Multilevel Regression and Post-stratification Approach for Small-Area Estimation of Population Health Outcomes at Local Geographic Areas using State BRFSS via Parametric Bootstrapping

Presented to
Northeast Epidemiology Conference
October 19, 2017

By
Xi Zheng, MPH, MS
BRFSS Epidemiologist
Connecticut Department of Public Health
Keeping Connecticut Healthy

Challenge of Small Area Estimation (SAE) in Connecticut

- Total population 3,574,097, 12.01% live in rural area (62.28% of total area).

Census 2010

- Eight counties, but no county health departments in CT.

- The next smallest common unit of analysis is the town.

Source: U.S. Census Bureau
Census 2010 Summary File 1
population by census tract
Challenge of Small Area Estimation (SAE) in Connecticut

**BRFSS**
- Designed for state and national level
- Direct estimates

**Small Area Estimation (SAE)**
- Statistical techniques involving estimation of reliable estimates for small areas using health survey.

- **Six** SAE papers published from 2016-2017, at county level.
- CDC/RWJF 500 Cities project provided SAEs at city- and census tract-level.
What Can the CT BRFSS Do for Local Health?

- Health District Oversampling, with existing state weighting;
- Reweight BRFSS responses at the local level;
- Post-BRFSS focus surveys in sub-state geographic areas; and
- Synthetic estimates (500 Cities project www.cdc.gov/500cities)
Four Steps of Multilevel Regression and Post-stratification (MRP) Framework

1. Construct and fit multilevel prevalence models using BRFSS data.
2. Apply multilevel prediction models to the census population.
3. Generate model-based SAEs via post-stratification.
4. Validate model-based SAEs.

Developed by Dr. Zhang, using all 50 states plus the District of Columbia (DC) 2014 BRFSS data.

Limitations of Current MRP

1) Use of single year BRFSS might produce temporally inconsistent SAEs
2) Use of national BRFSS data including 50 states and DC
3) Predicted standard errors (SEs) and confidence intervals (CIs) of SAEs could be substantially underestimated or overestimated.

Extended MRP Approach

1) Use five year state BRFSS (2011-2015) data to produce SAEs for counties and towns in CT
2) Apply parametric bootstrapping approach to estimate the predicted SEs and CIs of SAEs
Methodology

Data Sources
- CT BRFSS (2011-2015)
- US Census 2010
- American Community Survey (ACS)

- Population health indicators: current asthma, depression, diabetes, obesity, high blood pressure, uninsured (18-64 yrs old);
- Population data sex*age*race;
- Poverty level data

Multilevel Logistic Model

\[
\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = X\beta + r_{\text{county}} + r_{\text{town}}
\]

- Individual Demographics (age, sex, race/ethnicity)
- County and Town Context

SAEs

SAS GLIMMIX

w/wo survey weights
SAEs Validation

**Internal Validity**

- Compare SAEs with direct survey from single year & aggregated 5-years CT BRFSS, at town, county and state levels.

**External Validity**

- Compare SAEs with direct estimates from other survey: ACS (uninsured adults 18-64 years old).
- Compare SAEs with estimates from re-weighted 5-year CT BRFSS.
At State Level
PCC: 0.991 (unweighted), and 0.990 (weighted); MAD: 0.775 (unweighted), and 0.892 (weighted).

At County Level
Unweighted SAEs (89.6%) and weighted SAEs (87.5%) were within direct survey estimates 95% confidence interval.

At Town Level

<table>
<thead>
<tr>
<th>Indicators</th>
<th>No. of Units</th>
<th>PCC</th>
<th>MAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asthma</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAE&lt;sup&gt;a&lt;/sup&gt;</td>
<td>68</td>
<td>0.769</td>
<td>1.42</td>
</tr>
<tr>
<td>SAE&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td>0.831</td>
<td>0.58</td>
</tr>
<tr>
<td>Depression</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAE&lt;sup&gt;a&lt;/sup&gt;</td>
<td>107</td>
<td>0.562</td>
<td>1.49</td>
</tr>
<tr>
<td>SAE&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td>0.826</td>
<td>0.66</td>
</tr>
<tr>
<td>Diabetes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAE&lt;sup&gt;a&lt;/sup&gt;</td>
<td>77</td>
<td>0.769</td>
<td>1.29</td>
</tr>
<tr>
<td>SAE&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td>0.548</td>
<td>0.26</td>
</tr>
<tr>
<td>HBP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAE&lt;sup&gt;a&lt;/sup&gt;</td>
<td>132</td>
<td>0.44</td>
<td>1.66</td>
</tr>
<tr>
<td>SAE&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td>0.539</td>
<td>1.1</td>
</tr>
<tr>
<td>Obesity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAE&lt;sup&gt;a&lt;/sup&gt;</td>
<td>130</td>
<td>0.89</td>
<td>1.13</td>
</tr>
<tr>
<td>SAE&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td>0.766</td>
<td>0.89</td>
</tr>
<tr>
<td>Uninsured</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAE&lt;sup&gt;a&lt;/sup&gt;</td>
<td>32</td>
<td>0.905</td>
<td>3.16</td>
</tr>
<tr>
<td>SAE&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td>0.856</td>
<td>3.73</td>
</tr>
</tbody>
</table>

<sup>a</sup> SAEs without using BRFSS final survey weights, <sup>b</sup> SAEs based on using BRFSS final survey weights.

Abbreviation: Pearson correlation coefficient (PCC); Mean absolute difference (MAD)
Compare Model-based SAEs and Direct Survey Estimates Using Single Year CT BRFSS, at the State Level

Asthma

Prevalence (%)

Depression

Prevalence (%)

Diabetes

Prevalence (%)

HBP

Prevalence (%)

Obesity

Prevalence (%)

Uninsured

Prevalence (%)
Compare Model-based SAEs and Reweighted Direct Survey Estimates

- Re-weighted 53 local area estimates using combined 5 year CT BRFSS data (2011-2015), each area requires at least 500 interviews from the responses (Stone et al., 2017).

- Heat mapping were used to show the geographic clustering, by comparing SAEs and re-weighted estimates.
Current asthma prevalence: overlap of risk areas using SAE and reweighted estimates
Uninsured prevalence: overlap of risk areas using SAE and reweighted estimates
Compare of SAEs and ACS estimates for uninsured adults

At State Level from Year 2011-2015:
PCC=0.824 (p=0.04)

At County Levels in Year 2015:
PCC=0.978 (p=0.0002)

At Town Levels in Year 2015:
PCC=0.755 (p=0.03)
Summary

Extended MRP methodology,
  √ Use state BRFSS data
  √ Produce sensible SEs and CIs of SAEs
  √ With flexibility to use single or multiple years data
  √ With flexibility to incorporate survey weights
  √ Estimates are valid

Acknowledgement

Dr. Xingyou Zhang, Ph.D

for methodology development.

Thank you!

Xi Zheng, MPH, MS
CT BRFSS Epidemiologist
410 Capitol Avenue, Hartford, CT
860-509-7670
Xi.Zheng@ct.gov