Web Exhibit 11: Small Area Methodology

Estimating Uninsured Rates for California's Legislative Districts Using Small Area Methodology

Abstract

The objective of this study was to develop estimates of uninsured rates for child and nonelderly adult residents in California's 80 Assembly and 40 Senate districts. Small area or "synthetic" estimates are developed when direct estimates are not possible due to the absence of data. Thus, we utilized a "small area" methodology to estimate uninsured rates since no information exists on uninsurance at the legislative district level. We developed the estimates using logistic regression models and data from the 2001 California Health Interview Survey (CHIS 2001), 2000-2002 Current Population Surveys (CPS), 2000 Census, and other data sources. There were three major components of the methodology: 1) modeling poverty at the census block level; 2) modeling uninsurance at the census block level; and 3) calculating insured rates by legislative district level using the California Health Interview Survey (CHIS 2001) as a calibration. The final estimates for California's legislative districts were calculated by aggregating the census blocks within the boundaries of each district using a file supplied by the State Senate Office of Demographics.

Key words: Logistic regressions, synthetic estimates, calibration, CHIS 2001, insurance coverage, legislative districts

Introduction

The assessment of insurance coverage among California's 80 Assembly and 40 Senate districts is valuable for policy makers and advocates to address health care access issues at the local level. The average population size is approximately 400,000 in each Assembly district and 800,000 in each Senate district. However, most national surveys currently available, such as the CPS, typically have data at limited state or sub-state

levels, and data at smaller geographic units are virtually non-existent. Therefore, a small area methodology must be used to assess rates of health insurance coverage for California's legislative districts. The small area methodology utilized multiple data sources and involved three major components: 1) modeling poverty at the census block level; 2) modeling uninsurance at the census block level; and 3) calculating insured rates by legislative district level using the California Health Interview Survey as a calibration.

Data Sources

No single data source provides direct measures of health insurance coverage at the legislative district level in California. The March CPS contains rich information about insurance coverage, but it is not intended to provide sub-state level estimates. Although the CPS provides some sub-state geographic identifiers, there is no identification at legislative district level. Therefore in order to estimate the uninsured rates in Assembly and Senate districts, information was combined from various sources (Table 1.)

Table 1. Data Sources Used in Estimating Uninsured Rates

Variables	Data Source					
	2000 Census SF1	2000 Census SF3	2000-2002 CPS	2001 CHIS	2002 Department of Finance	State Senate Office of Demographics
MSA	X	X	X	X		
County	X	X	X	X	X	
Census Tract	X	X				X
Census Block	X					X
Legislative District						X
Age	X	X	X	X		
Sex	X	X	X	X		
Race/Ethnicity	X	X	X	X		
Income-to-Poverty Ratio		X	X	X		
Insurance Status				X		

Methods

Modeling Poverty

The first major component of the methodology was modeling poverty at the census block level. The Census 2000 SF1 data provide population counts for each census block in California by age, sex, and race/ethnicity, and these data are the base layer in our estimation algorithm. Since insurance coverage status is highly related with family income, we add poverty into the base layer by merging in the percentage of below 100% poverty from the 2000 Census SF3 data at the census tract level. However, only one indicator of below 100% poverty level is not sufficient in our small area's estimation procedure. Therefore, a set of poverty levels was derived from predictive models using the March Current Population Survey, stratified by seven regions in California (see Appendix A).

Using a 3-year combined CPS (2000-2002), concatenated to increase sample size, we modeled the distribution of poverty levels and then applied the estimated probabilities to the 2000 Census data. We preserved the percentage of below 100% FPL from the 2000 Census SF3 file by census tract, and adjusted the estimated poverty distributions for 100% FPL and above. The following are detailed steps our the statistical procedures:

- In each of the seven regions, five logistic regressions were applied to a 3-year combined CPS on five poverty levels: <100%, <133%, <200%, <251%, and <300%.

$$\text{logit}(p_j < povll_j) = age + sex + race/ethnicity \qquad (1)$$

j: 1…5. index of poverty level

The predicted probabilities were calculated and noted as $\hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}_4, \hat{p}_5$.

The distributions of poverty by age, gender, race/ethnicity at region level were then calculated as follows:

$p(1) = \hat{p}_1,$ Poverty < 100%

$p(2) = \hat{p}_2 - \hat{p}_1,$ Poverty 100%-133%

$p(3) = \hat{p}_3 - \hat{p}_2$, Poverty 134%-199%

$p(4) = \hat{p}_4 - \hat{p}_3$, Poverty 200%-250%

$p(5) = \hat{p}_5 - \hat{p}_4$, Poverty 251%-299%

$p(6) = 1 - \hat{p}_5$, Poverty 300% +

- To preserve the variation in percentages of below 100% FPL at the census tract level from Census 2000 SF3 data, the predicted percentages of above 100% FPL modeled from the CPS were adjusted by age, gender, race/ethnicity at the region level.
- The percentage of below poverty from the Census 2000 were merged into the base layer by age, gender, and race/ethnicity at the census tract level, and the predicted poverty distributions (above 100% FPL) estimated from the CPS data were also merged into the base layer by age, gender and race/ethnicity at the region level.
- The estimated numbers of the population in each level of poverty was calculated by multiplying *p*(1), … *p*(6) to total population in the 2000 Census data.

Modeling Uninsurance

The second major component of the methodology was modeling uninsurance at the census block level. The analytic file for estimating rates of uninsurance was the 2000 Census data with poverty distributions derived from modeling poverty. The predicted probabilities of being uninsured were derived as follows:

- Stratified logistic regressions were applied with a 3-year pooled CPS data. The dependent variable was the insurance status with 1 being uninsured and 0 otherwise. The independent variables included age, sex, race/ethnicity and poverty levels.
- The predicted probabilities of being uninsured from the logistic regressions were then merged into the 2000 Census data by age, gender, race/ethnicity, and poverty level at the region level.

Calculating Uninsured Rates in Assembly and Senate Districts

The third major component of the methodology was the calculation of uninsured rates by Assembly and Senate districts, and it involved three steps: calibration of uninsured rates, aggregation of population data by district, and calculation of confidence intervals. Predicted uninsured rates were calibrated to the 2001 California Health Interview Survey (CHIS 2001) uninsurance rates by age at the county or county group level (see Appendix A). We used CHIS 2001 for calibration instead of the CPS since the CHIS data on health insurance coverage are more reliable due to larger sample sizes at the county level compared to the CPS. A second calibration was made using Department of Finance data at the county level to update population numbers to the year 2002. The population without insurance coverage was calculated by multiplying predicted probabilities of being uninsured to total population at the block level. The total numbers of uninsured were aggregated for each census block using a file supplied by the California State Senate Office of Demographics that cross-referenced district to block. The uninsured rates for legislative districts were then calculated by multiplying predicted probabilities of being uninsured to the total population at the census block level.

Calculating confidence intervals for district uninsured rates was the last step of the analyses. The variances of the estimates were calculated in similar ways to that in Hosmer and Lemeshow (1995), scenario two (S2). There were two major procedures to estimate the variance of our estimates. The first procedure is to estimate the variance of estimated uninsured rate in the Census data. The second procedure is to estimate the variance of estimated uninsured rate after calibration, which has to take account of variance of uninsured rate in the CPS and in CHIS.

References

Hosmer, D., Lemeshow, S., (1995) "Confidence Interval Estimates of an Index of Quality Performance Based on Logistic Regression Models", Statistics in Medicine, 14, 2161-2172.

Appendix A. Regions of California Used in Stratified Analyses

Northern and Sierra Counties
Butte
Shasta
Humboldt/Del Norte
Siskiyou/Lassen/Trinity/Modoc
Mendocino/Lake
Tehama/Glenn/Colusa
Sutter/Yuba
Nevada/Plumas/Sierra
Tuolumne/Calaveras/Amador/Inyo/Mariposa/Mono/Alpine
Greater Bay Area
Santa Clara
Alameda
Contra Costa
San Francisco
San Mateo
Sonoma
Solano
Marin
Napa
Sacramento Area
Sacramento
Placer
Yolo
El Dorado
San Joaquin Valley
Fresno
Kern
San Joaquin
Stanislaus
Tulare
Merced
Kings
Madera
Central Coast
Ventura
Santa Barbara
Santa Cruz
San Luis Obispo
Monterey/San Benito
Los Angeles
Other Southern California
Orange
San Diego
San Bernardino
Riverside
Imperial