State Health Access Data Assistance Center
Bridging the gap between research and policy

A Two-sample Approach for State Estimates of a Chronic Condition Outcome

Peter F. Graven

2010 National Conference on Health Statistics

August 17, 2010 Washington, DC

Acknowledgements

- Research funded by a grant from the Robert Wood Johnson Foundation (RWJF) to the State Health Access Data Assistance Center (SHADAC) at the University of Minnesota, School of Public Health
- Travel funded by the National Center for Health Statistics

Objective

- To produce state estimates of health information in the NHIS (chronic condition, pre-existing condition, etc.)
  – NHIS does not include state in its public use file
  – restricted access file provides some opportunity but sample design not intended for state-level estimates
- To calculate appropriate errors for the estimates for comparison between states or as inputs in other analyses
- Predict/impute an outcome measure (condition status) in a survey with state-level sample design (CPS, ACS) using NHIS data

Method

- Applying elements of method used in (Schenker, N., Raghunathan, T., Bondarenko, I., 2010) *
  - imputed clinical values of hypertension, diabetes and obesity in NHIS with self-reported values and both clinical and self-reported values from NHANES
  - self-reported rates were lower than clinical values
  - requires multiple imputation techniques and propensity scores

*Schenker, N., Raghunathan, T., Bondarenko, I., "Improving on analyses of self-reported data in a large-scale health survey by using information from an examination-based survey". Statistics in Medicine, Volume 29, Issue 5, pages 533–545, February 2010

Method-Data

- National Health Interview Survey (NHIS) 1997-2001, 2004-2008
  - Minnesota Population Center and State Health Access Data Assistance Center, *Integrated Health Interview Series: Version 2.0.* Minneapolis: University of Minnesota. http://www.ihis.us
  - Harmonizes the data and documentation for the NHIS
  - 1,000's of vars, 38 years, linkable to NHIS data supplements
- Current Population Survey (CPS) 1999, 2006
  - Miriam King, Steven Ruggles, J. Trent Alexander, Sarah Flood, Katie Genadek, Matthew B. Schroeder, Brandon Trampe, and Rebecca Vick. *Integrated Public Use Microdata Series, Current Population Survey: Version 3.0*. [Machine-readable database]. Minneapolis: University of Minnesota, 2010.

Method-Primary Steps

1) Assemble Data
2) Identify outcome status in NHIS
3) Create identically coded covariates in NHIS and CPS
4) Predict survey of observation using covariates, create subgroups for model
   4b) Predict key variable using covariates
5) Impute missing CPS values using predicted survey, covariates (or predicted key variable) and interactions
6) Produce estimates of outcome using imputed data

Method- Data Assembly

CPS_obs	CPS	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
133,710	1999	103,477	98,785	97,059	100,618	100,759							
	2000		1	1	1	1	1						
	2001			1	1	1	1	1					
	2002				1	1	1	1	1				
	2003					1	1	1	1	1			
	2004						1	1	1	1	1		
	2005							1	1	1	1	1	
206,639	2006								94,460	98,649	75,716	75,764	74,236
	2007									1	1	1	1
	2008										1	1	1
	2009											1	1

*Reference period of survey not the year it was conducted

Method-Data Assembly

					Condition indicator			
Data	State	Chronic	Covariates	Pr(survey)	imp1	imp2	...	imp10
NHIS 2004		x	x	x	x	x		x
		x	x	x	x	x		x
		x	x	x	x	x		x
NHIS 2005		x	x	x	x	x		x
		x	x	x	x	x		x
		x	x	x	x	x		x
NHIS 2006		x	x	x	x	x		x
		x	x	x	x	x		x
		x	x	x	x	x		x
NHIS 2007		x	x	x	x	x		x
		x	x	x	x	x		x
		x	x	x	x	x		x
NHIS 2008		x	x	x	x	x		x
		x	x	x	x	x		x
		x	x	x	x	x		x
CPS 2006	x		x	x	x	x		x
	x		x	x	x	x		x
	x		x	x	x	x		x

Same as original

Impute these values

Method-Identify Outcome status

- Chronic condition
  - Limitation of activity due to chronic condition
  - asked of all persons
  - ~12% of population nationally

Method- Create Identical Covariates

	CPS	NHIS*		CPS	NHIS*		CPS	NHIS*
Age			**Education**			**Health Status**		
0-17	25.0	25.0	No HS Diploma	11.4	12.1	Excellent	33.1	35.5
18-34	23.0	23.1	HS Diploma	23.6	22.3	Very good	32.0	30.8
35-54	29.0	29.1	Some college/associates	20.4	21.3	Good	23.6	24.2
55-64	10.8	10.6	Bachelors or more	19.6	19.3	Fair	7.9	7.2
65-74	6.4	6.4	NIU	25.0	25.0	Poor	3.4	2.4
75+	5.7	5.7	**Wages**			**Insurance Status**		
Sex			0-10K	8.0	9.0	Uninsured	15.8	14.8
Male	49.1	48.9	10k-25K	12.6	13.3	Insured	84.2	85.2
Female	50.9	51.1	25k-50K	16.1	17.5	**Region**		
Marital Status			50K+	13.2	12.3	Northeast	18.2	18.1
Married	42.0	43.3	0/NIU	50.1	47.8	Midwest	22.1	23.4
Not married	33.0	31.7	**Poverty**			South	36.4	36.1
NIU	25.0	25.0	0-99	13.7	13.4	West	23.3	22.3
Race/Ethnicity			100-199	18.1	19.3	**Birthplace**		
White-NH	66.1	67.5	200-299	17.1	16.9	US born	86.1	86.7
Black-NH	12.1	12.5	300-399	13.4	13.8	Born outside US	13.9	13.3
Hispanic	15.1	14.9	400-499	10.5	10.2			
Other-NH	6.6	5.1	500+	27.3	26.5			

*NHIS missing values imputed using sequential hotdeck

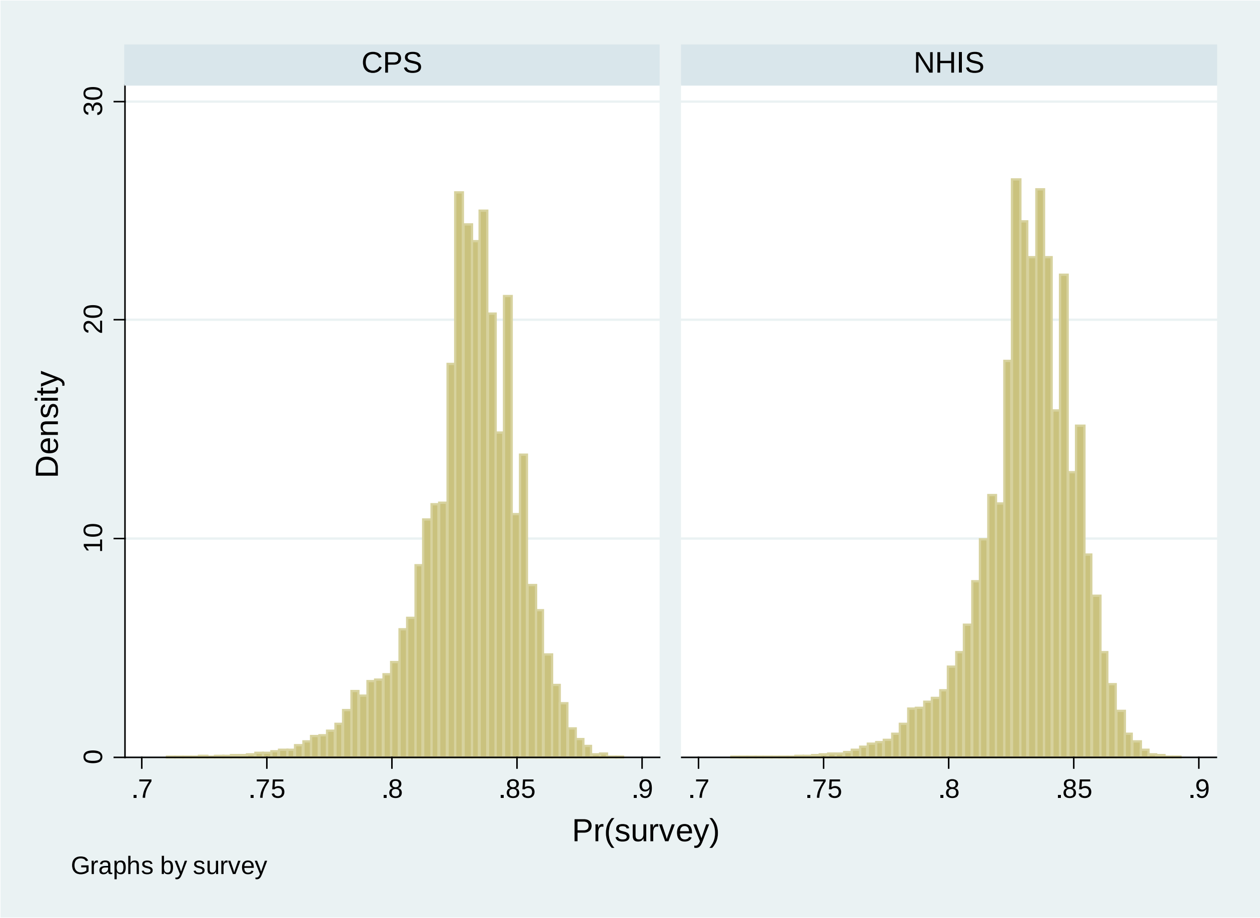
Method-Predict Survey Propensity

- Survey propensity: this predicts which survey an observation is from based on its covariates.
  – Ideally, you would have very similar distributions implying observations are similarly likely in either dataset.
  – This strengthens the case for using NHIS observations to impute CPS observations

Method-Predict Survey Propensity

- Why do you predict the survey?
  - although the values of covariates are coded the same, the responses in two surveys may not truly be identical.
  - Therefore, by predicting the survey there is a single dimension to assess how likely the observations are to be similar.
- Why predict propensities? Isn't that used for matching studies?
  - for the imputation we are looking for similar observations in different surveys to predict a likely values
- Survey propensity model: age, sex, race, education, marital status, birthplace insurance status, wages, poverty, region

Method-Predict Survey Propensity



1= NHIS, 0=CPS

mean>50% because there are more NHIS obs than CPS obs in the imputation sample

Similar and narrow shape indicates coding is similar between surveys.

Results- Imputation Models

- Full Imputation Model
  - All covariates interacted by propensity group
  - Possible due to large sample size
- Parsimonious Imputation Model
  - predicted health status interacted by propensity group
  - very similar results due to strength of common health status variable
- Two-Step Model
  - Fit all covariates on NHIS, predict on CPS
  - Standard errors of state means too small

Results-Chronic Prevalence by Survey and Covariates

	CPS	NHIS
Overall	13.3	12.0
Age		
0-17	7.2	7.1
18-34	5.1	4.6
35-54	11.5	10.2
55-64	22.2	20.0
65-74	28.7	25.5
75+	48.2	42.3
Sex		
Male	12.9	11.5
Female	13.7	12.4
Marital Status		
Married	12.5	11.0
Not married	19.1	17.2
NIU	7.2	7.1
Race/Ethnicity		
White-NH	14.7	13.2
Black-NH	14.5	13.0
Hispanic	8.1	7.3
Other-NH	8.9	7.3

	CPS	NHIS
Overall	13.3	12.0
Education		
No HS Diploma	25.1	23.7
HS Diploma	18.2	15.3
Some college/associates	13.4	11.8
Bachelors or more	8.5	7.2
NIU	7.2	7.1
Wages		
0-10K	9.9	9.9
10k-25K	6.3	6.2
25k-50K	4.6	4.4
50K+	3.2	3.2
0/NIU	21.1	19.0
Poverty		
0-99	21.5	19.6
100-199	19.4	17.0
200-299	14.3	13.1
300-399	11.7	10.2
400-499	8.7	8.4
500+	7.2	6.1

	CPS	NHIS
Overall	13.3	12.0
Health Status		
Excellent	3.1	2.9
Very good	5.5	5.7
Good	16.0	15.6
Fair	49.6	48.2
Poor	82.6	81.7
Insurance Status		
Uninsured	11.7	7.8
Insured	13.6	12.7
Region		
Northeast	13.1	11.7
Midwest	13.7	12.6
South	13.7	12.4
West	12.5	11.0
Birthplace		
US born	14.1	12.7
Born outside US	8.4	7.4

Results- Chronic Prevalence by State

State	Mean	SE	State	Mean	SE	State	Mean	SE
Alabama	17.4	1.14	Kentucky	18.6	0.98	North Dakota	13.2	1.18
Alaska	12.4	0.82	Louisiana	15.3	1.26	Ohio	14.8	0.74
Arizona	12.6	0.95	Maine	15.4	1.01	Oklahoma	15.1	1.08
Arkansas	16.3	1.01	Maryland	11.4	0.63	Oregon	15.4	1.01
California	11.9	0.39	Massachusetts	12.5	0.89	Pennsylvania	15.9	0.82
Colorado	11.2	0.83	Michigan	14.2	0.65	Rhode Island	13.0	1.01
Connecticut	11.4	0.72	Minnesota	11.4	0.63	South Carolina	16.0	1.15
Delaware	12.6	0.88	Mississippi	17.5	1.40	South Dakota	14.0	1.05
District of Columbia	11.4	0.85	Missouri	15.5	0.91	Tennessee	16.2	1.14
Florida	12.6	0.60	Montana	15.5	1.08	Texas	12.0	0.47
Georgia	11.6	0.72	Nebraska	11.6	0.91	Utah	11.1	0.85
Hawaii	11.3	0.72	Nevada	11.9	0.90	Vermont	13.1	1.10
Idaho	12.9	1.10	New Hampshire	11.4	0.73	Virginia	11.6	0.71
Illinois	12.8	0.70	New Jersey	10.8	0.77	Washington	13.0	0.86
Indiana	14.2	1.06	New Mexico	14.0	1.20	West Virginia	21.3	1.28
Iowa	13.0	0.76	New York	13.1	0.48	Wisconsin	12.8	0.77
Kansas	13.1	0.89	North Carolina	14.2	0.87	Wyoming	14.2	1.11

Results-Selected State Means of Chronic by Model

	State	Full		Parsimonious		Two-Step	
		Mean	SE	Mean	SE	Mean	SE
Highest 10 States	West Virginia	21.3	1.28	20.7	1.34	21.4	0.80
	Kentucky	18.6	0.98	17.5	1.08	18.3	0.68
	Mississippi	17.5	1.40	18.0	1.31	17.4	0.76
	Alabama	17.4	1.14	17.3	1.14	17.0	0.76
	Arkansas	16.3	1.01	16.3	1.09	16.1	0.65
	Tennessee	16.2	1.14	16.2	1.44	16.2	0.59
	South Carolina	16.0	1.15	16.4	1.07	16.1	0.62
	Pennsylvania	15.9	0.82	14.4	0.74	15.6	0.40
	Missouri	15.5	0.91	13.8	0.91	15.2	0.52
	Montana	15.5	1.08	12.6	1.02	15.0	0.65
	....						
Lowest 10 States	Nebraska	11.6	0.91	10.7	0.85	11.8	0.45
	District of Columbia	11.4	0.85	12.9	1.14	11.3	0.50
	New Hampshire	11.4	0.73	10.0	0.76	11.2	0.36
	Minnesota	11.4	0.63	10.2	0.66	11.4	0.37
	Connecticut	11.4	0.72	11.0	0.77	11.5	0.36
	Maryland	11.4	0.63	12.2	0.88	11.3	0.36
	Hawaii	11.3	0.72	12.4	0.74	11.1	0.37
	Colorado	11.2	0.83	10.5	0.65	11.3	0.38
	Utah	11.1	0.85	10.2	0.96	11.0	0.50
	New Jersey	10.8	0.77	11.5	0.64	10.9	0.38

Results-Region vs. State

- Using state instead of region results in 21 states with significantly different rates

State	Region	State Mean	State SE	Region Mean	Region SE	Difference
West Virginia	South	21.3	1.28	13.7	0.17	7.5 *
Kentucky	South	18.6	0.98	13.7	0.17	4.9 *
Mississippi	South	17.5	1.40	13.7	0.17	3.8 *
Alabama	South	17.4	1.14	13.7	0.17	3.7 *
Montana	West	15.5	1.08	12.5	0.19	2.9 *
Pennsylvania	Northeast	15.9	0.82	13.1	0.24	2.8 *
...						
Nebraska	Midwest	11.6	0.91	13.7	0.21	-2.1 *
New Jersey	Northeast	10.8	0.77	13.1	0.24	-2.3 *
District of Columbia	South	11.4	0.85	13.7	0.17	-2.3 *
Minnesota	Midwest	11.4	0.63	13.7	0.21	-2.3 *
Maryland	South	11.4	0.63	13.7	0.17	-2.4 *

Limitations/Future Research

- Limitations
  - Model selection not optimized, primarily an exercise
  - Testing effects would require common variables that should also be included in imputation model
- Future Research
  - Investigate why national means are different
  - Identify more commonly coded variables between surveys
  - Create outcomes that align with pre-existing condition definitions
  - Consider applications for other surveys (MEPS, BRFSS)

Summary

- Newer accessible methods allows for creative integration of data sources with appropriate uncertainty incorporated
- The ability to make valid state estimates is valuable for health policy
  - could be used to develop state estimates of those with pre-existing conditions eligible for the temporary high risk pool

Contact Info

Peter Graven

SHADAC

2221 University Ave SE, Suite 345

Minneapolis, MN 55414

peter@graven.com

612-624-2083

Extra: Results-Potential Upper bound of Unexplained Error at State level

- From 1997-2001 NHIS released MSA names
- Data assembled for 1999 CPS with matching MSA and non-MSA (region) codes.
  - only MSAs with over 3,000 observations per year of NHIS were used. Others grouped into their region
- The MSA portion of the MSA/Non-MSA(Region) code have smaller populations than states
  - provides an approximate upper bound on the error in the state imputation

Extra: Results-MSA/Region Comparison

	CPS		NHIS	
Name	Mean	SE	Mean	SE
Northeast-Non MSA	14.3	0.53	13.4	0.30
Midwest-Non MSA	14.3	0.39	12.8	0.33
South-Non MSA	14.5	0.33	13.0	0.20
West-Non MSA	12.7	0.34	11.4	0.26
Los Angeles-Long Beach, CA	12.0	0.61	9.8	0.36
New York, NY	13.9	0.76	9.6	0.32
Chicago-Gary-Kenosha, IL-IN-WI	11.6	0.76	9.9	0.50
Houston-Galveston-Brazoria, TX	10.6	1.20	7.8	0.47
Detroit-Ann Arbor-Flint, MI	13.7	1.08	12.3	0.39
Boston-Worcester-Lawrence, MA	12.0	1.12	11.5	0.31
Washington, DC-MD-VA-WV	8.7	0.89	8.2	0.49
Philadelphia, PA-NJ	14.5	1.02	10.9	0.54
Miami, FL	10.7	1.12	6.9	0.58
Phoenix-Mesa, AZ	11.0	1.21	9.6	1.38

Extra: Results-Survey Correlation for MSA/Region

- Scatterplot of CPS vs NHIS MSA/Region observations