Covariate-Adjusted Rates May Not Be Better: Thresholds of Rate Ratios for MSE Reductions

Wenjun Li
University of Massachusetts Medical School

Edward J. Stanek III
University of Massachusetts at Amherst
Background

• Public health studies often rely on sample surveys

• Behavioral Risk Factor Surveillance System (BRFSS)

• Major source of information for US public health policy making
BRFSS

Covers many behavioral risk factors:
  Cigarette smoking,
  Sexual behavior & SIV,
  Cardiovascular diseases,
  Drunk driving,
  Physical activity, …,
BRFSS Sampling Design

• Annual random digit dialing (RDD) telephone surveys

• Multi-stage stratified/cluster SRS on community-dwelling adults
BRFSS: Simple Example

• Estimating smoking prevalence rate in a town in Massachusetts (e.g., Wayland) in 2003

• A case of SRS, with a relatively small sample size (n~100)

• Since the population gender proportions are known, should the rates be adjusted for gender proportions?
Rate Estimators

Crude rate = $\bar{Y}$

Post-stratified rate estimator

$$\hat{\pi}_{y[post]} = \pi_x \times \bar{Y}_{\text{male}} + \left(1 - \pi_x\right) \times \bar{Y}_{\text{female}}$$
Population Parameters

- True population smoking rate = 0.20
- 50% male and 50% females
- Male smoking rate is 0.25
- Female smoking rate 0.15
Examples of better and worse performance of post-stratified rate estimates in comparison to crude rate estimates

<table>
<thead>
<tr>
<th>Samples</th>
<th>M</th>
<th>F</th>
<th>Crude</th>
<th>Post-stratified</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.6</td>
<td>0.4</td>
<td>0.22</td>
<td>0.20</td>
</tr>
<tr>
<td>Gender proportion</td>
<td>Gender-specific smoking rate</td>
<td>0.3</td>
<td>0.1</td>
<td>0.22</td>
</tr>
<tr>
<td>2</td>
<td>0.6</td>
<td>0.4</td>
<td>0.22</td>
<td>0.225</td>
</tr>
<tr>
<td>Gender proportion</td>
<td>Gender-specific smoking rate</td>
<td>0.2</td>
<td>0.25</td>
<td>0.22</td>
</tr>
<tr>
<td>3</td>
<td>0.4</td>
<td>0.6</td>
<td>0.18</td>
<td>0.20</td>
</tr>
<tr>
<td>Gender proportion</td>
<td>Gender-specific smoking rate</td>
<td>0.3</td>
<td>0.1</td>
<td>0.18</td>
</tr>
<tr>
<td>4</td>
<td>0.4</td>
<td>0.6</td>
<td>0.18</td>
<td>0.175</td>
</tr>
<tr>
<td>Gender proportion</td>
<td>Gender-specific smoking rate</td>
<td>0.15</td>
<td>0.2</td>
<td>0.18</td>
</tr>
</tbody>
</table>
Is it just circumstantial?

Who cares?
General form of adjusted rate estimator:

\[ \hat{\pi}_y = \bar{Y} - \beta_{yx} \left( \bar{X} - \mu_x \right) \]

\[ \text{var} \left( \hat{\pi}_y \right) = \left( 1 - \rho_{yx}^2 \right) \left( \frac{1 - \frac{f}{n}}{n} \right) \sigma_y^2, \quad \rho_{yx}^2 = \frac{\sigma_{yx}}{\sigma_y \sigma_x} \]

\[ \beta_{yx} = \pi_y[\text{male}] - \pi_y[\text{female}] \]
Estimator depending on known variance components

\[ \beta_{yx} = \sigma_{yx} / \sigma_x^2. \]

\[ \sigma_x^2 \quad \text{May be known, We have to estimate } \sigma_{yx} \]

If we let \( \hat{\beta}_{yx} = \hat{\sigma}_{yx} / \hat{\sigma}_x^2 \) then \( \hat{\beta}_{yx} = \hat{\pi}_{y[male]} - \hat{\pi}_{y[female]} \)

\[ \hat{\pi}_{y[post]} = \pi_x \times \bar{Y}_{male} + (1 - \pi_x) \times \bar{Y}_{female} \]
Examples of better and worse performance of post-stratified rate estimates in comparison to crude rate estimates

(True population smoking rate = 0.20)

<table>
<thead>
<tr>
<th>Samples</th>
<th>Gender proportion</th>
<th>Gender-specific smoking rate</th>
<th>Crude</th>
<th>Post-stratified</th>
<th>Sample beta</th>
<th>Adjusted with pop. Beta = 0.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.6</td>
<td>0.4</td>
<td>0.22</td>
<td>0.20</td>
<td>0.2</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.6</td>
<td>0.4</td>
<td>0.22</td>
<td>0.225</td>
<td>-0.05</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>0.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.4</td>
<td>0.6</td>
<td>0.18</td>
<td>0.20</td>
<td>0.2</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.4</td>
<td>0.6</td>
<td>0.18</td>
<td>0.175</td>
<td>-0.05</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>0.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
How does the uncertainty of $\sigma_{yx}$ affect $MSE(\hat{\pi}_y)$?
Simulations

- Small populations
  - N= 200, 400, 800, 1600, 3200
  - Finite correction may be important

- Imbalanced gender distributions:
  - % males in population: 30%, 35%, 40%, ..., 70%
  - % females: 70%, 65%, 60%, ..., 30%

- Smoking prevalence ranging from low to high
  - Overall rate: 15%, 20%, ..., 35%
  - RR (male rate / female rate): 1.0, 1.1, 1.2, ..., to 4.0

Rate Estimators

Crude rate

Adjusted rate: using known gender proportion and sample estimates of the correlation between Y and X:

\[ \sigma_x \text{ and } \hat{\sigma}_{yx} \]
Simulations

- 10,000 samples of sizes 25, 50, 100, 150, 200
- Sampling fraction: 1.5% - 50%.
- Estimating crude and adjusted rates
- Comparing MSEs of the two rate estimates
Relative Performance

• Ratios of MSEs of the adjusted rates to MSEs of the crude rates.

• The MSE ratios were plotted against % males in population, by population and sample sizes.

• Thresholds of RRs, above which adjusted rates have smaller expected MSEs, were estimated graphically.
Conclusions (I)

The MSE reductions due to gender adjustment depend on:

- Sample size,
- Relative risk,
- Gender ratio.
Conclusions (II)

- In populations with balanced gender ratios, the adjusted rates had smaller MSEs when RRs were above 1.6, 1.5, 1.4, 1.3 and 1.2 for sample sizes of 25, 50, 100, 150 and 200, respectively.

- In populations with unbalanced gender ratios, the RR thresholds were higher.
Take-home Message

Adjusted (post-stratified) rates should not be used in all settings, and in particular, not when both RRs and sample sizes are small.
Message to MA BRFSS

The male to female rate ratio is around 1.15 in Massachusetts, when reporting town-level or subgroup smoking rates with n<200,

* Forget about gender adjustment, &
* Use the crude rates!