The Hard-To-Interview in the American Community Survey

Gregg Diffendal
U.S. Census Bureau, Demographic Statistical Methods Division
Room 3784 -3, Washington, D.C. 20233

1. Introduction

There has been a recent decline in the interview rates for many surveys, even in the government sponsored surveys. In the Netherlands, they have a very large nonresponse rate in their major survey (De Heer 1999). Even the Census Bureau's major surveys, the Current Population Survey have seen some decline in the interview rates, where they were over 95% or higher up to 1992 and have now dropped to about 93% (De Heer 1999). Is this a one time occurrence or are the United States population becoming over burdened with too many surveys and are now more likely to refuse to respond to even official government surveys? And what is the impact of a lower response rate? We examine this issue by using the currently collected data to examine the potential impact of a lowering of the interview rate by assuming that the 'late' interviews will become noninterviews. Then we can test to see if these cases will change the survey estimates. To keep interview rates high, more intensive field procedures may be needed to convert these cases. However if the impact on the estimates is not a significant difference, we may want to save our limited budget resources and apply them to other areas.

In this paper, we examine the American Community Survey (ACS) estimates and the impact of the late or hard-to-interview cases. The ACS is a fairly new survey and has plans to further expand in the future, budget allowing. The ACS collects essentially the same information as the Census long form. The data includes the basic demographic characteristics, such as age, race, sex, relationship, marital status, and Hispanic origin as well as items on education, veteran status, employment, occupation and income. The current plans are for the ACS to replace the Census long form for the 2010 Census.

The paper is divided into four sections. Section 2 gives further details about the ACS and the data collection methodology. Included will be some details on the future of the ACS. Section 3 will discuss the definition of the hard-to-interview cases used here in the ACS. Section 4 will present the hard-to-interview data and their results from the 1999 ACS. Section 5 will summarize the results for the ACS.

2. ACS Methodology

The ACS is a monthly survey that collects similar data as the Census long form. The data is reported yearly for large areas (currently places over 65,000 in population) and aggregates data over multiple years for smaller areas. For example for the smallest areas which would include census tracts and block groups, the data would be aggregated over a 5 year time span to produce sufficiently reliable estimates.

The ACS sample is a systematic sample within four strata. The sampling frame is the master address file (MAF), a list of addresses for the entire country kept up to date by matching this list to the current address files (the delivery sequence file) supplied by the US Postal Service. The four strata are: blocks in places with less than 800 housing units, blocks in places with 800- 1200 housing units, blocks in large tract with over 2000 housing units and all other blocks. This mimics the oversampling for small places to increase the reliability of the estimates for these places. The basic sampling rate is 5% for the most counties. Some of the large counties had lower sampling rates. Table 1 contains a list of the 36 counties in the 1999 ACS and the nominal sampling rates.

The ACS collects the data using three modes of data collection, mail, telephone and personal visit. The data is collected for each monthly sample over a three month time span. The mail mode uses a pre-notice letter, the ACS questionnaire and a reminder post card. For those who do not mail their form back in about three weeks, a replacement questionnaire is mailed. For all mail returns, a computer edit checks the completeness of the questionnaire and if it is incomplete, a telephone follow up is conducted to try to complete the missing information. Currently we are getting about a 50% mail response.

If the questionnaire is not received by the end of the first month, all cases that have a telephone number (obtained from a private vendor matching on the mailing address) are sent to our telephone centers. About 40% of the cases that do not mail their forms back have telephone numbers and a telephone interviews are attempted for these cases. Mail returns continue to arrive in the mail as the telephone interviewing is occurring. Telephone interviewing occurs for a little less than a month.

Finally all cases that have not mailed a form back or obtained an interview from the telephone call are eligible for personal visit interviewing. Because of the large expense of conducting a personal visit interview, only 1 in 3 cases are sent to personal visit interviewing. The personal visit interviewing starts at the beginning of the third month and continues to the end of the month.

A few special remarks about the three modes of data collection. A small number of cases have no mailing address and cannot be mailed. For these cases, 2 out of 3 are sent to personal visit interviewing at the same time as the other cases in the third month. Mail returns are checked in during the whole three month time period. Since very few mail or telephone cases are noninterview housing units or vacant housing units or not a housing unit (delete), most of these cases occur during the personal visit interviewing. Typically we have seen about a 50% mail response rate and another 10% telephone interviews. Large cities generally have a smaller mail response rate, down around 40% while some other areas have mail response rates over 60%. The remaining 40% is subsampled with about 13% being eligible for personal visit interviewing.

In 2000, the ACS will also include a national sample component called the Census 2000 Supplementary Sample (C2SS). In 2003 the ACS (if the full budget is approved) will have sample in every county in the country. We expect to mail out about 250,000 forms per month for a 3 million Housing Unit sample per year. We will be producing estimates for large places yearly (over 65,000 in population) and for smaller places after three years of data collection (places over 30,000 in population) and for all places, tracts and block groups after 5 years of data collection in 2008 for data collected in 2003-2007.

The ACS web site, [www.census.gov/acs,](www.census.gov/acs) contains additional details about the ACS and how the data was collected and produced. Summary tables and public use microdata samples are available at this web site.

3. Defining the Hard-To-Interview in the ACS.

The hard-to-interview cases are defined as the last cases to be checked in for each month of sample. The hard-to-interview cases were defined for each mode of data collection, mail, telephone and personal visit.

The cutoff dates to define the hard-to-interview for each mode were: mail cases are the 17th of the month after mail out (so for January mail out, all late mail cases were those received after February 17th), telephone cases are the 25th of the month of interviewing, and personal visit cases are the 27th of the month of interviewing.

Other alternate definitions for the hard-to-interview may

have been possible. For example defining the hard-to-interview by the amount of missing data would be easy to define. But then it would not be possible to look at missing data rates for these cases. Also for the mail returns, the ACS mail form only has room for 5 persons to be included. An additional contact by telephone is needed to complete the information on persons 6 or higher. Of course some of these cases are never contacted (because no phone number available or incorrectly recorded) so we never complete information on the persons over 5. These may not be hard-to-interview though.

If we view hard-to-interview as an operational issue where we need to cut costs of the survey, then defining the hard-to-interview by time would be a good measure. For example if the ACS needed to cut costs and cut the time for personal visit cases to 25 days rather than the current 30 days, then the procedure used here would be a good method to gauge the impact of this decision on the ACS estimates.

I would expect that cases that are 'late' mail returns would roll into telephone (and personal visit) and most would become interviews. Likewise with the telephone late cases, they would roll over into personal visit phase and many would become interviews.

One better option for defining hard-to-interview cases for the telephone mode are those who refuse to answer the ACS questions on the phone. Given the time, I will include an analysis of these cases as well.

4. Data on Late Returns by Mode

I will examine each variable across modes rather than examining all of the variables within each mode. Some interesting results seem to appear when examining things this way. We have much more sample for the mail returns than for the telephone or personal visit cases.

To detect differences between late and not late cases, I used a chi-square statistic. Since the sample design is a systematic sample, this approximates a random sample especially within the different modes of data collection as used here. For the person data however there is some clustering due to all persons within a housing unit being included. I adjusted the chi-square statistic by dividing it by 2 to adjust for this clustering.

Sex is not different for late and not late cases for any of the modes.

Age is significantly different for all three modes, but shows a different pattern across modes. For the mail late cases, only age 55+ is less likely to be late. For telephone late cases, the ages 16-24 and 25-54 appear to be more

likely to be late and age 55+ is less likely to be late. The effect for under 16 is very small. For the personal visit late cases, the higher numbers only seem to show up for ages 25-54. The age 55+ effect is almost gone (very little difference between late and not late cases).

Hispanic Origin is significantly different for mail and telephone, but not different for personal visit cases. For mail and telephone cases, Hispanic persons are more likely to be late than are persons who are not Hispanic.

Race is significantly different across all three data collection modes. The effect does appear to shrink when going from mail to telephone to personal visit modes. Generally Blacks are more likely to be late across modes. Other race, which is mostly Hispanic persons, is different mostly for mail and telephone and not for personal visit. Multiple race persons have no difference for mail cases, but are more likely to be late for telephone and even more likely for personal visit cases. Asians have a higher late response for mail, but decreases for telephone and have a lower late response for personal visit. As an aside, I recoded race by including Hispanics as a race group and collapsing Other, Asian, Hawaiian and American Indian into one race group. The results are similar for mail and telephone, but the effect is not significant for personal visit mode.

Poverty is significantly different across all three data collection modes. More persons in poverty are late for the mail and telephone modes, but persons not in poverty are more likely to be late for the personal visit mode.

School is significant only for mail and personal visit. The effect is a complete reversal from mail where less than high school education is more likely to be late, no effect for telephone and higher than high school education more likely to be late for personal visit. The effects noted are the same if we limit the cases to be those age 25 or higher (not shown).

Amount of allocation is significantly different across all three modes. The effect is consistent across modes where late cases have more data allocated than the not late cases. The definition of the amount of allocation is sum of indicator variables on whether the variable was imputed (allocated). The variables used for the sum are age, race, sex, Hispanic Origin, marital status, relationship, poverty, and schooling. The variable was top coded to 4+ because very few cases were observed over this amount.

Tenure is significantly different all three modes. Renters are always more likely to be late.

Number of persons in the household is significantly different for all three modes. For mail and telephone large households are more likely to be late, while for personal visit small households (1 and 2 person households) are more likely to be late.

Structure is significantly different for mail and personal visit modes. Apartments generally are more likely to be late, with apartment 2-9 being higher for mail and all apartment higher for personal visit.

5. Conclusion

The results for the mail returns is that minorities (including Hispanics), people in poverty, households with children, larger households, renters, persons with less than a high school education, persons living in apartment are all more likely to mail their ACS forms in late. These effects were generally smaller for telephone cases. These effects are completely gone or even reversed for late cases with personal visit interviews. The Hispanics, renters and persons in poverty are no longer more likely to be late, single person households and persons with greater than a high school education are the late cases for personal visit. I find this to be a striking feature and not what I would have expected. I will point out that a higher percentage of late cases from mail made up a larger percentage of the universe of cases for personal visit. For example white were 80% of the mail returns, 62% of the telephone interviews and only 58% of the personal visit cases. Blacks show almost a reversal of this pattern with 8% for mail, 15% for telephone and 20% of the personal visit interviews.

Future work on this project will look at the last 5% overall and compare these results for the variables examined. I will try to rerun the estimates assuming that the last 5% are noninterviews to examine the impact on the final estimates. Also I will examine the results for refusals from telephone to see how likely they are converted into interviews during personal visit interviewing. These would clearly be considered hard-to-interview cases and may have implications for any all telephone surveys.

References
De Heer, W. (1999) "International Response Trends: Results of an International Survey", Journal Of Official Statistics, Vol. 15, No. 2, pp 129-142.

Bates, N. and Creighton, K. (2000) "The Last Five Percent: What Can We Learn from Late Interviews?" Paper presented at the 2000 American Statistical Association Meeting.

Table 1. ACS Counties and Sample Rates

County	Sample Rates
Rockland Co. NY	5%
Fulton Co. PA	5%
Multnomah Co. OR	5%
Harris Co.. TX	1%
Fort Bend Co. TX	1%
Otero Co. NM	5%
Franklin Co. OH	3%
Douglas Co. NE	5%
Broward Co. FL	3%
San Francisco Co. CA	3%
Tulare Co. CA	5%
Black Hawk Co. IA	5%
Hampden Co. MA	5%
Bronx Co. NY	3%
Schuykill Co. PA	5%
Yakima Co. WA	5%
Lake Co. IL	3%
Jefferson Co. AR	5%

County	Sample Rates
Upson Co. GA	5%
Miami Co. IN	5%
DeSoto Parish LA	5%
Calvert Co. MD	5%
Madison Co. MS	5%
Reynolds Co. MI	5%
Iron Co. MI	5%
Washington Co. MI	5%
Flathead Co. MT	5%
Lake Co. MT	5%
Sevier Co. TN	5%
Starr Co. TX	5%
Zapata Co. TX	5%
Petersburg Co. VA	5%
Ohio Co. WV	5%
Oneida Co. WI	5%
Vilas Co. WI	5%
Pima Co. AZ	5%

Table 2. Last 5 % of the Mail, Telephone, and Personal Visit Responses -Person data

		Mail			Telephone			Personal Visit		
		Not Late (N=184136)	Late (N=11595)	P-value (Chi-Square)	Not Late (N=32779)	Late (N=2664)	P-value (Chi-Square)	Not Late (N=37459)	Late (N=1924)	P-value (Chi-Square)
Sex	Male	47.4%	48.2%	>.10 ns (1.2)	49.4%	49.9%	>.10 ns (0.1)	48.5%	48.0%	>.10 ns (0.1)
	Female	52.6%	51.9%		50.6%	50.1%		51.5%	52.0%	
Age	0-11	14.3%	17.2%	<.001 (375)	19.1%	18.9%	<.001 (37.0)	23.2%	19.5%	<.01 (14.0)
	12-15	5.1%	6.6%		7.1%	6.9%		6.5%	6.3%	
	16-24	9.4%	11.8%		11.6%	14.8%		14.5%	13.3%	
	25-54	43.6%	48.1%		43.9%	46.9%		43.8%	49.7%	
	55+	27.7%	16.3%		18.3%	12.5%		12.0%	11.3%	
Hispanic Origin	Not Hispanic	89.1%	83.3%	<.001 (178)	80.5%	75.1%	<.001 (22.7)	69.6%	69.0%	>.10 ns (0.2)
	Hispanic	11.0%	16.7%		19.5%	24.9%		30.4%	31.0%	
Race	White only	80.9%	71.9%	<.001 (314)	63.4%	53.6%	<.001 (57.6)	58.5%	55.7%	>.10 ns (9.0)
	Black only	7.6%	11.5%		14.9%	17.5%		19.6%	20.7%	
	American Indian only	0.7%	0.8%		0.8%	1.2%		1.4%	1.8%	
	Asian only	4.1%	5.5%		4.0%	4.3%		4.1%	3.6%	
	Hawaiian, Pac.Is.Only	0.06%	0.03%		0.11%	0.15%		0.17%	0.00%	
	Other only	5.2%	8.7%		14.1%	19.9%		13.5%	14.4%	
	Multiple Race	1.6%	1.6%		2.7%	3.4%		2.7%	3.8%	
Poverty	Not in Poverty	90.9%	89.4%	<.001 (15.7)	87.9%	86.0%	<.10 (4.4)	74.4	78.4%	<.01 (7.7)
	In Poverty	9.1%	10.6%		12.1%	14.0%		25.6	21.6%	
School	Less than High School	33.5%	37.8%	<.001 (68.5)	44.8%	44.6%	>.10 ns (0.02)	53.7%	46.4%	<.001 (24.2)
	High School	20.4%	16.6%		23.4%	23.4%		19.4%	20.2%	
	Greater Than High School	46.1%	45.6%		31.8%	32.0%		26.9%	33.5%	

Relation-ship	Householder	42.8%	36.9%	<.001 (220)	35.2%	32.6%	<.001 (31.3)	36.6%	41.4%	<.10 (14.1)
	Spouse	22.0%	18.7%		18.7%	16.6%		13.9%	12.4%	
	Child	26.3%	31.8%		33.7%	34.7%		34.3%	30.7%	
	Parent	0.9%	1.2%		1.1%	1.0%		1.2%	0.8%	
	Sibling	0.8%	1.1%		1.1%	1.5%		1.5%	1.3%	
	Other Relative	2.3%	2.9%		4.1%	4.0%		3.2%	2.8%	
	Nonrelative	5.0%	7.4%		6.2%	9.8%		9.4%	10.7%	
Amount of Allocation	0	89.5%	85.6%	<.001 (110)	95.4%	83.9%	<.001 (318)	96.8%	91.2%	<.001 (113)
	1	7.7%	9.9%		3.7%	13.3%		2.5%	6.1%	
	2	1.5%	2.2%		0.6%	1.7%		0.3%	1.7%	
	3	0.6%	0.9%		0.2%	0.7%		0.2%	0.9%	
	4+	0.7%	1.4%		0.1%	0.4%		0.2%	0.1%	

Table 3. Last 5 % of the Mail, Telephone and Personal Visit Responses - Housing Questions

		Mail			Telephone			Personal Visit		
		Not Late (N=78788)	Late (N=4281)	P-value (Chi-square)	Not Late (N=11525)	Late (N=868)	P-value (Chi-square)	Not Late (N=13690)	Late (N=796)	P-value (Chi-square)
Tenure	Own	70.6%	66.1%	<.001 (40.1)	70.8%	67.3%	<.10 (4.9)	39.8%	35.1%	<.01 (7.0)
	Rent	29.4%	33.9%		29.2%	32.7%		60.2%	64.9%	
Number of Persons	1	30.1%	23.9%	<.001 (334.3)	22.6%	16.6%	<.001 (25.4)	26.8%	34.6%	<.001 (31.4)
	2	36.1%	30.0%		28.2%	27.5%		26.8%	27.6%	
	3	14.6%	17.5%		17.4%	18.3%		17.8%	16.1%	
	4	11.9%	15.8%		17.0%	21.4%		14.6%	11.3%	
	5	5.3%	9.3%		8.7%	8.6%		7.8%	6.3%	
	6+	2.1%	3.7		6.2%	7.5		6.3%	4.2	
Type of Structure	Mobile Home, Other	4.2%	2.9%	<.001 (42 .7)	4.4%	3.6%	>.10 ns (6.4)	5.6%	3.4%	<.001 (19.0)
	Single Family Detached	59.1%	57.4%		67.5%	64.6%		40.5%	35.8%	
	Single Family Attached	7.1%	7.4%		6.1%	6.5%		5.2%	5.5%	
	Apartment 2-9 Units	11.8%	14.5%		10.2%	12.0%		22.4%	23.7%	
	Apartment 10+ Units	17.8%	17.8%		11.8%	13.4%		26.4%	31.5%	