DRAFT

IMPACTS OF SAMPLE SIZES IN THE AMERICAN COMMUNITY SURVEY

by Ronald Eash
Northwestern University Transportation Center
600 Foster Street
Evanston, Illinois 60205

INTRODUCTION

The Census Bureau no longer intends to collect long-form data from a large sample of housing units during decennial censuses; however, roughly comparable long-form data will be available from the American Community Survey (ACS). The ACS is a continuous Census Bureau project that will annually survey a smaller national sample (approximately 3,000,000 or 2.5 percent) of housing units. Full implementation of the ACS is now underway with the initial monthly mailing of 250,000 questionnaires to housing units completed in January 2005. Surveying of persons residing in group quarters is anticipated in FY 2006 (*1*).

Tables of population and housing unit estimates from the ACS are expected to be similar in content to Census 2000 Summary File 3 (SF 3) tables. Yearly ACS estimates will be available for geographic area summary levels with populations greater than 65,000 in 2006 and beyond. Annual ACS estimates for areas between 20,000 and 65,000 population will be three-year averages and initially available in 2008. Small area census tract and block group ACS estimates will be five-year averages beginning in 2010.

If the ACS is implemented as envisioned by the Census Bureau (*2*), then the ACS estimates will generally be of lower quality than past decennial census long-form estimates due to smaller housing unit samples, even with three and five year sample accumulations. Reduced housing unit samples will increase the standard errors and confidence intervals associated with census long-form estimates used in socioeconomic forecasts, travel model calibration, and model validation by Metropolitan Planning Organization (MPO) and state Departments of Transportation (SDOT) staff. Transportation planners will be less confident in the future that long-form sample estimates are close to population values, and may find it difficult to attribute year-to-year changes in ACS estimates to either actual changes or sampling errors. Introduction of the ACS will also impact the Public Use Microdata Samples (PUMS) and the Census Transportation Planning Package (CTPP) special tabulation of long-form data, two census products widely used by MPO and SDOT staff for travel model development and validation.

Census 2000 Long-Form and ACS Housing Unit Samples

In Census 2000, the great majority of housing units were identified by address. A sample of these housing units received a long-form questionnaire through the mail or in-person from a census enumerator. The fraction of housing units receiving the long-form varied so that housing

[1] Census Bureau. President's FY 06 Budget Requests $169.9 Million for ACS. American Community Survey Alert, Number 30. Available at http://www.census.gov/acs/www/. February 25, 2005.
[2] Census Bureau. *American Community Survey Operations Plan.* Release 1, March 2003.

units in lightly populated areas were sampled at a higher rate (*3*). Overall, approximately one out of six housing units was included in the 2000 national long-form sample. One in six persons residing within group quarters also received long-form questionnaires in Census 2000.

Table 1 compares the Census 2000 long-form and ACS housing unit sampling procedures. The ACS rates in this table are based on the housing unit sampling outlined in the ACS Operations Plan (*4*), later modified to increase the sampling rate in the very smallest governmental units, tracts, and American Indian reservations (*5*). The fully implemented ACS will also obtain long-form data from approximately 2.5 percent of the individuals who reside in group quarters. It should be emphasized that the housing unit samples in Table 1 equal Census 2000 census questionnaires distributed and addresses in the ACS housing unit sample, not the completed questionnaires contributing to the sample estimates.

Table 1 Census 2000 and American Community Survey Housing Unit Samples

Census 2000		American Community Survey	
Area Type	Sample	Area Type	Sample
Blocks in Smallest Governmental Units (Less Than 800 Occupied Housing Units)	1 in 2	Blocks in Smallest Governmental Units (Less Than 200 Occupied Housing Units)	1 in 10
		Blocks in Smaller Governmental Units (Between 200 and 800 Occupied Housing Units)	1 in 13.3
Blocks in Small Governmental Units (Between 800 and 1200 Occupied Housing Units)	1 in 4	Blocks in Small Governmental Units (Between 800 and 1200 Occupied Housing Units)	1 in 26.7
Blocks in Large Tracts (More Than 2000 Occupied Housing Units)	1 in 8	Blocks in Large Tracts (More Than 2000 Occupied Housing Units)	1 in 53.3
All Other Blocks	1 in 6	All Other Blocks	1 in 40

When data from the ACS are accumulated over five years, the smallest governmental units (those with less than 200 occupied housing units) will have roughly the same number of sampled housing units as the decennial census. In most remaining areas, a three-year accumulation amounts to 45 percent of the decennial census sample, while a five-year accumulation roughly equals 75 percent of the decennial census sample.

ACS Housing Unit Data Collection and Processing

Annual ACS data is collected in twelve consecutive three-month cycles, as shown in Figure 1 reprinted from an ACS implementation report (*6*). Each of these data collection cycles includes housing unit sample selection, questionnaire mail-out and mail-back, Computer Assisted Telephone Interviewing (CATI), and Computer Assisted Personal Interviewing (CAPI).

---

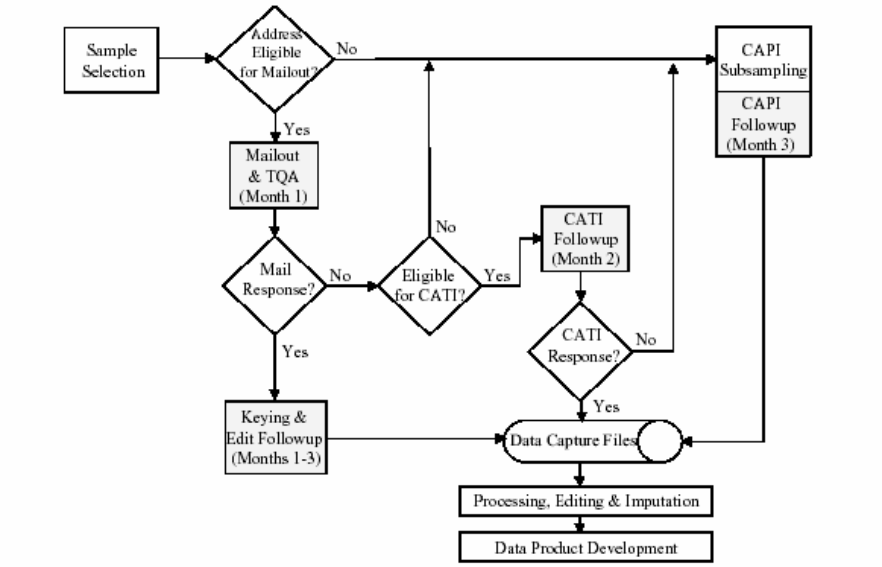
[3] Census Bureau. *2000 Census of Population and Housing: Summary File 3 Technical Documentation.* August 2004, pp 8-3,8-4.

[4] Ibid. *American Community Survey Operations Plan,* p 12.

[5] Census Bureau. Accuracy of the Data (2003). Available at http://www.census.gov/acs/www/Downloads/ACS/accuracy2003.pdf, p 4. (Accessed February 2005).

[6] Census Bureau. *Meeting 21st Century Demographic Data Needs—Implementing the American Community Survey. Report 2: Demonstrating Survey Quality,* May 2002, p 9.

FIGURE 1 ACS Data Collection, Capture, and Processing



Housing unit samples are selected county by county from a Master Address File using the previously noted sampling rates. Questionnaires are sent to housing units with usable mailing addresses, while those without mailing addresses go directly to the CAPI follow-up. Returned questionnaires - self-completed long-forms and long-forms completed with the aid of Telephone Questionnaire Assistance (TQA) – are then edited for completeness. Non-responding housing units with a known phone number pass to the CATI follow-up. All remaining non-responding households after CATI follow-up are sent to the pool of housing units for CAPI follow-up.

Table 2 CAPI Housing Unit Sample

CAPI Eligible Housing Unit	Sample
Without Mailing Address	2 in 3
In Tracts with Response Rate	
Less Than 35 Percent	1 in 2
Between 35 and 50 Percent	2 in 5
Rate Between 50 and 60 Percent	1 in 3
Greater Than 60 Percent (Initial Mail-Out Reduced by 8%)	1 in 3

CAPI sampling rates are listed in Table 2. These rates were recently adjusted from those published in the ACS Operation Plan (*7*,*8*). Two-thirds of the housing units without mailing addresses and one-third to one-half of housing units that remain after mail-back and CATI are sampled by CAPI. The CAPI phase completes the ACS data collection.

Results from the three prototype ACS surveys completed in 2000-2002 (the Census 2000 Supplementary Survey and similar surveys in 2001 and 2002) were used to estimate completion rates for the three phases of ACS data collection and the proportions of completed interviews from each of the phases. Although the sampling procedures in these three test ACS surveys are slightly different from the fully implemented ACS sampling procedures shown in Table 1 and Table 2, these prototype surveys are similar enough for general estimates of mail-back, CATI, and CATI completion rates.

[7] Ibid. *American Community Survey Operations Plan,* p 14.
[8] Dave Hubble. American Community Survey Research Report: Differential Sub-Sampling in the Computer Assisted Personal Interview Sample Selection in Areas of Low Completion Rates. 2005 ACS Documentation Memorandum Series, ACS-DOC-2, Internal census memorandum.

Table 3 lists the number of housing units in the sample, mailed questionnaires, returned questionnaires, and housing units selected for CATI and CAPI follow-ups for the prototype ACS surveys (*9*). Since the 2002 survey was curtailed due to budget limitations (one of the three-month data collection cycles was omitted and an additional cycle was truncated by eliminating CATI and CAPI follow-ups) the last two columns in Table 3 expand the 2002 results to approximate annual results by maintaining the same completion rates for the three survey phases.

Table 3 Data Collected in 2000, 2001, and 2002 Prototype ACS

	2000		2001		2002		2002 (12 Month)	
	Number	%	Number	%	Number	%	Number	%
Housing Unit Sample	891,000	100.0%	858,000	100.0%	742,000[b]	100.0%	810,000	100.0%
Mailed Questionnaires	850,000	95.5%	817,000	95.3%	706,000	95.1%	770,000	95.1%
Mailed Back Responses	441,000	49.5%	414,000	48.3%	364,000	49.0%	397,000	49.0%
Not Useable[a]	51,000	5.7%	19,000	2.2%	16,000	2.2%	18,000	2.2%
CATI Workload	228,000	25.6%	253,000	29.4%	229,000[c]	30.8%	275,000	33.9%
Eligible for CATI	*49.8%*		*47.3%*		*43.6%*		*43.6%*	
CATI Completion Rate	*57.3%*		*65.0%*		*63.5%*		*63.5%*	
CATI Interviews	65,000	7.3%	78,000	9.1%	63,000	8.5%	76,000	9.3%
CAPI Workload	144,000	16.1%	136,000	15.9%	107,000[c]	14.4%	128,000	15.8%
CAPI Completion Rate	*92.4%*		*94.5%*		*95.4%*		*95.4%*	
CAPI Interviews	133,000	14.9%	129,000	15.0%	102,000	13.7%	122,000	15.1%
Completed Interviews	588,000	66.0%	602,000	70.1%	513,000	69.1%	577,000	71.2%

a.. Estimated incomplete due to lack of edit follow-up, errors, and other reasons
b.. 11 months
c. 10 months

Even though the characteristics for all three prototype ACS surveys are similar, some additional points need to be brought out regarding the Table 3 figures. Publicity surrounding Census 2000 undoubtedly helped mail-back results in the 2000 survey, which are slightly higher than average. Increased CATI workloads in 2002 are due to additional telephone numbers obtained from Census 2000 datasets. CAPI workload figures are not the total number of housing units eligible for CAPI after completion of the mail-back and CATI phases, but the CAPI sampled housing units. Completed questionnaires for the three surveys were estimated using published CATI and CAPI completion rates (*10*). Based on these results, roughly two-thirds of all completed ACS questionnaires can be expected to be mail-back questionnaires; about 20 percent will be CAPI follow-up interviews and the remaining 13 percent CAPI follow-up interviews.

Estimation of Standard Errors in ACS

Sampling and non-sampling errors are inherent in the long-form estimates from the decennial census and ACS. Sampling errors occur because reported values in SF 3 and ACS tables are

[9] Census Bureau. *Meeting 21st Century Demographic Data Needs-Implementing the American Community Survey. Report 6; The 2001-2002 Operational Feasibility Report of the American Community Survey.* May 2004, p 4.
[10] Ibid. *Meeting 21st Century Demographic Data Needs-Implementing the American Community Survey. Report 6: The 2001-2002 Operational Feasibility Report of the American Community Survey*, pp. 13-16.

estimates of true values calculated from a sample of housing units. Random and biased non-sampling errors are caused by methodological problems in selecting the sample of housing units, questionnaire design, data collection, and processing of the questionnaires into datasets. The Census Bureau attempts to minimize, if not eliminate, non-sampling errors through rigorous quality control procedures.

The standard error of an estimate from a sample measures how far the estimate deviates from the mean value of the estimate computed from all possible samples. Since the expected value of the sample mean is the population value, standard errors quantify the sampling error in an estimate from a sample. Large standard errors mean that the sample estimate may be quite different from the true population value, while small standard errors imply that the sample estimate is probably quite close to the population value.

The effect of sample size on standard errors for totals can be computed as follows (*11*):

$$SE(\hat{Y}) = \sqrt{S\hat{Y}\left(1 - \frac{\hat{Y}}{N}\right)}.$$

In this equation, $SE(\hat{Y})$ is the standard error of the estimated total quantity $\hat{Y}$ (employed persons estimated to live within a census tract, for example); N equals the total count of people, housing units, households, or families depending on whether the estimated total quantity is expressed in persons, housing units, households, or families (N would equal persons living in the census tract in this example), and; S is the inverse of the sample rate minus one.

The long-form housing unit sample rate for Census 2000 is approximately one in six and the annual sample rate for ACS is roughly one in forty; therefore, S in the above equation is equal to five (6 minus 1) for Census 2000 long-form data and thirty-nine (40 minus 1) for an annual ACS. Given these sample rates, standard errors for annual ACS estimates of totals are about 2.8 times Census 2000 long-form standard errors (calculated by $\sqrt{\frac{39}{5}}$) due to the reduced sample (*12*).

If three years of ACS data are accumulated and averaged, the resulting standard error for the averaged three-year ACS estimate of a total equals:

$$SE(\hat{Y})_{3\ Year} = \frac{\sqrt{SE(\hat{Y})^2_{Year\ 1} + SE(\hat{Y})^2_{Year\ 2} + SE(\hat{Y})^2_{Year\ 3}}}{3}.$$

If one assumes roughly equal standard errors for each year, then standard errors for three-year average estimates of ACS totals are slightly more than 1.6 times Census 2000 standard errors (substituting into the above equation, $\frac{\sqrt{3*(2.8)^2}}{3}$). Repeating the calculation for a five-year average

[11] Ibid. *2000 Census of Population and Housing: Summary File 3 Technical Documentation.* p 8-22.
[12] This ratio understates relative standard errors for the two surveys by about 10 to 15 percent since it is based on distributed questionnaires rather than completed interviews.

estimate of ACS totals ($\frac{\sqrt{5*(2.8)^2}}{5}$) indicates that five-year average ACS estimate standard errors will be about 25 percent greater than Census 2000 standard errors due to reduced sample sizes.

These relative standard errors are unadjusted for the effects of expanding survey results to population totals and reconciliation of small and large geographic area results. Census 2000 long-form estimates for population and housing units were weighted to equal 100 percent enumeration figures for most areas that returned two hundred or more questionnaires (*13*), and these estimates have no standard errors since by definition they equal population values. Expansion of ACS sample estimates to population totals is not described in detail in the ACS Operations Plan (*14*), but ACS estimates are expected to be similarly weighted to population and housing unit control totals directly from the decennial census or from mid-decade estimates. However, far fewer geographic area estimates were controlled in the prototype ACS surveys than in Census 2000 and it seems probable that the fully implemented ACS will also have fewer controlled estimates than the decennial census because of the smaller sample.

Alternative ACS Housing Unit Sampling Scenarios

One can hypothesize possible circumstances that would disrupt annual ACS data collection, such as the earlier noted elimination of portions of the 2002 ACS due to restricted funds. Sampling impacts from the following scenarios are considered in this section:

- **Severe Funding Reduction Reduces ACS Housing Unit Sample by 50 Percent**. The scenario is that half the monthly ACS questionnaire waves would be eliminated, while remaining months would continue unaffected.
- **Moderate Funding Reduction Reduces ACS Housing Unit Sample by 25 Percent**. This is a second reduced funding scenario with three months of surveying eliminated.
- **Restricted Funding Eliminates CAPI Follow-Up**. In this scenario, all CAPI follow-up surveys are eliminated due to funding cuts. Based on Table 3 figures, the annual housing unit sample would be lowered by 20 percent.
- **Restricted Funding Eliminates Both CATI and CAPI Follow-Ups**. The scenario is that all CATI and CAPI follow-up surveys are eliminated reducing the housing units in the annual sample by about a third.
- **Unable to Carry Out ACS for One Year**. It is impossible to conduct an ACS during one of five consecutive years.
- **Voluntary Participation ACS Scenario 1**. This is the first of two ACS voluntary participation scenarios. The Census Bureau's testing of a voluntary ACS compared to mandatory participation indicated that about 68 percent of the mail back questionnaires, 82 percent of the CATI interviews, and 95 percent of CAPI interviews would be

[13] Census Bureau. Comparing SF 3 Estimates with Corresponding Values in SF 1 and SF 2. Available at http://www.census.gov/Press-Release/www/2002/sf3compnote.html. (Accessed March 2005)
[14] Ibid. *American Community Survey Operations Plan,* pp 34-38.

completed (*15*). In this scenario, the same rates of CATI and CAPI follow-ups are maintained for increased workloads due to fewer returned questionnaires, resulting in a housing unit sample approximately 85 percent of the sample obtained when ACS compliance is mandatory.

- **Voluntary ACS Scenario 2**. In this second voluntary participation scenario, the same numbers of CATI and CAPI interviews are completed as under mandatory participation in spite of increased workloads. The scenario housing unit sample is then about 80 percent of the mandatory participation ACS sample.
- **Government Accounting Office Proposal**. This scenario features a 60 percent increase in the ACS sample in the years preceding, during, and immediately after decennial censuses (*16*). The housing unit sample would be 4.8 million in these three years, but remain at 3.0 million in other years. The argument for this increased sampling is that it would permit small geographic area estimates from three-year averages within one year of the decennial census.
- **Seven-Year Accumulation and Averaging for Small Area Estimates.** This methodological change would increase the housing unit sample for small area estimates to very nearly the same number as in the decennial census long-form sample.

Table 4 summarizes the relative standard errors and confidence intervals for estimates developed under these alternative sampling scenarios. Confidence intervals are statistical measures of how close the estimate is to the actual value for the universe of housing units. This table shows 90 percent confidence intervals meaning there is a 90 percent probability that the population value lies within the interval on either side of the estimate.

The first line in the table corresponds to Census 2000 long-form estimates from a one in six sample. For the purposes of the comparisons, the standard error of the Census 2000 long-form estimate is expressed as unity and the 90 percent confidence interval as plus or minus 1.65 (*17*). The second line summarizes the fully implemented ACS survey characteristics, a one in forty housing unit sample, and one-year, three-year average, and five-year average standard errors and confidence intervals for estimates relative to Census 2000 long-form estimates. It should again be emphasized that the Table 4 figures just reflect differences in standard errors and confidence intervals caused by smaller ACS samples and averaging.

Remaining lines in Table 4 are relative results for the ACS sampling scenarios. For example, the first severe funding reduction scenario reduces the annual ACS housing unit sample by 50 percent to a one in eighty sample, which increases the standard errors and confidence intervals of the estimates by more than 40 percent compared to the fully implemented ACS. Remaining reduced funding scenarios are less onerous in their impact on the quality of the ACS estimates.

---

[15] Census Bureau. *Meeting 21st Century Demographic Data Needs—Implementing the American Community Survey. Report 3: Testing the Use of Voluntary Methods.* December 2003, pp 7-8.

[16] Government Accountability Office. *American Community Survey: Key Unresolved Issues.* Improving Timeliness and Quality of Small Geographic Area Data Would Increase Costs. October 2004, pp. 21-22.

[17] Ninety percent confidence intervals are plus or minus 1.65 standard errors from the sample estimate.

Table 4 Relative Impacts from Alternative ACS Sampling Scenarios

	Housing Unit Sample		1-Year		3-Year Average		5-Year Average	
	Scenario/ ACS	1/ Rate	Std. Err.	90% CI	Std. Err.	90% CI	Std. Err.	90% CI
Census 2000 Long-Form		6.0	1.00	±1.65				
Fully Implemented ACS	1.00	40.0	2.79	±4.61	1.61	±2.66	1.25	±2.06
Restricted Funding ACS								
50% Decrease	0.50	80.0	3.97	±6.56	2.29	±3.79	1.78	±2.93
25% Decrease	0.75	53.3	3.24	±5.34	1.87	±3.08	1.45	±2.39
No CAPI	0.80	50.0	3.13	±5.17	1.81	±2.98	1.40	±2.31
No CATI or CAPI	0.67	59.7	3.43	±5.65	1.98	±3.26	1.53	±2.53
Impact of Missing Year								
3-Year Average	0.67	59.7			1.98	±3.26		
5-Year Average	0.80	50.0					1.40	±2.31
Voluntary ACS								
Larger CATI and CAPI	0.85	47.1	3.04	±5.01	1.75	±2.89	1.36	±2.24
Smaller CATI and CAPI	0.80	50.0	3.13	±5.17	1.81	±2.98	1.40	±2.31
Other								
GAO Proposal	1.60	25.0	2.19[a]	±3.61[a]	1.26[b]	±2.09[b]		
7-Year Averaging	1.00	40.0					1.06[c]	±1.74[c]

a. Within one year of decennial census
b. Year Average of three years around decennial census
c. Seven-year average

Missing one year in the three-year average estimates is roughly equivalent to eliminating CATI and CAPI follow-ups. Eliminating a year from the five-year average estimates would affect the housing unit sample to nearly the same degree as eliminating CAPI follow-up. Differences between the two voluntary ACS scenarios are modest and generally similar to the scenario without CAPI follow-up.

The GAO proposal to increase the housing unit sampling rate for three years around the decennial census would allow small geographic area ACS estimates within one year of the census that are very nearly equal in quality to the five-year average estimates from the fully implemented ACS. Seven-year averaging of ACS data would yield estimates that are only slightly worse in quality than Census 2000 long-form estimates.

ACS Test Site and Census 2000 Comparisons

Thirty-six ACS test sites in thirty-one counties were surveyed in 1999, 2000, and 2001 generally following the methodology of the fully implemented ACS. Extensive analyses comparing ACS and Census 2000 long-form estimates for the test sites were completed by the Census Bureau and are published on the Census Bureau website (*18*). The comparisons in this section are based upon Census Bureau analyses, but prototype ACS survey figures are first averaged and then factored to match sample rates in the fully implemented ACS.

[18] Census Bureau. *American Community Survey 1999-2001 and Census 2000 Comparison Study*. Available at http://www.census.gov/acs/www/AdvMeth/acs_census/report.htm. (Accessed February 2005).

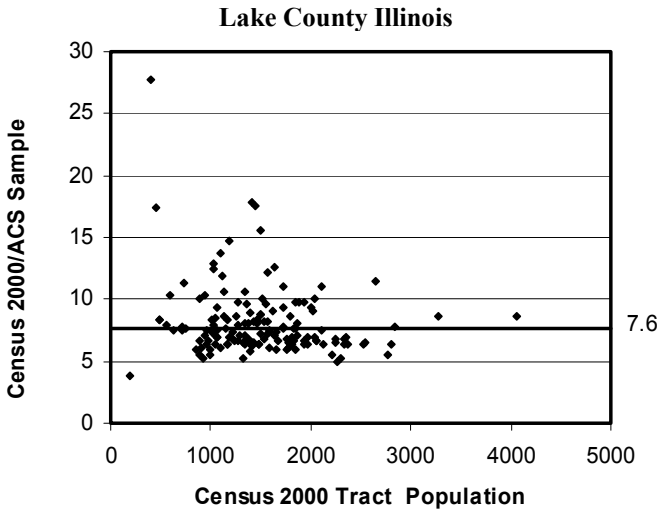
Table 5 Housing Unit Samples in ACS Test Site Counties

State	County	Census 2000	Average ACS[a]	Ratio	State	County	Census 2000	Average ACS[a]	Ratio
Arkansas	Jefferson	5,120	615	8.3	Nebraska	Douglas	26,733	3,354	8.0
Arizona	Pima	46,022	5,233	8.8	New Mexico	Otero	4,957	382	13.0
California	San Francisco	40,588	5,590	7.3	New York	Bronx	55,455	6,550	8.5
	Tulare	20,828	1,893	11.0		Rockland	13,048	1,573	8.3
Florida	Broward	87,048	12,135	7.2	Ohio	Franklin	66,513	8,046	8.3
Georgia	Upson	1,529	178	8.6	Oregon	Multnomah	40,658	4,868	8.4
Iowa	Black Hawk	8,393	1,139	7.4	Pennsylvania	Fulton	3,237	125	25.9
Illinois	Lake	32,195	4,262	7.6		Schuylkill	15,417	1,428	10.8
Indiana	Miami	2,319	322	7.2	Tennessee	Sevier	4,710	469	10.1
Louisiana	DeSoto	2,690	229	11.7	Texas	Fort Bend	16,487	1,938	8.5
Massachusetts	Hampden	24,998	2,934	8.5		Harris	162,930	20,168	8.1
Maryland	Calvert	3,898	442	8.8		Starr	2,443	220	11.1
Missouri	Iron	2,059	197	10.5		Zapata	761	61	12.6
	Reynolds	1,433	136	10.5	Virginia	Petersburg City	1,998	254	7.9
	Washington	2,176	225	9.7	Washington	Yakima	10,745	1,205	8.9
Mississippi	Madison	3,615	445	8.1	Wisconsin	Oneida	6,925	400	17.3
Montana	Flathead	7,551	708	10.7		Vilas	8,284	334	24.8
	Lake	2,983	312	9.6	West Virginia	Ohio	4,247	459	9.3

a. Average of the three years factored to match ACS housing unit sample rates

Table 5 compares the Census 2000 long-form and ACS prototype survey housing unit samples for the test sites. The ratio column is the Census 2000 sample divided by the ACS average annual sample. These figures show that it requires seven or more years of ACS data collection to accumulate a sample of housing units equal in number to the Census 2000 long-form housing unit sample, which is consistent with the overall housing unit sample rates and ACS completion rates discussed previously. However, the ratios between Census 2000 and ACS housing unit sample sizes are very much larger in lightly populated counties and in counties with large numbers of seasonal residents.

FIGURE 2 Tract Housing Unit Samples

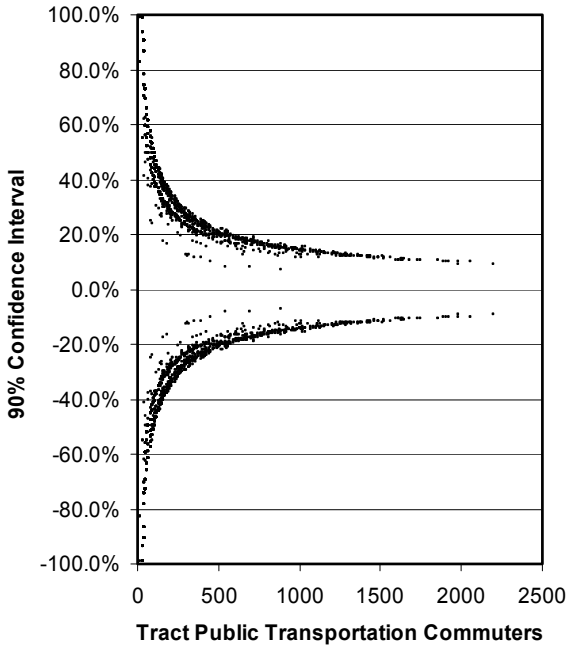


A similar tract level analysis is shown in Figure 2 for the Lake County Illinois test site. The Y-axis is ratio calculated in Table 5, and the X-axis is the Census 2000 tract population. The average value for all tracts in the county is 7.6 years to accumulate a housing unit sample equal to Census 2000. The tract level ratios are, of course, distributed around the average value, but for many tracts the average annual ACS sample is less than a tenth of the Census 2000 sample. For one tract in Lake County, more than twenty-five years of ACS samples would be needed to equal the Census 2000 sample.

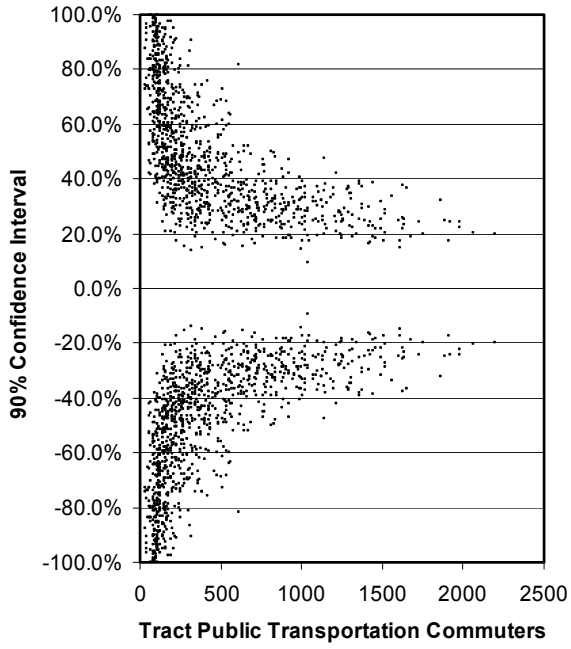
Figure 3 contains two plots of the 90 percent confidence intervals – measured in percentages - for estimates of workers commuting by public transportation at the tract level (2483 observations) in the ACS test site counties. Comparing tracts with approximately 500 public transportation commuters, the confidence intervals around the Census 2000 estimate are about ±20 percent or ±100 commuters. Comparable values for the average five-year ACS estimate range from ±25 to ±50 percent. The lower ranges of these ACS confidence intervals generally agree with earlier Table 4 figures, but some tracts have much larger relative confidence intervals due to differences in the ACS and Census 2000 interview and weighting procedures.

FIGURE 3 Confidence Intervals Around Estimated Public Transportation Commuters

a. Census 2000 Long-Form Tract Estimates



b. 1999-2001 ACS Estimates[a]



a. Standard errors for the three surveys are first adjusted to match fully implemented ACS sample rates, averaged, and then factored to account for five-year averaging.

ACS Public Use Microdata Samples (PUMS) Data Sets

Census Bureau plans are to annually release ACS based PUMS data sets comparable to the five percent decennial census PUMS (*19*). It is not clear whether this means that the number of PUMS records accumulated from five years of the ACS would be equal to the PUMS records available from a decennial census or less than a five-percent sample due to the reduced ACS sample. Since the five-year accumulation of sampled housing units from the ACS is three-quarters the size of the Census 2000 long-form housing unit sample, then ACS PUMS data sets might also be three-quarters the size of Census 2000 PUMS data sets.

[19] Mary Ellen Davis and Charles H. Alexander, Jr. *The American Community Survey: The Census Bureau's Plan to Provide Timely 21st Century Data.* Available at http://www.census.gov/acs/www/AdvMeth/Papers/ACS/Paper8.htm. (Accessed March 2005).

It is assumed that place of residence in ACS five-year accumulated PUMS data sets will be coded to Public Use Microdata Areas (PUMAs) containing more than 100,000 as was done for the Census 2000 PUMS. Since the PUMS is a sample, any tabulation of PUMS data for these PUMAs contains standard errors, which are larger than standard errors for the entire sample of housing units due to the reduced number contributing to the estimate. Table 6 measures the relative quality of Census 2000 and ACS PUMS tabulations due to housing unit sampling rates.

Table 6 ACS PUMS Standard Errors, and Confidence Intervals

	1-Year		5-Year Total	
	Std. Err.	90% CI	Std. Err.	90% CI
Census 2000 PUMS 5% Sample	1.00	±1.65		
Annual ACS PUMS				
1% Sample	2.28	±3.77	1.02	±1.68
0.75% Sample	2.64	±4.35	1.18	±1.94

Standard errors and confidence intervals for the Census 2000 five percent sample PUMS are base values in Table 6. Relative figures for ACS PUMS are listed for one year and five-year accumulations assuming either a one percent or 0.75 percent annual sample. Five-year total PUMS data sets with an annual one percent sample are very similar in quality to the PUMS data sets from Census 2000, while the reduced sample PUMS has about 20 percent larger standard errors and confidence intervals.

Census Transportation Planning Package (CTPP) Case Studies

The CTPP is a census special tabulation containing a number of tables developed from long-form data requested by transportation planners and may include a locally defined Transportation Analysis Zone (TAZ) geographic area summary level. The CTPP has three parts for tabulations by place of residence, workplace, and residence-workplace flow. Part 3 place of work by place of residence tables are of particular interest to transportation planners since they are similar to trip tables used in model applications.

Table cells in Part 3 of the CTPP contain many small values, and in the Census 2000 CTPP, Tables 3-03 through 3-07 - poverty status of workers, minority status of workers, workers by household income, workers by means of transportation to work, and workers by household income by means of transportation to work – are frequently suppressed due to the Census Bureau's disclosure avoidance policies. In most cases, this means that a residence-workplace combination must have three or more commuting workers to avoid having the cells in these five tables zeroed out. Suppression is not an issue for remaining tables in Part 3 since only handfuls of records with observations (56 records in CATS study area) have all data suppressed.

How substituting the ACS for decennial census long-form data impacts Part 3 of the CTPP was investigated through three MPO case studies. The large MPO case study is the Chicago Area Transportation Study (CATS), the MPO for six counties and part of a seventh in northeastern Illinois (the CTPP for CATS study area covers eight counties). A mid-sized MPO case study analyzes impacts for the TriCounty Regional Planning Commission's study area, the MPO responsible for Peoria, Tazewell, and Woodford counties (including the cities of Peoria and Pekin) in central Illinois. The last case study is a small MPO, the Kankakee Area Transportation Study (KATS), which covers Kankakee County in northeastern Illinois.

Table 7 MPO Tract and TAZ Characteristics

eography	CATS[a]	TriCounty	KATS
Study Area			
Square Miles	4495	1829	680
2003 Estimated. Pop.	8,397,771	346,758	105,625
Tracts			
Number in CTPP	1843	87	26
Median Area (square miles)	0.55	2.81	7.68
Median Population (Census 2000)	4169	4117	3417
TAZs			
Number in CTPP Part 1	6167	526	195
Median Area (square miles)	0.34	0.16	1.01
Median Population (2000 CTPP Part 1)	900	405	380

a. Eight county CTPP area

Tracts and TAZs in the CTPP in MPO study areas are listed in Table 7. Study area populations are totaled from the current Census Bureau county estimates (*20*), but Census 2000 population figures are used to rank tracts and TAZs to determine median tract population and land area. Tracts are sized to include roughly similar number of households, confirmed by the median population per tract figures. Since TAZx are locally specified by the MPO, they vary substantially from region to region. The small TAZ areas created for the TriCounty RPC study area are of some interest in this example. Some TAZs in the city of Peoria's central area are smaller than TAZs in the Chicago central area.

Table 8 summarizes characteristics of CTPP Part 3 for the three case study MPO areas. This table only considers internal residence-workplace work flows where both the residence and workplace are located in the study area. The maximum possible interchanges are, therefore, the number of tracts or TAZs in a study area squared.

Table 8 Case Study Areas: CTPP Journey-to-Work Interchanges

	2000 CTPP Part 3			Simulated ACS Part 3		
Journey-to-Work Interchange	CATS[a]	TriCounty	KATS	CATS[a]	TriCounty	KATS
Internal Tract to Tract						
Maximum Possible	3,396,649	7569	676	3,396,649	7569	676
Records in CTPP Part 3	220,063	4294	576	182,178	3894	547
CTPP/Maximum Records	6.5%	56.7%	85.2%	5.4%	51.3%	80.9%
Records with Tables 3-03 through 3-07 Suppression	179,687	2135	161	152,266	2099	177
Unsuppressed/CTPP Records	18.3%	50.3%	72.0%	16.4%	46.1%	67.6%
Internal TAZ to TAZ						
Maximum Possible	38,031,889	276,676	38,025	38,031,889	276,676	38,025
Records in CTPP Part 3	345,496	14,532	3674	276,584	11,970	3108
CTPP/Maximum Records	0.9%	5.3%	9.7%	0.7%	4.3%	8.2%
Records with Tables 3-03 through 3-07 Suppression	322,427	12,800	3072	260,896	10,765	2691
Unsuppressed/CTPP Records	6.7%	11.9%	16.4%	5.7%	10.1%	13.4%

a. Eight county CTPP area

[20] Census Bureau. Population Finder. Available at http://www.census.gov/. (Accessed April 2005).

CTPP Part 3 records are residence-workplace combinations with at least one worker respondent. As study areas increase in size, the proportions of possible residence-workplace combinations with a commuting worker decrease primarily due to excessive travel distances for many movements. This is followed in the table by the Part 3 records affected by the Table 3-03 through 3-08 disclosure requirements and the percentage of the total records with suppression. The great majority of interchanges in the case studies have data suppression in these five tables. For the CATS study area, nearly 82 percent of all tract residence to tract workplace records and more than 93 percent of TAZ level records have zeroed tables. Reduced but still substantial percentages of Part 3 records have data suppression in the two smaller MPO study areas.

A simulation was devised to estimate the impacts of an ACS-based CTPP. The simulation first converts a Census 2000 CTPP Part 3 work flow table into worker observations taking into account rounding of work flows and sampling rates, and then randomly samples the worker observations to match ACS sampling rates. The results shown in Table 8 under the "Simulated ACS Part 3" heading are average values for 100 simulations.

As expected, the smaller ACS housing unit sample reduces the numbers of CTPP interchanges with observations. More records are eliminated for TAZ level worker flows, and for the larger CATS case study. In both instances, this is due to large numbers of small cell values. The interchanges subject to data suppression generally decrease in the simulations because many of these interchanges are eliminated by the reduced sample.

Table 9 weights the values in Table 8 by the number of Census 2000 CTPP workers in each interchange cell. Even though the ACS simulation figures do not reflect the weighting of workers that would occur for ACS sample rates, they do point out that an ACS-based CTPP would have far less effect on the worker flows than on interchanges.

Table 9 Case Study Areas: Worker Weighted CTPP Journey-to-Work Interchanges

	2000 CTPP Part 3			Simulated ACS Part 3		
Journey-to-Work Flow	CATS[a]	TriCounty	KATS	CATS[a]	TriCounty	KATS
Internal County to County	3,711,570	152,290	36,935	3,711,570	152,290	36,935
Internal Tract to Tract						
Worker Flows in CTPP Part 3	3,641,716	151,583	36,917	3,337,314	148,568	36,691
CTPP/Max. Worker Flows	98.1%	99.5%	100.0%	89.9%	97.6%	99.3%
Worker Flows Suppressed in Tables 3-03 through 3-07	1,633,561	17,679	1359	1,602,366	22,409	2063
Unsuppressed/CTPP	55.1%	88.3%	96.3%	52.0%	84.9%	94.4%
Internal TAZ to TAZ						
Worker Flows in CTPP Part 3	3,586,462	147,901	36,585	3,064,969	131,181	32,941
CTPP/Max. Worker Flows	96.6%	97.1%	99.9%	82.6%	86.1%	89.2%
Worker Flows Suppressed in Tables 3-03 through 3-07	2,701,104	92,479	21,466	2,367,390	85,735	21,182
Unsuppressed/CTPP	24.7%	37.5%	41.3%	22.8%	34.6%	35.7%

a. Eight county CTPP area

General Findings and Transportation Planning Implications

The following sections organize findings in the paper under general headings and offer some conclusions based on these findings.

Evaluation of ACS Small Area Standard Errors: Standard errors for ACS small area estimates will be at least 25 percent (five-year accumulation) greater than standard errors for Census 2000 long-form estimates based on reduced sample size alone. This must be considered a conservative estimate, however, since it does not take into account successful interview completion rates, methodological differences between the two surveys, and expansion of fewer ACS small areas to control totals. Comparisons between the ACS test site results and Census 2000 estimates indicate that the standard errors of many small area ACS estimates will be more than twice Census 2000 long-form standard errors.

These increased standard errors are most important for ACS small area estimates that are a fraction of total persons, workers, households, or families. For example, the quality of estimates for workers who work at home or commute by transit, non-motorized modes, or in large carpools will consistently be much poorer than estimates for workers who drive alone. The low and high ends of distributions – examples are workers in households, household size, and household income - will frequently be problematic. Transportation studies that employ long-form estimates to locate and size selected subpopulations – environmental justice and specialized transit studies come to mind – should be most impacted.

Tracking Regional Socioeconomic and Demographic Changes: Changes in population and housing characteristics are difficult to measure when annual estimates have large standard errors. The standard error for the difference between two annual estimates is approximately forty percent larger than the standard error for a single year (the standard error for the difference between two estimates equals $\sqrt{SE(\hat{Y})^2_{Year\ 1} + SE(\hat{Y})^2_{Year\ 2}}$). Consider this example, a one-year ACS survey based county estimate of 10,000 public transportation commuters would typically have a confidence interval of about ±20 percent around the 10,000 estimate (a confidence interval from 8000 to 12,000 workers commuting by public transportation). The 90 percent confidence interval for the difference between two annual estimates is, therefore, about ±2800 workers. The difference between two annual estimates would have to be more than this number in order to reject with 90 percent confidence the hypothesis that the two estimates are equal.

ACS Methodology and Sample Size: The housing unit sample size of the ACS heavily depends on the success of the first phase of the survey; the distribution of questionnaires by mail to housing units and the willingness of individuals to fill out and return questionnaires. With respect to questionnaire distribution, the housing unit sample for an area with an inaccurate address file (lightly populated, rural, rapidly developing, and so forth) will generally be reduced because larger numbers of interviews will be passed to the CAPI phase, where they are subject to additional sampling. Even though the ACS is mandatory, the level of participation will likely vary from year to year, and one would expect questionnaire mail-backs to decline between decennial censuses and over time.

Alternative Sampling Scenarios: The evaluation of different sampling scenarios showed that any interruption of the ACS or reduction in housing unit sample due to restricted funds would have major consequences for the quality of estimates. The GAO proposal is appealing since: (1) the ACS would benefit from the publicity surrounding the decennial census; (2) the ACS long-form data would be available close to the decennial census, and could be utilized with SF 1 data, and; (3) fully enumerated population and housing unit figures would be available to weight the ACS long-form data.

PUMS and CTPP Special Tabulation: The PUMS based on the ACS will be quite similar to the decennial census long-form PUMS. It is perhaps the long-form product that will be least affected by introduction of the ACS.

The investigation of the CTPP dealt only with the Part 3 journey-to-work flows. It appears that the impact of the ACS on the CTPP Part 3 data will be marginal for most MPO regions if census data suppression policies do not change. The reasoning behind this conclusion is based on the following:

- Census 2000 CTPP Part 3 TAZ level tables that are subject to suppression (minimum of three observations) are already generally of little use, substitution of the ACS has no effect unless more tables are made subject to suppression.
- Similar comments apply to suppressed Part 3 tract level tables for larger MPO areas.
- Part 3 tract level tables for smaller MPO areas should continue to be useful, even with ACS sampling and data suppression.
- Unsuppressed Part 3 tract and TAZ level tables will be modestly reduced in quality by ACS sample sizes, but larger MPO areas will typically be impacted more than smaller areas.

The paper's analysis of the CTPP Part 3 does point out the need for more extensive preplanning for locally determined TAZs and evaluation of alternative TAZ geographic levels for future CTPP tabulations in order to ameliorate the loss of data due to suppression. The form of this evaluation could be similar to the simulation analyses in the paper; aggregate existing CTPP tables to different geographies and then resample to match ACS rates. Alternately, more general research projects could be undertaken through Census Bureau Research Data Centers to evaluate journey-to-work flow tables built from actual home and workplace locations.

Implications for MPO and State DOT Planners: How important are differences between the ACS and decennial census long-form estimates for transportation planners who have used previous census estimates for observed socioeconomic and demographic data, for future socioeconomic and demographic estimates, and for model calibration and validation? Few transportation planners (including the author) have been overly concerned about the statistical quality of past census long-form estimates, and have readily used small area estimates without questioning their statistical properties. They will only be impacted only if ACS estimates for small areas and traffic analysis zones are discontinued.

It is proposed that research needs to be undertaken on the use of small area ACS estimates as model inputs or for model calibration and validation. The objectives of this research would be to improve practice by developing guidelines and training opportunities for modelers and data analysts. This could be a research area of interest for the Travel Model Improvement Program.

MPO and state DOT planners may have to deal with discontinuities in forecasts as models are recalibrated and databases are updated with ACS estimates. Differences in how residences are defined - the ACS defines occupied housing units using current residence rather than the usual residence definition of past decennial censuses - could greatly affect occupied housing units in some zones, and estimates of average household sizes may be affected by ACS telephone editing of all large household interviews.

Agency staffing may be affected by the continuous nature of the ACS, dedicated staff may be required to annually assemble and maintain the more regularly updated databases. The related question is whether continuous annual updates from the ACS are needed – particularly small area estimates - or even wanted by MPO staff. Processing of census data for updating models typically takes place during multiple year planning cycles, and a constant base year is maintained throughout the planning cycle. One anticipates that annual ACS estimates between major planning cycles will largely go unused by MPO planners. Annual ACS county and large city estimates will, however, be regularly examined by MPO staff. They will use these larger area estimates to tract regional trends, structure their models so that they are sensitive to significant regional changes in socioeconomic characteristics, and help determine the issues to be dealt with in major planning cycles.

Acknowledgements

The author gratefully acknowledges the effort and knowledgeable comments of the paper reviewers. Their willingness to review the lengthy and quite preliminary draft is appreciated.