Imbedding Model-Assisted Estimation into ACS: The Impact on Users

Robert E. Fay[1]

U.S. Census Bureau
U.S. Census Bureau, 4600 Silver Hill Rd., Robert Fay, SRD, Washington, DC 20233; robert.e.fay.iii@census.gov

Key Words: American Community Survey, calibration estimators, GREG, survey estimation

Abstract. The American Community Survey (ACS) began full implementation in 2005 as a replacement for the decennial census long form. In 2010, ACS estimates will be released for the 5-year period 2005-2009; this release will be the first to offer ACS results at the geographic detail previously provided by Census 2000. A model-assisted approach has been proposed to reduce the variances of estimates for geographic units below the county level for both 3- and 5-year period estimates. The approach has been investigated as part of the Multiyear Estimates Study, based on an ACS test in 34 counties during 1999-2005. The anticipated variance reductions were empirically confirmed, although substantial reductions occurred for a few key variables and more modest ones for others. Given the beneficial variance impact, the primary focus of this paper will be to address some of the remaining concerns potential ACS users may have about the model-assisted methods, such as their impact on bias and whether their introduction complicates analysis.

1. Introduction

By the time of this conference, the Census Bureau will have published estimates for 2006 from the American Community Survey (ACS), for its second year of full implementation. The conference program includes several papers on the ACS, so arguably the survey requires little general introduction. In short, the ACS, as a replacement for the census long form, is based on a designated sample of approximately 3 million housing units per year, divided into essentially equal monthly samples. The questionnaire content combines items generally identical or similar to those in the long form for Census 2000. Each monthly sample receives questionnaires in the mail, and sample households in this first phase respond by mail at approximately 50%. Nonresponding units are followed up by a second phase of telephone interviewing when possible. A subsample of the remaining units is selected for the third and final phase of computer-assisted personal interviewing (CAPI). In 2005, about 2/3 of the originally designated sample, just under 2 million housing units, were interviewed, less than the designated sample primarily because of the subsampling for the CAPI phase. In 2005, the ACS achieved an overall weighted response rate of 97%.

Because the designated sample for the ACS in a single year is far smaller than the previous samples for the census long form—at approximately 18 million sampled households in 2000—ACS will publish 5-year period estimates based on accumulating data over 5 years to provide the same detailed geographic levels as the census. In 2010, the first 5-year period estimates from the full ACS will be published for the period 2005-2009. The 1-year ACS estimates for 2005 and 2006 were published only for counties, places, and other geographic areas of population 65,000+, a threshold that will continue to be used for 1-year estimates. In 2008, a first set of 3-year period estimates will be published for 2005-2007, for geographic units above a population threshold of 20,000.

A panel of the Committee on National Statistics, National Research Council of the National Academy of Sciences—the Panel on the Functionality and Usability of Data from the American Community Survey—recently released the advanced copy of their report (Citro and Kalton, 2007) in spring 2007 and published it to the Internet. Their report, *Using the American Community Survey: Benefits and Challenges* commended the ACS as a replacement for the long form in terms of timeliness and quality of the data. But it noted that users are likely to find the logic of the relationships and use of the 1-, 3-, and 5-year period estimates more complex than their previous experience in working with decennial data. Among other challenges, the report also noted the ACS estimates would be subject to appreciably higher sampling variances than the respective decennial long-form results they were replacing.

This paper concerns one aspect of the ACS variance problem, namely, the variances of sub-county estimates, particularly of estimated totals. Although ACS 5-year period estimates can be expected to have generally higher variances than corresponding decennial census estimates because of the sample size differences, in a previous ACS test the variances were particularly disproportionately large for sub-county estimates.

The problem with sub-county variances was first observed in two of the 34 counties selected as ACS test sites. During 1999-2001, the 34 counties were sampled for the ACS at rates above those for the production ACS, resulting in sufficient sample sizes for the 3-year period 1999-2001 to approximate the ACS sample sizes available in 5 years during full production. County- and tract-level estimates were produced and released for these test sites as part of the ACS 1999-2001 and Census 2000 Comparison Study. (Census tracts are areas of roughly 4000 people, on the average.) Paul Voss and his collaborators (Van Auken *et al.* 2004, 2006) observed that the tract-level variances considerably exceeded the projections in the two Wisconsin test counties they studied, whereas the county-level variances fell closer to the predictions. Starsinic (2005) confirmed the finding for the remaining test counties and demonstrated that much of the discrepancy could be attributed to the absence of sub-county weighting controls in ACS. In Census 2000, weights were computed for weighting areas that were frequently equivalent to census tracts, often achieving consistency between the 100% count and sample at the tract level. During the decade, however, there are no available weighting controls comparable to the detail available from the census during a decennial year.

This paper is one of a series (Fay 2005a, 2005b, 2006, 2007) describing and investigating a weighting method to reduce the variance of ACS sub-county estimates. Since the first paper in the series, both the outline of the problem and the basic approach have remained largely the same. The 2005 papers presented a case for combining model-assisted estimation, specifically Generalized Regression Estimation (GREG), with administrative record data to reduce the variance of tract-level estimates without introducing appreciable bias. The two papers included the most extensive literature reviews of the series but presented only indirect evidence, based on strong correlation coefficients between the ACS and administrative record data, that such an approach could work. Fay (2006) then presented preliminary GREG calculations using the 1999-2001 data in the ACS test counties and administrative record data for the year 2000. The preliminary results showed promising variance reductions, but they left open several questions, including how the results would generalize beyond the census year. The most recent paper (Fay 2007) in the series analyzed some of the findings from a full-scale test of the method in the 34 counties. The GREG research was incorporated as part of the ACS Multiyear Estimates Study. The new findings and those of Starsinic and Tersine (2007) showed varying levels of variance reduction from the inclusion of the GREG step. The results met a standard of "do no harm," that is, the variance reductions ranged from modest to substantial, but the observed increases were negligible for any of relatively few variables showing any increase at all.

This paper will summarize the variance comparisons from Fay (2007) and present additional ones. A primary focus, however, will be to address a set of other questions that can be raised about the impact of a possible inclusion of GREG in the 3-year ACS period estimates beginning with 2005-2007 and in the 5-year period estimates beginning in 2005-2009. Although there is a theoretical basis to argue that the GREG will not introduce appreciable systematic bias, a number of outside potential users of ACS data, including members of the Census Advisory Committee from the Population Association of America, have expressed some tentative skepticism about possible bias effects in the ACS context. Consequently, this paper will analyze data from the Multiyear Estimates Study to check for evidence of bias. The results to be presented here do find evidence of a small level of systematic bias for some variables. In general, 3-year estimates appear subject to less bias than the 5-year ones. For example, one of the most pronounced results for the 5-year estimates, a 2.0% drop in the estimated number of vacant housing units, corresponds to a drop of only 0.3% in the 3-year estimates. The majority of the 3- and 5-year variables appear to have no detectable bias, but this paper will attempt to quantify the degree to which some variables exhibit small biases.

The next section summarizes the GREG estimation, the use of administrative record data, and the Multiyear Estimation Study. Following that, the variance results from Fay (2007) and Starsinic and Tersine (2007) are summarized, with additional results presented for counties and places. General sources of potential bias in the ACS estimates will be reviewed to form the basis for a discussion of possible effect of introducing GREG. This section is followed by the analysis of data from the Multiyear Estimates Study for evidence of possible bias. The discussion section places the bias results in some perspective, and it argues that the inclusion of the g-weighting in the ACS Estimation to reduce the subcounty variances will not impose additional challenges on the ability of users to interpret the results.

2. Methods

General Approach
Starsinic (2005) showed that if housing and population controls were available at the tract level and assumed free from error, then the ACS data could be reweighted to achieve approximately the same relative variance reductions as the decennial census achieves by weighting the long-form estimates to 100% count control totals within census weighting areas. But, except for the census year itself, no such controls currently exist for ACS to use, and any attempt to produce such controls through implicit or explicit modeling would likely be subject to relatively large and difficult-to-measure errors. In the absence of such controls, Fay (2005a) pointed out a possible alternative by combining the following:

1. Data from administrative records for the same collection year as the ACS sample in question;
2. A match of the administrative data to the ACS sampling frame, that is, the Master Address File (MAF);
3. A step of model-assisted estimation, specifically GREG, imbedded as an additional step early in the multi-step ACS weighting.

A key feature of the argument was that the application of model-assisted estimation would result in some level of variance reduction, but not introduce appreciable new bias. Thus, the relative success or failure of the approach could be assessed directly through standard variance estimation techniques.

Administrative Records. The Census Bureau now has a central staff responsible for the integration and management of administrative records for statistical purposes. One product of this effort, updated annually, is the Person Characteristic File, which is based on a combination of Social Security and census information. Another series of files uses administrative records to identify the possible location of these persons. For this project, an extract file was created merging the demographic information on age, sex, race, and Hispanic origin from the Person Characteristic File with a likely address for the person. The likely address was expressed in the form of a MAFID, an identification number used on the Census Bureau's Master Address File. For purposes of this study, the administrative records staff provided extracted annual files for 2000 through 2005 for the 34 ACS test counties. The extracted files did not include any income information from the administrative sources.

As an identifier, the MAFID is stable across time. Although the administrative records are at the level of individual persons, when they are combined into households on the basis of common MAFIDs, the records present a census-like picture of most of the current household population. The population coverage of the administrative records, roughly in the range of 80 to 85%, varies over time and geographic areas.

The Master Address File (MAF). Leading up to Census 2000 and subsequently, the Census Bureau has maintained the MAF as an inventory of housing units and commercial buildings in the U.S. The ACS uses extracts from updated versions of the MAF as its sampling frame, and ACS sample households can be linked back to the corresponding MAF records that had been selected for the ACS sample. Both U.S. Census Bureau (2006) and Citro and Kalton (2007) describe the MAF in detail.

Because the administrative record data were linked on the basis of MAFID to the MAF, it is possible to tabulate, without reference to the ACS, the *unweighted* count of matched administrative records at various levels of demographic and geographic detail, such as the total number of administrative record persons or the number of persons age 0-17. At the same time, because of the linkage, it is also possible to estimate from the MAFIDs in the ACS sample and the ACS weights, the *weighted* estimate of administrative record persons for the same geographic universes as the unweighted counts. In other words, without considering the data collected in ACS, the ACS sample hits can be viewed as denoting a *sample* of administrative records matched to the MAF from the *universe* of administrative records matched to the MAF. By comparing the unweighted counts for the administrative record universe, which are not subject to sampling variability, to the ACS sample estimates for the same characteristics, the resulting difference is a source of auxiliary information about the ACS sample.

Model-Assisted Estimation. Model-assisted estimation represents a set of approaches to using the auxiliary information to improve survey estimates. There is an extensive literature on model-assisted estimation, including the 1992 book of Särndal, Swensson, and Wretman. One form, generalized regression estimation (GREG), has been applied to a wide variety of problems, including the weighting of the census long questionnaire data for the Canadian censuses of 1991, 1996, 2001, and 2006 (Bankier, Rathwell, and Majkowski, 1992; Bankier, Houle, and Luc, 1997; Bankier and Janes, 2003).

Ratio estimation, used extensively in survey research, can be viewed as a special case of GREG. As an introduction to GREG, a description of a hypothetical application of ratio estimation to ACS tract-level estimation will illustrate several features of GREG. In the notation of Rao (2003), let U represent the population in the scope of the ACS, and $y_1,\ldots,y_N$ the values in the population for a characteristic the ACS is to measure. ACS is to produce an estimate of the population total $Y=\sum_U y_j$ based on the ACS sample s drawn with probability $p(s)$. (Although the ACS weighting process is complex, the GREG step will come relatively early in the process, just after the noninterview adjustments.) Consequently let $w_j(s)=w_j$ denote preliminary weights, which are close to the ACS base weights based on the inverse probability of selection, $w_j=\pi_j^{-1}$ or, more specifically, the weights based on π_j^{-1} adjusted by the ACS noninterview adjustments. (The notation $w_j(s)$ incorporates s to recognize that the weight for j might depend on the realized sample s.) Then $\hat{Y}=\sum_s w_j(s)\,y_j$ denotes an initial estimate of Y.

Suppose as a matter of notational convention, we let $x_1,\ldots,x_N$ denote auxiliary information in the population. To continue with the example of ratio estimation, suppose the $x_1,\ldots,x_N$ denote scalars identically 1. Thus the sum of the sample weights $\hat{X}=\sum_s w_j(s)\,x_j=\sum_s w_j(s)$ estimates the frame total $X=\sum_U x_j$. A possible tract-level estimation step, based on ratio estimation, would be to compute

$$g_j(s)=X/\hat{X}=1+(X-\hat{X})/\hat{X} \qquad (1)$$

and revised weights $w_j^*=w_j^*(s)=w_j(s)\,g_j(s)$. As a consequence of this simple form of ratio estimation, the revised weights sum to the total for the frame within tract. Estimates that are correlated with the frame total, such as the number of occupied housing units in the tract, would generally have lower variance after this simple ratio adjustment.

Technically, ratio estimation often involves a small bias, but in the usual applications the effect of the bias is of lower order than the variance, so the effect of the bias in ratio estimation is typically negligible. For example, in the 1963 edition of his book, Cochran (p. 157) in discussing the bias of the ratio estimate, advises

> These results amount to saying that there is not difficulty if the sample is large enough that (a) the ratio is nearly normally distributed and (b) the large-sample formula for its variance is valid. As a working rule, the large-sample results may be used if the sample size exceeds 30 and is also large enough that the coefficients of variation of $\bar{x}$ and $\bar{y}$ are both less that 10%.

Survey practice appears to extend application of ratio estimation somewhat beyond the bounds suggested by Cochran. For example, Särndal, Swensson, and Wretman (1992, p. 251) remark "For the validity of the confidence intervals, the bias may be a factor of some concern in very small samples, but for samples of size 20 or more, the consequence of the bias is usually negligible." For a tract of 1600 housing units, the designated ACS 5-year sample size would reach about 200, with a realized sample of about 130 (based on 2005 rates), comfortably large by the standards of textbook advice. For a small tract of 300 housing units, however, the designated sample would be about 37 and the realized sample 24, approaching the boundary where some researchers have advised caution.

Starsinic (2005) investigated a similar tract-level ratio estimate, although his version assumed the availability of an independent estimate, X, of the number of housing units rather than simply the frame count. (Not all MAF units correspond to housing units in the universe, and determination is made during the followup of the CAPI sample of whether a MAF unit is either an occupied or vacant housing unit, or instead is not in the universe, such as a house under construction that is not yet ready for habitation. Hence, the ACS frame count is known, but the estimated number of valid housing units is based on a sample.)

The preceding example illustrates how ratio estimation incorporates the auxiliary information available from a single variable. Generalized regression estimation (GREG) allows the use of multivariate auxiliary information, that is, the $x_1,\ldots,x_N$ can denote column vectors of auxiliary information in the population. Let the population total of the auxiliary information be denoted by $X=(X_1,\ldots,X_p)^T$, where the superscript T denotes transpose. The vectors $x_1,\ldots,x_N$ can continue to include 1 as a first element, to achieve the same consistency with the frame total as the ratio estimation example, but the other components will pertain to administrative record values. As with ratio estimation, the sample estimate $\hat{X}=\sum_s w_j\,x_j$ estimates X. For an individual ACS characteristic, Y, the GREG estimate may be expressed as

$$\hat{Y}_{GR} = \hat{Y} + (X - \hat{X})^T \hat{B} \qquad (2)$$

where for constants $c_j > 0$

$$\hat{B} = (\hat{B}_1, ..., \hat{B}_P)^T = \left(\sum_s w_j x_j x_j^T / c_j \right)^{-1} \sum_s w_j x_j y_j / c_j \qquad (3)$$

(In the ACS application, c_j was chosen to be 1 identically, and alternatives have not yet been systematically investigated.) In spite of the apparent dependence of *(2)* on the choice of *Y,* GREG can be written as a weighting adjustment similar to *(1),*

$$g_j(s) = 1 + (X - \hat{X})^T (\sum_s w_j x_j x_j^T / c_j)^{-1} x_j / c_j \qquad (4)$$

in order to compute revised weights $w_j^* = w_j^*(s) = w_j(s)\, g_j(s)$. In other words, GREG can be incorporated as a weighting step, thus making its incorporation into the existing ACS weighting system quite feasible.

The *X* Variables Based on Administrative Records. To achieve the same calibration of the GREG-adjusted ACS weights $w_j^*(s) = w_j(s)\, g_j(s)$ to the tract-level frame totals, the first component of each vector x_j is set identically to 1. Depending on the characteristics of the tract, the remaining components of x_j are based on counting the number of person-level administrative records falling into specific demographic categories within the household. For example, one component of x_j may be the count of administrative records classified in the household classified as 0-17, and another the count of males 30-44. In the application to be described here, a maximum of 7 age/sex categories were created: females age 30-44 and 45-64, males age 30-44 and 45-64, and cells for ages 0-17, 18-29 and 65+. In some tracts, the number of variables was reduced, always collapsing categories of age and sex so that the sum of the age/sex cells was the total number of administrative record persons. Fay (2007) provides further details of the collapsing. Collapsing was performed whenever the matrix inversion in *(4)* did not exist because of singularity, or when the result of *(4)* produced a negative $w_j^*(s)$ within the tract.

Separate race/Hispanic origin categories were also included in some tracts where there were sufficient numbers in the universe. Components of x_j could correspond to the administrative record number of Hispanic, non-Hispanic Black, or non-Hispanic other races (including Asian and Pacific Islander and American Indian or Native Alaskan, but excluding White only). Further details may also be found in Fay (2007).

Steps of the Implementation. The following five steps, outlined previously (Fay 2005a), form the basic elements of the strategy to improve tract-level estimates:

1. ***Link administrative records to the ACS sampling frame (the MAF), dropping administrative records that cannot be linked.*** This step simply uses the MAFIDs supplied on the extract file for linking, and it does not entail a new linking algorithm.
2. ***Form unweighted tract-level totals of the linked administrative record characteristics*.** This operation computes the unweighted *X* appearing in eq. *(4).*
3. ***Apply ACS sampling weights at the housing-unit level to the linked administrative record data that fall into the ACS sample. The weighted estimates at this step represent unbiased (or essentially unbiased) estimates of the unweighted totals in step 2.*** This operation computes the estimate $\hat{X}$ in eq. *(4).* In the abstract, $\hat{X}$ should be an unbiased estimate of *X.*
4. ***Using generalized regression estimation (GREG), calibrate the ACS sample weights so that the weighted administrative totals from the sample match the unweighted totals from step 2. (The number of constraints is allowed to vary with the size and other characteristics of each tract.)*** This operation refers to the computation of the remaining parts of *(4)* and then its use to compute revised weights $w_j^*(s) = w_j(s)\, g_j(s)$.
5. ***Use the new housing-unit weights in subsequent stages of the ACS weighting, which includes ratio and raking/ratio estimation. Although the subsequent estimation steps adjust the new weights, the argument is that most of the variance reduction at the tract level will be retained in the final weights.*** This last point can be empirically tested in the analysis of the Multiyear Estimates Study, because the estimates include all steps of estimation.

A final complication should be noted. Although for simplicity the value X has thus far been described as an unweighted total, in fact it is a time-weighted total. The calculation of X occurs over a 5-year period, during which the MAF extracts can both drop some units and add others. The administrative record data is time weighted for the number of years that the MAFID is included in the frame, so that a MAFID in all 5 years receives a full weight of 1, but MAFID included only 2 years during the period receives a time-weight of 2/5.

Application to the Multiyear Estimates Study

In addition to the oversampling of the 34 test counties during 1999-2001 to support production of tract-level ACS estimates for the 1999-2001 ACS and Census 2000 Comparison Study, the counties were sampled during 2002-2004 at a rate approximating ACS sampling rates. In 2005, the ACS began full national production for the household population. The Multiyear Estimates Study (see http://www.census.gov/acs/www/AdvMeth/Multi_Year_Estimates/overview.html) was undertaken to produce 3- and 5-year period estimates in these 34 counties in order to anticipate the features of multiyear period estimates, which aren't scheduled for national release until 2008 and 2010, respectively. Specifically, the study produced 3-year period estimates for 1999-2001, 2000-2002, 2001-2003, 2002-2004, and 2003-2005; and 5-year period estimates for 1999-2003, 2000-2004, and 2001-2005, in addition to releasing revised or new 1-year estimates for 2000 through 2005, all for the household population. For this study, GREG was incorporated in all five sets of published 3-year period estimates and all three sets of 5-year period estimates. The estimates are available from the Census Bureau's web site, www.census.gov.

Of the 2270 tracts included in the 34 counties in the 2001-2005 comparison, the smallest 45 had less than 300 units in the frame. Because the expected ACS sample size would be small, GREG was not attempted in these tracts. Since the purpose of the GREG was to reduce variance where possible, omitting the GREG step appeared to be a rational solution to small sample sizes in these few tracts.

A goal of the 5-year GREG application was to improve tract-level estimates, and the weighting for the 5-year period estimates implemented GREG at this geographic level. For the 3-year period estimates, however, tract-level weighting is both (1) less feasible because of the smaller available ACS sample sizes at the tract level and (2) unnecessary, because 3-year estimates were published only for areas of population 20,000 or more. Instead, weighting areas were employed composed either of a place of 10,000 or more or of the intersection of a place and a minor civil division (MCD) of 10,000 or more. For example, Lake County, IL, is divided geographically into both places (cities and towns) and MCDs (townships). In Lake County and other counties where Census 2000 published MCD estimates, intersections of places with townships (or equivalent MCDs) were used in the 3-year GREG weighting.

As an imbedded experiment in the Multiyear Estimates Study, the GREG step was omitted for a second set of 2003-2005 and 2001-2005 period estimates to produce unpublished comparison sets for internal analysis. Starsinic and Tersine (2007) and Fay (2007) compared the published estimates using GREG to the unpublished ones omitting GREG to assess the variance improvements from GREG.

The results from the Multiyear Estimates Study were published in the form of four "profiles," a statistical product mixing selected estimated totals, proportions, and derived statistics such as median age and income. The products are close analogues to four of the *quick tables* available from the Census Bureau's American FactFinder for Census 2000 (the quick table/profile for selected demographic characteristics from SF-1 and the quick tables/profiles for selected social, economic, and housing characteristics from SF-3). Similarly, American FactFinder offers data profiles of demographic, social, economic, and housing characteristics from the 2005 ACS. The Multiyear Estimates Study published four profiles that include 397 lines with estimated totals, 378 of which were unique and 19 of which duplicated other lines, as well as a set of medians, means, and proportions.

3. The Variance Impact of GREG

Variances are estimated in the ACS using a series of 80 replicate weights. The variance impact of GREG, a smooth function of the sample data, was readily assessed by replicating GREG for each of the replicate samples. The variances compared in this paper were all computed as part of the Multiyear Estimates Study and are available on internal Census Bureau files.

In their analysis of the Multiyear Estimates Study, Starsinic and Tersine (2007) used a single measure to summarize the impact of GREG estimation, the median of the ratios of coefficients of variation. They computed this measure for various sets, S, of geographic levels and characteristics in the profiles. Their measure is the square root of the following measure, the median of the ratios of relative variances over a set S

$$ratio_{med} \quad = \quad median_S\,((\operatorname{var}(\hat{Y}_{GREG})/\hat{Y}_{GREG}^{\;2})/(\operatorname{var}(\hat{Y}_{noGREG})/\hat{Y}_{noGREG}^{\;2})) \qquad (5)$$

Their paper examined variance comparisons over a variety of different subsets S of characteristics and geographic levels. Their findings supported the gains from using GREG, but they found substantial gains for the ACS variables directly related to the administrative records variables used in GREG estimation, such as age and sex, and generally modest gains for other variables.

Fay (2007) analyzed *(5)* and two other measures, one of which, $mean_S\,(\operatorname{var}(\hat{Y}_{GREG})/\operatorname{var}(\hat{Y}_{noGREG}))$, tended to give results similar to *(5)*, and the ratio of the sum of variances,

$$ratio_{tot} \quad = \quad \sum\nolimits_S \operatorname{var}(\hat{Y}_{GREG}) / \sum\nolimits_S \operatorname{var}(\hat{Y}_{noGREG}) \qquad (6)$$

Measure *(6)* produced results similar to those from *(5)* in many cases, but in others tended to be much lower, and therefore more favorable towards GREG. Both measures *(5)* and *(6)* will be computed for the variance comparisons to be presented here.

Although Starsinic and Tersine (2007) categorized the ACS characteristics in the published profiles in finer detail, this paper will follow the coarse classification of Fay (2007): (1) total housing units, (2) total households, (3) total population, (4) age/sex variables (33 lines), (5) race/Hispanic origin variables (48 lines), and (6) all remaining totals (294 lines). The first five variables are directly or closely related to the administrative record variables used in the GREG estimation, although the ACS age, race, and Hispanic origin variables include several far more detailed than the administrative record data used in the GREG application. For example, Hispanic origin is treated as a dichotomy on the administrative record files, but the profile includes separate estimates for Mexicans, Puerto Ricans, and Cubans.

Fig. 1 summarizes measures *(5)* and *(6)* at the tract level for 2001-2005. Individual comparisons at the tract level are classified according to the size of the estimate. Although the estimates with and without GREG, $\hat{Y}_{GREG}$ and $\hat{Y}_{noGREG}$, are typically quite close, the average of the two was used to classify the observations along the x-axis for the sake of symmetry.

As expected, in Fig 1a-1f there is almost no change in variance for the 45 small tracts that were excluded from GREG.

For Fig. 1a-1c, corresponding to total housing units, total households, and total population, measures *(5)* and *(6)* essentially agree, both for total and across the categories defined by the size of the estimate. The average variance reduction is quite substantial—approximately 90% for housing units, 70% for households, and 60% for total population.

A different pattern appears in Fig. 1d-1f for age/sex, race/Hispanic origin, and the remaining variables. Consistent with Fig 1a, 1b and particularly 1c, GREG has a greater effect on larger estimates and little impact on estimates below 300. Within size categories, the measures *(5)* and *(6)* approximately agree. When aggregated to an overall summary measure, however, *(6)* based on totals indicates a much greater variance reduction than *(5)* based on medians. This difference can be attributed to the greater weight that *(6)* effectively places on larger estimates, which tend to have larger variances but smaller relative variances. The distinction has greater impact in measuring the overall variance impact than in measuring the impact within size classes.

Although tracts are the direct beneficiaries of GREG estimation for the 5-year estimates, the results of Fig. 2 show that places benefit indirectly. The 5-year estimates include a number of small places with less than 1000 housing units or less than 3000 persons. For these small places, Fig 2a, 2b, and 2c show that the improvements from GREG are far less than those for larger places. In contrast to Fig. 1a, 1b, and 1c, measures *(5)* and (*6)* differ in Fig 2a, 2b, and 2c for total housing units, total households, and total persons. As the case elsewhere, *(6)* based on totals weights the outcome for larger estimates more than *(5),* so *(6)* shows more improvement from GREG when the results are summarized.

Fig. 2d, 2e, and 2f show results similar to—although somewhat less strong than—their analogues in Fig. 1. Again, measure *(6)* based on totals indicates greater gains from GREG than measure *(5).*

As previously noted, places or place/MCD intersections were used as weighting areas for the 3-year period estimates. Fig. 3a-3f present results analogous to those in 2a-2f, but the population threshold of 20,000 limits the number of published places for 3-year period estimates. Weighted by size, *(6)* based on total for the 3-year period estimates in Fig. 3a, 3b, and 3c is similar to the outcome for the 5-year analogues in Fig. 2a, 2b, and 2c.

Table 1 summarizes the variance comparisons in the three figures, without stratifying by size of estimate. The measures *(6)* based on total are quite similar for places between the 5-year period estimates in Fig. 2 and the 3-year period estimates in Fig. 3, and both are fairly similar to Fig. 1 for tracts. In contrast, the median measures vary much more within each row of the table.

In summary, both measures support the introduction of GREG in the ACS. Measure *(6)* weighted by estimate size, however, is more favorable to the outcomes from GREG than measure *(5)* based on medians.

The last two columns of Table 1 show county-level results. County-level improvements in variance were not a specific objective of the GREG, so the appearance of modest improvements is an unexpected benefit. The explanation can be offered, however, that GREG estimation, implemented at the tract level, achieves a more representative geographic distribution of the ACS weights, which in turn may modestly improve the estimates of county-level characteristics.

Table 1 Comparison of overall variance summary measures from the Multiyear Estimates Study, for periods 2001-2005 and 2003-2005. Total housing units and total population are controlled at the county level, resulting in directly estimated variances of 0.

	5-yr tract		5-yr place		3-yr place		5-yr county	
	Total (6)	Median (5)	Total (6)	Median (5)	Total (6)	Median (5)	Total (6)	Median (5)
Total HU	0.11	0.09	0.25	0.70	0.15	0.15	-	-
Total Hhld	0.32	0.29	0.51	0.78	0.46	0.40	0.95	0.88
Total Pop	0.39	0.39	0.47	0.78	0.42	0.41	-	-
Age/sex	0.49	0.66	0.56	0.89	0.52	0.68	0.96	0.97
Race/Hisp.	0.55	0.94	0.61	0.97	0.58	0.95	0.93	0.99
Other	0.67	0.93	0.75	0.97	0.76	0.95	0.93	0.96

4. Can GREG Introduce Appreciable Bias?

Previous papers (Fay 2005a, 2005b, 2006) focused on the primary research question of whether, and to what extent, GREG could reduce ACS subcounty variances for multiyear estimates. The primary question has now been empirically answered in the affirmative: the results of the previous section, combined with the analyses by Starsinic and Tersine (2007) and by Fay (2007), all support the basic finding that GREG indeed lowers variance, at least for many of the ACS estimates.

The issue of bias has been a secondary concern, and the earlier papers each argued that GREG, when applied in the proposed manner, would not introduce *appreciable* bias. Särndal (1984) made an early argument of that form in advocating the view that model-assisted estimators should be applied to some small domain problems. Indeed, GREG's near unbiasedness would seem particularly appropriate for the ACS, as a successor to the decennial census long form. The long-form estimates based on raking-ratio estimation were also not entirely free from bias, but the biases were theoretically generally small relative to the long-form sampling variances. But bias is a natural concern to users, so the time has come to address concerns over bias in more detail.

In this section, the problem of bias will be approached from two perspectives: (1) guidance from the theoretical and empirical literature on GREG estimation, and (2) the specific circumstances of the proposed incorporation of GREG in ACS estimation. An empirical investigation using the data from the Multiyear Estimates Study then follows.

A General Perspective on the Question of Bias

Section 2 noted general remarks from Cochran (1963, p. 157) and Särndal, Swensson, and Wretman (1992, p. 251) concerning bias of the ratio estimator. Like the ratio estimator, general claims can be made for GREG that its bias will tend to be of lower order than the sampling variance for moderate to large samples.

The empirical evidence, although scattered, both supports the general claims but indicates that small biases do occur. An early simulation of Hidiroglou and Särndal (1985, p. 73) reported relative biases (=bias/standard error) to 2 decimal places based on 500 simulations of sampling from a known population; reported values were 0.00 for the 3 domains with mean sample takes of 10 or more, but non-zero relative biases appeared for most other domains. (A second study involved even smaller mean sample takes, with none over 10.)

Bankier, Houle, and Luc (1997) compared a preliminary and refined version of the 1996 Canadian Census weighting. For both versions, their Table 3 (p. 74) indicated national discrepancies with the population count generally on the order of tenths of 1 percent for the control variables that were candidates for the GREG estimation. Because of the large size of the sample involved, the small differences were presumably primarily a result of bias. They interpreted the observed differences primarily as arising from differential coverage in the census sample, but possible bias from the estimator could have also contributed.

Applications such as censuses and the ACS share the uncommon features of (1) applying an estimation approach at a fine-grained level to improve small domain reliability, and (2) indirectly estimating characteristics at higher levels through summation. Thus, small biases at the weighting area level potentially can aggregate to high levels where, although still relatively small, they grow in relation to the sampling variance.

Another feature of Canadian and U.S. census applications, and to a lesser extent the GREG approach studied here, is the use of data-dependent rules to select the specific regression variables or controlled variables within each weighting area. The application of data-dependent rules could be a source of small biases that would depend on the choice of rules. Hence, the literature provides general guidance that biases from appropriate application of GREG are generally small, but it may lack specific guidance on the size of the biases in complex applications such as the GREG implementation in the Multiyear Estimates Study.

Bias in the ACS Context

In contrast to the theoretical results, the proposal to imbed GREG in ACS weighting involves a specific placement in a multi-step estimation process. Because the actual variance and bias properties of an estimator depends on the context of its application, a more complete assessment of the proposed use of GREG requires a detailed account of its placement in the ACS weighting. Two previously cited sources provide particularly useful overviews of ACS weighting for this purpose. An ACS Technical Paper (U.S. Census Bureau 2006, ch. 11) summarizes ACS weighting and estimation in some detail, both for single- and multi-year estimates. Chapter 5 of the recent National Academy report (Citro and Kalton, 2007) reviews ACS weighting for single-year periods, but the same ideas are the basis for multiyear weighting as well. The two sources both overlap in their descriptions and complement each other: In many instances the National Academy report provides more detail or comments in more depth on the underlying reasoning than the technical paper, but on other aspects the reverse is true.

The two sources take similar perspectives on survey weighting, namely that weighting serves multiple purposes but usually only approximately achieves its goals. The Census Bureau report remarks (p. 11-1),

> The weights compensate for differences in sampling rates across areas, for differences between the full sample and the interviewed sample, and for differences between the sample and independent estimates of basic demographic characteristics (Alexander, Dahl, and Weidman, 1997).

Similarly, the National Academy report observes (p. 185)

> Survey sampling weighting methods are applied to the respondents for the given period in order that valid estimates can be produced. These methods include weights to compensate for unequal selection probabilities, weighting adjustments for nonresponse, and calibration adjustments that compensate for noncoverage and can improve the precision of some survey estimates.

Note that both sources use "compensate" to characterize the role of the estimation steps. The choice of the word "compensate" suggests an effort to improve without claiming perfection. Primarily, ACS weighting attempts to compensate for unequal probabilities of selection, the undercoverage of the MAF as a sampling frame, the effect of household nonresponse, and person undercoverage within sampled ACS households.

The Academy report organizes the Census Bureau's weighting procedures for ACS into 9 steps (Citro and Kalton, 2007, p. 186)

Step 1: Base weights.

Step 2: Variation in monthly response factor.

Step 3: Noninterview factors 1 and 2.

Step 4: Mode bias noninterview factor.

Step 5: Housing unit control factor 1.

Step 6: Population control factor.

Step 7: Housing unit control factor 2.

Step 8: Adjustments to eliminate extreme weights.

Step 9: Rounding of weights.

Step 1 computes the inverse probability of selection as the initial weight, including the effect of subsampling for CAPI followup. Step 2 attempts to compensate for variation in the monthly ACS interview sample size, which varies because it comprises a mixture of respondents who were initially selected in three different months, depending on the mode of response. Both steps 1 and 2 can be viewed as compensating for unequal probabilities of selection and observation, but there is also a contrast between the generally controlled application of probability sampling in step 1 and the respondent-governed pattern of response affecting step 2.

The noninterview adjustments in steps 3 share features with unit nonresponse adjustments commonly used in survey practice. The adjustment in step 4 is less standard; after summarizing the apparent intention of the step, the National Academy report (Citro and Kalton, 2007, p. 193) specifically questioned the underlying logic and choice of estimator for this step.

Steps 1 and 2 compensate for unequal probabilities of selection from the ACS frame, and steps 3 and 4 compensate for noninterviews or the unequal monthly distribution of interviewed cases. All four steps accept the size of the ACS frame as given. In contrast, Step 5 controls the ACS weights to agree with independent information about the number of housing units, which partially compensates for omissions from the ACS frame. For the Multiyear Estimates Study, the GREG step was imbedded after step 4 and before step 5. This placement of the GREG puts it at the last step where the general level of the ACS weights approximates the frame total, so that $\hat{X}$ can be viewed as an essentially unbiased estimate of X in *(2)* and *(4)*. After step 5, the weights attempt to represent the underlying population of housing units represented by the independent housing unit estimate. If in a particular county the ACS frame has a shortfall of housing units, the weights will generally be raised in step 5 to compensate.

Steps 6 and 7 attempt to compensate for undercoverage of the population within housing units and to achieve a consistent set of housing unit and individual weights. The final two steps refine the weights.

In the abstract, GREG estimation is nearly unbiased. In the context of the proposed use in ACS, GREG estimation is imbedded among estimation steps intended to compensate for different sources of potential bias. In this context, the proposal is to place GREG at the last step where the weights are consistent with the ACS frame, before ratio estimators adjust the weights toward auxiliary information from external sources.

5. Empirical Evidence of Bias

As noted in the introduction, the findings to be presented here detect some small patterns of bias from GREG. To outline the argument, the overall investigation begins with an initial analysis that parallels the variance analysis of section 3, comparing estimates with and without GREG for the 2001-2005 and 2003-2005 periods. The comparisons show larger differences apparently due to GREG in the 5-year period totals than in the 3-year period totals.

To provide additional evidence that the differences stem from bias, a second analysis compares the 3- and 5-year period estimates to the 1-year estimates using the same data. The second comparison provides somewhat limited evidence on the question, because there are differences between the 3- and 5-year period estimates and the 1-year estimates that occur even if GREG is not used. The analysis provides perspective on how the ACS estimates can be affected by other subtle changes in estimation procedures or controls.

A third analysis compares the consistency of the three 5-year period estimates and the five 3-year period estimates with each other. Linear regression can be used to estimate equivalent 1-year totals, which may be thought of as a form of pseudo-values. The third analysis produces independent evidence supporting the conclusions from the first.

Two primary measures will be used to summarize differences over a set *S* of estimates. One, the *percent absolute difference rate* (*PAD),* will be computed over a set *S* as

$$PAD_S = \frac{100}{n}\sum_S \frac{|\hat{Y}_{GREG} - \hat{Y}_{noGREG}|}{(\hat{Y}_{GREG} + \hat{Y}_{noGREG})/2} \qquad (7)$$

where *n* denotes the number of elements in *S.* The PAD is a simple average over *S* of the individual absolute percent differences. A second, termed here the *weighted percent absolute difference (WPAD),* resembles *(6)*

$$WPAD_S = 100\frac{\sum_S |\hat{Y}_{GREG} - \hat{Y}_{noGREG}|}{\sum_S (\hat{Y}_{GREG} + \hat{Y}_{noGREG})/2} \qquad (8)$$

Comparisons for 2001-2005 and 2003-2005

Table 2 presents measures *(7)* and *(8)* comparing estimates with and without GREG for the 5-year period 2001-2005 at different geographic levels, including the difference for the total over all 34 test sites. Differences at the tract level are substantially diminished at the county level. Paralleling the patterns from section 3, there is little difference between *(7)* and *(8)* for housing units, households, and total population, but weighting the absolute percent differences has more effect on the remaining sets of variables.

Table 2. Average absolute percent change with and without GREG, computed on an unweighted and weighted basis, for the 2001-2005 5-year period estimates, ACS Multiyear Estimates Study

	Tracts	Counties	Total for Test Sites
PAD (Unweighted)			
Housing units	4.56	0	0
Households	4.66	0.33	0.22
Population	5.64	0	0
Age/sex	8.53	0.24	0.01
Race/Hispanic origin	7.61	5.12	0.82
Other estimated totals	8.82	2.67	0.51
WPAD (Weighted)			
Housing units	3.92	0	0
Households	4.00	0.23	0.22
Population	4.96	0	0
Age/sex	5.30	0.03	0.00
Race/Hispanic origin	6.58	0.20	0.06
Other estimated totals	6.17	0.45	0.27

The small values of PAD, *(7),* and particularly WPAD, *(8),* when summed to the test site level, seem to invite the conclusion that there is no evidence of systematic bias. Yet, further analysis is warranted. For households, WPAD is .23 for counties but almost the same, .22, when summed over counties. In fact, GREG raises the estimated number of households in 29 out of the 34 counties, a simple non-parametric indication of systematic shift. This increase in estimated households occurs without any shift in total housing units or total population.

An associated shift appears for vacant housing units. Households, that is, occupied housing units, and vacant housing units sum to the total number of housing units. Because ACS weighting forces the estimated total number of housing units to agree with the independent control total, the estimated number of vacant units must decrease. Indeed, GREG lowers the estimated total vacant units by 2.03%, making vacant units one of the most affected variables of its size.

Other variables stand out with particularly strong non-parametric evidence, such as the number of persons who live in a different house in the U.S. year ago, which drops in 31 out of 34 counties; the number of persons who live in a different house in the same county, which also drops in 31 counties; and the number of renter occupied units, which drops in 30 counties. GREG increases other variables increase, such as the number of owner-occupied housing units (31 out of 34), and the number of households with 1 or more children under 18 (27 out of 34).

Table 3 presents a similar analysis to the one in Table 2 for the 30 counties with published 3-year estimates, omitting the column for tracts (available only for the 5-year estimates). Comparison of the two tables indicates that the 3-year period estimates show less change than the 5-year. Although the 3-year period estimates are based on fewer sample cases than the 5-year, there were far fewer weighting areas, so 3-year weighting areas were each based on more data.

Table 3. Average absolute percent change with and without GREG, computed on an unweighted and weighted basis, for the 2003-2005 3-year period estimates, ACS Multiyear Estimates Study

	Counties	Total for Test Sites
Unweighted		
Housing units	0	0
Households	0.18	0.03
Population	0	0
Age/sex	0.20	0.02
Race/Hispanic origin	3.37	0.56
Other estimated totals	1.68	0.15
Weighted		
Housing units	0	0
Households	0.07	0.03
Population	0	0
Age/sex	0.02	0.00
Race/Hispanic origin	0.09	0.04
Other estimated totals	0.24	0.07

Unlike the 5-year data, simple sign tests do not strongly confirm the presence of systematic bias for vacant units (18 out of 30 decreases). Rather than interpret the null finding as an absence of bias in the 3-year period estimates, however, it is possible that the sign test has insufficient power to detect the presence of small differences. Instead, two further analyses were attempted.

Comparisons to 1-Year Data

Out of the 34 counties in the Multiyear Estimates Study, 19 met the threshold of 65,000 population for publication of 1-year estimates. An analysis was performed at the county level, where county-level estimates are essentially unaffected by geographic changes, while sub-county geographic shifts are far more common. Although 3- and 5-year period estimates are not the simple average of 1-year estimates, a second analysis measured differences between the county-level 3-year period estimates for 2000-2002, 2001-2003, 2002-2004, and 2003-2005 and estimates that can be constructed as the simple averages of the corresponding 1-year estimates during the period 2000-2005. A similar approach was taken to the 2000-2004 and 2001-2005 period estimates. Because cost of living adjustments complicate the analysis of income and other monetary variables, all variables affected by these adjustments were dropped from the second analysis.

As previously implied, none of the 1-year estimates incorporates GREG estimation. If the 5-year period estimates of totals for 2001-2005 at the county level without GREG matched the corresponding simple average of the 1-year estimates for this period, then all differences between the estimates with and without GREG could be attributed to the GREG. With this outcome, the 1-year data could also be used to analyze the 2000-2004 5-year estimates, as well as the four sets of 3-year estimates just listed.

In fact, the findings are more complex. Even without GREG, some large differences can be noted between the 5-year period estimate for 2001-2005 and the simple average of the 1-year estimates. For example, there is a difference of about 1.75% for age 20-24, a cell where the 2001-2005 period estimates with and without GREG are virtually identical. Similarly, the average of the 1-year estimates differ from the 2001-2005 estimates without GREG by about 0.44% for Black (single race) and 0.45% for total Hispanic. The estimates for total population, however, differ by a mere 0.03%. As noted in the documentation for the MYES (U.S. Census Bureau, undated), the controls used for the multiyear estimates are not the simple average of the controls used for the corresponding 1-year estimates; instead, the multiyear controls average the population estimates (possibly revised) available during the last year of the period. These inconsistencies, although minor, may be typical of discrepancies that users will face in the future in trying to compare 1-year to multiyear estimates.

Regression Analysis

Although the use of the 1-year estimates does not yield a definitive analysis of systematic differences introduced by GREG, the attempt to anchor period estimates back to 1-year values motivates the third analysis. Consider a set β_{1999}, β_{2000}, …, β_{2005}, of 1-year estimates such that any 3-year estimate could be expressed as the appropriate average. For example,

$$Y_{2003-2005} \quad = \quad X_{2003}\,\beta_{2003} + X_{2004}\,\beta_{2004} + X_{2005}\,\beta_{2005}\ ,$$

where $X_{2003} = X_{2004} = X_{2005} = 1/3$ and (not shown) $X_{1999} = X_{2000} = X_{2001} = X_{2002} = 0$. Suppose that 5-year period estimates are similarly expressed, but with an added term to represent the difference in bias between 3- and 5-year period estimates,

$$Y_{2001-2005} \quad = \quad X_{2001}\,\beta_{2001} + X_{2002}\,\beta_{2002} + X_{2003}\,\beta_{2003} + X_{2004}\,\beta_{2004} + X_{2005}\,\beta_{2005} + I_5\,\beta_5\ ,$$

where $X_{2001} = X_{2002} \ldots = X_{2005} = 1/5$ and (not shown) $X_{1999} = X_{2000} = 0$, and $I_5 = 1$, as an indicator variable for 5-year periods. Using regression through the origin, the five sets of 3-year period estimates with GREG combined with the three sets of 5-year period estimates with GREG provide 8 points, enough to estimate all 8 coefficients, although leaving no residual degrees of freedom. Of course, the estimation of the coefficients in this instance could also be accomplished as standard linear algebra. The regression analysis was applied to the sums, *Y,* for the 19 counties above the threshold of 65,000 population. As in the previous analysis, items based on income or other monetary amounts were omitted.

For purposes of analysis, the estimates were divided into 3 broad strata: (1) 99 with sums over the 19 counties of 1,000,000 or greater, (2) 137 with sums 100,000 or greater but below 1,000,000, and (3) the remaining 65. From the first stratum, the following had the algebraically smallest values of $\beta_5/\ \beta_{2005}$ (less than –0.0030), in other words, where the 5-year GREG estimation appears to depress the estimate relative to the 3-year (algebraically increasing order, beginning at –0.0078)

1. 1 year ago, lived in same county, different house
2. 1 year ago, lived in different house in the U.S.
3. Foreign-born, not U.S. citizen
4. Year householder moved into unit: 2000 or later
5. Foreign born, Latin America
6. Speaks a language at home other than English, speaks English less than very well
7. Speaks Spanish at home

The following had the highest values of $\beta_5/\ \beta_{2005}$ (algebraically decreasing from 0.0056 to 0.0030)

1. Housing units with a mortgage
2. Owner-occupied
3. Households with 1 or more people under 18
4. Family households with own children under 18
5. Married couple families
6. Households with 2 vehicles available
7. Gas heating
8. Family households
9. Persons age 25+ with some college education, no degree

Without providing a comparable list for the second stratum, it is worth noting that the largest algebraic difference in the second stratum for $\beta_5/\ \beta_{2005}$ occurs for vacants (-0.0162).

The preceding two lists of variables are offered as clues to patterns of possible bias, stopping short of a claim of statistical significance for each of the individual entries. The first list generally suggests high mobility and perhaps an increased risk of not being represented by an administrative record. Similarly vacant units would be less likely to be associated with administrative records. The second list generally suggests stability.

A remarkable consistency can be observed between the results from the first analysis and the third. The results of the 2001-2005 comparison of with GREG vs. without GREG were similarly stratified (with the same boundaries, but by the size of the 2000-2004 period estimate) and classified by direction (GREG raises vs. GREG lowers relative to without GREG). In all cases, the variables in the first list all had estimates with GREG lower than without. In a few cases, the variables were classified in the second size stratum, which had a total of 64 variables where GREG lowered the estimate, rather than the first stratum, which had a total of 31 variables where GREG lowered the estimate. In fact, the variables were among the largest absolute percentage changes. Table 4 reports the results from the 2001-2005 comparison of GREG vs. no GREG for the set of variables selected on the basis of the third analysis. The signed direction of the shift is entirely consistent between the third and first analysis. In most but not all cases, the variables identified in the third analysis are associated with large relative ranks in the first.

Table 4 also shows the results for the 2003-2005 comparison. With only one exception, the signs are again consistent and in the same direction as the 5-year period comparison. In other words, the biases for the 3-year estimates are in the same direction, but smaller.

Table 4 Comparison of the regression coefficient β_5 in the third analysis with comparisons of GREG vs. no GREG for the 2001-2005 and 2003-2005 period estimates.

	Regr.	2001-2005 GREG vs. no GREG		2003-2005 GREG vs. no GREG	
	β_5	% diff.	Rank in group	% diff.	Rank in group
1 year ago, lived in same county, different house	-0.78	-1.14	61/64	-0.20	57/73
1 year ago, lived in different house in the U.S.	-0.63	-1.12	31/31	-0.23	44/44
Foreign-born, not U.S. citizen	-0.52	-0.44	37/64	-0.06	22/73
Year householder moved into unit: 2000 or later	-0.48	-0.59	30/31	0.02	14/51
Foreign born, Latin America	-0.39	-0.26	25/31	-0.15	46/73
Speaks a language at home other than English, speaks English less than very well	-0.37	-0.38	28/31	-0.03	26/44
Speaks Spanish at home	-0.34	-0.26	24/31	-0.08	38/44
…					
Housing units with a mortgage	0.30	0.35	56/66	0.10	43/50
Owner-occupied	0.32	0.37	58/66	0.01	18/50
Households with 1 or more people under 18	0.35	0.45	59/66	0.08	41/50
Family households with own children under 18	0.40	0.51	61/66	0.06	40/50
Married couple families	0.45	0.58	63/66	0.05	37/50
Family households with own children under 18	0.45	0.52	62/66	0.00	14/50
Households with 1 or more people under 18	0.46	0.48	60/66	0.00	5/50
Owner-occupied	0.53	0.78	66/66	0.13	48/50
Housing units with a mortgage	0.56	0.76	65/66	0.17	50/50

To summarize the evidence, in the first analysis the consistency of the direction of bias at the county level provides strong evidence that the observed differences are due to bias, although the evidence does not immediately address the question of whether the bias should be attributed to GREG or not. In fact, the third analysis shows that GREG produces a bias in the same direction for both the 5- and 3-year data, but the bias is generally quite small and less consistent for the 3-year data.

The first and third analyses both employ the 2003-2005 and 2001-2005 estimates with GREG, but they measure bias by comparison to different sources. The first analysis employs the alternative estimates for the same period without GREG. The third analyses uses only estimates with GREG, but it looks at all 8 available sets and contrasts the 3-year period estimates with the 5-year. Table 4 summarizes the convergence of this evidence on the existence and direction of bias from GREG, so that the combination of the two analyses is stronger than either of them separately.

6. Discussion

The Multiyear Estimates Study provides a clear indication of the impact of GREG estimation in reducing variance for ACS subcounty estimates for multiyear periods. The improvements are far from uniform but rather are more pronounced for larger estimates than small ones. This uneven distribution of improvements is evident from differences between the two summary measures. Thus, the improvements from GREG estimation are not enough to overturn a basic point made by the recent National Academy report (Citro and Kalton, 2007), namely, that users of ACS 5-year period estimates will face larger sampling variances than previously offered by recent decennial censuses. The GREG estimation will partially mitigate but not eliminate this issue.

Although the author had hoped not to find any practical traces of bias resulting from GREG, the analysis presented here does so. The quantified measures suggest that the bias is not large—generally less than 1% for almost all variables—and it is far more evident in the 5-year than the 3-year period estimates. The detail presented in the previous section was intended both to present the evidence for the existence of a bias and to provide a basis for a possible attack, either by the author or subsequent researchers. There is a strong possibility the bias may be reduced to negligible levels through further diagnosis of its source followed by a modification of the estimator. On the other hand, the actual effect of the bias may be sufficiently small that its continued presence may not be detrimental for any practical use of the ACS data.

For ACS statistical products to enjoy the widespread use among the American public previously accorded the decennial census data, they should be relatively accessible to those with modest degrees of numerical literacy. It is therefore an advantage of GREG that it can be imbedded into the ACS without requiring any additional understanding of its properties beyond what the ACS generally requires. Users who can work with reported confidence intervals or margins of error may continue to do so without attending to the presence of GREG estimation in the background—the variance reductions from GREG can simply be built into the confidence intervals or margins of error. Although it would be desirable to reduce the bias that can be detected in the 5-year estimates, if this did not occur, the level of bias is at a sufficiently low level that for most applications it can also be ignored.

[1] *This report is released to inform interested parties of research and to encourage discussion. Any views expressed on statistical and methodological issues are those of the author and not necessarily those of the U.S. Census Bureau. The author thanks Michael Starsinic and Tommy Wright for their comments on an earlier version of the paper.*

References

Bankier, M.D., Rathwell, S., and Majkowski, M. (1992), "Two Step Generalized Least Squares Estimation in the 1991 Canadian Census," *Proceedings of the Survey Research Methods Section,* American Statistical Association, pp. 764-769.

Bankier, M., Houle, A.-M., and Luc, M. (1997), "Calibration Estimation in the 1991 and 1996 Canadian Censuses," *Proceedings of the Survey Research Methods Section,* American Statistical Association, pp. 66-75.

Bankier, M. and Janes, D. (2003), "Regression Estimation of the 2001 Canadian Census," *Proceedings of the 2003 Joint Statistical Meetings on CD-ROM,* American Statistical Association, pp. 442-449.

Citro, C.F. and Kalton, G. (eds.) (2007), *Using the American Community Survey: Benefits and Challenges, Panel on the Functionality and Usability of Data from the American Community Survey,* National Research Council, Committee on National Statistics, Division of Behavioral and Social Sciences and Education, The National Academies Press, Washington, DC.

Cochran, W.G. (1963), *Sampling Techniques,* second edition, John Wiley & Sons, Inc., New York, NY.

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

______ (2005b), "Potential Applications of Model-Assisted Estimation to Demographic Surveys in the U.S.," paper presented at the Federal Committee on Statistical Methodology Research Conference, available from www.fcsm.gov/05papers/Fay_IIIC.pdf.

______ (2006), "Using Administrative Records with Model-Assisted Estimation for the American Community Survey," *Proceedings of the 2006 Joint Statistical Meetings on CD-ROM,* American Statistical Association, Alexandria, VA, pp. 2995-3001.

______ (2007), "Imbedding Model-Assisted Estimation into ACS Estimation," presented at the Joint Statistical Meetings, Salt Lake City, UT, 29 July-2 Aug. 2007.

Hidiroglou, M.A. and Särndal, C.E. (1985), "An Empirical Study of Some Regression Estimators for Small Domains," *Survey Methodology, 11,* 65-77.

Rao, J.N.K. (2003), *Small Area Estimation,* John Wiley, New York.

Särndal, C.-E. (1984), "Design-Consistent Versus Model-Dependent Estimation for Small Domains," *Journal of the American Statistical Association, 79,* 624-631.

Särndal, C.-E., Swensson, B., and Wretman, J. (1992), *Model Assisted Survey Sampling,* Springer-Verlag, New York, NY.

Starsinic, M. (2005), "American Community Survey: Improving Reliability for Small Area Estimates," *Proceedings of the 2005 Joint Statistical Meetings on CD-ROM,* American Statistical Association, pp. 3592-3599.

Starsinic, M. and Tersine Jr., A. (2007), “Analysis of Variance Estimates from American Community Survey Estimates,” prepared for presentation at the Joint Statistical Meetings, Salt Lake City, UT, 29 Jul-2 Aug 2007.

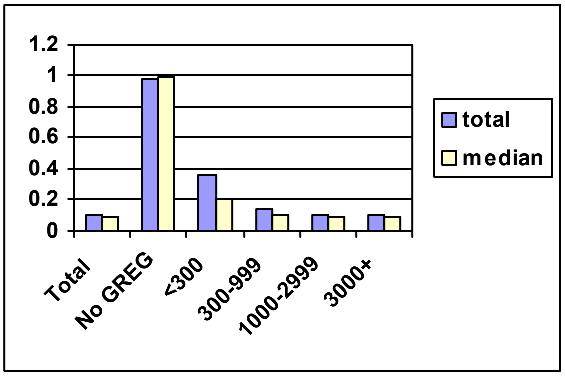
U.S. Census Bureau (2006), “Design and Methodology—American Community Survey,” Washington, DC: U.S. Department of Commerce, issued May 2006, from www.census.gov/acs/www/Downloads/tp67.pdf.

______ (undated), “Accuracy of the Data (Multi-Year Estimates Study),” downloaded 10 Oct. 2007, available at http://www.census.gov/acs/www/Downloads/Accuracy_of_the_Data_MYE.pdf.

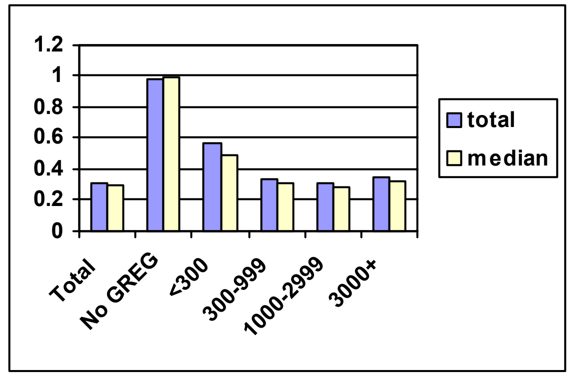
Van Auken, P.M., Hammer, R.B., Voss, P.R., and Veroff, D.L. (2004), “American Community Survey and Census Comparison, Final Analytical Report, Vilas and Oneida Counties, Wisconsin; Flathead and Lake Counties, Montana,” unpublished report dated March 5, 2004, available at http://www.census.gov/acs/www/AdvMeth/acs_census/lreports/vossetal.pdf.

______ (2006), “The American Community Survey in counties with ‘seasonal’ populations,” *Population Research and Policy Review*, *25,* 275-292.

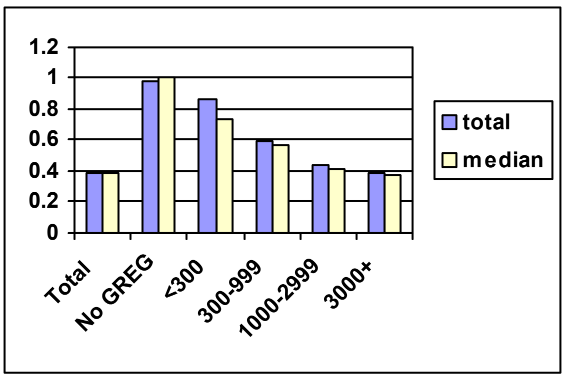
a. Total housing units



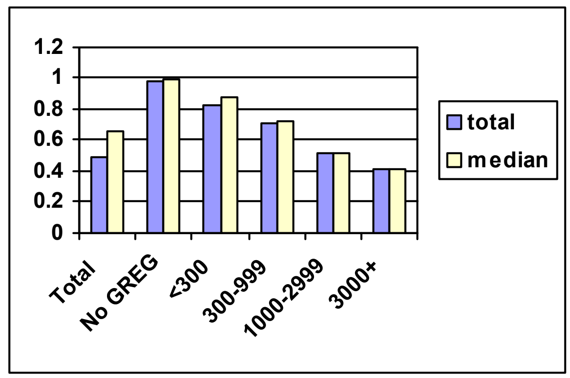
b. Total households



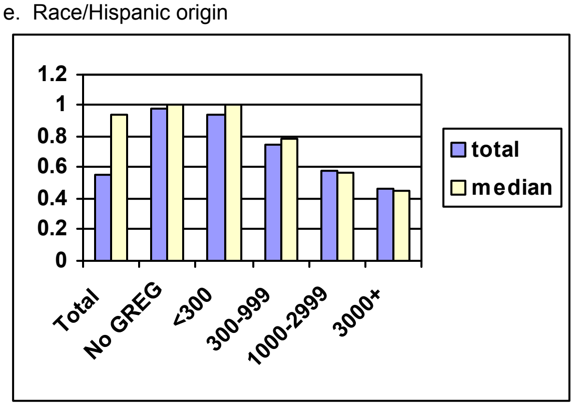
c. Total persons



d. Age/sex



e. Race/Hispanic origin



f. Remaining estimated totals

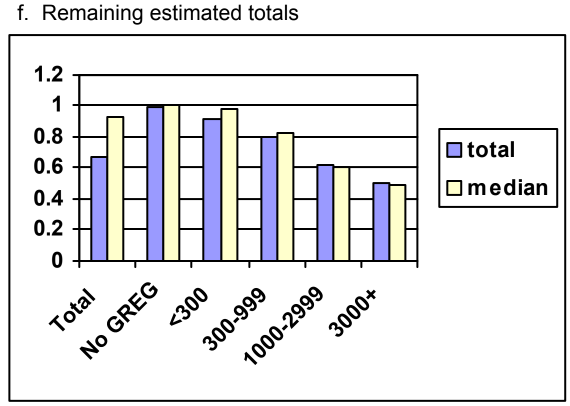


Fig. 1 Variance comparisons at the tract level in 34 ACS test counties, 2001-2005. In each case, results are shown separately for the 45 small tracts where GREG was not implemented ("No GREG"). The x-axis groups tracts by the size of the tract-level estimate, with the result for all tracts on the left. Two measures, $ratio_{tot}$ (total) and $ratio_{med}$ (median) are compared. The results are for (a) total housing units, (b) total households, (c) total population, (d) age/sex, (e) race/Hispanic origin, (f) all remaining totals published in the ACS profiles.

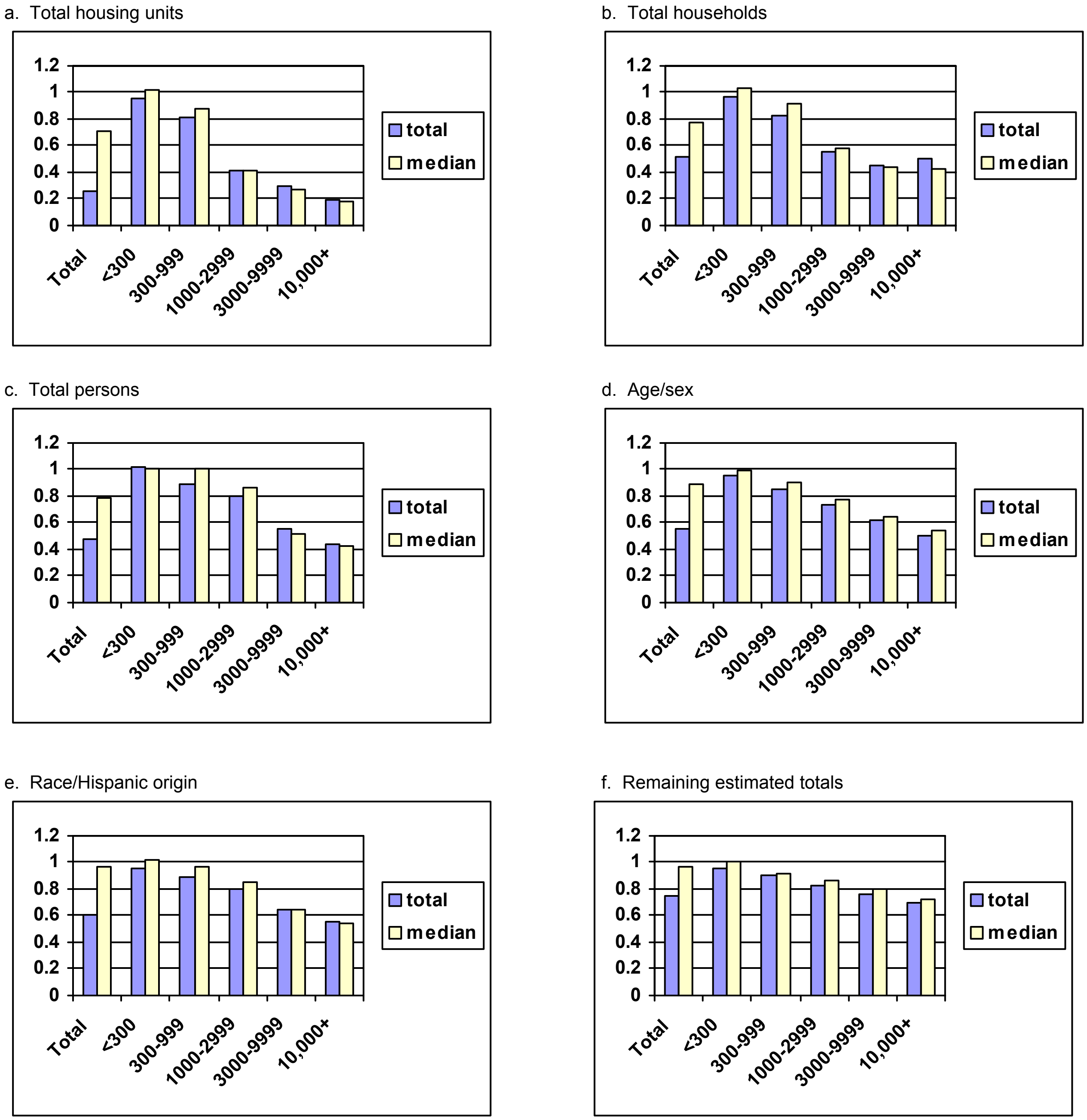
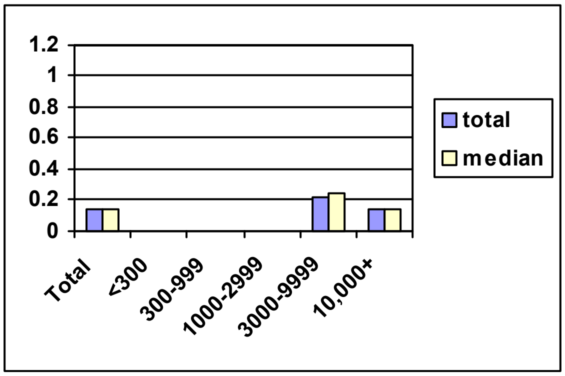
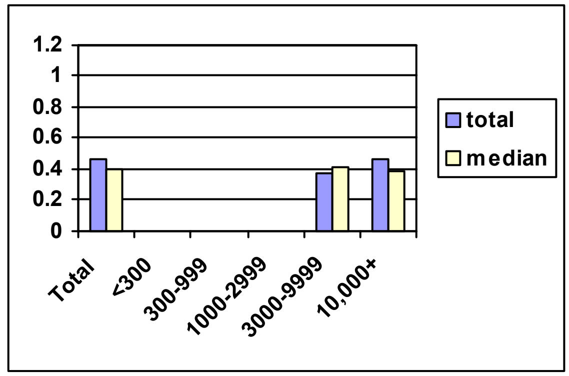


Fig. 2 Variance comparisons at the place level in 34 ACS test counties, 2001-2005. The x-axis groups places by the size of the place-level estimate. Two measures, $ratio_{tot}$ (total) and $ratio_{med}$ (median) are compared. The results are for (a) total housing units, (b) total households, (c) total population, (d) age/sex, (e) race/Hispanic origin, (f) all remaining estimated totals published in the ACS profiles.

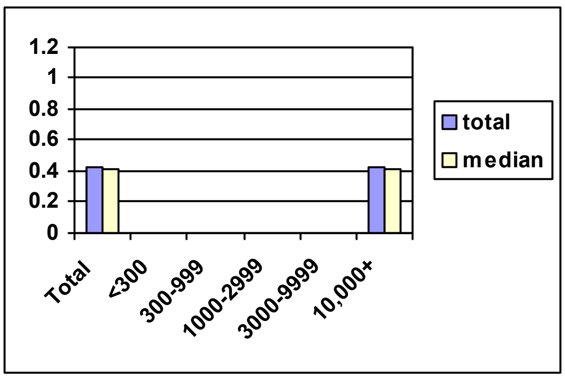
a. Total housing units



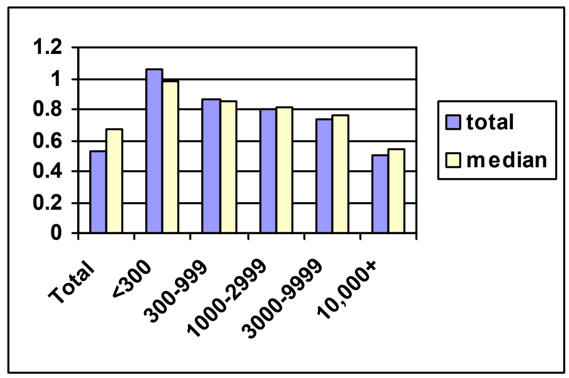
b. Total households



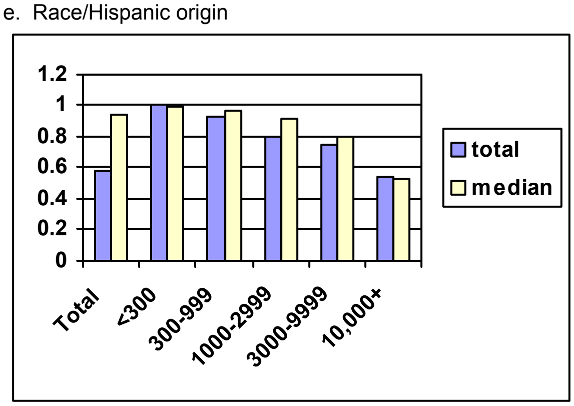
c. Total persons



d. Age/sex



e. Race/Hispanic origin



f. Remaining estimated totals

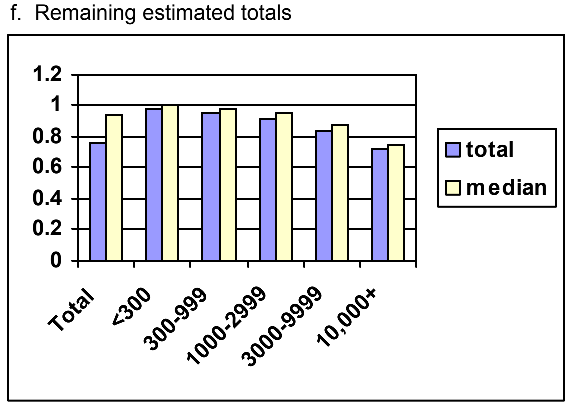


Fig. 3 Variance comparisons at the place level in 34 ACS test counties, 2003-2005. The x-axis groups places by the size of the place-level estimate. Two measures, $ratio_{tot}$ (total) and $ratio_{med}$ (median) are compared. The results are for (a) total housing units, (b) total households, (c) total population, (d) age/sex, (e) race/Hispanic origin, (f) all remaining estimated totals published in the ACS profiles.

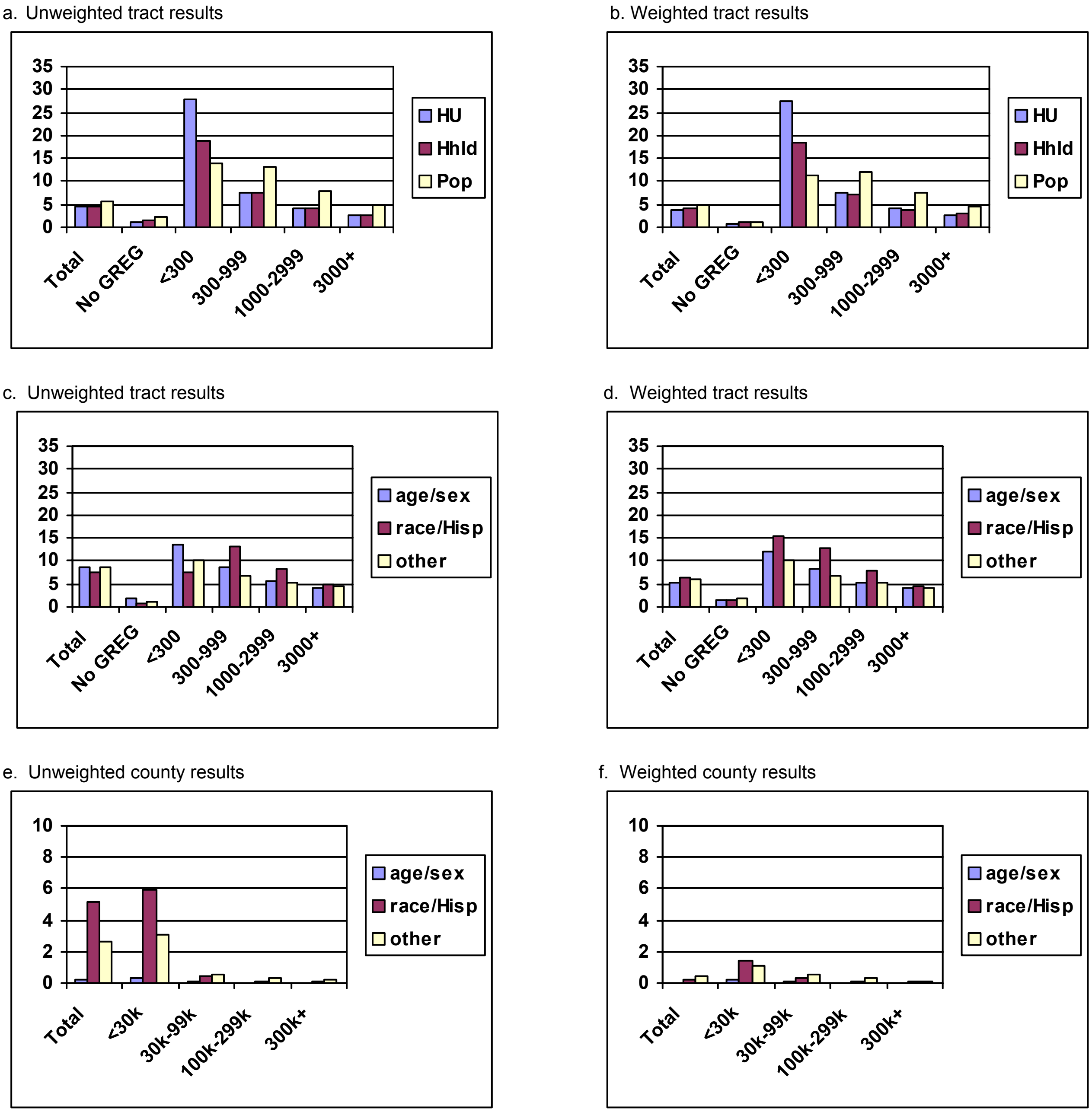


Fig. 4 Effect of GREG step on estimates of total at the tract and county levels. (a), (c):Average absolute percent changes for tracts are shown for (a) total housing units, total households, and total persons, and for (c) age/sex, race/Hispanic origin, and the remaining 294 estimated totals from the profiles. (b), (d): Weighted average of absolute percent changes for the same universes as (a) and (c). Weighting has the most notable effect on the measures for total in (d) compared to (c), while not substantially changing the outcome within individual size categories. (e)-(f): Estimates at the county level show far less relative change (note change of scale on y-axis). Distributions by age/sex show virtually negligible changes. An unweighted analysis (e) suggests moderately large changes overall and for estimates below 30k (30,000), but weighting (f) shrinks these substantially.