Medicaid Under-reporting in the Current Population Survey and One Approach for a Partial Correction

Michael Davern, Jacob Alex Klerman and Jeanette Ziegenfuss[i]

REVISED: November 10, 2008

Survey estimates of public health insurance program enrollment tend to be lower than those compiled from administrative enrollment data for those same programs. This discordance is particularly apparent for Medicaid and has become known as the "Medicaid undercount". The crude Medicaid undercount in the Current Population Survey's Annual Social and Economic Supplement (CPS), the most prominently used survey for policy research that measures health insurance coverage, was 32 percent in both 2000 and 2001.[ii]

Given the important uses of the CPS data, improved understanding of the undercount in the CPS is crucial. Towards that end, this paper reports preliminary results from an interagency project (funded by the Robert Wood Johnson Foundation and the Offices of the Assistant Secretary for Planning and Evaluation) that involves researchers from the State Health Access Data Assistance Center (SHADAC), The National Center for Health Statistics (NCHS), Office of the Assistant Secretary for Planning and Evaluation (ASPE), the Center for Medicare and Medicaid Services (CMS), and the US Census Bureau (Census) (taking the first letter of each abbreviation lead to the acronym "SNACC").

This paper presents parameters from two logistic regression models. These parameters can be used to predict the probability that the Medicaid administrative data imply that a CPS sample member actually had Medicaid. We report the coefficients from two logistic regression models that can be used to adjust CPS responses towards their Medicaid status in the prior calendar year. Finally, we present a sample STATA program using the parameters and what we refer to as "partially corrected" estimates of the number of uninsured Americans based on these parameters. With caveats we discuss below, these models will help researchers better understand the magnitude of the underlying survey response error, and their implications for policy research using the CPS.

Data Linking:

For the SNACC project, CMS provided the Census Bureau with a Social Security Number (SSN)-identified version of its MSIS files for 2000, 2001, and 2002. To protect confidentiality, Census replaced the SSNs on the MSIS files with the PIKs/Personal Identification Keys that Census maintains on its internal CPS file (in place of the SSN). Census then used these PIKs to

match the MSIS data to the corresponding CPS records for survey years 2001 and 2002 (corresponding to health insurance coverage in calendar year 2000 and 2001). Twenty-four percent of the CPS records were missing PIKs (18 percent refused to let the Census Bureau link their data and for an additional 6 percent the Census Bureau as unable to identify a unique PIK). To make the resulting analysis file representative of the full CPS survey, we reweight the 76 percent of CPS cases with SSNs to align with the full population of the CPS. Specifically, we post-stratify to adjust for age, race, sex, Hispanic ethnicity, and poverty status.

In addition, after we dropped MSIS records that we identified as duplicates, living in institutional group quarters, or partial benefit Medicaid enrollees, about 6 percent of the remaining MSIS Medicaid full benefits enrollees were missing PIKs. The analysis reported here makes no attempt to correct for these missing MSIS full benefit cases that were missing PIKs (i.e., SSNs). Below, we discuss the implications of that feature of this analysis.

Logistic Regression Models:

Using the pooled data set for the CPS years of 2001 and 2002 (covering the health insurance coverage calendar years of 2000 and 2001), we predict the probability that any given CPS record was successfully matched to the MSIS and had at least one day of "full benefit Medicaid" enrollment during the survey reference period. Our approach explicitly treats the CPS response as corresponding to the question as asked--i.e., referring to Medicaid at any time in the previous calendar year; this is unlike some other analyses that have treated the CPS responses as though they referred to insurance status of the interview date.

We estimate two logistic regression models. For people who are coded as not having Medicaid in the CPS (Model 1), one regression models the probability that the MSIS implies that they have Medicaid. For people who are coded in the CPS as having Medicaid in the CPS (Model 2), the other regression models the probability that the MSIS implies that they do not have Medicaid. The predictors in these models all come from the CPS public use file. They are coded in the following fashion:

Medicaid Coded on the CPS

This variable is used to separate Model 1 from Model 2. Being coded as having Medicaid is determined by using a combination of several variables in the CPS. First, if the person is coded as having Medicaid in the CPS variable asking explicitly about Medicaid are considered to have Medicaid, or if they were coded as having "Medicaid" in the "other please specify" sections of the other public health insurance survey item or the verification survey item in the CPS.[iii] Model 2 is estimated on those people coded in the CPS to have Medicaid. Model 1 includes everyone else not coded to have Medicaid.

Age

An independent ordinal variable recoded into six categories from the CPS age variable.

Health Insurance Allocation Status

This variable is used in Model 1 to separate those people with reported health insurance coverage from those with imputed coverage. In model 2 there is an additional variable entered for whether people coded to have Medicaid in the survey were edited to have Medicaid. This information was derived from the allocation flag variables in the CPS.

CPS Health Insurance Codes

Only another Public Insurance are those people coded to have public insurance but do not have Medicaid and do not have some other type of private coverage. People coded to have only private health insurance coverage were not coded to have any type of public coverage. People who were coded as being uninsured in the CPS were given the uninsured variable. People who were recorded to only have Medicaid and not other type of coverage are in the only Medicaid Reported on the CPS variable (this variable's reference category is having Medicaid and at least one other type of coverage recorded on the CPS).

Race/Hispanicity

An independent nominal variable constructed from the CPS items for race and ethnic origin. Categories include: Hispanic of any race, non-Hispanic black, non-Hispanic American Indian/Aleut Eskimo, non-Hispanic Asian or Pacific Islander and non-Hispanic white. This variable was mutually exclusive in 2001 and 2002 CPS and did not use the new Office of Management and Budget's new race and ethnicity coding scheme. To use this data on more recent CPS files requires a change in coding.

Male

An independent dummy variable indicating that the CPS person is male. Females are the reference category.

Relationship to Survey Reference Person

An ordinal independent variable from CPS indicating the person's relationship to the survey reference person (usually the respondent). Categories include: Child, Self, Parent, Spouse, and Other. Odds ratios are the effect for the given category relative to average.

Income and Poverty

An ordinal independent variable representing the ratio of the CPS person's family income to the federal poverty level. There are eight categories ranging from 0-49 percent of the poverty level (for the person's family size) to greater than 200 percent of the level. An independent dummy variable indicating that the CPS shows the person's family with zero income. CPS persons with positive family income are the reference category.

State

A nominal independent variable from CPS state of residence. For each CPS person this variable comes from the CPS coefficients are the effect of being in a given state relative to average.

Table 1: Logistic Regression Coefficients for Those CPS Cases Without Medicaid Reported on the CPS (Model 1) and Those Cases Without Medicaid Reported (Model 2) Predicting the Probability of Being Linked to the Medicaid Statistical Information System (MSIS)

Variable	Model 1	Model 2
Intercept	-0.7001	0.6674
Age		
Age 0 - 05	1.1846	0.3242
Age 06 - 14	0.7502	0.3446
Age 15 - 17	0.5352	0.1002
Age 18 - 44	0.0626	0.176
Age 45 - 64	-0.8833	-0.1864
Age 65 +	-1.6492	-0.7587
Health Insurance Allocation Status		
Medicaid Status Edited	.	-0.3425
Health Insurance Status Imputed	0.358	-0.787
Health Insurance Status Reported	-0.358	1.1295
CPS Health Insurance Codes		
Only Another Public Insurance Program Reported on CPS	1.1862	.
Only Private Insurance Reported on CPS	-1.0566	.
Other Public and Private Insurance Reported on CPS	0.1047	.
Uninsured Reported on CPS	-0.2343	.
Only Medicaid Reported on CPS	.	0.0833
Race and Ethnicity		
Hispanic	0.1171	-0.0532
Black	0.4959	0.1209
American Indian	0.1761	0.0811
Asian or Pacific Islander	-0.2219	0.0254
White	-0.5672	-0.1742
Sex		
Male	-0.5252	-0.33
Relationship to Reference Person		
Parent	0.9465	0.5378
Spouse	-0.5023	-0.6079
Child	-0.0723	0.0155
Other	-0.0748	0.218
Self	-0.2972	-0.1633
Income		
Zero Family Income Reported	0.2918	-0.3157
Ratio to Poverty Level 0-49%	0.47	0.5193
Ratio to Poverty Level 050-75%	0.6374	0.5367
Ratio to Poverty Level 075-99%	0.4517	0.5316
Ratio to Poverty Level 100-124%	0.1911	0.1985
Ratio to Poverty Level 125-149%	0.0329	-0.1593
Ratio to Poverty Level 150-174%	-0.1671	-0.2296
Ratio to Poverty Level 175-199%	-0.2912	-0.5188
Ratio to Poverty Level >200%	-1.3248	-0.8785

Variable	Model 1	Model 2
State		
Alabama	-0.1594	-0.0449
Alaska	-0.1092	-0.071
Arizona	0.0898	0.128
Arkansas	0.1346	-0.3122
California	-0.1289	0.3439
Colorado	-0.3336	-0.3836
Connecticut	-0.1959	-0.7136
Delaware	0.2288	0.2928
District of Columbia	-0.00658	-0.0619
Florida	-0.1398	-0.0376
Georgia	-0.3804	-0.2233
Hawaii	0.2023	0.043
Idaho	-0.1961	-0.1433
Illinois	0.1063	-0.1021
Indiana	0.1954	-0.0759
Iowa	0.0636	0.3585
Kansas	-0.3113	0.2179
Kentucky	0.0113	-0.2076
Louisiana	-0.203	-0.5873
Maine	1.1895	0.854
Maryland	-0.4332	-0.8844
Massachusetts	0.2376	-0.1753
Michigan	-0.1762	0.0409
Minnesota	0.2457	0.2035
Mississippi	-0.3595	-0.9459
Missouri	0.4285	0.3915
Montana	-0.9728	-1.6859
Nebraska	0.1689	0.6795
Nevada	-0.6973	-0.6105
New Hampshire	-0.1639	0.7828
New Jersey	-0.3905	-0.6305
New Mexico	0.1241	0.0584
New York	-0.148	0.0412
North Carolina	0.2153	0.4068
North Dakota	-0.0921	0.352
Ohio	-0.0647	0.2491
Oklahoma	0.0728	-0.0665
Oregon	-0.00899	-0.0122
Pennsylvania	0.2923	0.5112
Rhode Island	0.3565	0.2481
South Carolina	0.1664	0.1238
South Dakota	-0.1507	-0.26
Tennessee	0.9234	0.9372
Texas	-0.6294	-0.16
Utah	-0.3238	0.0123
Vermont	1.2055	0.8642
Virginia	-0.6028	-0.4457
Washington	0.6575	0.3998
West Virginia	0.3304	0.2473
Wisconsin	-0.0839	0.1057
Wyoming	-0.1846	-0.0526

Source: 2001 and 2002 Expanded Sample CPS ASEC data files Linked to the 2000 and 2001 MSIS
Note: Effect coding (as opposed to dummy coding) was used for all categorical variables except for "Sex" (reference category for sex is female), "Only Medicaid Reported on the CPS" recorded on the CPS (the reference category was Medicaid and at least one other type of coverage reported on the CPS) and the Variable "Zero Family Income Reported" (the reference category was having at least some income --or loss of income reported).

How to Use Coefficients:

These coefficients can be used by researchers to predict the probability that each person in the CPS sample can be linked to an MSIS administrative data that shows him/her receiving full benefit Medicaid for at least one day in the CPS survey reference period (the preceding calendar year). These probabilities can then be used to impute a new estimate of Medicaid coverage that partially corrects for misreporting. See Appendix A for two tables showing results from worked example using the 2005 and 2006 CPS data files as well as the example Stata code for generating the numbers.

Caveats:

This information sheet reports intermediate results from ongoing research. This new estimate of Medicaid coverage is likely to be an improvement over using only the self-reported Medicaid coverage variable. However, this analysis is also likely to be far from perfect. Among the outstanding issues is that this analysis uses data from Calendar Year 2000 and 2001 (CPS survey years 2001 and 2002). Response error patterns in those years may be substantially different from those in more recent CPS survey years. For example CPS coding of race and ethnicity changed substantially in the mid-2000s. . We will continue to address this issue using current MSIS data as it becomes available (MSIS fiscal year 2005 data is expected to become available September 2007).

In addition, for several reasons we are more confident about the models for those who, according to the CPS, do not have Medicaid, than for the models for those who according to the CPS do have Medicaid. First, because we could not link MSIS records without an SSN to the corresponding CPS record (and because these models do not otherwise adjust for such missing MSIS SSNs), the imputational model implicitly treats these people's CPS reports that they have Medicaid as response errors. This is clearly incorrect and yields an estimate of the number of people with Medicaid that is too small. The magnitude of the bias will depend on the extent to which MSIS records without an PIK (i.e., SSN) do not also appear in the file as individuals with an PIK (where due to the lack of an PIK, we could not unduplicate the records).

Second and relatedly, some of the people who in the CPS report that they have Medicaid, but do not appear in the MSIS files as having Medicaid, may be in combined Medicaid/SCHIP programs. For the purposes of MSIS (and therefore for the purposes of our imputation model), these individuals are not classified as in Medicaid.

Third, there is considerable evidence that many of the people who report having Medicaid in the CPS but are not found in the MSIS actually have some other form of health insurance (e.g., a Separate SCHIP program). Failure to adjust for the fact that some of these people probably have other (non-Medicaid) insurance will upwardly bias the estimate of the number of uninsured Americans.

Fourth, our analysis only considers response errors in the Medicaid variables. Just as our linked analysis has found response errors in the Medicaid variables, it seems clear that there are also response errors in the other health insurance variables (i.e., Medicare, other public health

insurance programs, private health insurance). Our linked data files provide no direct information on non-Medicaid errors. For example, it is possible that people without health insurance coverage at all incorrectly report that they have some type of coverage in the CPS (falsely claim to have private coverage for example when they are uninsured). We are unable to correct for this problem through the current SNACC project and it remains an unknown issue that should be investigated.

Ongoing analyses as part of the SNACC effort are investigating each of these issues. In particular, as part of our broader project, we are building more complicated statistical models to explore the implications of our findings for estimates of the uninsured and the sensitivity of such inferences to auxiliary assumptions. These models use all of the data (including those unidentified in the CPS and those with imputed or allocated responses) and explicitly model the effects of non-identification in the MSIS data.

For more information on the SNACC effort please contact Michael Davern at the University of Minnesota/SHADAC. daver004@umn.edu or by phone 612-624-4802.

Notes

[i] This analysis was made possible by the efforts of a great team of Census Bureau employees (especially Dean Resnick and Victoria Lynch), Office of the Assistant Secretary for Planning and Evaluation (ASPE) employees (George Greenberg and Rob Stewart), and the Center for Medicare and Medicaid Services (CMS) employees (Dave Baugh and Gary Ciborowski). The work was funded by the Robert Wood Johnson Foundation and the Office of the Assistant Secretary for Planning and Evaluation (ASPE) and headed by the State Health Access Data Assistance Center (SHADAC). Special thanks to Karen Soderberg at SHADAC for her help in editing and formatting this paper. All remaining errors and the opinions expressed are the solely the possession of the authors only.

[ii] Based on CPS estimates of the number of people with Medicaid and MSIS administrative data counts tabulated in the second phase of our research project. Our revised MSIS estimates for 2000 indicated that 38.2 million people had full benefits Medicaid compared with 26.1 million estimated by the CPS. In 2001 our revised MSIS estimates indicated that 40.5 million people were enrolled in full benefits Medicaid versus 27.7 million in the CPS.

[iii] . A Person has Medicaid if CPS variable CAID=1 or othstyp(1-6)=2 or ahityp(1-6)=2.

Appendix A: Tables from a Worked Example and Example Code

Table A-1: Comparing Medicaid Enrollment Estimates from our Partially Corrected Model to the Regular CPS Estimates by Selected Characteristics and State: Calendar Year 2004 and 2005 Average

	Medicaid Enrollment Estimate - CPS		Medicaid Enrollment Estimate - Model Adjusted	
	Percent	Number	Percent	Number
Alabama	12.9%	584,795	15.7%	709,775
Alaska	11.0%	71,874	12.7%	82,837
Arizona	15.4%	909,034	17.8%	1,051,249
Arkansas	14.3%	393,211	15.6%	429,035
California	14.5%	5,198,527	17.0%	6,109,544
Colorado	6.7%	305,448	8.3%	382,285
Connecticut	8.0%	278,325	9.1%	316,961
Delaware	9.8%	81,843	13.6%	113,561
District of Columbia	19.2%	104,489	21.6%	117,578
Florida	9.2%	1,626,431	12.0%	2,115,047
Georgia	11.4%	1,013,587	13.5%	1,195,717
Hawaii	8.9%	112,941	12.3%	155,380
Idaho	10.8%	151,601	11.5%	162,601
Illinois	8.5%	1,071,473	12.8%	1,607,336
Indiana	11.4%	702,588	13.5%	830,882
Iowa	10.4%	302,383	11.6%	336,221
Kansas	8.0%	216,075	9.9%	265,840
Kentucky	12.5%	506,291	13.7%	556,077
Louisiana	12.8%	543,804	15.6%	662,758
Maine	19.8%	258,685	23.3%	304,251
Maryland	7.3%	404,320	8.6%	476,109
Massachusetts	13.4%	849,966	13.1%	831,176
Michigan	12.3%	1,223,613	12.9%	1,287,374
Minnesota	8.3%	424,733	10.3%	530,499
Mississippi	17.9%	511,144	16.6%	474,579
Missouri	12.7%	718,953	16.1%	913,089
Montana	9.7%	89,219	6.1%	56,435
Nebraska	9.4%	164,837	12.6%	220,374
Nevada	5.6%	136,833	7.3%	176,045
New Hampshire	5.0%	64,682	6.7%	87,071
New Jersey	7.1%	619,255	7.8%	674,935
New Mexico	16.5%	317,544	19.6%	376,611
New York	15.4%	2,933,238	16.0%	3,051,811
North Carolina	11.8%	1,002,308	16.2%	1,379,612
North Dakota	7.0%	44,006	9.3%	58,078
Ohio	11.1%	1,259,101	12.9%	1,461,253
Oklahoma	11.9%	412,661	14.6%	507,547
Oregon	10.7%	386,896	12.3%	442,432
Pennsylvania	8.8%	1,073,654	13.0%	1,592,858
Rhode Island	16.2%	171,177	16.6%	175,441
South Carolina	13.8%	573,418	17.3%	718,658
South Dakota	10.8%	82,448	11.3%	85,882
Tennessee	15.8%	927,840	23.5%	1,376,352
Texas	10.7%	2,417,884	13.0%	2,933,890
Utah	8.4%	205,676	9.9%	243,551
Vermont	18.5%	114,523	21.4%	132,280
Virginia	6.5%	483,442	8.0%	590,495
Washington	10.2%	627,965	15.0%	926,844
West Virginia	12.1%	216,875	15.1%	271,236
Wisconsin	11.1%	606,197	11.7%	640,839
Wyoming	9.1%	46,095	9.9%	49,894
Total - United States	11.5%	33,543,900	13.8%	40,248,189

Table A-2: Comparing Uninsured Rates Based on our Partially Corrected Model to the Regular CPS Estimates by Selected Characteristics and State: Calendar Year 2004 and 2005 Average

	Uninsured Estimate - CPS		Uninsured Estimate - Model Adjusted	
	Percent	Number	Percent	Number
Alabama	13.5%	609,652	11.6%	522,544
Alaska	16.9%	110,235	14.9%	97,457
Arizona	18.1%	1,071,869	14.8%	875,279
Arkansas	16.8%	461,939	14.4%	394,823
California	18.4%	6,598,937	15.6%	5,599,237
Colorado	16.3%	745,308	14.2%	651,663
Connecticut	10.9%	380,711	9.5%	331,101
Delaware	12.7%	106,123	10.5%	87,833
District of Columbia	12.8%	69,328	10.3%	55,891
Florida	19.8%	3,499,008	17.0%	3,013,556
Georgia	17.6%	1,561,543	15.3%	1,357,974
Hawaii	8.5%	107,196	7.2%	90,924
Idaho	14.7%	206,762	13.2%	185,458
Illinois	13.4%	1,685,087	10.8%	1,366,986
Indiana	13.7%	839,701	11.8%	721,940
Iowa	8.7%	253,540	7.6%	219,901
Kansas	10.5%	281,884	9.3%	250,774
Kentucky	13.0%	526,699	11.3%	459,191
Louisiana	16.9%	715,848	14.1%	597,702
Maine	9.6%	125,510	7.5%	98,268
Maryland	13.4%	743,905	11.6%	647,335
Massachusetts	10.3%	651,546	8.9%	565,179
Michigan	10.7%	1,070,776	9.3%	928,962
Minnesota	8.2%	420,887	7.0%	361,374
Mississippi	16.8%	480,505	14.2%	405,169
Missouri	11.8%	669,149	9.6%	545,721
Montana	16.9%	155,199	15.8%	145,483
Nebraska	10.5%	182,658	8.8%	153,069
Nevada	17.7%	428,840	16.0%	386,655
New Hampshire	9.9%	128,243	9.1%	118,232
New Jersey	14.2%	1,232,984	12.6%	1,092,982
New Mexico	20.1%	385,235	16.2%	310,181
New York	12.8%	2,438,112	11.0%	2,093,914
North Carolina	15.1%	1,282,045	12.2%	1,039,916
North Dakota	10.5%	65,805	9.2%	57,489
Ohio	11.0%	1,240,446	9.5%	1,072,828
Oklahoma	18.5%	644,016	15.9%	552,094
Oregon	15.9%	574,479	13.9%	500,773
Pennsylvania	10.3%	1,263,457	8.5%	1,039,423
Rhode Island	10.9%	115,401	9.2%	96,850
South Carolina	16.0%	664,939	13.3%	553,636
South Dakota	11.4%	86,852	9.9%	75,399
Tennessee	13.3%	781,982	9.5%	558,673
Texas	23.9%	5,399,198	21.0%	4,732,408
Utah	14.9%	366,211	13.2%	326,079
Vermont	11.0%	68,323	9.1%	56,471
Virginia	13.1%	968,774	11.8%	875,925
Washington	12.8%	793,502	10.5%	649,271
West Virginia	16.5%	296,477	14.4%	258,822
Wisconsin	9.7%	530,106	8.6%	466,926
Wyoming	13.7%	69,153	12.4%	62,499
Total - United States	15.1%	44,156,084	12.9%	37,708,243

Example Stata Code for Creating Variables Used in Model Using 2006 and 2007 CPS Data

```
*** Medicaid Under-reporting in the
*** Current Population Survey and
*** One Approach for a Partial Correction

* October 2007
* Updated November 2008 - see model2 variable creation

** Combine 2005 and 2006 CPS

use "P:\CPS Stata files\cpsmar06.dta", clear
append using "P:\CPS Stata files\cpsmar05.dta"

save "C:\CPS\cps05_06.dta"

**** Generate CPS weights
gen geocode=.
replace geocode=gtcbsa if gtcbsa>0
replace geocode = (gestfips*100000) if geocode==.
svyset [w=marsupwt], psu (h_seq) strata (geocode)

** Health insurance recode

** first designate which model individuals are in
** In model2 if state medicaid (or imputed or edited to medicaid), else in model1

gen model2 = .
# d;
replace model2 = 1 if caid == 1 | othstyp1 == 2|othstyp2 == 2|othstyp3 ==2|othstyp4 == 2|
othstyp5 == 2|othstyp6 == 2 |ahityp1 == 2|ahityp2 ==2|ahityp3 ==2|ahityp4 == 2|
ahityp5 ==2|ahityp6 ==2;
# d cr
replace model2 = 0 if model2 == .

gen model1 = .
replace model1 = 1 if model2 == 0
replace model1 = 0 if model2 == 1

gen caidplus = model2
label var caidplus "caid and othstyp and ahityp"

***********************************************
*****                                     *****
*****   INSURANCE STATUS VARIABLES        *****
*****                                     *****
***********************************************

gen insurance = .
replace insurance = 1 if mcaid == 2 & mcare == 2 & cov_hi == 2 & champ == 2
replace insurance = 2 if (mcare == 1 | (mcaid == 1 & caidplus == 0)) & (cov_hi == 2 & champ == 2)
replace insurance = 3 if (cov_hi == 1 | champ == 1) & (mcaid == 2 & mcare == 2)
replace insurance = 4 if caidplus == 1 & (champ == 2 & cov_hi == 2 & insurance != 2)
replace insurance = 5 if insurance == .
label define insurance 1 "uninsured" 2 "public_nocaid" 3 "private_only" 4 "medicaid_only" 5
"public_private"
label values insurance insurance

tab insurance, gen(ins)

rename ins1 uninsured
rename ins2 public_nocaid
rename ins3 private_only
rename ins4 medicaid_only
rename ins5 public_private
```

```
************************************************************
/* DUMMY VARIABLE FOR MEDICAID IMPUTATION STATUS
        THE VARIABLE "i-caid" IMPUTATION FLAG HAS THREE VALUES:
                0 - NO
                1 - ALLOCATED
                2 - LOGICAL IMPUTED  */
************************************************************
tab i_caid, gen(medicaid_flag)
        rename medicaid_flag1 medicaid_explicit
        rename medicaid_flag2 medicaid_imputed
        rename medicaid_flag3 medicaid_edited

*** Add additional code for imputed from fl_665, othstyp, or ahityp

replace medicaid_imputed = 1 if (fl_665== 2|fl_665 == 3) | i_oth == 1 | iahityp == 1
replace medicaid_edited = 0 if medicaid_imputed == 1
replace medicaid_explicit = 0 if medicaid_imputed == 1

********************
/* RACE VARIABLES */
********************
* make as consistent with earlier years as possible - hisp, white only, then any american indian,
any black, and any asian
* the non-identified multiple races are assigned the modal category - white
gen race =  .
replace race = 1 if pehsp == 1
replace race = 2 if (prdtrace == 1|prdtrace == 20|prdtrace == 21) & race == .
replace race = 3 if (prdtrace == 3|prdtrace == 7|prdtrace == 10 |prdtrace== 13|prdtrace ==
15|prdtrace == 17|prdtrace == 19)  & race == .
replace race = 4 if (prdtrace == 2|prdtrace == 6|prdtrace == 11 |prdtrace == 12|prdtrace == 16) &
race == .
replace race = 5 if (prdtrace == 4 | prdtrace==5|prdtrace == 8 | prdtrace==9|prdtrace ==
14|prdtrace == 18) & race == .

        #d ;
        label define race
                3 "American Indian"
                4 "Black"
                5 "Asian/Pacific Islander"
                2 "White Only"
                1 "Hispanic", modify ;
        #d cr

        label values race race

tab race, gen(race_cat)
        rename race_cat1 phispanic
        rename race_cat2 pwhite
        rename race_cat3 pai
        rename race_cat4 pblack
        rename race_cat5 papi

*******************************
/* GENDER INDICATOR VARIABLE */
*******************************
gen male = .
replace male = 1 if a_sex == 1
replace male = 0 if a_sex == 2
label define male  1 "Male" 0 "Female"
label val male male


********************
/* AGE CATEGORIES */
********************
gen age_cat = .
replace age_cat = 1 if a_age < 6
replace age_cat = 2 if a_age >= 6 & a_age < 15
replace age_cat = 3 if a_age >= 15 & a_age < 18
replace age_cat = 4 if a_age >= 18 & a_age < 45
replace age_cat = 5 if a_age >= 45 & a_age < 65
```

```
replace age_cat = 6 if a_age >= 65

tab age_cat, generate(cat_ages)
        rename cat_ages1 age_005
        rename cat_ages2 age_614
        rename cat_ages3 age_1517
        rename cat_ages4 age_1844
        rename cat_ages5 age_4564
        rename cat_ages6 age_65up

*********************
/* INCOME VARIABLE */
*********************
generate ZERO = .
replace ZERO = 1 if ftotval <= 0
replace ZERO = 0 if ftotval > 0

*** dummy indicator variables for poverty ratios ***
replace rpl = .
replace rpl = 1 if povll == 1
replace rpl = 2 if povll == 2
replace rpl = 3 if povll == 3
replace rpl = 4 if povll == 4
replace rpl = 5 if povll == 5
replace rpl = 6 if povll == 6
replace rpl = 7 if povll == 7
replace rpl = 8 if povll > = 8

tab rpl, gen(rpl)

*****************************
/* RELATIONSHIP VARIABLES */
*****************************
gen rel = .
replace rel = 1 if a_exprrp == 3 | a_exprrp == 4
replace rel = 2 if a_exprrp == 5
replace rel = 3 if a_exprrp == 8
replace rel = 4 if a_exprrp == 6 | a_exprrp == 7 | (a_exprrp > 8 & a_exprrp <= 14)
replace rel = 5 if a_exprrp == 1 | a_exprrp == 2

tab rel, gen(rel)

rename rel1 rel_spouse
rename rel2 rel_child
rename rel3 rel_parent
rename rel4 rel_other
rename rel5 rel_self

****************
*** States *****
****************
# d;
label define state
        1       "Alabama"
        2       "Alaska"
        4       "Arizona"
        5       "Arkansas"
        6       "California"
        8       "Colorado"
        9       "Connecticut"
        10      "Delaware"
        11      "District of Columbia"
        12      "Florida"
        13      "Georgia"
        15      "Hawaii"
        16      "Idaho"
        17      "Illinois"
        18      "Indiana"
        19      "Iowa"
        20      "Kansas"
        21      "Kentucky"
```

```
        22      "Louisiana"
        23      "Maine"
        24      "Maryland"
        25      "Massachusetts"
        26      "Michigan"
        27      "Minnesota"
        28      "Mississippi"
        29      "Missouri"
        30      "Montana"
        31      "Nebraska"
        32      "Nevada"
        33      "New Hampshire"
        34      "New Jersey"
        35      "New Mexico"
        36      "New York"
        37      "North Carolina"
        38      "North Dakota"
        39      "Ohio"
        40      "Oklahoma"
        41      "Oregon"
        42      "Pennsylvania"
        44      "Rhode Island"
        45      "South Carolina"
        46      "South Dakota"
        47      "Tennessee"
        48      "Texas"
        49      "Utah"
        50      "Vermont"
        51      "Virginia"
        53      "Washington"
        54      "West Virginia"
        55      "Wisconsin"
        56      "Wyoming";

#d cr
label values gestfips state

* indicator variables for the states
dummieslab gestfips

*****************************
****** Prepare for models
*****************************
**** create constant variables that will save the coeff values for each model (from Table 1)
**** call each one varname1 or varname2 dependent on the model from which they are from

gen constant = 1

gen     constant1       =       -0.6089
gen     ZERO1           =       0.2475
gen     age_0051        =       1.3364
gen     age_6141        =       0.8797
gen     age_15171       =       0.6517
gen     age_18441       =       -0.0311
gen     age_45641       =       -1.0515
gen     age_65up1       =       -1.7853
gen     medicaid_edited1        =       .
gen     medicaid_imputed1       =       0.3617
gen     medicaid_explicit1      =       -0.3617
gen     public_nocaid1 =        1.1714
gen     private_only1  =        -1.0714
gen     public_private1         =       0.0936
gen     uninsured1      =       -0.1936
gen     medicaid_only1 =        .
gen     phispanic1      =       0.1155
gen     pblack1         =       0.5177
gen     pai1            =       0.1917
gen     papi1           =       -0.2467
gen     pwhite1         =       -0.5782
gen     male1           =       -0.5109
gen     rel_parent1     =       0.888
```

```
gen     rel_spouse1    =       -0.5062
gen     rel_child1     =       -0.2866
gen     rel_other1     =       0.1965
gen     rel_self1      =       -0.2917
gen     rpl11          =       0.3891
gen     rpl21          =       0.6237
gen     rpl31          =       0.45
gen     rpl41          =       0.1944
gen     rpl51          =       0.0504
gen     rpl61          =       -0.1552
gen     rpl71          =       -0.2717
gen     rpl81          =       -1.2808
gen     Alabama1       =       -0.1379
gen     Alaska1        =       -0.1272
gen     Arizona1       =       0.0813
gen     Arkansas1      =       0.1515
gen     California1    =       -0.124
gen     Colorado1      =       -0.3486
gen     Connecticut1   =       -0.1982
gen     Delaware1      =       0.2252
gen     DistrictofColumbia1    =       0.0206
gen     Florida1       =       -0.1452
gen     Georgia1       =       -0.3799
gen     Hawaii1        =       0.2828
gen     Idaho1         =       -0.2137
gen     Illinois1      =       0.1144
gen     Indiana1       =       0.1683
gen     Iowa1          =       0.0545
gen     Kansas1        =       -0.3241
gen     Kentucky1      =       0.0305
gen     Louisiana1     =       -0.1636
gen     Maine1         =       1.18
gen     Maryland1      =       -0.4281
gen     Massachusetts1 =       0.2211
gen     Michigan1      =       -0.1803
gen     Minnesota1     =       0.223
gen     Mississippi1   =       -0.3619
gen     Missouri1      =       0.4235
gen     Montana1       =       -1.0005
gen     Nebraska1      =       0.159
gen     Nevada1        =       -0.6962
gen     NewHampshire1  =       -0.1836
gen     NewJersey1     =       -0.3858
gen     NewMexico1     =       0.1199
gen     NewYork1       =       -0.1396
gen     NorthCarolina1 =       0.2104
gen     NorthDakota1   =       -0.0914
gen     Ohio1          =       -0.0658
gen     Oklahoma1      =       0.08
gen     Oregon1        =       -0.0195
gen     Pennsylvania1  =       0.3005
gen     RhodeIsland1   =       0.3507
gen     SouthCarolina1 =       0.174
gen     SouthDakota1   =       -0.1485
gen     Tennessee1     =       0.9171
gen     Texas1         =       -0.6106
gen     Utah1          =       -0.3107
gen     Vermont1       =       1.1751
gen     Virginia1      =       -0.5826
gen     Washington1    =       0.6428
gen     WestVirginia1  =       0.3519
gen     Wisconsin1     =       -0.0958
gen     Wyoming1       =       -0.1949

gen     constant2      =       0.7521
gen     ZERO2          =       -0.2862
gen     age_0052       =       0.396
gen     age_6142       =       0.4068
gen     age_15172      =       0.1538
gen     age_18442      =       0.1553
gen     age_45642      =       -0.2539
```

```
gen     age_65up2      =       -0.8579
gen     medicaid_edited2       =       -0.3439
gen     medicaid_imputed2      =       -0.7819
gen     medicaid_explicit2     =       1.1258
gen     public_nocaid2 =       .
gen     private_only2  =       .
gen     public_private2        =       .
gen     uninsured2     =       .
gen     medicaid_only2 =       0.1033
gen     phispanic2     =       -0.0447
gen     pblack2        =       0.1324
gen     pai2           =       0.0797
gen     papi2          =       -0.00385
gen     pwhite2        =       -0.1635
gen     male2          =       -0.3084
gen     rel_parent2    =       0.8454
gen     rel_spouse2    =       -0.6605
gen     rel_child2     =       -0.1461
gen     rel_other2     =       0.1694
gen     rel_self2      =       -0.2083
gen     rpl12          =       0.4922
gen     rpl22          =       0.5247
gen     rpl32          =       0.5368
gen     rpl42          =       0.1999
gen     rpl52          =       -0.1616
gen     rpl62          =       -0.21
gen     rpl72          =       -0.5104
gen     rpl82          =       -0.8718
gen     Alabama2       =       -0.0488
gen     Alaska2        =       -0.0857
gen     Arizona2       =       0.1248
gen     Arkansas2      =       -0.2814
gen     California2    =       0.3479
gen     Colorado2      =       -0.3851
gen     Connecticut2   =       -0.7219
gen     Delaware2      =       0.2802
gen     DistrictofColumbia2    =       -0.0589
gen     Florida2       =       -0.0341
gen     Georgia2       =       -0.2252
gen     Hawaii2        =       0.0564
gen     Idaho2         =       -0.1441
gen     Illinois2      =       -0.1066
gen     Indiana2       =       -0.0716
gen     Iowa2          =       0.348
gen     Kansas2        =       0.2111
gen     Kentucky2      =       -0.2099
gen     Louisiana2     =       -0.571
gen     Maine2         =       0.8533
gen     Maryland2      =       -0.8764
gen     Massachusetts2 =       -0.1872
gen     Michigan2      =       0.0434
gen     Minnesota2     =       0.2205
gen     Mississippi2   =       -0.9372
gen     Missouri2      =       0.3584
gen     Montana2       =       -1.6887
gen     Nebraska2      =       0.6703
gen     Nevada2        =       -0.6033
gen     NewHampshire2  =       0.7746
gen     NewJersey2     =       -0.6282
gen     NewMexico2     =       0.0559
gen     NewYork2       =       0.0361
gen     NorthCarolina2 =       0.4162
gen     NorthDakota2   =       0.3506
gen     Ohio2          =       0.2443
gen     Oklahoma2      =       -0.0671
gen     Oregon2        =       -0.0192
gen     Pennsylvania2  =       0.5203
gen     RhodeIsland2   =       0.2558
gen     SouthCarolina2 =       0.1124
gen     SouthDakota2   =       -0.2581
gen     Tennessee2     =       0.9406
```

```
gen     Texas2          =       -0.1475
gen     Utah2           =       0.0172
gen     Vermont2        =       0.853
gen     Virginia2       =       -0.4431
gen     Washington2     =       0.4109
gen     WestVirginia2   =       0.2588
gen     Wisconsin2      =       0.1001
gen     Wyoming2        =       -0.0608


* Create interaction terms for all variables * coefficients, as defined above.

#d cr
foreach i of varlist constant ZERO age_005 age_1517 age_4564 age_614 age_65up age_1844 medicaid_edited medicaid_imputed medicaid_explicit public_nocaid private_only public_private uninsured medicaid_only phispanic pblack pai papi pwhite male rel_parent rel_spouse rel_child rel_other rel_self rpl1 rpl2 rpl3 rpl4 rpl5 rpl6 rpl7 rpl8 Alabama Delaware DistrictofColumbia Florida Georgia Hawaii Idaho Illinois Indiana Iowa Alaska Kansas Kentucky Louisiana Maine Maryland Massachusetts Michigan Minnesota Mississippi Missouri Montana Nebraska Nevada NewHampshire NewJersey NewMexico NewYork NorthCarolina NorthDakota Ohio Arizona Oklahoma Oregon Pennsylvania RhodeIsland SouthCarolina SouthDakota Tennessee Texas Utah Arkansas Vermont Virginia Washington WestVirginia Wisconsin Wyoming California Colorado Connecticut {
        gen pr1_`i' = `i'*`i'1
        }

#d cr
foreach i of varlist constant ZERO age_005 age_1517 age_4564 age_614 age_65up age_1844 medicaid_edited medicaid_imputed medicaid_explicit public_nocaid private_only public_private uninsured medicaid_only phispanic pblack pai papi pwhite male rel_parent rel_spouse rel_child rel_other rel_self rpl1 rpl2 rpl3 rpl4 rpl5 rpl6 rpl7 rpl8 Alabama Delaware DistrictofColumbia Florida Georgia Hawaii Idaho Illinois Indiana Iowa Alaska Kansas Kentucky Louisiana Maine Maryland Massachusetts Michigan Minnesota Mississippi Missouri Montana Nebraska Nevada NewHampshire NewJersey NewMexico NewYork NorthCarolina NorthDakota Ohio Arizona Oklahoma Oregon Pennsylvania RhodeIsland SouthCarolina SouthDakota Tennessee Texas Utah Arkansas Vermont Virginia Washington WestVirginia Wisconsin Wyoming California Colorado Connecticut {
        gen pr2_`i' = `i'*`i'2
        }

* sum variables across case - model1 variables from above if in model1, model2 variables if in model2
egen model1var = rsum (pr1_constant - pr1_Connecticut) if model1 == 1
egen model2var = rsum (pr2_constant - pr2_Connecticut) if model2 == 1

gen model1exp = exp(model1var) if model1 == 1
gen model2exp = exp(model2var) if model2 == 1

* transform from log odds
gen pre_prob1= model1exp/(1+model1exp) if model1 == 1
gen pre_prob2 = model2exp/(1+model2exp) if model2 == 1

* Create one variable that creates two predicted probabilities, dependent on what model is applicable to case
gen preprob = .
replace preprob = pre_prob1 if model1 == 1
replace preprob = pre_prob2 if model2 == 1

* First create dummy variables for year

gen y2005 = .
replace y2005 = 1 if h_year == 2005
replace y2005 = 0 if h_year == 2006

gen y2006 = .
replace y2006 = 1 if h_year == 2006
replace y2006 = 0 if h_year == 2005
```

```
*** Create tables - table 1 is medicaid response by state and US - use caidplus

************************************************
******** Table One ********************************
************************************************

* 1.A) unadjusted medicaid rate, national and by state
quietly svy, subpop(y2005):mean caidplus
estat size, obs size

quietly svy, subpop(y2006):mean caidplus
estat size, obs size

quietly svy, subpop(y2005):mean caidplus, over(gestfips)
estat size, obs size

quietly svy, subpop(y2006):mean caidplus, over(gestfips)
estat size, obs size

* 1.B) adjusted medicaid rate - apply preprob to total population
quietly svy, subpop(y2005):mean preprob
estat size, obs size

quietly svy, subpop(y2006):mean preprob
estat size, obs size

quietly svy, subpop(y2005):mean preprob, over(gestfips)
estat size, obs size

quietly svy, subpop(y2006):mean preprob, over(gestfips)
estat size, obs size

************************************************
******** Table Two ********************************
************************************************

* 2.A) unadjusted uninsured rate
quietly svy, subpop(y2005):mean uninsured
estat size, obs size

quietly svy, subpop(y2006):mean uninsured
estat size, obs size

quietly svy, subpop(y2005):mean uninsured, over(gestfips)
estat size, obs size

quietly svy, subpop(y2006):mean uninsured, over(gestfips)
estat size, obs size

* 2.B) updated uninsured rate - Uninsured rate minus the number of uninsured who were switched to
medicaid (preprob_un)

***** preprob for unin
gen preprob_un = preprob * uninsured

quietly svy, subpop(if y2005 == 1):mean preprob_un
estat size, obs size

quietly svy, subpop(if y2006 == 1):mean preprob_un
estat size, obs size

quietly svy, subpop(if y2005 == 1):mean preprob_un, over(gestfips)
estat size, obs size
quietly svy, subpop(if y2006 == 1):mean preprob_un, over(gestfips)
estat size, obs size
```