



Fitting Square Pegs Into Round Holes: Linking Medicaid and Current Population Survey Data to Understand the 'Medicaid Undercount'

Center's for Medicare and Medicaid Services
Baltimore, MD February 15, 2007

Michael Davern, Ph.D.

Assistant Professor, Research Director
SHADAC, Health Policy & Management
University of Minnesota

Supported by a grant from The Robert Wood Johnson Foundation and the Office of the Assistant Secretary for Planning and Evaluation

“SNACC” Project collaborators and co-authors

- Collaborators:
 - Office of the Assistant Secretary for Planning and Evaluation:
 - Rob Stewart
 - George Greenberg
 - Kate Bloniarz
 - US Census Bureau Collaborators:
 - Sally Obenski
 - Ron Prevost
 - Dean Resnick
 - Marc Roemer
- Coauthors:
 - RAND
 - Jacob Klerman
 - Centers for Medicare and Medicaid Services
 - Dave Baugh
 - Gary Ciborowski
 - State Health Access Data Assistance Center
 - Kathleen Thiede Call
 - Gestur Davidson
 - Lynn Blewett



What is the Medicaid undercount ?

- Survey estimates of Medicaid enrollment are well below administrative data enrollment figures
- In the present study we compare the Medicaid Statistical Information System (MSIS) to the Current Population Survey (CPS).
 - Extent of Undercount Varies by Age, and Geography



3

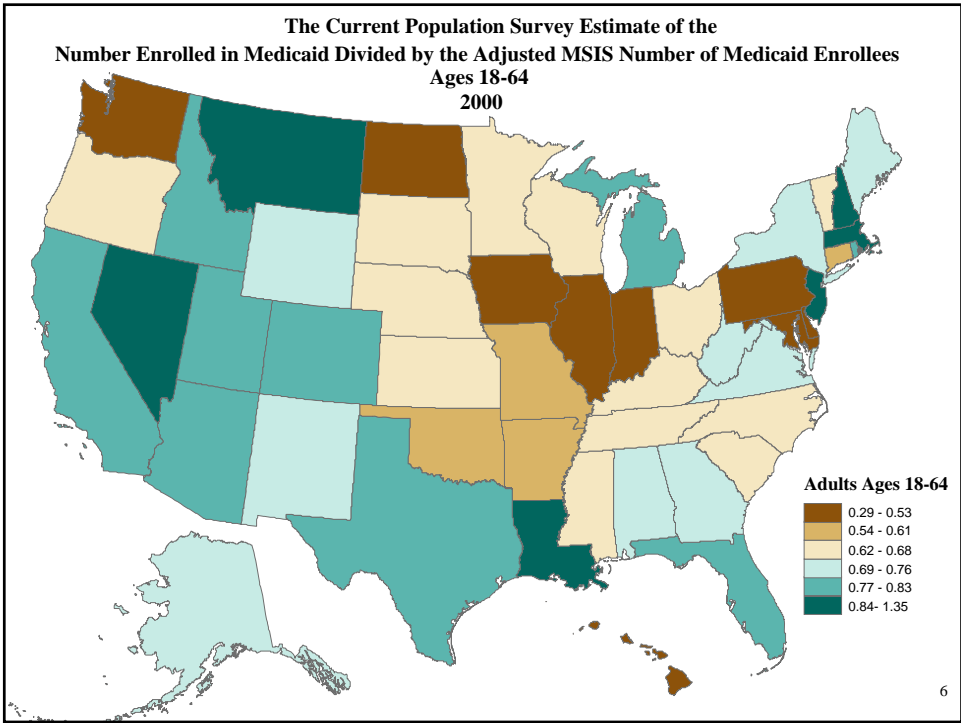
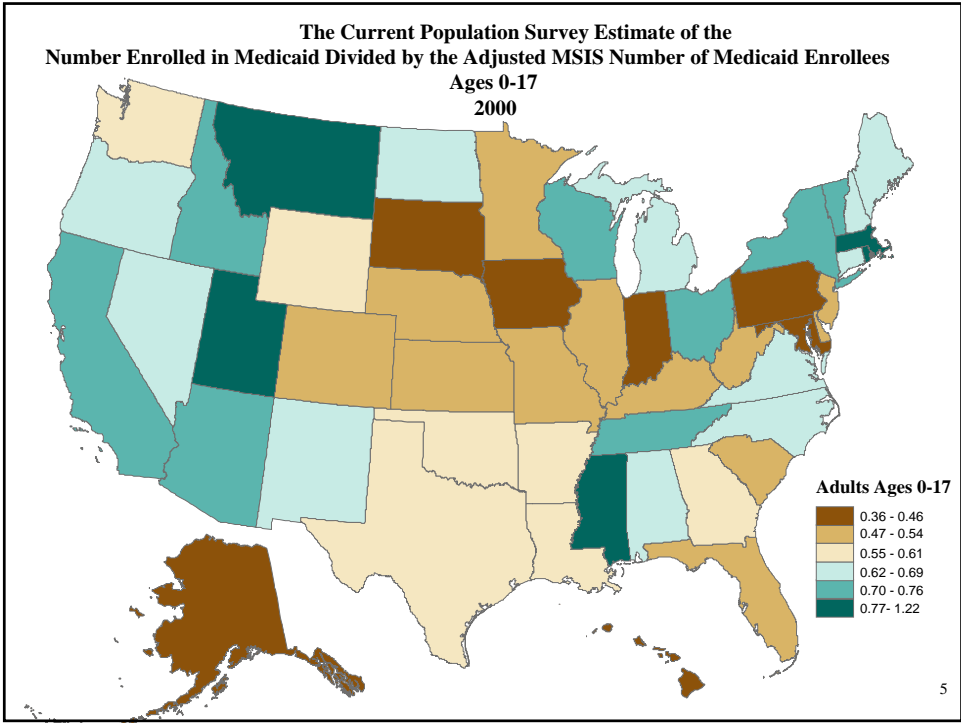
Raw ratios of enrollment counts

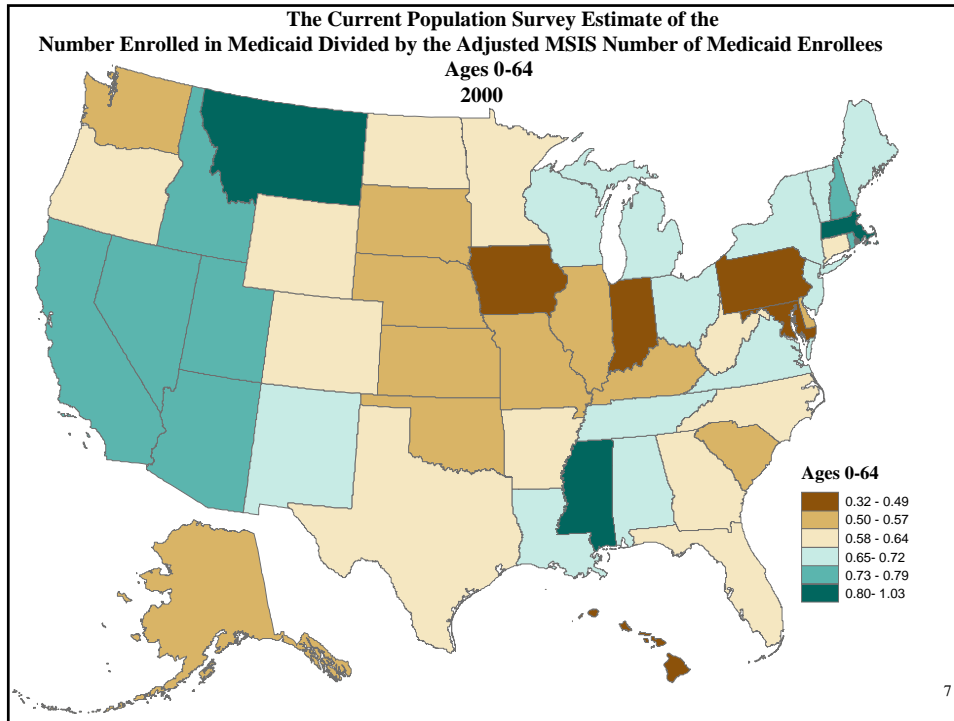
Table 1: Current Population Survey (CPS) Medicaid Enrollment Counts Relative to Enrollment Counts from the Medicaid Statistical Information System (MSIS): Ever Enrolled During Calendar Year 2000

Selected Characteristics	MSIS Total*	CPS Total	Ratio of CPS to
Age 0 - 5	8,840,115	5,063,124	0.573
Age 6 - 14	9,244,849	5,958,622	0.645
Age 15 - 17	2,228,258	1,465,603	0.658
Age 18 - 44	10,601,999	6,403,361	0.604
Age 45 - 64	3,570,495	3,249,252	0.910
Total	34,485,716	22,139,962	0.642

*MSIS is total enrolled for full benefits, excluding residents of institutional group quarters and duplicative records.

4





Why do we care?

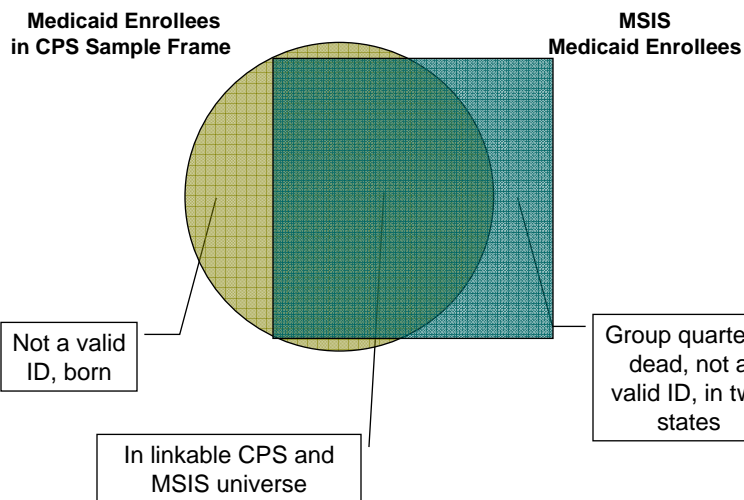
- CPS estimates are important to health policy research
 - Used for policy simulations by federal and state governments
 - Surveys like the CPS are the only sources for population estimates on the uninsured
 - Surveys are also the only source of the Medicaid/SCHIP eligible, but uninsured population
 - CPS is used in the SCHIP funding formula
 - CPS is often used to evaluate federal programs and state initiatives
- The magnitude of the undercount calls the validity of CPS survey insurance estimates into question

8

What could explain the undercount?

- Explanations I have some data on today:
 - Universe differences between MSIS and CPS survey data
 - Measurement error
- Explanations that are in progress:
 - Administrative and survey data processing, editing and imputation
 - Survey sample coverage error and survey nonresponse bias

Building a common 'linked universe'



Preparing MSIS Data for linking to CPS

- Removed MSIS cases defined as a “group quarter” by Census geography
 - These cases are not in the sample frame of the CPS
 - Our ability to remove group quarters addresses was limited
 - To limit the problem we focus on 0-64 year olds
 - A third phase of the project is under way to fix this issue
- Removed duplicate valid records
- Removed those MSIS enrollees not enrolled in “full benefits”
 - Not considered comprehensive health insurance coverage
- Ran the 2000 MSIS data through Census Bureau’s Person-ID validation system
 - A record is “valid” if it has a social security number in the appropriate format and demographic data is consistent

Fitting the MSIS records into the “linkable universe”

Number of MSIS Medicaid records in CY 2000 for persons Under 65:

39.2 M (total Medicaid MSIS records)

- .1 M (known group quarters)

- 1.3 M (duplicative records)

- 3.3 M (partial Medicaid benefits)

34.5 M (the target Medicaid total)

- 1.7 M (Not a valid ID on record)

32.8 M (records in the “Linkable Universe”)

Fitting the CPS into the “linkable universe”

Weighted total of CPS records 65 and under:

247.3 M

-19.0 M (respondent’s record not validated)

-37.6 M (respondent refused to provide an SSN)

190.7 M (total eligible to be linked to MSIS)

- We assumed the CPS missing and unvalidated records were like the validated records by age, race, ethnicity and poverty status and we re-weighted the validated data to represent 247.3 million (from 190.7 M)
- We are currently modeling other ways as well and that part of the project is being lead by RAND



13

Comparing linked universe totals

32.8 M Target MSIS total

30.8 M CPS reweighted linked total

2.0 M Mismatch between CPS and MSIS universe

- **94%** of total expected ‘linkable universe’ was matched between MSIS and CPS
- Why is the CPS 2.0 M lower?
 - Further universe issues we need to explore
 - Group quarters definitions
 - Survey sample frame coverage
 - Survey non-response
 - Problems with weighting (pop control totals, etc)

14

How well is the CPS classifying Medicaid enrollees?

- 19.8% of linked records are edited/imputed and they are left off the linked measurement analysis below.
- Focusing on only those with reported health insurance data
 - 57.6%** Respond Medicaid
 - 7.6%** Respond some other type of public
 - 17.0%** Respond some type of private coverage
 - 17.8%** Respond they were uninsured
 - 100.0%**
- Note: For those of you with calculators in hand this *crudely* equals 5.5 million less uninsured out of 39.6 million in CY 2000
- **More important note: This is far from our final answer. We are awaiting detailed modeling results**



15

What factors are associated with measurement accuracy/error?

- Length of time enrolled in Medicaid
- Recency of enrollment in Medicaid
- Poverty status impacts Medicaid reporting but does not impact the percent reporting they are uninsured
- Adults 18-44 are less likely to report Medicaid enrollment
- Adults 18-44 more likely to report being uninsured



16

How does the CPS compare to other studies of Medicaid reporting error?

Experimental and Matching Studies Reports of Insurance Coverage

Studies and Target Population	Percent of Medicaid Population Answering Correct Insurance Type	Percent of Medicaid Population Answering Wrong Insurance Type	Percent of Medicaid Population Answering They are Uninsured
Experimental Studies			
Adults on Medicaid in CA 2004	83.1%	6.4%	10.5%
Non-Elderly (<65) Persons on Medicaid in FL 2004	87.0%	8.6%	4.4%
Persons on Medicaid in PA 2004	79.9%	16.7%	3.4%
Children on Medicaid in MN 1999 ^a	79.5%	16.0%	4.5%
*Persons on Medicaid in MN 1999 ^b	54.0%	41.9%	4.1%
Adults on Medicaid and MNCare in Blue Cross in MN 2003 ^c	86.8%	12.8%	0.4%
Persons on Medicaid in MD 2004 ^d	87.5%	8.0%	4.5%
CPS Matching Studies			
Adults (age 15-64) on Medicaid in CA (pooled 1990-2000 data) ^e	72.3%	6.0%	21.7%
Persons on Medicaid/SCHIP in the US 2000	57.6%	24.6%	17.8%

17

Conclusions about the CPS estimates

- Overall CPS rate of those with Medicaid reporting that they are uninsured is much higher than other studies
- Overall CPS rate of those with Medicaid reporting Medicaid is much lower than other studies
- Why?
 - My opinion: Most of it has to do with the poor question design of the CPS
 - 17 month recall period leads people to forget spells of insurance coverage including Medicaid
 - Household level looping versus person level question design
 - CPS has highest ‘all year uninsured’ estimate of coverage in 2002 at 44 million with the Survey of Income and Program Participation coming in at 22 million

18

Explanations of the undercount revisited: work remaining to be done

- **Universe differences:**
 - Use 7 state Medicaid files with name and address information to understand the impact of MSIS non-validation (one of the states is CA)
 - Use 7 state Medicaid files to further analyze the CPS sample frame coverage (MN, LA, MD, CA, NJ, FL, PA)
- **Measurement error:**
 - Compare measurement error in the CPS to the National Health Interview Survey (NHIS) by linking the NHIS to the MSIS
- **Administrative and survey data processing, editing and imputation**
 - Evaluate how well the CPS edits and imputations work at both the micro level and the overall macro level
 - Evaluate additional state-level Medicaid data we receive for processing issues

Work to be done, continued

- **Survey sample coverage error and survey nonresponse bias**
 - Assess whether those addresses with a Medicaid enrollee are more likely to not participate in CPS survey
- **Model the impact of sample loss due to non-validated CPS and MSIS records on our estimates**
 - Being lead by RAND
- **Getting a handle on ‘false-positive’ Medicaid responses**
 - .6% (or crudely 1.3 M) of the CPS reweighted ‘linkable’ population answered having Medicaid only but did not link to the MSIS
 - .2% (or crudely .4 M) reported Medicaid in combination with something else without ‘linking’
 - This is a reason why the 5.1 million reduction is ‘crude’
 - ‘False positives’ for Medicaid are much harder to study

Conclusions

- These are preliminary results that are subject to change after further investigation
- Survey measurement error is playing the most significant role in producing the undercount
 - Some Medicaid enrollees answer that they have other types of coverage and some answer that they are uninsured
- ‘Linkable universe’ issues play a major role as well
- The overall goal of the project is to improve the CPS for supporting health policy analysis
 - At the moment the CPS has serious validity issues for its major policy uses in evaluation, simulations and allocations
 - Our goal is to work with census to improve the CPS



21

SHADAC contact information

www.shadac.org

State Health Access Data Assistance Center
University of Minnesota
2221 University Avenue, Suite 345
Minneapolis Minnesota 55414
(612) 624-4802
shadac@umn.edu



22