

Changes to the Imputation Routine for Health Insurance in the CPS ASEC

Michel Boudreaux, MS

State Health Access Data Assistance Center
University of Minnesota

Joint Statistical Meetings
Miami, Florida
August 3, 2011

Acknowledgements

- Co-Authors
 - Joanna Turner, SHADAC
 - Michael Davern, NORC
- Research Assistance
 - Shamis Mohamoud, NORC
 - Health and Disability Statistics Branch, U.S. Census Bureau
- Funding
 - Grant to SHADAC from the Census Bureau

Annual Social and Economic Supplement to the Current Population Survey (CPS ASEC)

- CPS is a monthly labor survey
- ASEC fielded in Feb-April
- Questions on work, income, migration and health insurance
- State representative (n~200,000)

Health Insurance in CPS ASEC

- Measures coverage in previous calendar year
- Detailed information for each person
- Widely used...
 - Surveillance
 - Projecting costs of proposed legislation
 - Evaluating impact of enacted policy
 - Historically used to allocate federal funds to states for public health insurance programs

Quality Improvement to Health Insurance

- Census Bureau dedicated to improving the quality of health insurance data
 - Conceptual definitions (1998)
 - Verification Question (2000)
 - Sample Expansion (2002)
 - Addition of premium costs and medical out-of-pocket information (2010)
 - Improvements to missing data imputation (2011)

Background of ASEC Imputation

- Approximately 10% of monthly CPS sample does not respond to ASEC
 - All data for these cases are imputed
 - ‘Full Supplement Imputations’ (FSI)
- Additionally, 2-3% of responders are missing data on health insurance items

Imputation Method

- Hot deck randomly draws values for missing cases (recipients) from similar, non-missing records (donors)
- Donors are organized into matrices consisting of variables that define “similar”
 - E.g. Age, marriage, work
- Assumes missing is random within cells
 - Maintains correlations within complete data

Background of Imputation Problems

- Davern et al., (2007) discovered errors in the hot deck specification...
 - Instrument allows any household member to be a private plan dependent
 - Interviews can press a single key to apply coverage to entire household
 - Allocation routine assigned dependent coverage only to nuclear family members of a policy holder
 - Did not consider other coverage the case may have had

Methods in Davern et al., (2007)

- Compared Non-Elderly coverage rates by FSI
 - Hierarchical coverage variable
 - Any public, only private, uninsured
 - Multinomial logit
 - Controlled for variables in the hot deck
 - Relative Rate Ratios (RRR)
 - Alternative estimates
 - Removed FSI and re-weighted
 - Model based prediction

Effect of Imputation Problem (2004 Data)

<i>Independent Variables</i>	<i>RRR</i>	<i>Standard Error</i>
Full supplement imputation	2.20	0.10***
Full supplement imputation × one person family	0.67	0.06***
Full supplement imputation × under 19 years	1.05	0.07

<i>Estimate Type</i>	<i>Private Insurance Rate (%)</i>	<i>Public Insurance Rate (%)</i>	<i>Uninsured Rate (%)</i>
<i>Total population</i>			
Unadjusted CPS	68.1	14.3	17.6
Reweighted CPS	69.1	14.3	16.6
Model based CPS (if none imputed)	69.0	14.3	16.7

Source: Table 3 from Davern et al. (2007)
HSR: Health services Research 42:5 (October 2007)

Effect of Imputation Problem (2004 Data)

<i>Independent Variables</i>	<i>RRR</i>	<i>Standard Error</i>
Full supplement imputation	2.20	0.10***
Full supplement imputation × one person family	0.67	0.06***
Full supplement imputation × under 19 years	1.05	0.07

<i>Estimate Type</i>	<i>Private Insurance Rate (%)</i>	<i>Public Insurance Rate (%)</i>	<i>Uninsured Rate (%)</i>
<i>Total population</i>			
Unadjusted CPS	68.1	14.3	17.6
Reweighted CPS	69.1	14.3	16.6
Model based CPS (if none imputed)	69.0	14.3	16.7

Source: Table 3 from Davern et al. (2007)
HSR: Health services Research 42:5 (October 2007)

Response by the Census Bureau

- Switch order
 - Public coverage imputed first, followed by private coverage
- Include public coverage in the private coverage matrix
- Remove nuclear family restriction
- Data from the new routine will be published in fall of 2011

Directly Purchased Coverage

- Census discovered and corrected a coding error that undercounted directly purchased coverage for children
- This data reflects that correction
 - All estimates reflect the imputation change and the coding fix

Study Goals

- Document the effect of the new routine to health insurance estimates from the full file
- Determine if the new routine attenuates problem in full supplement imputation cases identified in previous work

Data

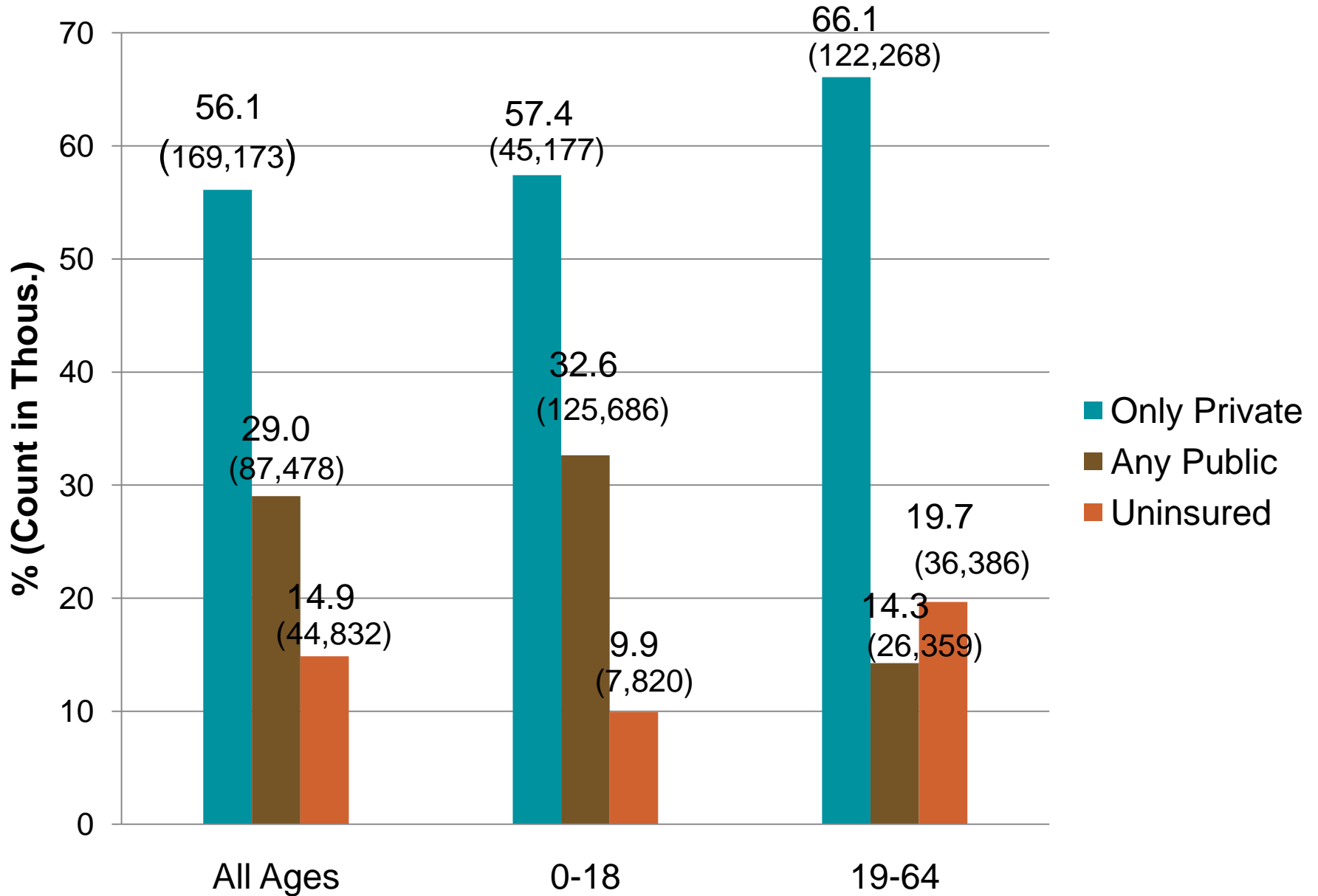
- 2009 CPS ASEC Research File
- 2009 SHADAC Enhanced CPS (ECPS)¹
 - FSI cases removed and data re-weighted
 - Developed by SHADAC

1. See Ziegenfuss, J. and Davern, M. “Twenty Years of Coverage: An Enhanced Current Population Survey: 1989–2008” Health Services Research 46:1, Part I (February 2011)

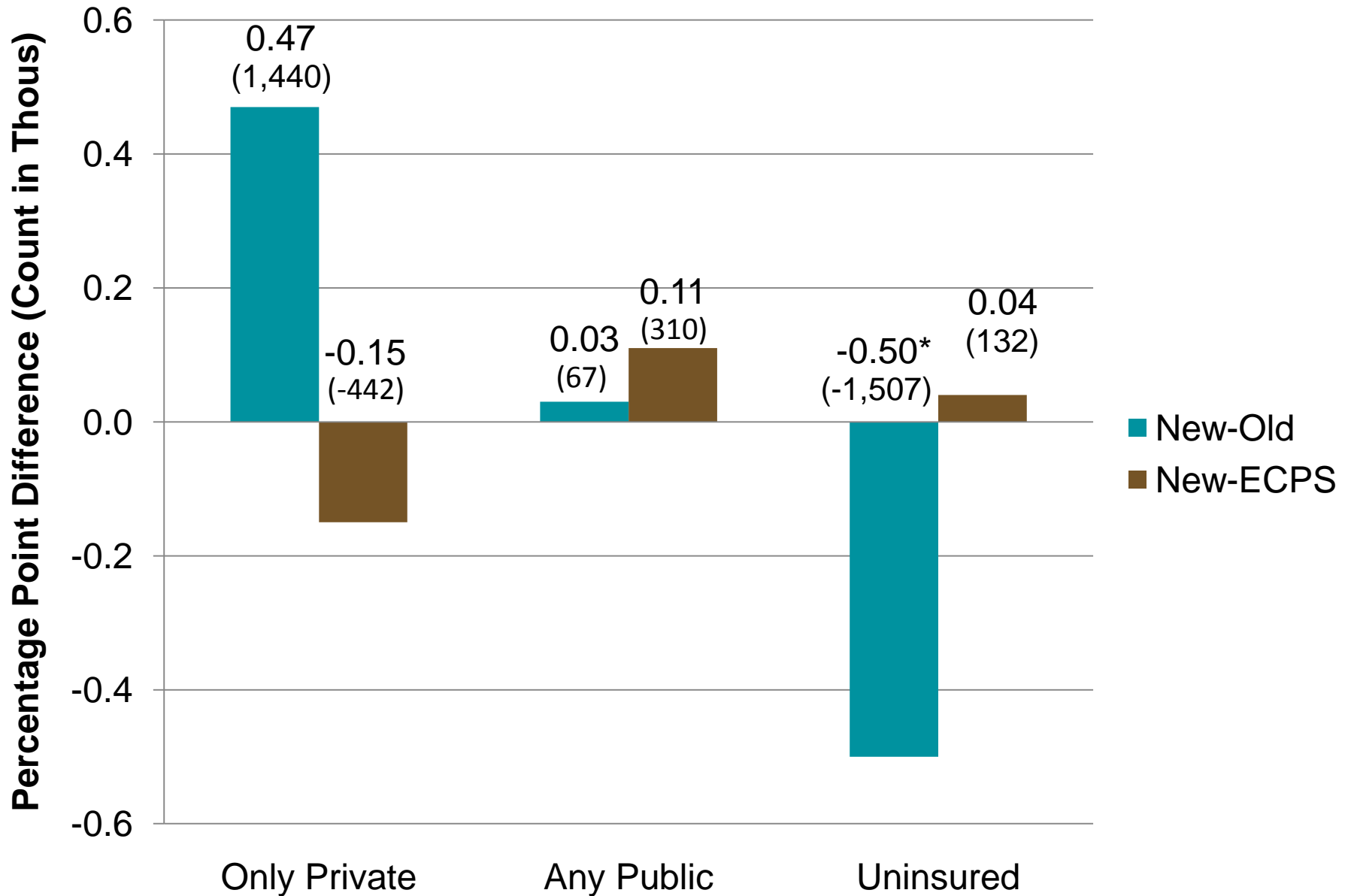
Methods

- Replicate Davern's study
- Bivariate comparisons
 - Hierarchical Coverage rates
 - Only private (private alone)
 - Any public (public alone or public and private)
 - Uninsured (no coverage in previous year)
 - Old Routine vs. New Routine vs. E-CPS
- Multinomial logistic regression to study impact of imputation change in FSI sample

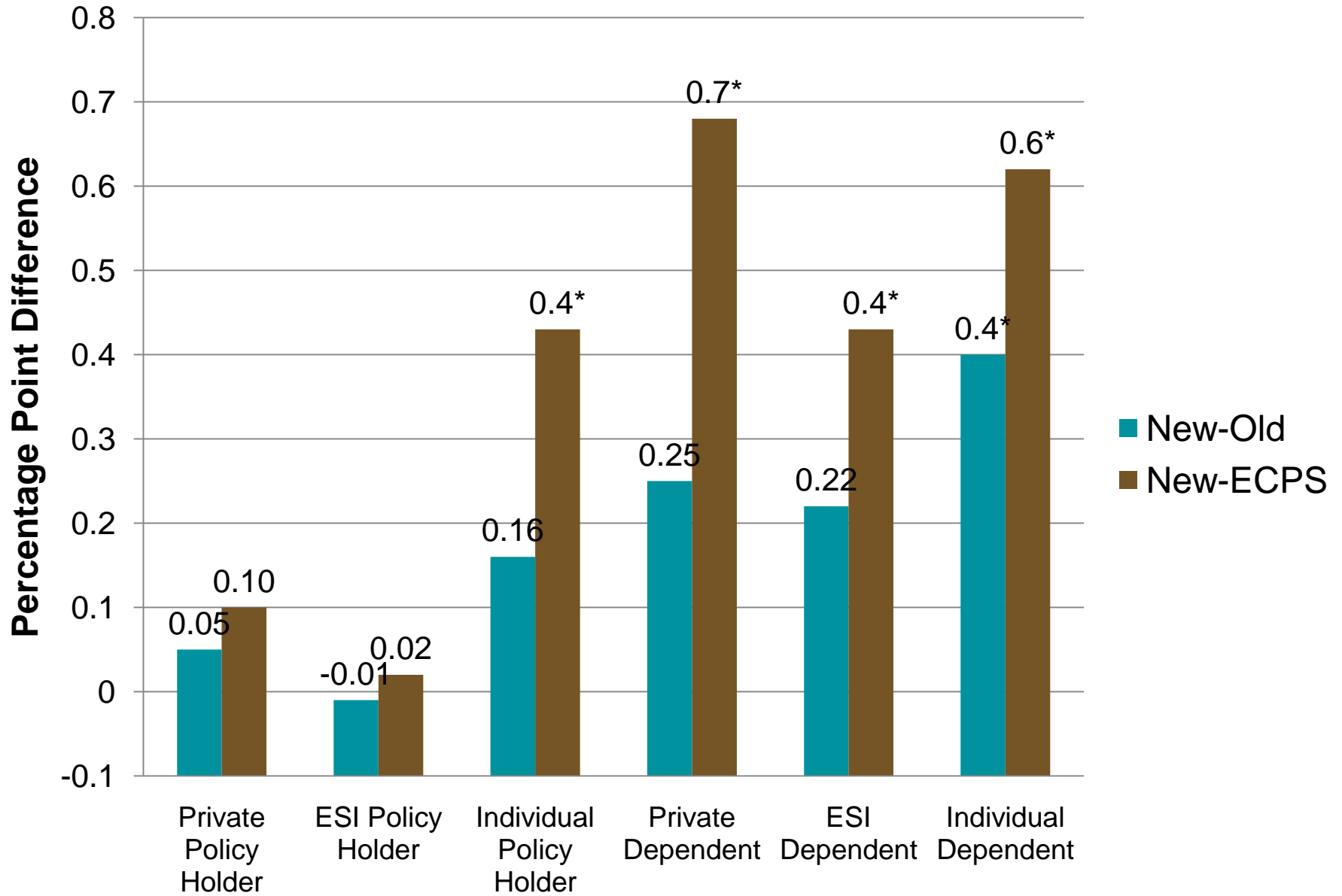
Health Insurance By Age, New Routine



Impact of Imputation Change, All Ages



Impact of New Routine by Private Plan Type, All ages



Model

$$P_{ijr} = \frac{\exp(FSI_i \beta_{jr} + x_i \lambda_{jr})}{\sum_{l=1}^3 \exp(FSI_i \beta_{lr} + x_i \lambda_{lr})}, j = 1, \dots, 3$$

- Person *i* with coverage *j* under routine *r*
- FSI: Full supplement status
- *x*: Covariates include hot deck variables and other important variables
- All Ages and no interactions
- Attenuation in new routine would indicate improvement

Selected Means by Supplement Status

	Not FSI		FSI	
	%	SE	%	SE
Only Private ¹	56.4	.20	52.7*	.63
Any Public ¹	29.0	.16	29.0	.54
Uninsured ¹	14.5	.14	18.3*	.46
<18 yrs	25.1	0.05	21.5*	0.36
< HS grad	14.7	0.11	16.2*	0.38
Unemployed	4.8	0.06	4*	0.19
White only	80.2	0.07	76.9*	0.60

1. From the new routine

* Significantly different at the $p < 0.001$ level

Selected Model Results

	Uninsured v. Private		Public v. Private	
	RRR	SE	RRR	SE
Old Routine				
FSI	1.83*	.077	1.40*	.057
New Routine				
FSI	1.24*	.059	1.23*	.051

The adjusted Wald test of FSI was significant in both equations.

* significant at $p < 0.001$ level.

Complete model controlled for gender, health, race/ethnicity, nativity, employment, poverty, family type, family size, education, veteran status, firm size, and self-employment.

Key Findings

- The new routine increases insurance coverage by 1.5 million people relative to the old routine
- Gain occurs mainly for dependent coverage
- In line with expectations Davern et al and ECPS

Key Findings

- Regression analyses showed that the undercount of private coverage in FSI cases attenuated
- While less substantial, FSI still significant
 - Missing other logical inputs (state & poverty)
 - Limited by sample size

Conclusions

- The imputation change appears to improve the quality of the ASEC health insurance data
- While a nuclear family restriction is conceptually appealing, the imputation routine is not the appropriate place to fix the problem

Thank You!

Michel Boudreaux
boudr019@umn.edu



**State Health Access Data Assistance Center
University of Minnesota, Minneapolis, MN
612-624-4802**

UNIVERSITY OF MINNESOTA

School of Public Health

Sign up to receive our newsletter and updates at www.shadac.org



www.facebook.com/shadac4states



Robert Wood Johnson Foundation



[@shadac](https://twitter.com/shadac)