately, evidence suggests that the size of this

population is small relative to the total population living either in households. Cork and Voss (2006: 225) report that in 1990 and 2000 only about 3 percent of the U.S. population resided in group quarters and that the number of homeless on a given day is on the order of 840,000 (Cork and Voss, 2006: 146).

Judson, Popoff, and Batutis (2001) have pointed out that there is a great deal of evidence to support the idea that administrative records systems have systematic biases and they found support for this in an empirical study they conducted. This means that the MAF and, hence, the proposed EMAF will be subject to systematic biases. Fortunately, however, Judson, Popoff, and Batutis (2001) also use their findings to make several recommendations regarding the reduction of these biases. Considering their research in conjunction with the experience being gained by U.S. Census Bureau in regard to the MAF/TIGER system, we believe that the accuracy of an EMAF would be sufficient for purposes of resource allocation, research, and planning.

Another obstacle is the need to have a set of unified identification codes in order to match and merge records from different systems using electronic processing. As noted by Statistics Finland (2004), if there is no unified system of identification codes then it is extremely difficult and laborious, if not impossible, to link records across different systems. In particular, a unique code will be needed for every dwelling in the register, including those in multi-unit structures. In this regard, we point out that Finland has developed such a coding system and that it includes all structures – commercial, residential, and seasonal (Statistics Finland, 2004: 58-60).

VII. Is EMAF Feasible?

With the exception of the issues of confidentiality and privacy, all of the challenges facing the development of a national housing register are in the form of costs, technical problems, or a combination of both. We agree with Wang (1999) that the major technical tasks of developing a “National Address and Housing Inventory” come down to two areas - Address data collection and MAF/TIGER update. We also agree with Wang (1999) that a feasible way to effect a solution to these problems is to enhance the federal-state-local cooperative programs already part of U.S. Census Bureau activities such that local entities are compensated for helping to maintain the system. This is how Statistics Finland (2004) maintains its register system and there are data collection activities in the U. S. that already follow this model (Wang, 1999).

EMAF goes beyond what was envisioned by Wang, who viewed it largely as a basis for doing population estimates using the Housing Unit Method. As such, we believe that his suggestions are necessary but not sufficient for this purpose. There are many political, administrative, and technical obstacles that would need to be overcome. How exactly would researcher access be reconciled with confidentiality and privacy? What would EMAF cost to build and maintain and what savings elsewhere would be gained, if any? How would ACS data be combined with individual housing units – are they sufficient to provide the household level estimates that we are proposing (e.g., age, race, sex, household relationships, household size, vacancy rates, and socio-economic characteristics) or would that stretch imputation, microsimulation, and related modeling techniques, as well as other capabilities too far?⁸ We have sketched an outline for answering these questions, but ultimately only further thought informed by empirical studies is likely to fully answer them. The question that the U.S. Census

Bureau needs to answer at this point is if it appears our recommendation is sufficiently interesting to considering giving it the “thought” test before considering any small empirical studies (e.g., studies similar to the Administrative Records Census Experiment reported by Judson and Bauder, 2002) before proceeding further. In regard to such a “thought” test, we offer a quote from Wang’s (1999: 15) paper on developing the MAF into a resource for making post-censal population estimates:

“Is the development of the National Accounting of Addresses and Housing Inventory feasible? The ideas presented in the paper may cause many people to say that it is impossible because there are so many problems. This is exactly the same reaction we saw in the late 80s when the Census Bureau was developing the TIGER to digitize the nation’s geography from coast to coast. Now we can see how useful and powerful the TIGER is today.”

In closing, we would like to believe that if Ching-Li Wang were still alive, he would be willing to make a similar statement on behalf of the proposed EMAF. We also believe that the idea of EMAF, the Enhanced Master Address File, could be of interest to other countries with MAF files and strong administrative records systems that, like the United States, are facing the challenge of producing good population information in the face of increasing census costs.

Endnotes

1. One can also construct estimates for a point in time that predates a census. We have not run across the term “pre-censal,” however, and so do not use it 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, we do not cover the estimation methods associated with “statistically underdeveloped areas” and wildlife populations
2. The MAF is already being used for “direct estimation” because it forms the sample frame for the Census Bureau’s “American Community Survey.” Liu (2007) discusses the Census Bureau’s evaluation work that is being used to support the goal of using a MAF-based frame to replace the current multiple frames for the 2010 Demographic Survey Redesign.
3. 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). We also note that the “composite” method (Bryan, 2004b: 550-551) is a type of synthetic estimation.
4. 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.
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 # 8). 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. Confidentiality is the idea that there should be restrictions on how information is collected and used and that no data should be disclosed about a respondent that would allow him or her to be either identified or harmed; privacy is the idea that it is the right of an individual to decide

whether and to what extent he or she will divulge thoughts, opinions, feelings, and facts to the government (Mayer, 2002).

7. Statistics Finland (2004) has a measure of oversight over its data users while the U.S. Census Bureau assumes no responsibility for what users do with its data. El-Badry and Swanson (2007) argue that the U.S. Census Bureau's stance serves to decrease public trust in the Census Bureau. This is not a trivial issue because public trust has been identified as a major contributing factor to conflict over census results (El-Badry and Swanson 2007; Walashek and Swanson, 2006), an activity that requires the consumption of Bureau resources.

8. In regard to the capabilities of imputation and modeling, Swanson and Knight (1998) developed four model-based procedures for estimating household income using SIPP data statistically matched to Metromail's proprietary database. The procedures were developed with a random sample ($n=6,559$) from the data base and tested with the remaining "out of sample" portion of it ($n= 7,048$). The results were found to be sufficiently accurate and the procedures sufficiently tractable for use by the client. Given this personal experience, it is difficult for us to believe that the U.S. Census Bureau is not technically capable of developing accurate and tractable procedures for purposes of developing the demographic and socio-economic information we propose for the national housing register. we also note here that subsequent to the project reported by Swanson and Knight (1998), Metromail was acquired by Experian, a subsidiary of GUS, which holds numerous databases containing public and proprietary information on consumers and also engages in direct mailing lists and other forms of marketing (The Motley Fool, 2000).

References

- Alvey, W. and Scheuren, F. (1982). "Background for an Administrative Records Census." pp. 47-65 in *Statistics of Income and Related Administrative Record Research*. Washington, DC: U.S. Department of the Treasury, Internal Revenue Service.
- Anderson, M. (1988). *The American Census: A Social History*. New Haven, CT: Yale University Press.
- Bousfield, M. (2002). "Population Estimation for Census Tracts Using Dynamic Models." Paper presented at the Annual Meeting of the Population Association of America, Atlanta, GA.
- Brandon, P. and D. Hogan. (2004). "Impediments to Mothers Leaving Welfare: The Role of Maternal and Child Disability." *Population Research and Policy Review* 23 (4): 419-436.
- 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.
- CACPA/PAA. (2005). "Recommendation 10c" in *Recommendations from the Population Association of America Advisory Committee*. Meeting of the Census Advisory Committee of Professional Associations. Arlington, VA: October 21st - 22nd.
- Clark, W. A. V. (1986). *Human Migration*. Volume 7, Scientific Geography Series. Beverly Hills, CA: Sage Publications.
- Clemetson, L. (2004a). "Homeland Security Given Data on Arab-Americans: Census Bureau Complies with Request." *New York Times*, 30 July. Late Edition – Final: A14
- Clemetson, L. (2004b). "Coalition Seeks Action on Shared Data on Arab-Americans." *New York Times*, 13 August: Late Edition –Final: A12.
- Clemetson, L. (2004c). "Census Policy on Providing Sensitive Data is Revised." *New York Times*, 31 August: Late Edition – Final: A14

Coale, A. and P. Demeny. (1966). *Regional Model Life Tables and Stable Populations*. 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. and P. Voss (eds.). (2006). *Once, Only Once, and in the Right Place: Residency Rules in the Decennial Census*. National Research Council. Washington D.C.: National Academies Press.

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).

Dharmalingam, A. (2004). "Reproductivity." pp. 429-453 in J. Siegel and D. Swanson (eds.) *The Methods and Materials of Demography, 2nd Edition*. New York, NY: Elsevier Academic Press.

El-Badry, S. and D. Swanson. (2007). "Providing Special Census Tabulations to Government Security Agencies in the United States: The Case of Arab-Americans." *Government Information Quarterly* 24(2): 470-487.

Fay, R. (2005). "Model-Assisted Estimation for the American Community Survey." *Proceedings of the Joint Statistical Meetings*: 3016-3023. Alexandria, VA: American Statistical Association.

Freedman, D., (2004). "The Ecological Fallacy." p. 293 in M. Lewis-Beck, A. Bryman, and T. Liao (Eds.) *The Encyclopedia of Social Science Research Methods*. Beverly Hills, CA: Sage Publications.

Ghosh, M and Rao, J.N.K. (1994). "Small Area Estimation: An Appraisal." *Statistical Science* 9:55-93.

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

Hakanson, A. (2007). "Address Coverage Improvement and Evaluation Program – 2006 National Estimate of Coverage of the Master Address File." Decennial Statistical Studies Division, U.S. Census Bureau.

Happel, Stephen K. and Timothy D. Hogan. (2002). "Counting Snowbirds: The Importance of and the Problems with Estimating Seasonal Populations," *Population Research and Policy Review* 21: 227-240.

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. (2003). "The Statistical Administrative Records System and Administrative Records Experiment 2000: System Design, Successes, and Challenges." Invited Presentation Given at the University of Maryland, October 8.

Judson, D. H. and M. Bauder. (2002). "Evaluating the Ability of Administrative Records Databases to Replicate Census 2000 Results at the Household Level." Paper presented at the Annual Meeting of the American Statistical Association, New York, NY, August 11th - 15th.

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.

Judson, D., C. Popoff, and M. Batutis. (2001). "An Evaluation of the Accuracy of U.S. Census Bureau County Population Estimation Methods." *Statistics in Transition* 5: 185-215.

King, G., O. Rosen, and M. Tanner (eds.) (2004). *Ecological Inference*. Cambridge, England: Cambridge University Press.

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.

Kliss, B. and W. Alvey (eds). (1984). *Statistical Uses of Administrative Records: Recent Research and Present Prospects, Volumes I and II*. Washington, D.C.: Department of the Treasury, Internal Revenue Division, Statistics of Income Division.

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

Lee, E., and H. Goldsmith (eds.). (1982). *Population Estimates: Methods for Small Area Analysis*. Beverly, Hills, CA: Sage Publications.

Li, N. and S. Tuljapurkar. (2005). "A Formal Model of Age-Structural Transitions." pp. 91 – 105 in S. Tuljapurkar, I. Pool and V. Prachuabmoh (eds.) *Population Resources and Development: Riding the Age Waves- Volume I*. Dordrecht, The Netherlands: Springer.

Lipton, E. (2004). "Panel Says Census Move on Arab-Americans Recalls World War II Internments." *New York Times*, 10 November.

Liu, X. (2007). "Comparing the Quality of the Master Address File and the Current Demographic Household Surveys' Multiple Frames." Paper presented at the 2007 Research Conference of the Federal Committee on Statistical Methodology, November 7-10, Arlington, VA.

Liu, X. (2008). "Using a MAF-Based Frame for Demographic Household Surveys." *Proceedings of the Government Statistics Section, American Statistical Association*: 2864-2871.

Livingston, G. (2006). "Gender, Job Searching, and Employment Outcomes Among Mexican Immigrants." *Population Research and Policy Review* 25 (1): 43-66.

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

Longford, N. (2005). *Missing Data and Small-Area Estimation: Modern Analytical Equipment for the Survey Statistician*. Dordrecht, The Netherlands: Springer.

Lowenthal, T. (2009). "House Subcommittee Approves 2010 Census Funding." *Census News Brief*, 4 June.

Lowenthal, T. (2006). "House Cuts \$58.3M from Census Budget; Senate Panel Approved \$50M less than Bush Request." *Census News Brief*, 11 July.

Mayer, T. (2002). *Privacy and Confidentiality Research and the U. S. Census Bureau: Recommendations Based on a Review of the Literature*. *Research Report Series*. Survey Methodology Report #2002-01. Statistical Research Division, U.S. Census Bureau. Washington, D.C. U.S. Census Bureau (available online at <http://www.census.gov/srd/www/byyear.html>, last accessed March 2009).

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.

Mutchler, J. and L. Baker. (2004). "A Demographic Examination of Grandparent Caregivers in the Census 2000 Supplemental Survey." *Population Research and Policy Review* 23 (4): 359-377.

National Research Council. (2007). *Using the American Community Survey: Benefits and Challenges*. Washington, D.C.: National Academies Press.

National Research Council. (2005a). *Population, Land Use, and Environment: Research Directions*. Washington, D.C.: National Academies Press.

National Research Council. (2005b). *Expanding Access to Research Data: Reconciling Risks and Opportunities*. Washington, D.C.: 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. (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.

Pollard, J. (1973). *Mathematical Models for the Growth of Human Populations*. Cambridge, England: Cambridge University 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. (1999). "Design Alternatives for Building Block Estimates." Paper presented the Estimates Methods Conference, U. S. Bureau of the Census, Suitland, MD. (available online at <http://www.census.gov/population/www/coop/popconf/paper.html>, last accessed June 2009).

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 at <http://www.fcsm.gov/99papers/prevost.pdf>, last accessed, May 2009).

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.

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

Rao, J. (2003). *Small Area Estimation*. San Francisco, CA: Jossey-Bass.

Redfern, P. (1986). "Which Countries Will Follow the Scandinavian Lead in Taking a Register-based Census of Population?" *Journal of Official Statistics* 2 (4): 415-424.

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.

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.

Rogers, A. (1995). *Introduction to Multiregional Mathematical Demography*. New York, NY: John Wiley & Sons.

Rogers, R., R. Hummer, and C. Nam. (2000). *Living and Dying in the USA: Behavioral, Health, and Social Differentials of Adult Mortality*. New York, NY: Academic Press.

Ryan, S., J. Manlove, and S. Hofferth. (2006). "State-level Welfare Policies and Nonmarital Subsequent Childbearing." *Population Research and Policy Review* 25 (1): 103-126.

Scheuren, F. (1999). "Administrative Records and Census Taking." *Survey Methodology* 25(2): 151-160.

Schmitt, R. 1975. "Defacto Population Estimates and the Demography of Nonresident Population. *Asian and Pacific Census Newsletter* 2(2): 5-8.

Seltzer, W., & Anderson, M. (2000). "After Pearl Harbor: The proper role of population data systems in time of war." *Statisticians in history paper series*. American Statistical Association. (Available online at <http://www.amstat.org/about/statisticians/index.cfm?fuseaction=papers>, last accessed March, 2009).

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.

Siefert, J., and M. Reylea. (2004). Do You Know Where Your Information is in the Homeland Security Era? *Government Information Quarterly* 21(4): 399–405.

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

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

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

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

Statistics Canada. (2009). “Post Conference MODGEN Workshop.” 2nd General Conference of the Microsimulation Association. June 11th, Ottawa, Ontario, Canada (see <http://www.statcan.gc.ca/conferences/ima-aim2009/modgen-eng.htm>, last accessed June, 2009).

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.

Stockwell, E., F. Goza, and K. Balistreri. (2005). “Infant Mortality and Socioeconomic Status: New Bottle, Same Old Wine.” *Population Research and Policy Review* 24 (4): 387-399.

Subcommittee on Small Area Estimation. (1993). *Statistical Policy Working Paper 21 - Indirect Estimators in Federal Programs*. Federal Committee on Statistical Methodology, Statistical Policy Office, Office of Information and Regulatory Affairs, U.S. Office of Management and Budget.

Suchindran, C. (2004). “Part II: Model Life Tables.” pp. 662-675 in J. Siegel and D. Swanson (eds.) *The Methods and Materials of Demography, 2nd Edition*. New York, NY: Elsevier Academic Press.

Swanson, D. (2009). “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).” Unpublished Report prepared for the U.S. Census Bureau (Contract Order Number: YA132307SE0374).

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 M. Knight. (1998). *Metromail Wealth Estimation Project Final Report: Recommendations, Summary Findings, and Technical Documentation*. Madison, WI: Third Wave Research Group.

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 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., 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.

Swanson, D., T. Burch, and L. Tedrow. (1996). "What Is Applied Demography?" *Population Research and Policy Review* 15 (5-6): 403-418.

The Motley Fool. (2000). "Extracting Value from Experian" (By Maynard Paton). 16 March 2000 (Available online at www.fool.co.uk/aualiport/2000/qualiport000315.htm, last accessed May, 2009).

Treyz, G., D. Rickman, G. Hunt, and M. Greenwood. (1993). "The Dynamics of U.S. Internal Migration." *Review of Economics and Statistics* 75: 209-214.

U.S. Census Bureau. (2007). Housing Unit Based Estimates Research Project Draft Research Agenda. Unpublished Document. Population Division, U.S. Census Bureau.

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 at http://www.census.gov/geo/mod/maf_basics.pdf, last accessed January, 2008).

U.S. Census Bureau. (2004b). "Address List Development in Census 2000." Census Topic Report No. 8. (TR-8). U.S. Bureau of the Census (available online at <http://www.census.gov/pred/www/rpts/TR-8.pdf>, last accessed June, 2009).

U. S. Census Bureau. (No Date). "The U.S. Census Bureau's Intercensal Population Estimates and Projections Program: Basic Underlying Principles." Unpublished Document. U.S. Census Bureau.

U.S. GAO. (2003). *2000 Census: Coverage Measurement Programs' Results, Costs, and Lessons Learned*. GAO -03-287. Washington, DC: U. S. General Accounting Office (Note: Effective July 7, 2004, the GAO's legal name became the Government Accountability Office).

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, D.C. U. S. Government Accountability Office (Note: Effective July 7, 2004, the GAO's legal name became the Government Accountability Office).

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

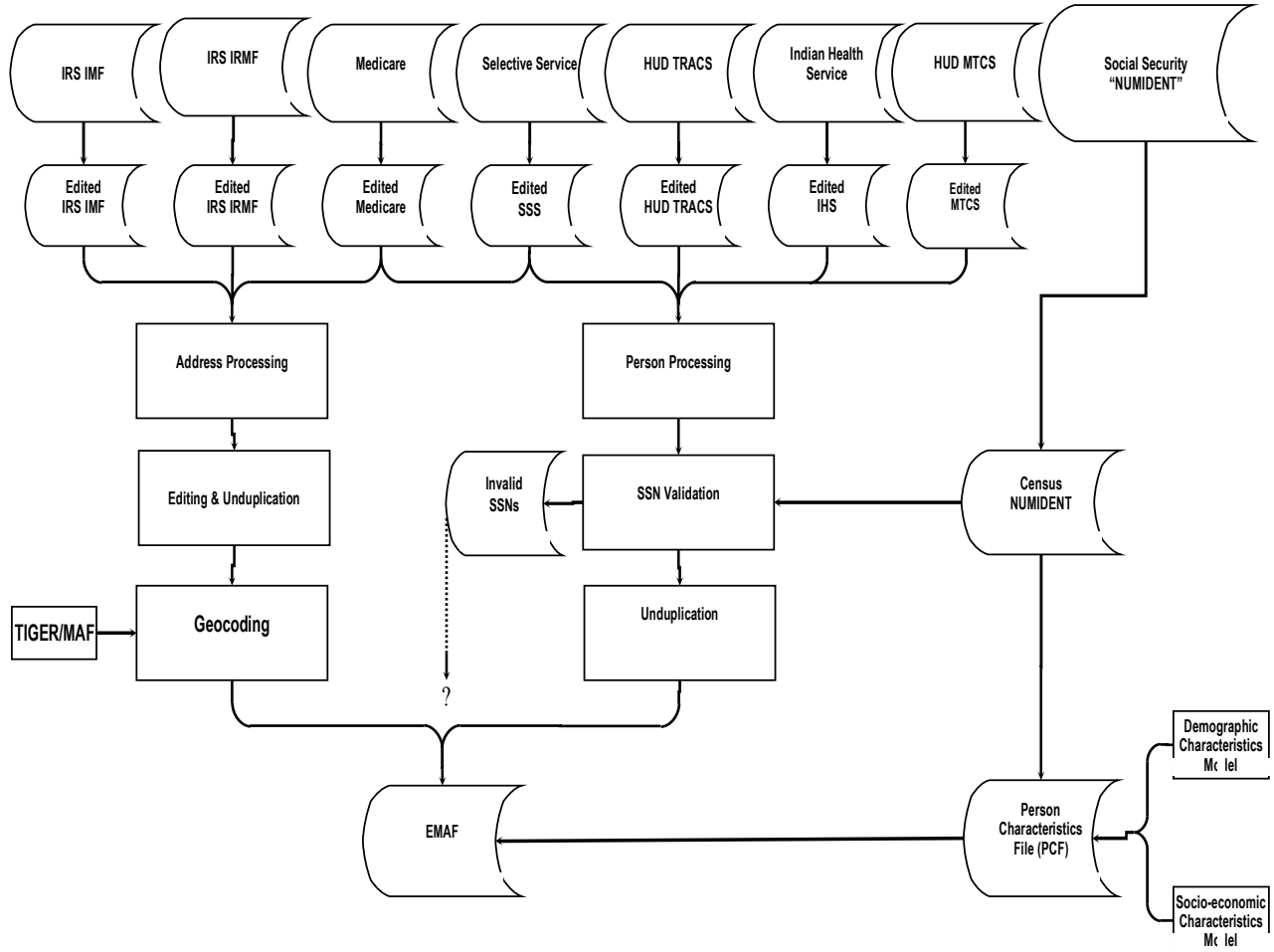
Waldrop, J. (1995). "Preface." pp. v-vi in N. Rives, W. Serow, A. Lee, H. Goldsmith, and P. Voss (eds.) *Basic Methods for Preparing Small-Area Population Estimates*. Madison, WI: Applied Population Laboratory, Department of Rural Sociology, University of Wisconsin.

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 July, 2008).

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.

Exhibit 1. SCHEMATIC VIEW OF EMAF*



*Adapted from Judson (2003).