



## *Linking Administrative and Survey Data for Health Policy Research in the US: The case of the 'Medicaid Undercount'*

Exploiting Existing Data for Health Research  
Saint Andrews, Scotland September 18, 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



## US Health Insurance and the Medicaid Undercount

- The US does not have universal coverage
  - Most recent estimate from 2006 is that 47 million people in the US lacked insurance coverage (out of 297 million) --15.8 percent
- Medicaid and the State Children's Health Insurance Program (SCHIP) are two means tested publicly financed health insurance programs aimed at insuring low income Americans
  - Tend to have more liberal eligibility criteria for children
- Survey estimates of health insurance coverage and enrollment in Medicaid and SCHIP are important to health policy research in the US
- Survey estimates of Medicaid enrollment are well below administrative data enrollment figures



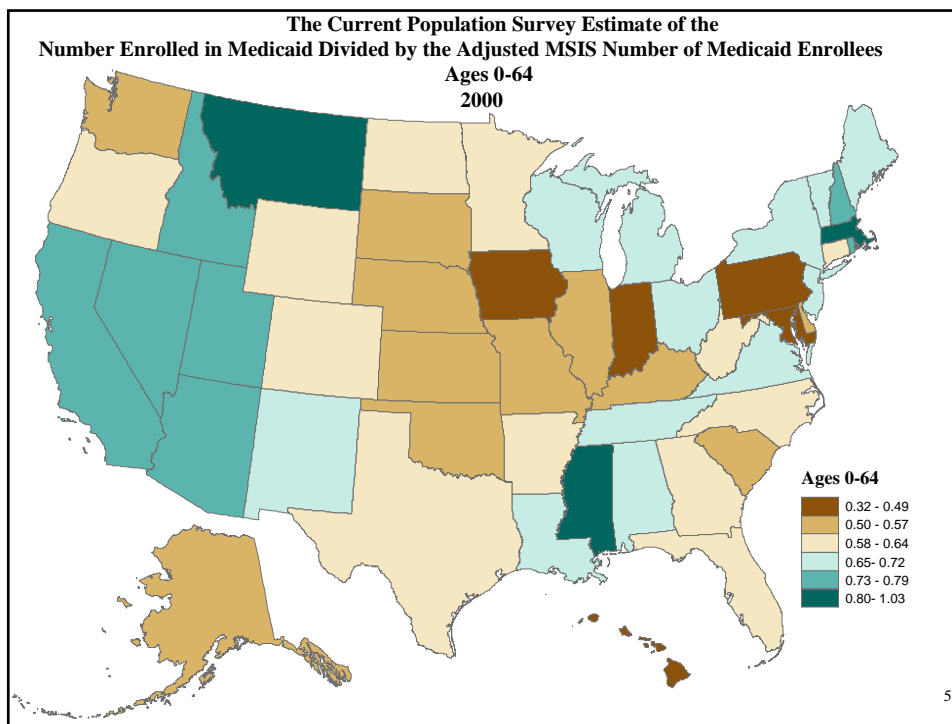
3

## Extent of the Medicaid undercount ?

- In the present study we compare the Medicaid Statistical Information System (MSIS) to the Current Population Survey (CPS).
- Raw Undercount numbers:
  - Calendar Year 2001 MSIS has 48.6 million records and CPS counts 27.8 million Medicaid enrollees (43% undercount)



4



## Why do we care?

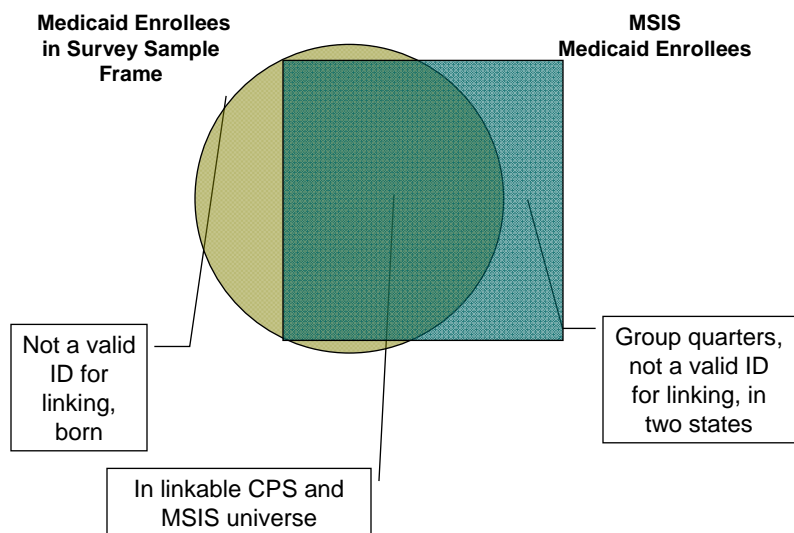
- CPS estimates of health insurance are important to US health policy research
  - Used for policy simulations by federal and state governments
  - Surveys like the CPS are the only sources for population estimates on the uninsured
  - Surveys are also the only source of the Medicaid/SCHIP eligible, 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

6

## What could explain the undercount?

- Explanations I have some data on today:
  - Universe differences between MSIS and CPS survey data
    - Paying special attention to sample loss (those cases with missing IDs that cannot be linked)
    - The data are linked through use of the Social Security Number (SSN).
      - This is replaced by a protected identification key or (PIK)
  - Survey Measurement error

## Building a common 'linked universe'



## The Basic Table

**Table 1: Counts from the MSIS, CPS and Linked Data Files: 2000, 2001**  
**[Numbers in Thousands]**

MSIS Data Counts	Calendar Year	
	2001	2000
A.1. All People In MSIS	48,556	45,039
A.2. Minus All SCHIP Only Enrollees	46,717	43,655
A.3. Minus Non-Full Medicaid Benefit Enrollees	42,184	39,736
A.4. Minus Those in Inst. Group Quarters	42,059	39,577
A.5. Minus Duplicate Enrollees	40,461	38,170
A.6. Minus Those Without PIKs (SSNs)	38,213	36,216
<b>CPS Counts</b>		
B.1. All People in the CPS	279,588	277,497
B.2. Recorded as Having Medicaid in CPS	27,749	26,170
<b>Linked Data File Counts</b>		
C.1. Raw Number of Linked Cases	24	22
C.2. Weighted Number of Linked Cases*	35,988	33,439

Source: 2000 and 2001 MSIS Calendar Year files

\* Weighted using the adjusted CPS person weight



9

## Fitting the MSIS and CPS records into the “linkable universe”

- Line A.6. shows how many cases from MSIS had Medicaid and were linkable to the survey data
  - 38.2 and 36.2 million in 2001 and 2000
  - Linking IDs were not missing at random
- In the CPS 24% of records are not linkable as they are missing linking IDs (i.e., SSN or PIKs).
  - Most of these refused to provide the linking data in the survey
  - We assumed IDs were missing at random and to fix this problem we simply re-weighted the remaining 76% of cases to equal the full 100% of cases using common post-stratification adjustment cells
- The re-weighted number of CPS survey cases in row C.2 was slightly below the row A.6 totals
  - 6% under in 2001 and 8% in 2000



10

## Universe Issues Explain Part of the Undercount but not all

- The original crude undercount was 43% in 2001
- After universe adjustments it is down to 31% (comparing line A.5 to B.2)

11

## The linked Survey and Medicaid data for Survey Response Error

- Focusing on only those with reported health insurance data
  - 58.9%** Respond Medicaid
  - 14.6%** Respond some other type of public
  - 9.5%** Respond some type of private coverage
  - 16.9%** Respond they were uninsured
  - 99.9%**
- 41% of the linked respondents report something other than Medicaid
  - This is major cause of the remaining undercount

12

## 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

## Work remaining to be done

- Linkable universe differences:
  - We know our initial group quarter adjustment is too small
    - We should be removing more MSIS cases from the count
    - Use 7 state Medicaid files to further analyze the CPS sample frame coverage
- Measurement error:
  - Compare measurement error in the CPS survey to other federal government surveys by linking the MSIS to them

## 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
- Getting a handle on ‘false-positive’ Medicaid survey responses

## 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



## SHADAC contact information

[www.shadac.org](http://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

