Title: The Official U.S. Census Bureau Population Estimates by Demographic Characteristics: Requirements, Evaluation, and Future Directions.
Author: Matthew Christenson

This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

1. Introduction

The Population Division of the U.S. Census Bureau recently instituted a project to recommend ways to improve the methods for the official sub-national population estimates with age, sex, race, and Hispanic origin detail. This project was necessary in order to take advantage of newly available data, to replace aspects of the current methods developed under the constraints of outdated technology, and to correct some irregularities that have been identified in the resulting estimates data.

In terms of new data, our primary interest is in a modified way of using individual income tax data from the Internal Revenue Service (IRS) to measure migration between states and counties. Up until now, the Census Bureau created internal migration rates by matching the individual records of filers in two consecutive years of IRS migration data. In so doing, the migration activity of the spouse and dependents claimed by the filer were assumed to be the same as the filer. Similarly, the age, sex, race, and Hispanic origin of the spouse and dependents were modeled based on the responses of the filer.[1] This method has been referred to as a "returns based" method. Now, the Census Bureau intends to measure migration activity and demographic characteristics by handling each individual on the file independently. This technique is being referred to as a "person-based" method. The goal of the current project is to determine the best way to incorporate the person-based migration data into the overall estimate system.

The second goal of the project is to replace certain aspects of the current methods developed under the constraints of outdated technology. The great majority of estimates at the U.S. Census Bureau are programmed in SAS. Some of these programs have legacies that go back to the 1980s when the Census Bureau first began developing estimates by characteristics. At this point, the ability to store and process very large datasets (often several gigabytes in size) and to use coding conventions such as multi-dimensional arrays allow for different approaches to estimates than was true even a few years ago. We are now able to spend more time thinking about the method by itself instead of the interaction of the method with the constraints of the computer resources. Thus, another goal of the current project is to determine the best way to approach the estimates given the advances in computer technology.

The final motivation for this project is to see if we can eliminate several anomalies we have identified in the estimates we currently produce. For example, the estimates for the population ages 18-24 in some states and counties appear unreasonable due to a pattern of internal migration for this population that is not reflected well in the administrative datasets we

[1] The demographic characteristics of the people in the IRS file are obtained by matching it to a micro-data file generated from information from the Social Security Admininstration and Census 2000.

currently have available. Similarly, the race and Hispanic origin distributions from our vital statistics data do not seem to match up well with the distributions from Census 2000. Thus, the final goal of the current project is to see if there are ways we can shape our estimates methods in order to take advantage of the strengths of the input data and minimize its weaknesses.

Working from these goals, the intended products of this project are four-fold: 1) Requirements with which to evaluate the suitability of the methods and data for the sub-national population estimates, 2) Competing proposals for how to improve the methods, 3) Test code and provisional estimates based on the proposed changes, and 4) Evaluations of the proposed methods and provisional data.

In an ideal world, the search for new methods with which to produce detailed subnational population estimates would simultaneously address the whole range of geographic, temporal, universe, and characteristic detail that are needed. It addition, it would operationalize multiple, diverse approaches to address the estimation problem. For better or worse, the current project does not have this luxury. Instead, it is constrained by both time and personnel considerations that limit the breadth of work that can be done before implementable recommendations must be made. In fact, the first deliverable that is required of this project is a set of recommendations to be considered for implementation in the vintage 2006 estimates series. Thus, although the long-run goal of this project is to test multiple, diverse methods that address the full scope of subnational estimates detail, a decision needed to be made to limit the scope of the initial research. At the present time, this decision is to first investigate the annual estimates for the state resident population by age and sex.

One of the requirements identified early on in this project was the need for its work and products to be transparently accessible to stakeholders and the demographic community. This paper attempts to further this transparency by summarizing the work on the project to date. First, it outlines the requirements that have been identified as important for the evaluation of the methods and data of the sub-national characteristics population estimates. Next, it uses these requirements to evaluate the current method used to produce state estimates by age and sex. Finally, it summarizes the ideas to date on a replacement for the current state age-sex method and comments on the improvements expected from the new method.

2. Method and Data Requirements for the U.S. Census Bureau Population Estimates

The Population Estimates and Projections area of the Population Division of the U.S. Census Bureau has recently begun formal discussions about what qualities the estimates methods and the resulting population estimates should possess. Although these discussions are not yet complete, it is insightful to look at the positions currently under consideration.

The first position that has been generally agreed upon is that there are no *a priori* preferred methods for producing the estimates. However, any population estimates method used by the Census Bureau must incorporate three sources of input data: The most recent decennial census results, all special censuses results collected since the last decennial census, and all accepted challenges to the estimates since the last decennial census. In addition, any method must produce output that can be used to satisfy the commitments we have made to existing customers (see Appendix A for a listing of the required output for the Vintage 2005 estimates).

Next, there are numerous data sources that are available on a consistent basis that can provide indications of either the current level or recent changes to the size and distribution of the population. All estimates methods need to consider these data either as direct input to the estimates or as a metric of comparison and evaluation. These include:

1. Births and deaths (e.g., as collected from the National Center for Health Statistics and the Federal State Cooperative for Population Estimates).
2. International migration (e.g., from sources such as the Department of Homeland Security and the American Community Survey).
3. Internal migration (e.g., from administrative sources such as income tax data received from the Internal Revenue Service and survey sources such as the American Community Survey).
4. Group Quarters Population (e.g., as collected by the Federal State Cooperative Program for Estimates, other administrative sources).
5. Military population.
6. Medicare enrollment.
7. Public and private school enrollment.
8. Households and housing units.

In addition, the ideal estimates method for the resident population would allow for the introduction of special input data that reflect changes to the level or distribution of a population that is not reflected in nationally available data sources (e.g., estimates of changes to a population as the result of natural disaster).

In terms of the procedures with which the input data is translated into population estimates, the methods will be 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]
2. Integrity: The strategy that consistently applies the declared method is preferred to the one that uses ad-hoc fixes to address particular challenges of the estimation task.
3. Parsimony: A more simple 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 products), thus promoting the integration of the Census Bureau estimates system, is preferred to a more product-specific remedy.
6. Transparency: The strategy that is more readily understandable and replicable by external parties is preferred. Moreover, the 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 us to meet current deadlines.

[2] http://www.project2061.org/publications/sfaa/online/chap2.htm- "Application."

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

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-institutional), 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 census.
    b. Time-Series Change: The reasonableness of the annual change in the estimates since the last census.
    c. Coherence: The input data, the values of the estimates, and the demographic rates that can be derived from them provide a cohesive and believable representation of the population.
    d. Comparability: The estimates appear realistic when compared to other indicators of the size and distribution of the population (Medicare enrollment, school enrollment, housing unit estimates, etc.).

3. Evaluation of the Current State Age-Sex Method

So why is a new method necessary for the official state estimates by age and sex? The current method is a legacy from the 1980s that has remained basically unchanged since then. Sometimes referred to as "Component-Method II," the state age-sex method utilizes different techniques to produce preliminary estimates of the populations 0 to 64 years of age and 65 years of age and older. For the former, a version of the cohort-component technique starts with the most recent decennial census and estimates population change with births and deaths from vital registration and net migration (internal and international together) from school enrollment data. For the population 65 years of age and older, the method starts with the most recent decennial census and estimates population change using change in Medicare enrollment. The resulting estimates for both age groups are then converted into ratios and used to distribute estimates from two other products; national estimates by age and sex, and state estimates for the age groups 0-64 and 65 and older.[3]

In general, the current method meets some of the most basic requirements for an acceptable method as detailed above. With respect to the input data:

- It begins with a Census 2000 base population.

[3] For a detailed description of this method, see the following URL: http://www.census.gov/popest/topics/methodology/State_AgeSex_v2004.html

- It incorporates some of the input data that have been identified as potentially useful for population estimates (base population, births, deaths, school enrollment, Medicare enrollment).
- It reflects the special censuses conducted and the challenges received since Census 2000 through the control to the state totals estimates.[4]

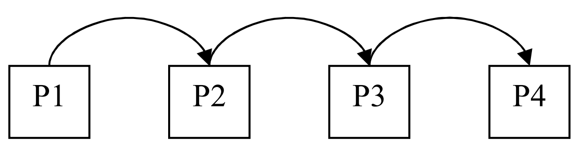
With respect to the procedures it employs, it is usable – i.e., the process for producing estimates is executable with current staffing levels in a way that allows us to meet current deadlines. Finally, in terms of output, the current state age-sex model produces results that are orthodox, consistent, and satisfy the commitments we have made to existing customers.

This being said, there are still numerous places where the estimates method and the resulting data have been criticized. While there is not room in this paper for an exhaustive treatment of the state age-sex method, three issues have surfaced that by themselves are of sufficient concern to warrant the search for substantial improvements to the method. They are:

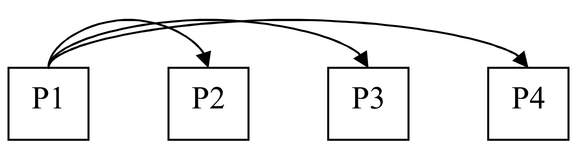
1) The atypical application of the cohort-component technique
2) The use of school enrollment data to estimate migration for the adult population under age 65, and
3) The strong impact of the "rake" to the previously developed estimates on the final values of the state age-sex estimates.

<u>The Application of the Cohort-Component Technique</u>: The first troublesome aspect of the state age-sex method is that the cohort-component method as applied estimates each of the panels (i.e., years) in the time series independently. As a result, the cumulative change to the population over time is not additive.

In a typical application of the cohort component technique, one panel of estimates is produced and then used as the base population for the next panel of estimates. This can be represented graphically in the following manner:



The method employed by the state age-sex method, in contrast, goes back to the census base each time it estimates a panel in the time series. This approach can be represented graphically as follows:



In effect, then, each panel in the time series is independent from all the others except for that it has the same base and shares part of the input data with the other panels.

[4] Challenges and special censuses are direct inputs to the State and County Totals estimates process. The State Age-Sex estimates reflect the challenges and special censuses by being forced to the levels in the state and county totals.

The main source of the non-additivity in the method comes from the way that net-migration is calculated for the population ages 17-65.[5] Using the vintage 2004 estimates as an example, the state age-sex estimates started with a base of April 1, 2000 (the census date). To estimate the first panel in the time series (July 1, 2000), a set of age-specific net-migration rates were created for the period from 4/1/00 to 7/1/00. Then, a proxy for the midperiod population was calculated (i.e., the base population minus ½ of the deaths in the period) and multiplied by the net migration rate to arrive at the estimates of net migration for the period.

In order to estimate the next panel (for July 1, 2001), another set of age-specific net-migration rates was calculated -- this time for the period again starting at the census date (4/1/00) but now going to 7/1/01. Then, a proxy for the population at the middle of this period was estimated and multiplied by the net migration rate to arrive at the estimates of net migration for the period. In like fashion, the net migration for the rest of the panels in the time series was as follows: 4/1/00 to 7/1/02, 4/1/00 to 7/1/03, and 4/1/00 to 7/1/04. Consequently, the net-migration for each panel in the time series was estimated independently from all others except that it has the same base and shares part of the input data with the other panels.

While an argument can be made that these calculations are at least partially sound as defined above (i.e., the formulas are mathematically valid), they suffer from other problems. First, this embodiment of the cohort-component technique invalidates one of the cited strengths of the method – that it parallels actual demographic changes (i.e., changes to the population are additive).[6] In addition, the difficulties associated with keeping track of the cohorts in the calculation of multi-year rates, the corresponding midperiod populations, and the base estimates to which the net-migration must be added makes the technique quite unparsimonious. Finally, both of these things, along with the fact that this method is so different from what is typically envisioned for a cohort-component technique, greatly reduce the transparency of the method. As such, it has been determined that this aspect of the method threatens the overall desirability of the approach.

The use of school enrollment to estimate migration for the adult population: A second troublesome feature of the state age-sex method is that school enrollment data, specifically the enrollment of students between 6 and 14 years of age, are used to estimate net migration for the population ages 0-64. In the illustration above, we saw that the actual way in which these data are used has its own problems, but there are also more general problems with the approach. To begin, members of the Federal State Cooperative Program for Estimates report to us that the coverage and quality of school enrollment data varies widely by state, particularly for students who are enrolled in private institutions or are home schooled. In addition, the use of these data alone to estimate migration neglects the information available from data sources used in the other estimates products (i.e., estimates of international migration from ACS data and internal migration from IRS migration data). Finally, the assumption that change in school enrollment captures the movement of the entire population ages 0 to 64 is likely unsound.

Evidence for this last point can be seen by examining a set of 1990-based state age-sex estimates for April 1, 2000 that were prepared for this project. These estimates were produced by converting the vintage 1999 state age-sex estimates for July 1, 1999 into proportions by age and sex and then performing an iterative two-way distribution of two sets of previously prepared estimates for 4/1/2000: National estimates by age and sex, and state totals estimates. These new

---

[5] The net migration for the population ages 0-16 is calculated in terms of the numeric change in the school enrollment population (grades 1-8) minus the net effect of natural increase and so does not suffer from this problem.
[6] The Methods and Materials of Demography, p 527.

state age-sex estimates for 4/1/2000 were then pitted against comparable data from Census 2000 and summary measures of the "error" in the estimates (the Mean Algebraic Percent Error [MALPE] and Mean Absolute Percent Error [MAPE]) were prepared. These comparisons are presented in Appendix B.

From these data, it can be seen that the average error in the estimates for the population below age 16 is, on average, less than the corresponding error for the population ages 17 to 64. In fact, the average of the MAPEs for the age groups 17 to 64 is 50 percent larger than the average of the MAPEs for the population ages 0 to 16 (4.6 and 3.0, respectively). While the relative size of these errors may not be due exclusively to the use of school enrollment as a measure of net-migration,[7] it certainly is an indication that this is a likely problem.

A further indication of this likelihood can be seen in a comparison of the 1990-based population estimates for 4/1/2000 for Washington D.C. to the data from the 1990 and 2000 censuses. When graphed by age (appendix C), it is striking to note the similarity between the age-structure of the 1990 and 2000 censuses, in particular for the population ages 17 to 19. If the population were closed (i.e., there were no migration into or out of D.C.), the approximate doubling of the proportion of the population between ages 17 and 19 in the 1990 Census would have moved to ages 27 to 29 in Census 2000 due to the aging of the population. Yet, since this approximate doubling of the proportions is again present between the populations 17 to 19 in the Census 2000 data, it is clear evidence of a large, annual influx of people at these ages into D.C., presumably to attend college, participate in military service, or work in the federal government. As such, it should be present in any post-censal estimates for D.C. However, the 1990-based estimates for 4/1/2000 show the increase in the age-proportions peaking at age 29 – a pattern more consistent with the notion of a closed population than with the actual age pattern of migration reflected in the two censuses. Hence, this is fairly direct evidence that the estimates of migration based on school enrollment are inadequate for the adult population.

<u>The Impact of the "Rake" on the State Age-Sex Estimates</u>. The final potentially troublesome feature of the state age-sex estimates is the "rake" that occurs at the end of the process – i.e., the conversion of the estimates into ratios which are then used in an iterative proportional distribution of national estimates by age and sex and state total estimates. Presumably, having a method that goes to the lengths of the current state age-sex method to model changes in the age distribution of the estimates implies that this distribution should be reflected to a large extent in the final estimates. The level of the estimates might change, but the distribution should be preserved.

To calculate the relative impact of the rake on the estimates, we extracted the values for July 1, 2003, and July 1, 2004 from the vintage 2004 estimates both prior to and after the proportional distribution routine. Then, to calculate the impact of the components of change from the age-sex model, we subtracted the unraked population values for July 1, 2003 from the unraked population values for the corresponding cohorts on July 1, 2004. Subsequently, we calculated the impact of the "rake" by subtracting the unraked population values on July 1, 2004 from the corresponding "raked" values. Finally, we calculated the number of times for which the impact of the rake was greater than the impact of the components of change and converted this number into a percentage of the corresponding total cases. The results of this evaluation are presented in Appendix D.

---

[7] Other possible contributors to the error in these estimates are the differences in coverage between the 1990 Census and Census 2000, errors in the reporting of deaths, and distortion in the estimates from the 2-way rake to the National Age-Sex and State Totals estimates.

Clearly, the rake of the state age-sex estimates to the national age-sex and state totals estimates had a disproportional effect on the estimates. Overall, the average absolute change to the estimates indicated by the components of change was 2.2 percent, while the subsequent average absolute change forced on the estimates by the rake was 3.8 percent. In addition, the impact of the rake was larger than the impact of the components of change in over 70 percent of the age-sex cells. When broken down by age, this pattern holds generally as well. The average absolute impact of the components of change on the estimates was larger than the impact of the rake for only 18 age groups. Hence, the rake of the estimates does have a disproportional affect on the overall estimates and, consequently, raises questions about the desirability of this method.

In summary, the atypical application of the cohort-component technique, the impact of the use of school enrollment data to estimate migration for the adult population under age 65, and the strong impact of the final "rake" on the estimates all lead to serious concerns about the desirability of the state age-sex method for the official estimates of the U.S. Census Bureau. These things limit the soundness, integrity, robustness, and transparency of the method and the reasonableness of the resulting estimates. Indeed, the research for a new method, the beginnings of which are described in the remainder of this document, is well warranted.

4. Thoughts About a New State Age-Sex Method

The process of thinking about what changes to make to the current subnational estimates methodologies is multifaceted. The identification of the requirements we wish our methods to have is the first step. Likewise, a thorough review of our current estimates methods and the strengths and weaknesses in the resulting data can help to shed light on possible alternatives. In the end, however, one has to pick a starting point and begin to make suggestions. At this stage, the endeavor boils down to a series of questions, the answers to which begin to shape the proposal for the revised method. While the answers to these questions are best made somewhat simultaneously, it is instructive to discuss them independently.

What should be the base method? There are several possible approaches that could be used as the base method for our estimates. The following have been identified in the literature:[8]

1. Mathematical Extrapolation: Typically applied in situations where there are no or inadequate data that reflect changes to the population, the process of mathematical extrapolation updates a base year population by making assumptions about the rate of change in the population. Since there has been a wealth of potential data sources to indicate population change in the United States, this category of methods has not been used for the official estimates of the United States Census Bureau.
2. Censal-ratio methods: Censal-ratio methods produce population estimates by computing ratios of the population from a census to data "symptomatic" of the size or distribution of the population and then applying this ratio to subsequent symptomatic data for the estimate date.
    a. One example of this method is the housing unit method, which estimates the size of the population by obtaining the current number of housing units in a given area,

[8] Siegel, Jacob S., and David A. Swanson, eds. 2004. The Methods and Materials of Demography. San Diego: Elsevier Academic Press, p. 527.

adjusting that number to reflect vacancy, and then multiplying the resulting occupied housing units (i.e., households) by a persons-per-household rate – typically from the census -- to arrive at the population estimate. This method was also used in the 1980s at the Census Bureau to estimate the population of counties (often averaged with the results of other methods) and is still used by the United States Census Bureau in the population estimates for places in the Unites States.

b. A second example of this method is a characteristics ratio method. It was more widely used at the United States Census Bureau in the late 1980s and early 1990s. It functions by using a set of ratios, or proportions, to distribute previously generated estimates at a higher level of aggregation down to the desired lower levels of aggregation. It was the method used to generate county characteristics estimates at the end of the 1990s.

c. A third example of this method that has been advanced more recently is an administrative records ratio method. It was made possible when the United States Census Bureau began to acquire massive micro-record databases that began to approach national coverage. These include the full micro-data files from sources such as the Social Security administration, the Internal Revenue Service, Medicare, etc. These data are used to produce estimates by combining them into a single file with unduplicated records, observing the coverage of the data by comparing a file with a reference date close to that of a census to that census, and then adjusting future files by the observed coverage rates.

3. Component methods: Potentially the most widely known and used method, the component method starts with a base population and estimates population change using fertility, mortality, and migration. If done with age-detail, it is known as the cohort-component method because it pays special attention to the effect of people aging over time. The component method is currently used for the state and county totals estimates, while different incarnations of the cohort-component method are used for the national, state age-sex, state characteristics, and county characteristics estimates.[9]
4. Statistical methods: Statistical methods for producing population estimates, typically couched in a regression framework, are similar to the censal-ratio methods described above in that they produce prediction algorithms for population estimates by modeling the relationship between symptomatic data and a decennial census. They differ from censal-ratio methods, however, in that they allow for the use of multiple symptomatic indicators to inform the final prediction algorithm and permit more complex relationships to be modeled between the input data and the final estimates. At the Census Bureau, one example of a statistical method that has been used in the official population estimates is the ratio-correlation method. It was used during the 1980s to produce state and county totals estimates.

[9] In reality, the four methods cited here are not true examples of the component or cohort component models. While they use these techniques in an intermediate step during the production of the estimates, the end product is actually a ratio, which is then used to distribute a previously developed set of estimates. Technically speaking, therefore, these estimates use a ratio method.

At the present time, we have decided to start my testing of improved methods with a traditional cohort-component approach[10] applied to estimates of the state resident population by age and sex. We chose the traditional cohort-component method because, while the other methods may have specific advantages, this technique has the broadest appeal. In particular, [11]

- The data available for use in a cohort component technique (births, deaths, migration) are of relatively high quality. In addition, those data for which the quality may be in question (particularly with respect to domestic and international migration) are the focus considerable research both inside and outside of the Census Bureau.
- The method is widely viewed as a sound method, particularly because the traditional embodiment of the technique uses direct indicators of population change, models change in a way that parallels actual demographic change, and permits the use of existing theoretical knowledge about population growth.
- The cohort-component technique is recognized as having equal legitimacy for both estimates and projections, thus facilitating the production of robust estimates in situations where the input data from administrative records is unavailable or deemed inadequate.
- The method facilitates transparency in the estimates because it is widely used and understood. Moreover, it gives some indication of the causal mechanisms behind change in the population, at least with respect to the basic components and the age composition of the population.
- The method is adaptable to many levels of aggregation, thus having the potential to promote the integration of the Census Bureau's estimates system.

Nevertheless, we remain open to the possibility of drawing from the other methods by either creating a "hybrid" method (i.e., using a different method for the estimates of a specific subgroup – e.g., Medicare for the population ages 65+) or by changing gears altogether if the estimates from the cohort-component method are unacceptable and/or if research into the other methods are proven to substantially improve on the results.

At what level of detail should the estimates process start? For dissemination purposes, estimates data are needed for multiple levels of geography (national, state, and county), time-intervals (annual and monthly), characteristics (single years of age 0 to 100+, sex, 31 race groups, and Hispanic origin), and universes (resident, civilian and military, and household-institutional GQ-non-institutional GQ categories).[12] Developing a method that produces data to satisfy all these requirements at one time is unrealistic. Rather, the estimates will likely need to be produced in stages. This being the case, the challenge is to pick the starting point which will maximize the desirableness of the methods according to the requirements set forth above while

---

[10] Although it may go without saying, the cohort-component method being proposed here will use the estimates produced for each period as the base for the estimates for the next period as opposed to the current state age-sex estimates that goes back to the census base each time it estimates a panel in the time series.

[11] Many of the arguments in favor of the cohort-component method have been taken from chapters 20 and 21 in Siegel, Jacob S., and David A. Swanson, eds. 2004. The Methods and Materials of Demography. San Diego: Elsevier Academic Press, p. 527.

[12] See appendix A for a listing of the specific requirements.

minimizing the number of steps so it remains as implementable as possible in terms of time and staff resources.

In terms of the level of aggregation of the estimates, we have chosen to concentrate on annual resident estimates by age and sex for states. This has the following advantages:

1. Resident Universe: Currently, population estimates with military/civilian and household/institutional GQ/non-institutional GQ detail are produced by subtracting information from population registries for the military and GQ populations from the resident population. Changing to estimating these subgroups with a cohort-component method would require modeling migration between these statuses. Since there is no data with which to do this, the resident universe is preferred.
2. Age and Sex: Research into the robustness of data by age, sex, race, and Hispanic origin has shown that while the reporting of age and sex are fairly robust, that of race and origin may be unstable. The latter has been shown in research on the consistency of race and origin reporting in Censuses and Surveys and has become evident in our estimates through observed incompatibility between census and vital registration data. Thus, we believe it is preferable to estimate population with race and origin in a separate step. Estimating population by age and sex, however, is particularly important due to the challenges inherent in tracking the cohorts.
3. State Geography: Traditionally, estimates at the Census Bureau have been produced at the national level first and then at the subnational level. Starting at the national level has the advantage of being a quite manageable task. Yet, it can hide a wealth of heterogeneity that can cause problems when estimates at the state and county levels are forced to be consistent with them. It also reflects an area so large so as to preclude the use of intuitive or local knowledge in the evaluation of the estimates. Finally, the need to "control" to national estimates sometimes creates a situation in which states must compete against one another for population. At the state and county levels, conversely, local knowledge can be brought to bear, particularly through our formal relationship with the states in the Federal-State Cooperative Program for Estimates. However, the small size of many counties makes the application of a cohort-component a more complicated endeavor. Thus, we have decided to begin with estimates at the state level.
4. Annual Time Increments: Estimating the population by month is a difficult endeavor since it requires monthly input data that is generally unavailable, particularly for estimating migration. Moreover, applying a cohort-component method by month would require cohorts to be tracked by month of age, for which input data is even less available and of questionable quality. Annual estimates, on the other hand, are the most common and familiar type of estimates that are produced. For these reasons, we have chosen to begin with annual estimates.

What are the preferred estimation procedures? Taking as given that the method is cohort-component and the level of aggregation is annual resident estimates by age and sex for states, we have made the following initial choices in terms of estimation procedures.

The basic specification of the cohort component model, often referred to as the balancing equation, is commonly represented in the following formula:

$$Pop_{a+1,t+1} = Pop_{a,t} - Dth_a + InMig_a - OutMig_a \quad \text{(Eq. 1)}$$

where,

$Pop_{a+1,t+1}$	= The population aged [a+1] at time [t+1]
$Pop_{a,t}$	= The population aged [a] at time [t]
Dth_a	= Deaths to the population age [a] at time [t] during the period from [t] to [t+1].
$InMig_a$	= In-migrants to the population age [a] at time [t] during the period from [t] to [t+1].
$OutMig_a$	= Out-migrants to the population age [a] at time [t] during the period from [t] to [t+1].

For the population born between time [t] and [t+1], the formula is adjusted as follows:

$$Pop_{0,t+1} = Bir - Dth_B + InMig_B - OutMig_B \quad \text{(Eq. 2)}$$

where,

$Pop_{0,t+1}$	= The population aged 0 at time [t+1]
Bir	= Births during the period from [t] to [t+1]
Dth_B	= Deaths during the period from [t] to [t+1] to the population born during that period
$InMig_B$	= In-migrants during the period from [t] to [t+1] that were born during that period
$OutMig_B$	= Out-migrants during the period from [t] to [t+1] that were born during that period

There are several ways that the cohort-component method can be applied depending on the nature of the input data available for the task. We have decided to use the following approaches in order to maximize the advantages of the input data currently available. For the time being we will be relying on the data available from some of the other methods for producing population estimates except in the modeling of internal migration. The current data include the most recent unadjusted decennial census,[13] births and deaths from vital registration,[14] and international migration.[15]

One of the first decisions regarding the base population for the cohort-component technique is to determine a reference date for the annual estimates. There are two real possibilities: April 1 and July 1. On the one hand, an April 1 reference date has the following advantages:

- Since the census date in the United States is April 1, estimates for this date each year would preserve the whole-year cohort structure from the census. If a July 1 date were used, special care would need to be exercised in moving from April 1 to July 1 in order to preserve the cohort structure – in particular for cohorts such as the beginning of the baby boom.
- Estimates for April 1 of each year would also allow the estimates to stay closest to the reporting of residence in the census, in particular for those who are seasonal migrants.
- The reference month associated with the address in the IRS tax migration data file is April (with the 15th being as close to the "income-tax filing moment" as April 1 is to the

[13] The decision to use unadjusted data is an organizational position based on the Decision on Intercensal Population Estimates of March 12, 2003. See http://www.census.gov/dmd/www/dipe.html.
[14] Provided to us by National Center for Health Statistics and the members of the Federal State Cooperative Program for Estimates
[15] Produced for us by the International Migration Statistics Staff of the U.S. Census Bureau from data on the foreign born population from the American Community Survey.

"census moment"). Thus, an April 1 date would more closely respect the nature of these data.

However, a July 1 reference date also has advantages:

- Most of our customers require data for July 1 of each year. Thus, having an April 1 reference date would require an additional step in the processing of the estimates before we could deliver these data. While we need to produce monthly state estimates in the long run, having a July 1 reference data would simplify the steps for many of our deliveries.
- With a July 1 reference date, the extra step of moving the estimates from April to July before deliveries could be made would need to be done just once. Arguably, doing this in the base year would preserve the cohort structure more closely because of the information available from the census in this year.

For the time being, we have decided to produce estimates with a reference date of July 1 for the reasons cited above and in line with current practices.

With respect to the data on births and deaths, we am fortunate to have the date of birth on the files for both of these indicators from the National Center for Health Statistics. This will allow for the tabulation of births and deaths according to age as of July 1 of each year as opposed to the more typical tabulation by age at time of occurrence. Consequently, births and deaths will be associated directly with the cohorts to which they pertain.

In the last years of the estimates time series for each vintage, birth and death data will not yet be available. In this situation, we plan to estimate these events using Age-Specific Fertility Rates for births and Age-Specific Central Death Rates for deaths. To arrive at initial values of these rates, estimates will be calculated from the existing data on births by age of mother, deaths by age and sex, and the corresponding population estimates from the current vintage. Then, these data will be used to projected future rates from the years for which vital event data is unavailable using the assumption of an exponential rate of change.

With respect to internal migration, one of the initial impetuses of this project was to explore the impact of incorporating person-based migration data from the IRS migration file. There are two different ways that are being considered to incorporate these data. In the first, the exemption data will be converted into in- and out-migration probabilities by state, age, and sex. The manner in which the net migration probabilities will be converted into net migration values will be discussed below.

In the second method, the fact that a full state-to-state migration matrix is available will be harnessed in a manner similar to that of the Census Bureau’s state and county characteristics methods as well as the state projections method. This will be done by first calculating out migration probabilities for each origin state, receiver state, age, and sex cell. Then, when these rates are applied to a population, they will produce the number of migrants by age and sex going from each state to each other state. The in-migration values for each receiver state will then be calculable by summing on the receiver-state variable.

One possible drawback to this method is that the data may be sliced too thin to produce valid results. If this were the case, an alternative method would be to specify the out-migration probabilities by state, age, and sex with respect to groups of destination states (e.g., U.S. regions) or to the nation as a whole.[16] Then, in-migration proportions could be calculated by dividing the number of in-migrants in the tax file by age and sex for a given state by the corresponding number of total in-migrants in the tax file for the state group. Finally, these proportions would be applied to the migration values that result from the application of the out-migration rates to the population.

The final topic to be discussed in this paper is the manner in which the population will be moved forward from time [t] to time [t+1] when the values for the components are not available for the period. This is a particularly important issue because the current methods used by the U.S. Census Bureau are sub-optimal. In general terms, the current approaches either multiply the event rates (e.g., central death rates, net-internal migration rates, etc.) by the population at the beginning of the period or by a "pseudo-midperiod" population calculated by surviving the population to the middle of the period using only the components for which event values are available. As both of these methods can be shown to create bias in the resulting estimates, it is important to develop more robust approaches.

Let us start with original balancing equation (equation 1 above) with some modifications to reflect the input data to be used in this particular application:

$$Pop_{a+1,t+1} = Pop_{a,t} - Dth_a + InDMig_a - OutDMig_a + IMig_a \qquad \text{(Eq. 3)}$$

where,

$Pop_{a+1,t+1}$ = The population aged [a+1] at time [t+1]
$Pop_{a,t}$ = The population aged [a] at time [t]
Dth_a = Deaths to the population age [a] at time [t] during the period from [t] to [t+1].
$InDMig_a$ = Domestic in-migrants to the population age [a] at time [t] during the period from [t] to [t+1].
$OutDMig_a$ = Domestic out-migrants to the population age [a] at time [t] during the period from [t] to [t+1].
$IMig_a$ = Net International migration to the population age [a] at time [t] during the period from [t] to [t+1].

The first thing to keep in mind is that calculation of the number of domestic out-migrants and the corresponding domestic in-migrants for the July to July estimation period can be calculated with the population at the beginning of the period and the out-migration probabilities. In addition, the projection of net international migration will be done in terms of absolute values using the the pattern observed in the ACS data and expert opinion about how this should change in the future. Thus, the only "unknown" in this equation that needs to be dealt with is the deaths.

Now, it is important to note that vital event data are tabulated and delivered to the Census Bureau on a calendar-year basis. Thus, in the final year for which vital event data are available, we will know the number of events for the first half of the July-to-July estimation period (July to December). Moreover, when it is necessary to project deaths using the age-specific central death rates, the rate itself will be centered on the July 1 date. Thus, when multiplied by the July 1 estimate, they will produce calendar-year events – again giving us the number of events for the first half of the July-to-July estimation period (July to December). Thus, equation 3 can be

[16] Specifying out-migration to regions is the method currently used in the Census Bureau's state projections. See http://www.census.gov/population/www/projections/projectionsagesex.html. Likewise, specifying out-migration to the U.S. as a whole is the method currently used in the Census Bureau's state characteristics estimates. See http://www.census.gov/popest/topics/methodology/2004_st_char_meth.html.

revised to reflect the fact that death data from two successive calendar years will be needed to produce the estimate:

$$Pop_{a+1,t+1} = Pop_{a,t} – [0.5 * Dth_{a,p1}] – [0.5 * Dth_{a+1,p2}] + InDMig_a - OutDMig_a + IMig_a \qquad \text{(Eq. 4)}$$

where,

$Pop_{a+1,t+1}$	= The population aged [a+1] at time [t+1]
$Pop_{a,t}$	= The population aged [a] at time [t]
$Dth_{a,p1}$	= Deaths to the population age [a] at time [t] during the period from [t – 0.5] to [t+0.5].
$Dth_{a+1,p2}$	= Deaths to the population age [a+1] at time [t+1] during the period from [t+0.5] to [t+1.5].
$InDMig_a$	= Domestic in-migrants to the population age [a] at time [t] during the period from [t] to [t+1].
$OutDMig_a$	= Domestic out-migrants to the population age [a] at time [t] during the period from [t] to [t+1].
$IMig_a$	= Net International migration to the population age [a] at time [t] during the period from [t] to [t+1].

Now, the only "unknown" on the right side of this equation that needs to be dealt with is the $Dth_{a+1,p2}$.

To get around this unknown, let us first stipulate that the values for $Dth_{a+1,p2}$ could be calculated if we had the population at [t+1] and central death rates for the period centered on [t+1]. Thus,

$$Dth_{a+1,p2} = Pop_{a+1,t+1} * Mx_{a+1,p2} \qquad \text{(Eq. 5)}$$

where,

$Dth_{a+1,p2}$	= Deaths to the population age [a+1] at death during the calendar year centered on [t+1]
$Pop_{a+1,t+1}$	= The population aged [a+1] at time [t+1]
$Mx_{a+1,p2}$	= Central death rate to the population age [a+1] at death during the calendar year centered on [t+1]

Substituting these values into equation 4, we get:

$$Pop_{a+1,t+1} = Pop_{a,t} – [0.5 * Dth_{a,p1}] – [0.5 * Pop_{a+1,t+1} * Mx_{a+1,p2}] + InDMig_a - OutDMig_a + IMig_a \qquad \text{(Eq. 6)}$$

where,

$Pop_{a+1,t+1}$	= The population aged [a+1] at time [t+1]
$Pop_{a,t}$	= The population aged [a] at time [t]
$Dth_{a,p1}$	= Deaths to the population age [a] at time [t] during the period from [t – 0.5] to [t+0.5].
$Mx_{a+1,p2}$	= Central death rate to the population age [a+1] at death during the calendar year centered on [t+1]
$InDMig_a$	= Domestic in-migrants to the population age [a] at time [t] during the period from [t] to [t+1].
$OutDMig_a$	= Domestic out-migrants to the population age [a] at time [t] during the period from [t] to [t+1].
$IMig_a$	= Net International migration to the population age [a] at time [t] during the period from [t] to [t+1].

Solving for $Pop_{a+1,t+1}$, we get a final equation of:

$$Pop_{a+1,t+1} = \frac{Pop_{a,t} - [0.5 * Dth_{a,p1}] + InDMig_a - OutDMig_a + IMig_a}{1 + [0.5 * Mx_{a+1,p2}]} \quad \text{(Eq. 7)}$$

At this point, the population estimates at every age for time [t+1] can be estimated. First, the births for the period centered on time [t+1] can be calculated by multiplying age-specific fertility rates by the population of women of childbearing age at time [t+1] that result from the application of equation 7. Then, equations similar to 7 can be used on the births. In the end, the number of deaths for year [t+1] can then be computed using equation 5.

5. Conclusions

In many respects, the research project to improve the methods for the official sub-national population estimates with age, sex, race, and Hispanic origin detail as described in this paper is still in its infancy. While we have begun the important task of identifying the requirements for our methods and estimates data (section 2 of this paper), these ideas still need to be vetted internally and then through our stakeholders. Likewise, much more work could be done to evaluate and document the strengths and weakness of the current methods and estimates data (section 3 of this paper). Similarly, the proposed techniques described in this research paper (section 4) go only part way toward arriving at an adequate method for producing state and county estimates with age, sex, race, and Hispanic origin detail that meet our requirements.

This last point deserves some elaboration, for there is a considerable amount of work left to be done. First, there are still several issues left outstanding with respect to the method for the annual resident state estimates by age and sex. For example, it is likely that the person-based IRS migration data will continue to inadequately capture the movement of the young adult population. In addition, a technique still needs to be proposed for incorporating the results from special censuses and challenges to the estimates in a way that produces a reasonable time series and maintains the consistency of the estimates.[17] Both of these issues will need to be addressed before we begin developing test-code, producing provisional estimates, and evaluating the results of our work.

Second, the approaches presented in this paper will allow us to produce only part of the output needed to meet the current commitments to our customers. At the state level, we still need a method for producing monthly estimates for the resident, civilian, and civilian-non-institutional universes with race and Hispanic origin detail. Moreover, we still need a method to produce annual estimates for the resident, household, and GQ universes with age, sex, race, and Hispanic origin detail at the county level.

Once we have what we believe to be robust methods for producing estimates that meet all the content requirements of our customers, we can then turn to exploring what changes we might make to our input data in order to make the estimates more acceptable. For example, some have argued that there are issues with the Census 2000 data for which we might be able to

[17] In particular if the estimates must be forced to be consistent with previously produced estimates at higher levels of aggregation (e.g., national characteristics estimates, state and county totals estimates, etc.).

compensate. Likewise, there are numerous places where the vital statistics data we use are not consistent with the Census 2000 population, especially when considering the data by race and Hispanic origin. Perhaps the biggest questions regarding the input data pertain to our measurements of migration – both internal and international. In the former, more research is needed into the adequacy of our current assumption that the migration-patterns of filers and their dependents are the same as those for non-filers and their dependents. For the latter, there is no doubt we can improve on the current method for distributing international migrants to the state and county level using information from Census 2000. All these issues pertaining to input data for the estimates deserve a close look.

Finally, it is important to remember that the scope of the current project leaves alone the national characteristic and state and county totals estimates. It is possible that some improvements tested here could be applied at the national level. It is also possible that the national data and state and county totals could simply be derived by summing the state and county data by characteristics.

Appendix A

Required Output to Meet Delivery Commitments for the Vintage 2005 Estimates

<u>National Estimates</u>

1. Resident population
    a. Time: Annual estimates for July 1st from last census year through the current vintage year.
    b. Characteristics:
        i. Age: Single years of age 0 to 99 and 100+.
        ii. Sex: Male and female.
        iii. Race: 31 races.
        iv. Origin: Hispanic and non-Hispanic.

2. Resident + military overseas population
    a. Time: Annual estimates for July 1st from last census year through the current vintage year.
    b. Characteristics:
        i. Age: <15, 15-64, and >65).
        ii. Sex: Male and female.

3. Civilian population:
    a. Time: Annual estimates for July 1st from last census year through the current vintage year.
    b. Characteristics: none.

4. Civilian non-institutional population:
    a. Time: Monthly estimates for the first of each month from April of census year through January of vintage year +2.
    b. Characteristics:
        i. Age: Single years of age 0 to 99 and 100+.
        ii. Sex: Male and female.
        iii. Race: 31 races.
        iv. Origin: Hispanic and non-Hispanic.

5. Births to resident population:
    a. Time: Monthly from January of vintage year – 1 to January 2006 vintage year +1.
    b. Characteristics:
        i. Sex: Male and female.
        ii. Race: 31 races.
        iii. Origin: Hispanic and non-Hispanic.

Appendix A (continued)

Required Output to Meet Delivery Commitments for the Vintage 2005 Estimates

National Estimates (continued)

6. Deaths to resident population:
    a. Time: Monthly from January of vintage year – 1 to January 2006 vintage year +1.
    b. Characteristics:
        i. Age: Single years of age 0 to 99 and 100+.
        ii. Sex: Male and female.
        iii. Race: 31 races.
        iv. Origin: Hispanic and Non-Hispanic.
7. Net international migrants to resident population:
    a. Time: Monthly from January of vintage year – 1 to January 2006 vintage year +1.
    b. Characteristics:
        i. Age: Single years of age 0 to 99 and 100+.
        ii. Sex: Male and female.
        iii. Race: 31 races.
        iv. Origin: Hispanic and non-Hispanic.

State Estimates

1. Resident population:
    a. Time: Annual estimates through the current vintage year and January 1st populations for vintage year and vintage year-1.
    b. Characteristics:
        i. Age: Single years of age 0 to 84 and 85+.
        ii. Sex: Male and female.
        iii. Race: 31 races.
        iv. Origin: Hispanic and non-Hispanic.

2. Civilian population:
    a. Time: Vintage year.
    b. Characteristics:
        i. Age: Age 17.
        ii. Sex: Males.

Appendix A (continued)

Required Output to Meet Delivery Commitments for the Vintage 2005 Estimates

State Estimates (continued)

3. Civilian non-institutional population:
   a. Time: Monthly estimates for the first of each month from April of census year through January of vintage year +2.
   b. Characteristics:
      i. Age: Single years of age 0 to 84 and 85+.
      ii. Sex: Male and female.
      iii. Race: 6 races.
      iv. Origin: Hispanic and non-Hispanic.

4. Civilian institutionalized population:
   a. Time: Vintage year.
   b. Characteristics:
      i. Age: Age 17.
      ii. Sex: Males.

5. Births to resident population:
   a. Time: Annual estimates through the current vintage year.
   b. Characteristics: none.

6. Deaths to resident population:
   a. Time: Annual estimates through the current vintage year.
   b. Characteristics: none.

7. Net internal migrants to resident population:
   a. Time: Annual estimates through the current vintage year.
   b. Characteristics: none.

8. Net international migrants to resident population:
   a. Time: Annual estimates through the current vintage year.
   b. Characteristics: none.

Appendix A (continued)

Required Output to Meet Delivery Commitments for the Vintage 2005 Estimates

County Estimates

1. Resident population:
    a. Time: Annual estimates through the current vintage year.
    b. Characteristics:
        i. Age: Single years of age 0 to 99 and 100+.
        ii. Sex: Male and female.
        iii. Race: 31 (Census 2000) races and 4 (1990 Census) races.
        iv. Origin: Hispanic and non-Hispanic.

2. Household population:
    a. Time: Annual estimates through the current vintage year.
    b. Characteristics:
        i. Age: Single years of age 0 to 84 and 85+.
        ii. Sex: Male and female.
        iii. Race: 31 races.
        iv. Origin: Hispanic and non-Hispanic.

3. Group quarters population:
    a. Time: Annual estimates through the current vintage year.
    b. Characteristics:
        i. Age: Single years of age 0 to 84 and 85+.
        ii. Sex: Male and female.
        iii. Race: 31 races.
        iv. Origin: Hispanic and non-Hispanic.

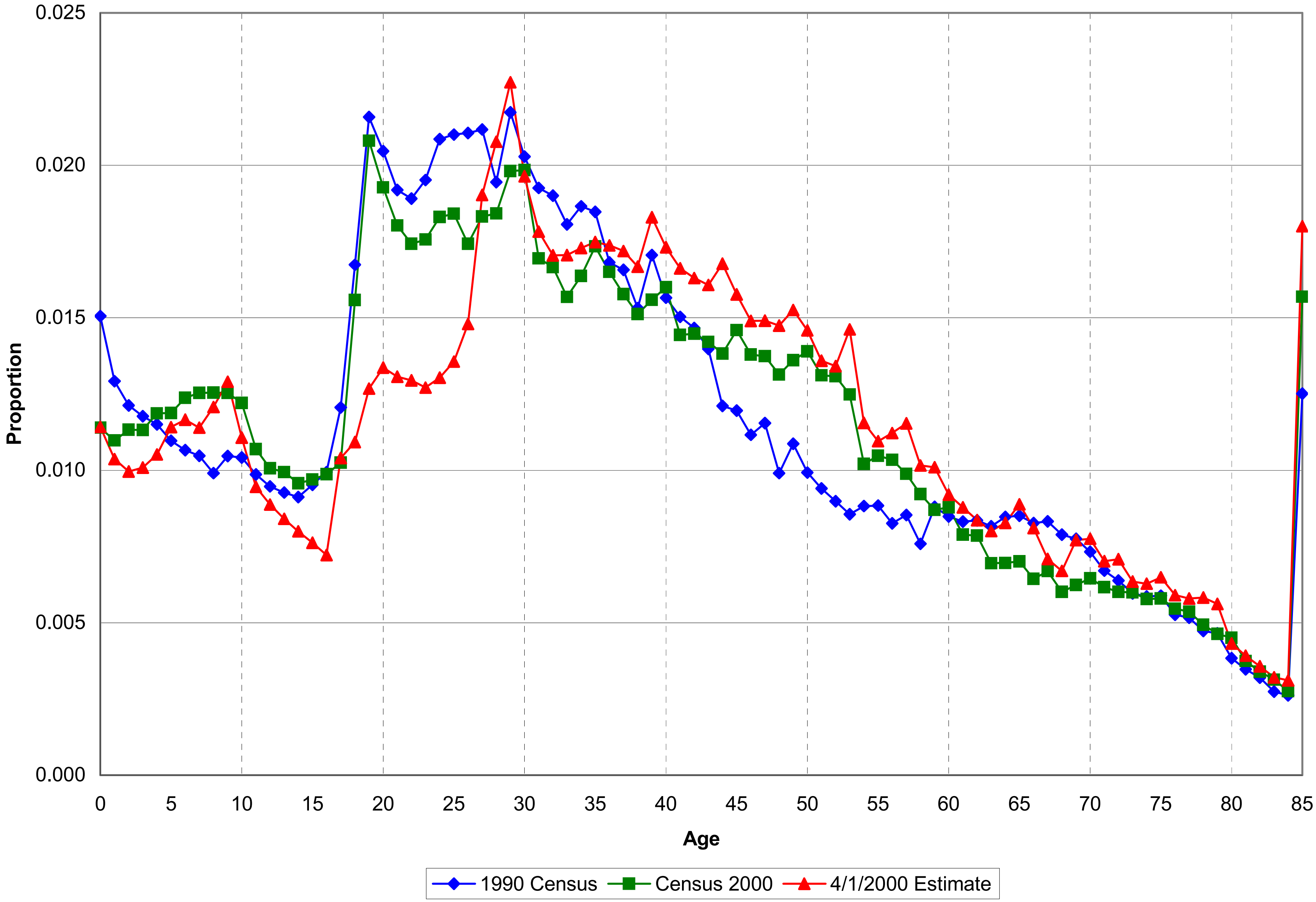
Appendix B

The Mean Algebraic Percent Error (MALPE) and Mean Absolute Percent Error (MAPE)

The Age Distribution of 1990-based State Age-Sex Estimates for 4/1/00 Compared to Census 2000

Age	MALPE	MAPE	Average of MAPES Ages 0-16	Age	MALPE	MAPE	Average of MAPES Ages 17-64
Total	0.3	4.1					
0	2.8	3.1		45	0.9	2.1	
1	0.7	2.6		46	0.6	2.0	
2	1.0	2.5		47	1.1	2.2	
3	-0.1	2.4		48	-3.5	4.4	
4	-0.7	2.5		49	4.5	4.7	
5	-0.9	2.4		50	-0.5	2.2	
6	-1.1	2.5		51	-0.8	2.4	
7	-2.3	3.2		52	-2.0	3.0	
8	-5.8	5.8		53	0.1	3.5	
9	1.4	2.4		54	0.5	3.1	
10	0.2	2.5		55	-1.3	3.2	
11	-0.7	2.7		56	2.0	3.5	
12	-1.1	2.8		57	3.9	5.5	
13	-0.2	2.8		58	-3.4	5.0	
14	0.1	2.9		59	3.2	4.8	
15	0.4	3.1		60	-0.4	4.4	
16	0.9	3.8	3.0	61	1.6	4.3	
17	2.8	3.1		62	1.0	4.8	
18	-0.6	4.2		63	1.6	4.7	
19	2.9	7.4		64	4.6	6.3	4.6
20	2.5	6.3		65	2.4	3.4	
21	1.1	5.3		66	1.2	2.7	
22	-1.0	4.5		67	-1.0	2.6	
23	-2.0	4.8		68	-0.9	2.9	
24	-1.0	5.1		69	2.3	3.5	
25	-4.7	6.3		70	1.2	3.2	
26	-3.9	5.6		71	2.2	3.5	
27	-4.2	7.3		72	3.2	3.7	
28	-8.1	10.8		73	-1.8	3.4	
29	0.9	7.9		74	-1.6	2.7	
30	-4.6	8.0		75	-0.7	3.7	
31	-3.1	6.4		76	-0.1	2.9	
32	-3.6	5.4		77	1.2	3.0	
33	-2.7	4.8		78	5.3	5.4	
34	0.4	3.8		79	5.6	6.0	
35	-1.8	3.5		80	-1.3	4.1	
36	-1.1	2.7		81	1.0	3.4	
37	1.3	3.3		82	-0.5	2.6	
38	-3.8	4.4		83	3.5	4.6	
39	6.9	7.2		84	3.1	4.8	
40	1.4	2.6		85	3.7	4.5	
41	3.7	4.2					
42	2.0	2.9					
43	1.5	2.5					
44	5.5	5.6					

Appendix C:
Proportional Distribution by Age of the 1990-Based Both-Sex Population Estimates for the District of Columbia vs. the 1990 and 2000 Censuses



Appendix D
Average Absolute Percent Change in the Population Estimates between 7/1/2003 and 7/1/2004 (Vintage 2004) Due to Components of Change and Final Rake of 7/1/2004 Estimates And Percent of Cases Where Change to Estimates from Rake is Greater Than Change from Components

Age	Components of Change	Rake	% Rake Change > Component Change	Age	Components of Change	Rake	% Rake Change > Component Change
Total	2.2	3.8	70.4				
0	n/a	n/a	n/a	45	0.9	2.1	99.0
1	0.8	0.8	58.8	46	0.9	3.1	97.1
2	0.8	1.0	70.6	47	0.9	3.2	89.2
3	0.8	0.2	11.8	48	1.0	3.6	88.2
4	0.8	0.6	42.2	49	1.0	3.1	97.1
5	0.9	1.0	64.7	50	0.9	3.3	94.1
6	0.9	0.2	12.7	51	1.0	3.3	92.2
7	0.9	0.2	16.7	52	1.0	3.2	90.2
8	0.8	0.4	24.5	53	1.1	3.1	92.2
9	0.8	0.2	17.6	54	1.1	3.6	92.2
10	0.8	0.2	7.8	55	1.1	3.4	88.2
11	0.8	0.4	28.4	56	1.1	3.3	83.3
12	0.8	0.1	4.9	57	1.1	2.9	94.1
13	0.8	0.6	42.2	58	1.2	6.4	80.4
14	0.8	0.7	50.0	59	1.4	4.2	82.4
15	0.8	0.4	28.4	60	1.4	3.6	72.5
16	0.8	0.3	25.5	61	1.5	3.1	89.2
17	1.0	3.2	93.1	62	1.6	4.6	67.6
18	1.2	3.0	94.1	63	1.7	3.0	67.6
19	1.5	3.6	93.1	64	1.8	3.3	71.6
20	1.5	3.1	86.3	65	1.9	3.7	35.3
21	1.7	3.5	78.4	66	8.4	7.9	2.0
22	1.8	3.5	74.5	67	10.5	3.0	86.3
23	1.9	3.2	76.5	68	3.2	6.9	96.1
24	1.9	3.2	72.5	69	4.4	10.1	94.1
25	2.0	2.9	72.5	70	3.5	9.8	1.0
26	2.1	2.8	70.6	71	13.6	3.6	92.2
27	1.9	3.0	75.5	72	2.7	10.0	43.1
28	1.6	2.8	77.5	73	5.0	4.7	73.5
29	1.6	3.1	80.4	74	2.2	4.2	73.5
30	1.5	2.9	85.3	75	2.4	4.6	0.0
31	1.4	3.0	86.3	76	14.6	4.4	78.4
32	1.3	3.1	83.3	77	3.0	6.1	72.5
33	1.2	3.1	86.3	78	2.3	5.0	70.6
34	1.1	3.0	91.2	79	3.2	5.4	91.2
35	1.0	2.9	90.2	80	3.6	9.1	0.0
36	1.0	3.1	93.1	81	17.4	2.9	74.5
37	1.0	3.2	87.3	82	4.2	6.2	90.2
38	0.9	3.2	92.2	83	2.7	7.7	100.0
39	0.9	3.3	94.1	84	2.4	18.8	100.0
40	0.9	3.3	92.2	85	2.1	20.8	n/a
41	0.8	3.2	92.2				
42	0.8	3.2	94.1				
43	0.9	3.3	91.2				
44	0.9	3.3	97.1				