Using Linked Survey and Administrative Records Studies to Partially Correct Survey Program Participation for Timely Policy Research Purposes

Michael Davern, Ph.D.
Assistant Professor, Research Director
SHADAC, Health Policy & Management
University of Minnesota

2009 Joint Statistical Meetings, Washington DC
August 2, 2009

Funded by a grant from the Robert Wood Johnson Foundation

Co-authors and the "SNACC" project team

- This presentation is co-authored with Jacob Klerman Jeanette Ziegenfuss and Michael Plotzke.
- SNACC Project Team:
  - National Center for Health Statistics
    - Chris Cox, Kim Lochner, Linda Bilheimer, Robin Cohen and Eve Powell-Griner
  - Office of the Assistant Secretary for Planning and Evaluation
    - George Greenberg, and Don Cox
  - U.S. Census Bureau Collaborators
    - Dean Resnick, Victoria Lynch, Amy O'Hara, Maryan Cammarata, and Chuck Nelson
  - Centers for Medicare and Medicaid Services
    - Dave Baugh, Gary Ciborowski
  - State Health Access Data Assistance Center
    - Michael Davern, Kathleen Thiede Call, Gestur Davidson, Lynn Blewett

Why is imputation or correction needed?

- Past research has shown a significant amount of Medicaid reporting errors in the Current Population Survey and in the National Health Interview Survey.
  - People get the type of coverage wrong quite often and some people with Medicaid are even coded as being uninsured.
- As a result survey estimates of Medicaid enrollment are below administrative data enrollment figures.
  - E.g., raw CPS count is 57% of the unadjusted MSIS count
- Survey estimates are important for health policy research.
  - Surveys are the only sources for population estimates on the uninsured.
  - CPS is used in the SCHIP funding formula.

Possible Approaches to Adjust Data for Reporting Errors:

- Create a linked data file and:
  1. Replace reported values with administrative data values:
     - Disclosure issues and it would not be timely.
  2. Estimate a regression model for being on Medicaid using the linked data (which tends to be dated by 5 - 7 years for a variety of reasons).
     - Then run the most recent public use microdata through the model to come up with predicted probabilities and use those to impute enrollment.
- In this paper we implement the second option on two surveys and we compare results.

Data

- The Census Bureau linked 2001 and 2002 CPS records with Medicaid Statistical information System (MSIS) data for CY 2000-2001.
  - Limitations of the linking
    - 9% of all full-benefit Medicaid cases in MSIS are missing linking keys.
      - Our analysis is limited to full-benefit Medicaid enrollees with linking identifiers.
    - In 2001 20% of CPS cases are missing linking keys (largely due to refusal to provide data).
      - Remaining CPS cases are reweighted to equal the whole population.

Data

- The Census Bureau linked 2001 and 2002 NHIS records with MSIS data.
  - There are important limitations of the linking:
    - MSIS are missing linking keys (as with the CPS link).
    - In 2001 48% (in 2002 it was only 31%) of NHIS cases are missing linking keys (largely due to refusal to have data linked).
      - Remaining NHIS cases are reweighted to equal the whole population in both years.

Analysis

- The imputational models use only predictors that are available in the public use file of the CPS and NHIS files.
  - This makes the results more useful to the wider health policy research community.
- Dependent variable in the models is whether the CPS case or NHIS case was linked to MSIS.

Analysis

- Survey cases are divided and two mutually exclusive logistic regressions:
  - One for people recorded as having Medicaid.
  - One for people not recorded as having Medicaid.
- For the CPS each of the 2008 and 2007 public use data files were run through these regression models to obtain their predicted probability of being linked.
- For the NHIS each of the 2006 and 2007 public use data files were run through these regression models.
- We then impute Medicaid enrollment based on the predicted probability which both gives Medicaid coverage to some and takes it away from others.

Selected covariates used in the regressions

- Covariates of being linked include:
  - Relationship to household reference person
  - Age
  - Imputation/editing
  - Poverty status
  - Sex
  - Race and ethnicity
  - State (CPS only)
  - Type of health insurance status in survey
- Model coefficients, and sample SAS and Stata CPS coding are available on SHADAC's web site in a technical paper.

NHIS imputed results versus regular survey estimates for Medicaid

Table 3 & 4: Comparing Medicaid Enrollment Estimates from our Partially Corrected Imputation Model to the NHIS Estimates by Selected Demographic Characteristics: Calendar Year 2006 and 2007 Average

Selected Characteristics	Medicaid Enrollment Estimate - NHIS			Medicaid Enrollment Estimate - Imputed NHIS		
	Percent	SE	Number	Percent	SE	Number
Sex						
Male	9.51%	0.22%	13,722,856	12.14%	0.19%	17,512,551
Female	11.68%	0.24%	17,558,144	14.34%	0.21%	21,554,265
Age						
0 - 17	23.09%	0.49%	16,970,192	32.94%	0.43%	24,210,492
18 -44	7.30%	0.20%	8,055,751	8.00%	0.17%	8,824,793
45 - 64	5.06%	0.17%	3,794,543	4.89%	0.14%	3,660,325
65+	6.88%	0.31%	2,460,515	6.63%	0.23%	2,371,207
Poverty (% FPL)						
0-50% FPL	40.24%	1.63%	4,084,482	46.36%	1.42%	4,705,421
50-74% FPL	47.99%	1.53%	3,724,195	53.42%	1.11%	4,145,704
75-99% FPL	37.29%	1.29%	3,492,379	42.42%	0.91%	3,972,265
100-124% FPL	26.92%	1.15%	2,715,190	33.46%	0.81%	3,374,555
125-149% FPL	20.46%	1.25%	2,013,491	24.81%	0.96%	2,441,691
150-174% FPL	14.99%	1.09%	1,457,541	18.84%	0.77%	1,831,753
175-199% FPL	11.87%	0.86%	1,208,934	14.80%	0.65%	1,506,375
>= 200% FPL	2.41%	0.10%	3,722,501	3.54%	0.08%	5,458,056
Unknown FPL	12.12%	0.44%	8,862,289	15.91%	0.37%	11,630,996

NHIS imputed results versus regular survey estimates for Medicaid

Table 3 & 4: Comparing Medicaid Enrollment Estimates from our Partially Corrected Imputation Model to the NHIS Estimates by Selected Demographic Characteristics: Calendar Year 2006 and 2007 Average

Selected Characteristics	Medicaid Enrollment Estimate - NHIS			Medicaid Enrollment Estimate - Imputed NHIS		
	Percent	SE	Number	Percent	SE	Number
Race/Ethnicity						
White	6.56%	0.22%	12,795,835	8.30%	0.19%	16,195,735
AA	21.89%	0.65%	7,921,316	26.76%	0.55%	9,684,046
AIAN	18.60%	2.79%	353,090	26.92%	2.31%	511,109
Asian	7.59%	0.52%	956,335	11.15%	0.46%	1,404,714
Hispanic	18.97%	0.46%	9,254,426	23.11%	0.38%	11,271,213
Relationship						
Self	6.83%	0.17%	8,190,952	7.20%	0.14%	8,629,533
Parent	13.55%	0.87%	701,233	10.78%	0.56%	557,927
Spouse	3.03%	0.13%	1,796,322	2.65%	0.09%	1,567,729
Young Child	21.03%	0.49%	14,024,815	31.13%	0.43%	20,757,449
Adult Child	9.80%	0.45%	1,962,826	11.55%	0.39%	2,313,014
Other	19.50%	0.68%	4,604,853	22.20%	0.56%	5,241,165
Insurance Status as Reported in the NHIS						
Medicaid only	100.00%	0.00%	24,106,505	83.97%	0.18%	20,243,373
Medicaid + Other	100.00%	0.00%	7,174,495	77.08%	0.43%	5,529,938
Public Only	0.00%	0.00%	0	21.39%	0.60%	4,936,189
Private Only	0.00%	0.00%	0	2.39%	0.03%	4,005,274
Public and Private	0.00%	0.00%	0	2.97%	0.15%	780,533
Uninsured	0.00%	0.00%	0	7.71%	0.15%	3,571,509
Total - Overall	10.62%	0.21%	31,281,000	13.26%	0.19%	39,066,816

CPS results: Selected state rates of Medicaid enrollment

Table 3: Comparing Medicaid Enrollment Estimates from our Partially Corrected Imputation Model to the Regular CPS Estimates by Selected Characteristics and State: Calendar Year 2006 and 2007 Average

State	Medicaid Enrollment Estimate - CPS		Medicaid Enrollment Estimate - Imputed	
	Percent	Number	Percent	Number
Montana	10.7%	100,137	6.6%	61,470
Massachusetts	14.7%	933,550	13.9%	882,257
Rhode Island	17.1%	179,941	16.7%	174,960
Mississippi	16.7%	484,803	16.5%	478,696
Wisconsin	11.5%	628,074	12.0%	654,742
New York	15.6%	2,966,617	16.2%	3,092,605
Michigan	11.9%	1,181,475	12.7%	1,261,259
Kentucky	13.6%	567,655	14.7%	610,185
Idaho	9.9%	147,320	10.9%	161,746
District of Columbia	18.5%	106,410	20.5%	117,797
Iowa	11.0%	322,924	12.2%	358,425
South Dakota	8.8%	68,402	9.8%	76,313
Ohio	12.0%	1,356,077	13.5%	1,521,382
Arkansas	15.3%	424,882	17.4%	482,660
Virginia	7.1%	539,975	8.1%	615,754

CPS results: Selected state rates of Medicaid enrollment

Table 3: Comparing Medicaid Enrollment Estimates from our Partially Corrected Imputation Model to the Regular CPS Estimates by Selected Characteristics and State: Calendar Year 2006 and 2007 Average

	Medicaid Enrollment Estimate - CPS		Medicaid Enrollment Estimate - Imputed	
State	Percent	Number	Percent	Number
North Dakota	8.0%	49,512	10.3%	63,651
Illinois	10.3%	1,302,901	13.4%	1,692,258
New Hampshire	5.6%	73,279	7.3%	95,309
Alaska	7.9%	52,717	10.3%	68,787
Hawaii	9.6%	120,995	12.6%	158,258
Georgia	9.8%	921,076	12.9%	1,213,240
Washington	11.1%	713,811	15.1%	969,239
Nevada	5.2%	131,723	7.1%	180,217
Delaware	10.0%	86,083	13.7%	117,787
Missouri	11.5%	665,376	15.8%	916,236
North Carolina	11.9%	1,075,839	16.8%	1,513,618
Florida	8.3%	1,492,133	11.7%	2,108,309
Pennsylvania	9.3%	1,141,641	13.3%	1,634,183
Nebraska	7.8%	137,848	11.6%	203,913
Tennesse	14.1%	852,853	22.0%	1,327,184
Total - United States	11.4%	33,943,913	13.8%	40,978,989

a2

CPS results: Selected demographic characteristics

Table 4: Comparing Medicaid Enrollment Estimates from our Partially Corrected Impuation Model to the Regular CPS Estimates by Selected Demographic Characteristics: Calendar Year 2006 and 2007 Average

Selected Characteristics	Medicaid Enrollment Estimate - CPS		Medicaid Enrollment Estimate - Imputed	
	Percent	Number	Percent	Number
Sex				
Female	12.3%	18,690,402	15.7%	23,796,451
Male	10.4%	15,253,511	11.7%	17,182,538
Age				
0 to 5	28.0%	6,940,135	35.2%	8,721,131
6 to 14	21.6%	7,806,389	27.0%	9,759,070
15 to 17	17.2%	2,298,373	21.3%	2,844,614
18 to 44	8.1%	8,927,380	11.2%	12,361,727
45 to 64	6.5%	4,960,652	5.9%	4,505,927
65 and older	8.3%	3,010,991	7.7%	2,786,526
Poverty (% FPL)				
0-49%	38.6%	6,143,220	48.3%	7,684,512
50-75%	44.7%	4,338,799	52.9%	5,132,264
75-99%	37.7%	4,393,453	45.7%	5,320,193
100-124%	28.8%	3,865,479	35.6%	4,773,715
125-149%	20.8%	2,860,185	27.1%	3,728,322
150-174%	16.8%	2,193,058	22.0%	2,877,823
175-199%	12.4%	1,687,144	16.7%	2,277,475
>200%	4.1%	8,462,580	4.4%	9,184,698

Slide 14

a2 "cases" not "cass" on pages 13-17. Also, which year was the "last year"? (2000, 2001, or 2007?)
alte0083, 11/12/2008

a3

Selected demographic characteristics

Table 4: Comparing Medicaid Enrollment Estimates from our Partially Corrected Impuation Model to the Regular CPS Estimates by Selected Demographic Characteristics: Calendar Year 2006 and 2007 Average

Selected Characteristics	Medicaid Enrollment Estimate - CPS		Medicaid Enrollment Estimate - Imputed	
	Percent	Number	Percent	Number
Race/Ethnicity				
Hispanic	18.9%	8,570,519	24.6%	11,196,304
White Only	7.8%	15,426,485	8.5%	16,712,533
American Indian	19.1%	755,026	23.9%	943,027
Black	21.0%	7,847,696	28.0%	10,463,993
Asian/Pacific Islander	9.1%	1,344,188	11.3%	1,663,139
Employment Status^				
Not working	16.6%	14,034,875	17.2%	14,563,697
Working	9.3%	19,909,039	12.4%	26,415,292
Insurance Status as Reported in the CPS				
Uninsured	0.0%	-	14.3%	6,635,088
Public, No Medicaid	20.1%	4,243,310	28.3%	5,984,334
Private Only	0.0%	-	3.2%	5,665,833
Medicaid Only	100.0%	23,445,465	80.6%	18,891,537
Public and Private	21.1%	6,255,141	12.8%	3,802,201

Discussion of adjusted results from the NHIS model

- These results do not take into account adjustments to the NHIS that improved Medicaid reporting in 2005.
- 25 percentage point increase in the Medicaid Enrollment with imputation in the U.S.
  - 7.8 million more enrolled than the unadjusted NHIS.
- Largest absolute increases were for kids, blacks, Hispanics, and low income-to-poverty ratios.
- Many people linked to Medicaid fail to report any other type of coverage (over 3.6 million).
  - About 8% of the 46.3 million people in the NHIS estimated to be uninsured.

Slide 15

a3 "cases" not "cass" on pages 13-17. Also, which year was the "last year"? (2000, 2001, or 2007?)
alte0083, 11/12/2008

Discussion of adjusted results from the CPS model

- 21 percentage point increase in the Medicaid Enrollment with imputation in the U.S.
  - 7 million more enrolled than the straight CPS.
- Bigger percentage adjustments for someone in the family working, women, blacks, Hispanics, lower income, etc.
- Many people linked to Medicaid fail to report any other type of coverage (over 6.6 million).
  - About 14% of the 46.3 million the CPS estimates to be uninsured.

What do we learn from both models?

- Similarities:
  - Over 20% increase in Medicaid count in both surveys.
    - Gets much closer to “adjusted” administrative data totals.
  - Similar characteristics of those who misreport in both surveys.

What do we learn from both models?

- Differences:
  - NHIS is a point in time estimate of 39 million enrolled in Medicaid using our model.
  - CPS is an ever enrolled in Medicaid estimate of 41 million using our adjusted model .
  - NHIS has far fewer people who were coded in the survey as being uninsured who were linked to Medicaid enrollment data (3.6 million versus 7 million in the CPS).
    - NHIS has a less error prone measure of insurance coverage than the CPS.
    - NHIS reporting of Medicaid has likely improved with the new coding scheme they introduced in 2005 than we see in our linked model.

Our two models could also be used to partially correct uninsurance estimates

- Need to adjust the surveys for those cases reported to be uninsured that actually link to Medicaid.
- Need to adjust the surveys for those cases who reported only Medicaid but who did not link to the Medicaid data.
  - Without this report of coverage (which could not be verified) they would have otherwise been uninsured.
    - In the CPS there are almost 4.5 million weighted cases where its there only type of insurance.
    - For many reasons that have to due with limitations of our model we believe many of these cases do have insurance but more research is needed.

Strengths of this approach

- Our approach reduces the survey undercount and comes closer to administrative data targets of enrollment.
- Can be used to develop improved estimates of the eligible but not enrolled populations for Medicaid.
- CPS model can be used to show how well various states do in informing their Medicaid enrollees they have coverage.
  - Some states have vastly different probabilities of reporting being uninsured even those the administrative data shows enrollment.

Limitations of our approach

- We treat the CPS as a "all year uninsured" concept as the question literally reads.
  - Many people think the CPS is a "point in time measure".
- We only validate Medicaid coverage and not other sources (SCHIP, Medicare, Private, etc.).
  - This is truly only a "partial adjustment" as there are many more factors for which we need better data.
- We use data from 2001 and 2002 to simulate findings for 2006, 2007 and 2008.
- Missing identifying information on the CPS, NHIS and MSIS are troubling.
- We could not take advantage of the improvements made to Medicaid reporting in the 2005 NHIS.

Next steps in our SNACC project plan

- Try to get a better handle on SCHIP and how it impacts reporting errors.
  - New project under way to use the limited SCHIP information reported in the MSIS to make projections.
- Use the 2005 data from the CPS in the linked model (the 2005 has a higher proportion of linked cases).
- Use the 2005 NHIS data and take advantage of the improved Medicaid variables.

Contact information

- Michael Davern
  State Health Access Data Assistance Center (SHADAC), University of Minnesota
  - daver004@umn.edu
  - 612-624-4802