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 AAPOR, Hollywood Fl.
May 15th, 2009

Co-authors and the "SNACC" project team

- This presentation is co-authored with Jacob Klerman Jeanette Ziegenfuss and Michael Plotzke

- SNACC Project Team
  - 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, and Chuck Nelson
  - Abt Associates
    - Jacob Klerman and Michael Plotzke
  - 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

www.shadac.org

2

Why is imputation or correction needed?

- In this paper we focus on Medicaid enrollment and our ability to partially correct the Current Population Survey
- Survey estimates of Medicaid enrollment are well below administrative data enrollment figures
  - Raw CPS count is 57% of the unadjusted MSIS count
- CPS estimates are important for health policy research
  - Surveys such as the CPS are the only sources for population estimates on the uninsured
  - CPS is used in the SCHIP funding formula
  - CPS is often used to evaluate federal programs and state initiatives

Possible Approaches to fix using a linked data file:

- Create a linked data file and:
  - 1. replace reported values with administrative data values:
    - Disclosure issues and it would not be timely
  - 2. Release the linked file to researchers?
    - Disclosure issues and it would not be timely
  - 3. Estimate a regression coefficients for being on the program using the older linked data and run the most recent set of microdata through the model to come up with predicted probabilities and use those to impute enrollment
  - In this paper we implement number 3 and we also
    - Discuss the strengths and limitations of this approach
    - Conclude with discussion of our next steps

Data

- The Census Bureau linked 2001 and 2002 CPS records with MSIS data for CY 2000-2001
  - There are important limitations of the linking
    - 9% of all full benefit Medicaid cases in MSIS are missing linking keys
      - Our analysis 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

5

Analysis

- The Imputational models use only predictors that are available in the public use file of the CPS
  - So they can be useful to the wider health policy research community
- Dependent variable in the models is whether the CPS case was linked to MSIS

6

Analysis

- CPS cases are divided and two mutually exclusive logistic regressions
  - One for people recorded as having Medicaid in the CPS
  - One for people not recorded as having Medicaid
- Each case in the 2008 and 2007 CPS data files are run through these regression models to obtain their predicted probability of being linked
- We then impute Medicaid enrollment which both gives Medicaid coverage to some and takes it away from others

Selected covariates used in the regressions

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

Demographics of linked CPS cases

	Percent of cases linked (row %)	Percent of total linked cases by characteristic (col %)
Total Unweighted Count	13.3%	22,869
Total Weighted Count	12.5%	100.0%
Age 0 - 5	31.7%	21.5%
Age 6 - 14	23.4%	25.5%
Age 15 - 17	18.1%	5.6%
Age 18 - 44	9.5%	30.0%
Age 45 - 64	5.3%	9.6%
Age 65+	8.2%	7.7%
Age Other*	1.6%	0.0%
White	9.9%	64.7%
Black	27.5%	28.8%
AIAN	27.5%	2.7%
API	10.8%	3.9%
Male	10.4%	40.6%
Female	14.4%	59.3%
Hispanic	23.8%	21.7%
Non-Hispanic	11.0%	78.3%

9

Demographics of linked CPS cases (continued)

	Percent of cases linked (row %)	Percent of total linked cases by characteristic (col %)
Ratio to Poverty Level 0 - 49%	48.5%	18.2%
Ratio to Poverty Level 50 - 74%	51.5%	12.5%
Ratio to Poverty Level 75 - 99%	43.7%	13.6%
Ratio to Poverty Level 100 - 124%	33.0%	11.2%
Ratio to Poverty Level 125 - 149%	24.6%	9.0%
Ratio to Poverty Level 150 - 174%	20.0%	7.5%
Ratio to Poverty Level 175 - 199%	15.6%	5.7%
Ratio to Poverty Level 200% or Greater	4.0%	22.3%
Relationship to Reference Person: Self	8.4%	26.7%
Relationship to Reference Person: Spouse	3.8%	6.2%
Rltnshp. to Ref. Pers.: Child (Non-Adult)	22.2%	42.0%
Rltnshp. to Ref. Pers.: Child (Adult)	11.7%	7.0%
Relationship to Refernce Person: Parent	19.8%	1.6%
Relationship to Refernce Person: Other	25.4%	16.3%

10

States rates of Medicaid enrollment (15 lowest percentage point changes)

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

11

States rates of Medicaid enrollment (15 highest percentage point changes)

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
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

12

6

a2

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

www.shadac.org

13

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

www.shadac.org

14

Slide 13

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

Slide 14

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

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

- Need to adjust the CPS for those cases reported to be uninsured that actually link to Medicaid
- Need to adjust the CPS 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

www.shadac.org

15

Partially adjusted uninsurance rate

Table 5: Comparing Uninsured Rates Based on our Partially Corrected Impuation Model to the Regular CPS Estimates by Selected Demographic Characteristics: Calendar Year 2006/2007

Selected Characteristics	CPS Uninsurance Rate		Adjusted Uninsurance Rate*	
	Rate	Number	Rate	Number
Sex				
Female	14.03%	21,287,345	13.03%	19,763,008
Male	17.12%	25,038,571	16.74%	24,481,743
Age				
0 to 5	10.89%	2,694,685	10.32%	2,553,109
6 to 14	11.03%	3,984,057	10.52%	3,799,896
15 to 17	12.91%	1,726,352	12.44%	1,663,292
18 to 44	23.95%	26,545,975	22.18%	24,590,437
45 to 64	14.08%	10,761,026	14.43%	11,029,919
65 and older	1.68%	613,814	1.67%	608,096
Poverty (% FPL)				
0-49%	35.49%	5,645,218	30.31%	4,821,839
50-75%	28.25%	2,741,516	24.60%	2,387,791
75-99%	27.80%	3,238,163	24.02%	2,797,406
100-124%	26.90%	3,609,416	24.42%	3,276,647
125-149%	26.82%	3,692,864	24.98%	3,439,870
150-174%	23.02%	3,014,626	21.86%	2,862,895
175-199%	23.36%	3,191,545	22.36%	3,054,900
>200%	10.25%	21,192,564	10.45%	21,603,403

www.shadac.org

16

8

Partially adjusted uninsurance rate

Table 5: Comparing Uninsured Rates Based on our Partially Corrected Imputation Model to the Regular CPS Estimates by Selected Demographic Characteristics: Calendar Year 2006/2007

Selected Characteristics	CPS Uninsurance Rate		Adjusted Uninsurance Rate*	
	Rate	Number	Rate	Number
Race/Ethnicity				
Hispanic	33.10%	15,032,840	30.15%	13,694,925
White Only	10.61%	20,857,496	10.75%	21,138,488
American Indian	22.23%	878,587	19.93%	787,618
Black	19.51%	7,280,289	17.18%	6,410,367
Asian/Pacific Islander	15.49%	2,276,702	15.06%	2,213,353
Employment Status^				
Not working	16.27%	13,764,860	15.88%	13,434,769
Working	15.26%	32,561,049	14.44%	30,809,978
Medicaid				
Explicit	14.72%	37,374,785	13.76%	34,956,410
Imputed	22.52%	8,951,122	21.18%	8,413,805
Edited	0.00%	0	20.68%	874,539
Total	15.5%	46,325,916	14.85%	44,244,749

17

Discussion of adjusted results from the model

- 21 percentage point increase in the Medicaid Enrollment with imputation in the US
  - 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 percent of the 46.3 million the CPS estimates to be uninsured

18

9

Discussion of adjusted results from the model (continued)

- Many people report Medicaid whom we can not link to MSIS
  - For almost 4.5 million weighted cases its there only type of insurance)
- The net change to the uninsured rate using both adjustments is 2.1 million less uninsured (or 5 percent of the uninsured)
  - For many reasons that have to due with limitations of our model we believe this adjustment is too low

19

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
- 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

20

Limitations of our approach

- We treat the CPS as a "all year uninsured" concept as the question literally reads
  - Program eligibility does not require that people are uninsured for the entire prior calendar year
  - 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 we need better data on
- We use data from 2001 and 2002 CPS to simulate findings for 2007 and 2008 CPS
- Missing identifying information on the CPS and MSIS are troubling

Next steps in our SNACC project plan

- Finalize similar analysis on the National Health Interview Survey
  - Basic regression models have been completed
- Adding more covariates to our model and more years of linked data
- 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

Contact information

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

www.shadac.org